Applications of residual dipolar couplings in structural studies of proteins

Kimmo Pääkkönen

Division of Biochemistry, Department of Biosciences,
Faculty of Science, University of Helsinki
Finland

Academic Dissertation

To be presented, with the permission of the Faculty of Science of the University of Helsinki, for public criticism in the auditorium XIII of the main building, Unioninkatu 34, Helsinki, on September 19th, 2003, at 12 o'clock noon.

ISBN 952-91-6136-0 (paperback)

ISBN 952-10-1296-X (PDF)

Dark Oy, Vantaa 2003

Abstract

Residual dipolar couplings (RDC) have recently been introduced to liquid state NMR spectroscopy in studies of protein structures. Although the phenomenon of partial alignment in a magnetic field has been known for a long time, only the recent discovery of dilute liquid crystals to enhance the orientation of proteins has allowed for quantitative measurements of residual dipolar couplings. By using residual dipolar couplings it is possible to determine angles of internuclear vectors with respect to the molecular coordinate system. This kind of information is highly complementary to traditional short-range distance restraints. It has opened many new possibilities to study proteins and their complexes some of which have been explored in the thesis work.

In the present work it was shown by studies of human cardiac troponin C that conformational changes are conveniently elucidated by measurements of residual dipolar couplings. Conformational changes are accompanied with changes in the shape of the protein domain and have to be taken into account in the analysis. Furthermore conformational isomerism that was examined earlier by relaxation measurements does also affect residual dipolar coupling data and will complicate the data analysis.

The rapid growth of known structures deposited in the protein data bank provides possibilities to build molecular complexes from subunit structures. The feasibility of this approach was investigated by studies of the calmodulin-trifluoroperazine complex. The relative orientations of calmodulin domains were determined by residual dipolar couplings and their mutual displacement were deduced from small-angle x-ray scattering data to result in a structure very similar to that obtained by x-ray crystallography.

Finally the use of residual dipolar couplings was extended to denatured proteins. Importantly, it was shown that even random coil structures give non-vanishing values. The study gives basis for understanding information content of residual dipolar couplings recorded from weakly or partially structured proteins.

Acknowledgements

The present study was carried out in the laboratories of Chemical Technology and Biotechnology of the Technical Research Center of Finland, and in NMR Laboratory at the Institute of Biotechnology. I wish to thank research directors Juha Ahvenainen from VTT and Mart Saarma from Institute of Biotechnology for providing excellent working facilities. My warmest thanks to my supervisor Professor Arto Annila, University of Helsinki. I am deeply grateful for his support, encouragement and guidance through all these years. I am also in much dept to Professor Torbjörn Drakenberg, whose vast knowledge in practical and theoretical issues has been invaluable, and I feel honored to have had the opportunity to work with him. I wish also thank Professor Ilkka Kilpeläinen for his support and optimistic attitude. My sincere thanks to Doctor Hannu Maaheimo for support and help.

I wish to thank my colleagues, past and present, with whom I have had the pleasure to work with: Aleksanteri, Anita, Anna, Anne Mari, Anu, Atro, Erja, Hanna, Harri, Helena, Ilona, Jaakko, Juho, Kai, Kari, Katri, Krista, Laura, Maarit, Maija-Liisa, Mari, Martti, Merja, Nana, Niina, Olli, Outi, Paula, Perttu, Piia, Reija, Ritva, Saara, Sami, Susanna, Tero, Tia and Vesa. I could not have done my work without your help, and you all have made my time in the group enjoyable. I want especially thank Helena, who has lightened my working days by her warm and positive character, and whose professional skills I truly appreciate.

I am grateful for Professor Erkki Kolehmainen, University of Jyväskylä, and Professor Mark Johnson, Åbo Akademi, for critically reviewing my thesis work. I also wish to thank Kaija Söderlund from the Graduate School of Informational and Structural Biology for funding and support. Special thanks go to my brother Harri Pääkkönen for language corrections and my sister Tarja Hyvärinen for much needed holiday during my last days of work with the thesis. My greatest thanks go to all my relatives and friends. You are the most important thing in my life. Without your support I would not have achieved this.

Helsinki, July 2003 Kimmo Pääkkönen

Contents

ABSTRACT 3

ACKNOWLEDGEMENTS 4

CONTENTS 4

ABBREVIATIONS 6

ORIGINAL PUBLICATIONS 7

1. INTRODUCTION 8

2. BASIC CONCEPTS 10

2.1 Biology 10
2.1.1 Muscle contraction 10
2.1.2 Troponin C 11
2.1.3 Troponin I 13

2.2 Studying protein structures with NMR 14
2.2.1 Assignment 15
2.2.1.1 Unlabeled samples 16
2.2.1.2 Labeled samples 18
2.2.2 Residual dipolar couplings 22
2.2.2.1 Theory 22
2.2.2.2 Liquid Crystals 25
2.2.2.3 Measuring residual dipolar couplings 27
2.2.2.4 Applications of residual dipolar couplings 28

3. AIM OF THE STUDY 30

4. APPLICATIONS 31

4.1 Troponin C: assignment and structure calculation 31

4.2 Dipolar couplings 32
4.2.1 Conformational changes 32
4.2.2 Building large structures from subunits 33
4.2.3 Denatured conformers 36

5. CONCLUSIONS 39

REFERENCES 40

Abbreviations

ATP	adenosine triphosphate
COSY	correlation spectroscopy
cTnC	cardiac troponin C
DHPC	dihexanoylphosphatidylcholine
DMPC	dimyristoylphosphatidylcholine
DNA	deoxyribonucleic acid
HSQC	heteronuclear single quantum coherence
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser enchancement spectroscopy
NTnC	N-terminal domain of troponin C
PKA	protein kinase A
PKC	protein kinase C
PKG	protein kinase G
RCD	residual dipolar coupling
RNA	ribonucleic acid
SAXS	small angle x-ray scattering
SVD	singular value decomposition
TFP	trifluoroperazine
TnC	troponin C
TnI	troponin I
TOCSY	total correlation spectroscopy
TROSY	transverse relaxation optimized spectroscopy

Original Publications

This thesis is based on the following original publications, which are referred to in the text by the accompanying Roman numerals.

I. Pääkkönen, K., Annila, A., Sorsa, Ti., Pollesello, P., Tilgmann, C., Kilpeläinen, I., Karisola, P., Ulmanen, I., Drakenberg, T. (1998) Solution structure and main chain dynamics of the regulatory domain (residues 1-91) of human cardiac troponin C. *J. Biol. Chem.* **273**, 15633 - 15638.

II. Pääkkönen, K., Sorsa, T., Drakenberg, T., Pollesello, P., Tilgmann, C., Permi, P., Heikkinen, S., Kilpeläinen, I., Annila, A. (2000) Conformations of the regulatory domain of cardiac troponin C examined by residual dipolar couplings. *Eur. J. Biochem.* **267**, 6665 - 6672.

III. Mattinen, M., Pääkkönen, K., Ikonen, T., Craven, J., Drakenberg, T., Serimaa, R., Waltho, J., Annila, A. (2002) Quaternary structure built from subunits combining NMR and small-angle x-ray scattering data *Biophys. J.* **83**, 1177 – 1183.

IV. Fredriksson, K., Pääkkönen, K., Louhivuori, M., Permi, P., Annila, A. (2003) On the origin of residual dipolar couplings from denatured proteins. *JACS, submitted*

1. Introduction

During the last decade our knowledge of genes has increased tremendously. The human genome is almost entirely sequenced and more than one hundred other genomes have been fully determined, and many more are being processed. The next major concerted effort is to study gene products i.e. proteins, their structures and functions at the level of atomic detail.

Proteins are the key molecules of life. They function as structural units, enzymes, receptors, ion channels, chaperons and in many other roles. Chemically proteins are long polypeptides, consisting of a string of 20 different amino acids. These polypeptides fold to unique three-dimensional structures. The structural complexity and diversity of proteins is the source of versatile functions found in nature. However, not much can be perceived from the function by examination of the amino acid sequence only. The detailed physico-chemical shape of the protein dictates its function.

There are basically two methods for determining the three-dimensional structures of proteins in atomic detail. X-ray crystallography is the most common method, and a majority of the known protein structures have been solved by it. X-ray crystallography is today a well-established method. It is fast and precise, and it can also solve structures of very large complexes. However, not all proteins do form crystals. Especially proteins with mobile parts tend to be hard to crystallize. Even when successful, the data from mobile parts of the protein remains often limited.

NMR spectroscopy is the other method that can resolve protein structures in atomic detail. It is not as widely practiced as X-ray crystallography. This is partially because the method has been feasible for biomolecular structure determination only during a relatively short time, from the beginning of 1980's [1]. Furthermore the solution state NMR method is restricted by the molecular size, although recent developments have managed to push the size limitation considerably [2]. Obviously, for NMR there is no need to crystallize proteins. Thus NMR is making important contribution to structural biology by solving those structures that are hard to crystallize.

NMR is a particularly well-suited method for studying protein dynamics and molecular interactions. Understanding these phenomena is increasingly important in order to comprehend how functions of life take place at the molecular level. Thus the NMR method can assess both the structure and dynamics in attempts to decipher the function of a protein, something that may not be obvious from the structure alone. Today, and even more so in the future, a structure that has already been determined by X-ray crystallography is the starting point for NMR studies. Indeed, X-ray crystallography and NMR spectroscopy can be used together to acquire abundant information from molecules and molecular systems.

NMR in structural biology is still a comparatively young discipline. Recent instrumental and methodological developments have increased the size limit of the molecules that can be examined by this method. The advancements in NMR have also allowed working with lower protein concentrations. Novel methods have emerged for studies of proteins and their interactions. Some of these new methods serve to improve the precision of structural models while others deliver completely new information. One of the recent breakthroughs is the discovery of residual dipolar couplings (RDC). Residual dipolar couplings are very important parameters that can be used to increase the precision of structure determination and to identify folds. They are also used to study dynamics and address questions about folding of proteins.

In our work we have used RDCs in studies of biological macromolecules. The work was initiated by studies of troponin C, for which traditional structure determination was carried out [I]. Structural changes in troponin C were further examined using residual dipolar couplings [II]. Subsequently the method for using RDCs and small angle x-ray scattering (SAXS) together to determine molecular complex was demonstrated [III]. Finally, the origins of RDCs from denaturated protein states was studied [IV].

2. Basic concepts

2.1 Biology

2.1.1 Muscle contraction

Muscle cells are highly specialized cells, containing long fibers that contract when the cell receives an electrical impulse. The impulse triggers a release of calcium ions inside the cell. Calcium ions bind specifically to a regulatory protein known as troponin C. The binding launches a cascade of molecular events that finally results in ATP driven muscle contraction. During the contraction two protein filaments slide with respect to each other, causing the entire fiber to contract.

The structural hierarchy of the contractile apparatus is as follows. Muscle fibers are long rod like structures about 50 μm in diameter. They consist of myofibrils, which lay orthogonal to the direction of the fibers. Myofibrils are approximately 1 μm in diameter, and can further be divided into shorter structural units referred to as sarcomeres.

Sarcomeres contain the protein filaments, actin and myosin, which slide in between each other upon muscle contraction. The typical striated look of sarcomeres comes from the various regions that contain either actin, myosin or both.

Actin is a multimeric protein complex, consisting of many globular actin monomers which polymerize to form a double helix structure. Actin is also used in cell cytoskeleton to provide cell motility. Actin forms the thin filaments of the sarcomeres.

Myosin is a long rod-like protein, with a thick head at the end. The rods are coiled with each other, forming a thick fiber. The myosin heads point to both directions from the center. The head of the myosin contains an actin binding part and an ATP binding part. When the myosin head binds to actin, it hydrolyses ATP to make a pull along the

actin fiber. This molecular motor thus moves the fibers between each other, as myosin heads “walk” along the actin fiber, constantly rebinding and pulling along actin fiber.

Binding of myosin to actin is controlled by tropomyosin, which is a 385 Å long fiber like protein along the actin fiber. Each tropomyosin also contains a troponin complex, which regulates tropomyosin’s location on actin. When calcium concentration in the cytosol is low, troponin complex shields myosin binding region from actin. As calcium concentration increases, it binds to the troponin complex, which then moves aside and exposes the myosin binding site. This allows myosin to initiate the walk along the actin and the muscle to contract.

The troponin complex consists of three proteins, troponin C, troponin I and troponin T. Troponin T is the structural protein that binds the complex to tropomyosin. Troponin I inhibits the ATPase activity of actomyosin. Troponin C binds calcium, which triggers conformational change that results in troponin I allowing ATPase activity.

2.1.2 Troponin C

Troponin C is the key regulatory protein in muscle contraction and the one investigated in the present work. Troponin C belongs to a large and diverse family of calcium binding proteins. Typical feature of calcium binding proteins is the EF hand [3] that coordinates Ca^{2+} binding. The calcium binding loop of the EF hand consists of twelve residues that coordinate Ca^{2+} with their side chains or main chain carbonyls. This loop is flanked from both sides by alpha helices.

Skeletal troponin C was the first isoform whose structure was determined [4, 5]. It consists of two globular domains that are connected by a linker region. Each domain binds two Ca^{2+} by the classical calcium binding loops. The C terminal domain is a structural domain, and has a high affinity for Ca^{2+} ($K_a \approx 2 * 10^7\ M^{-1}$), and can also bind Mg^{2+} ($K_a \approx 2*10^3\ M^{-1}$) [6]. Probably it binds either calcium or magnesium ions at all times under physiological conditions. The N-terminal domain of troponin C (NTnC) is the regulatory domain. It binds Ca^{2+} when cellular calcium concentration increases, however the binding is not as strong as in the C-terminal domain ($K_a \approx 10^6$

M^{-1}) [6]. When binding Ca^{2+}, skeletal NTnC undergoes a structural change from a closed state to an open form. The five helices of NTnC are commonly referred to as N, A, B, C and D, the latter is also the linker helix between the domains. The A and B helices flank the first calcium binding loop and C and D the second. Opening of the structure involves a relative reorientation of B and C helices, with respect to the N, A and D helices which remain in place. This opening exposes a hydrophobic patch for troponin I to bind.

Initially the mechanism in cardiac muscle was thought to be similar to the one in skeletal muscle. It was known that the first of the calcium binding sites in NTnC had diverged by mutations that had rendered it inactive so that cardiac NTnC binds only one calcium ion. Calcium binding loops consists of twelve residues, with six amino acids contributing to the binding of Ca^{2+}. In NTnC, there's one insertion (V28) and two mutations (D29L and D31A) compared to the skeletal form. Both mutated aspartate side chains normally participate in the Ca^{2+} coordination.

When the structure of cardiac NTnC was determined [7] it became apparent that cardiac NTnC was in the closed conformation also in the Ca^{2+} form. Later also the complex structure of cardiac NTnC with TnI peptide was solved [8], revealing an open conformation. Thus the cardiac troponin C is different from the skeletal isoform, and leaves the exact mechanism of the opening of the structure unclear.

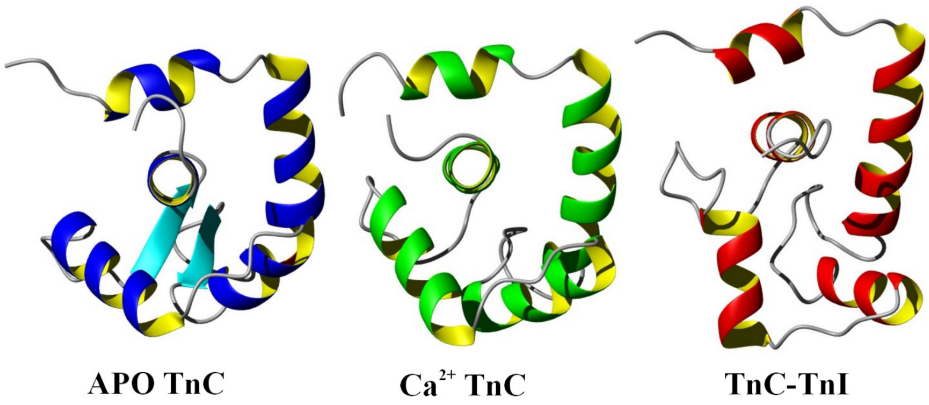

Figure 1. Conformations of cardiac troponin C in its different forms (1SPY, 1AP4, 1MXL). Structures were superimposed according to the structurally invariant NAD-frame. Only minor differences can be seen between apo and Ca^{2+} forms of TnC, whereas the TnC-TnI complex is clearly in the open conformation. The figure was made with the program MOLMOL [9].

2.1.3 Troponin I

Troponin I (TnI) inhibits actomyosin ATPase [10]. There are two isoforms of skeletal TnI, and one of the cardiac form. Actin binding sites overlap those that bind troponin C. When TnI binds to troponin C, it dissociates from actin. The whole troponin complex moves away from actin, allowing myosin to bind actin.

TnI interacts with TnC through several binding sites. The main chains of TnC and TnI are antiparallel [11, 12]. TnI has three sites that bind troponin C. The first binding site, residues 1-40, interact with the C-terminal domain of TnC [13, 14]. The two other binding sites, 96 – 116 [15] and 128 – 148 [13], bind central helix and NTnC, respectively.

Troponin I has several phosphorylation sites. TnI can be phosphorylated by three protein kinases, protein kinases C, G and A (PKC, PKG, PKA). Cardiac TnI has a 31 residue extension at its N-terminus. In this extension, serines 22 and 23 are both substrates for PKA phosphorylation. Phosphorylation of these residues increases the rate of Ca^{2+} dissociation of cardiac TnC [16]. The exact mechanism how the phosphorylation at the N-terminus of TnI is transmitted to N-terminus of TnC is not known.

2.2 Studying protein structures with NMR

Solution state NMR spectroscopy is a versatile method to study proteins. However, it is only applicable to proteins of a limited size. There are several reasons for the size limit. When using non-labeled samples, the complexity of the spectra (especially NOESY spectrum) increases rapidly, in effect making studies of proteins over 100 amino acid residues difficult and prone to errors. On the other hand, relaxation rates of NMR signals increase with correlation time. Hence for very large proteins the problem is not too many signals but too few signals or no signals at all. Also the requirement for high concentrations (approx. 1 mM) due to the intrinsic insensitivity of NMR may well be a limiting factor, because not all proteins are that soluble.

Fortunately recent developments have significantly alleviated the situation. Sensitivity of the measurements has increased as the magnetic field strengths have increased. Also the advent of cryoprobes, which have markedly lower thermal noise and consequently a better signal-to-noise ratio, have brought about a four-fold increase in sensitivity. Today studies can be conducted using lower protein concentrations or alternatively measurements at high concentrations can be carried out much faster than before. Finally the invention of transverse relaxation optimized spectroscopy (TROSY) [2] has significantly increased the size limit.

The number of signals that are observed in a particular NMR spectrum can be reduced by isotope labeling. The common isotope labels used are ^{13}C and ^{15}N labels. Perdeuteration, i.e. ^{2}H labeling, has also become widely practiced, in particular when using TROSY experiments. With isotope labels the complexity of spectra is brought down by increased dimensionality and enhanced selection of signals that are observed in the spectra. Based on isotope labeling signal identification strategies have been developed. Customarily alpha and beta carbon shifts are used to recognize amino acid type and sequential connectivities for building up the sequence specific assignments. For very large proteins amino acid specific labeling provides further simplification of the spectra.

Isotope labeling does impose certain requirements on the sample preparation. Usually the expression has to be carried out by bacteria in a minimum media in order to keep labeling costs reasonable. Consequently the yields tend to diminish.

^{2}H labeling is not primarily used to alleviate the assignment but to slow the relaxation processes. Relaxation rates are often dictated by the proton density, thus perdeuteration serves to reduce the signal decay. The other way to slow down the signal decay is TROSY, which basically makes the relaxation mechanisms of dipole-dipole and chemical shift anisotropy cancel each other out. TROSY works very well for amide groups where the optimal magnetic field strength is a little over one gigahertz field expressed in terms of proton Larmor frequency. This field strength should be achievable in the near future, as there are already 900 MHz instruments.

NMR data analysis has traditionally been carried out manually, using computers only to display the data and to keep record of the assignments. This process is usually very time consuming and tedious. This is somewhat surprising because the process itself is well understood and could in principle be automated. Data is however often noisy and sparse, making the automation task much more demanding than it initially seems. Even the first step, identification of peaks from noise, can be hard. Magnitude of the weak peaks are close to the noise level and their line shapes may be distorted. This makes the separation of the signals from the noise complicated. Also fast relaxation may diminish the intensity of an expected signal below detection. Furthermore spin diffusion, i.e. relay of magnetization in the spin system, in nuclear Overhauser enhanced spectra may give rise to signals that are not reporting from close spatial connectivities. Errors in peak identification make subsequent steps harder, thus peak identification and analysis of distance restraints is often carried out iteratively.

2.2.1 Assignment

The step of spectral interpretation where signals are identified to specific amino acids and their atoms is referred to as the assignment. It is usually separated in two overlapping phases. First amino acid type is determined, then its position in the sequence is concluded.

2.2.1.1 Unlabeled samples

The first NMR structures were determined using unlabeled samples. The use of non-labeled samples is attractive from the sample preparation point of view, as material costs are low and there is no need to grow host in the minimum medium. Unfortunately this approach severely limits the range of experiments that can be used and structures of only relatively small proteins can be determined. With increasing number of amino acids the proton spectra become very crowded, putting the practical limit for structure determination to around 100 amino acids. Spectral overlap makes both the assignment and extraction of reliable distance restraints ambiguous and difficult. Also the time to analyze spectra increases and the process becomes increasingly tedious.

COSY and TOCSY are the two experiments that are primarily used for amino acid type determination. Both experiments are based on through-bond scalar connectivities. COSY gives cross peaks for protons that are directly spin coupled to each other, in practice not separated by more than three bonds. TOCSY gives cross peaks to all protons in the same spin system, i.e. within the network of spin coupled protons (Figure 2). The connectivities displayed in TOCSY can be further modified by varying the mixing time in the experiment. The mixing time, i.e. coherence transfer time, governs the intensities of the cross peaks. For short mixing times, adjacent nuclei give correlations with large intensities, and for long mixing times nuclei that are remotely coupled give rise to peaks with increased intensities. While information in TOCSY contains information present in COSY, TOCSY is easier to use owing to its better dispersion. On the other hand COSY type experiments can be used to extract scalar couplings to determine dihedral angles.

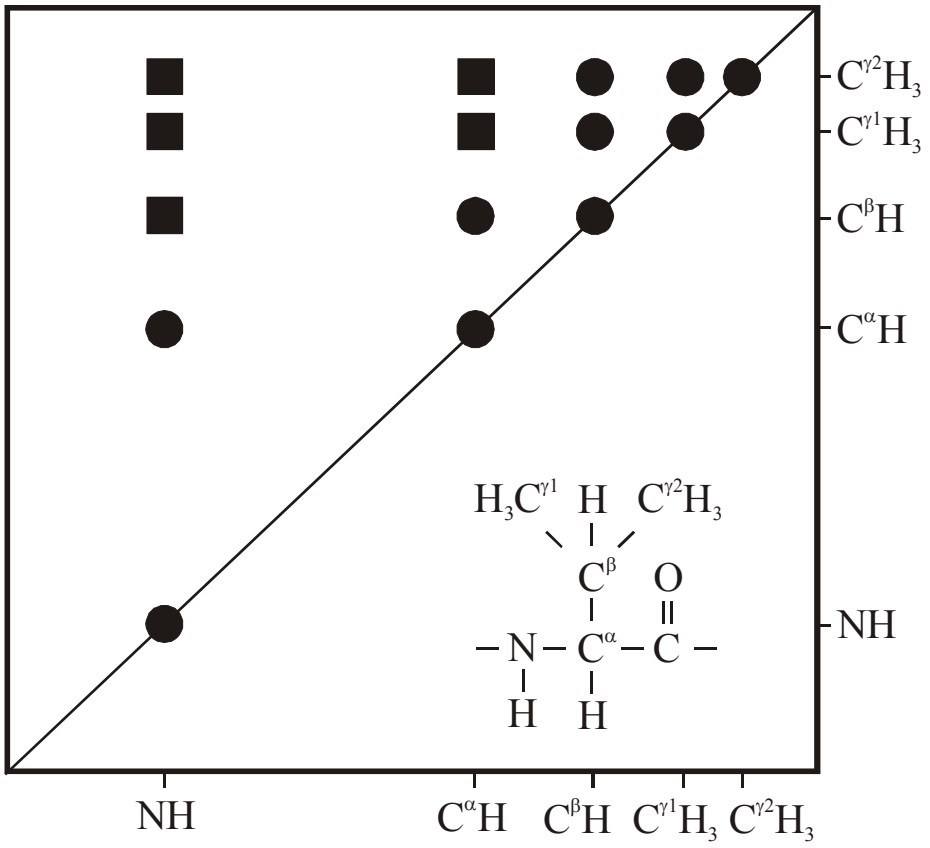

Figure 2. Peaks observed in COSY and TOCSY experiments, using valine as an example amino acid. Circles mark peaks seen in COSY and TOCSY spectra, squares mark peaks seen in only TOCSY spectrum. In COSY spectrum, cross peaks are seen if protons are separated by at most three chemical bonds, while in TOCSY all peaks in the same spin system are seen. For sake of clarity only peaks above the diagonal are marked.

Amino acid type is determined based on the number of signals and characteristic resonance frequencies i.e. chemical shifts. However, proton shifts are very sensitive to chemical environment. Especially the ring current of the aromatic ring may strongly influence nearby proton shifts. This may give raise to abnormal chemical shifts, complicating amino acid typing. The peak intensities are also affected by relaxation rates, which may make some peaks or even spin systems unobservable. Clearly due to these problems, together with increasing spectral overlap, NMR spectroscopy using non-labeled samples are only suitable for small proteins.

Amino acid sequence position is determined from a NOESY spectrum. The NH-NH region of the NOESY spectrum usually contains all sequential NH-NH peaks, which allows one to "walk" through the sequence. This is especially true for α-helices where these distances are short. Alternatively sequential walk via NH to preceding $C_\alpha H$ is a suitable strategy in β-sheets and extended structures. Prolines and residues with no signal due to fast relaxation produce gaps to the sequential walk as there is no amide signal in such cases. Often the sequential connectivity can be validated by more than one cross peak as neighboring amino acids are spatially close, giving rise to many sequential connectivities.

The last step in the analysis of the NOESY spectrum is to collect distance restraints. Complexity of this task is strongly dependent on the spectral overlap and the completeness of the assignment. The analysis of NOESY is an iterative process. The analysis commences from peaks that are strong, unambiguously identified and do not overlap. These are then used to generate an initial structure. After the initial structure has been obtained some of the ambiguous signals can be assigned by the help of the initial structure to rule out inconceivable assignments. This cycle goes on until most of the signals have been analyzed and further improvements on the structure are no longer obtained. Erroneous assignments in the early steps are especially unfortunate as they can lead the process to a wrong direction, so special care must be taken during the first iteration rounds.

2.2.1.2 Labeled samples

Labeling proteins by magnetic nuclei has become the standard way of carrying out protein structural work by NMR. The labeled samples provide several advantages over the non-labeled ones. In the structure determination, the assignment is simplified considerably. Labeling also allows for various types of experiments, e.g. relaxation studies of the main chain and the side chains, and measurement of residual dipolar couplings.

Most commonly employed labels are ^{15}N, ^{13}C and ^{2}H isotopes. ^{15}N labeling is the easiest and cheapest to introduce. *E. coli* is cultivated in a medium containing labeled

ammonia as the nitrogen source. Yields are usually good. ^{13}C labeling is achieved by growing *E.coli* in a minimum medium with labeled sugar as carbon source. ^{13}C labeling is more costly than ^{15}N labeling, and the *E.coli* must be adapted to grow in minimum medium, which usually lowers the yields. ^{2}H labeled samples are prepared by growing *E.coli* in $^{2}H_2O$. Often all three labels are used simultaneously.

Isotope labeling simplifies the assignment process by significantly lowering the spectral overlap. The experiments are usually three-dimensional, in contrast to the two-dimensional spectra that are most practical with non-labeled samples. The added dimensionality improves the signal dispersion by reducing the likelihood of spectral overlap. The second reason for the reduced overlap is that the heteronuclear experiments are more selective. Only some of the signals that are seen in, for example TOCSY spectrum, can be selected to appear with labeled samples. This lowers the number of signals seen in the spectrum and thus the overlap.

There are several experiments from which to derive the spin systems from. In each only a subset of the signals belonging to a spin system are present. The larger number of experiments result in many spectra that are simpler to analyze. Also there is considerable redundancy in various experiments. This helps to differentiate noise peaks from actual peaks by checking if the weak signals appear in several spectra.

Most common spectra used for the main-chain assignments are HSQC, HNCO, HNCA, HNCACB and HNCOCACB (Figure 3). HSQC serves as the reference spectrum and it is not strictly necessary for the assignment. However, HSQC displays a better resolution in the nitrogen dimension than the other experiments, which can help to unravel potential signal overlap. With these spectra it is possible to establish spin systems that have HN, N, CO, CA and CB chemical shifts. CA and CB chemical shifts are very dependent on chemical structure and thus well suited for the amino acid type determination. The sequential assignment can be gathered from the very same spectra. Connectivities can be verified by the CO, CA, CB and HA nuclei, which significantly reduces chance of overlap.

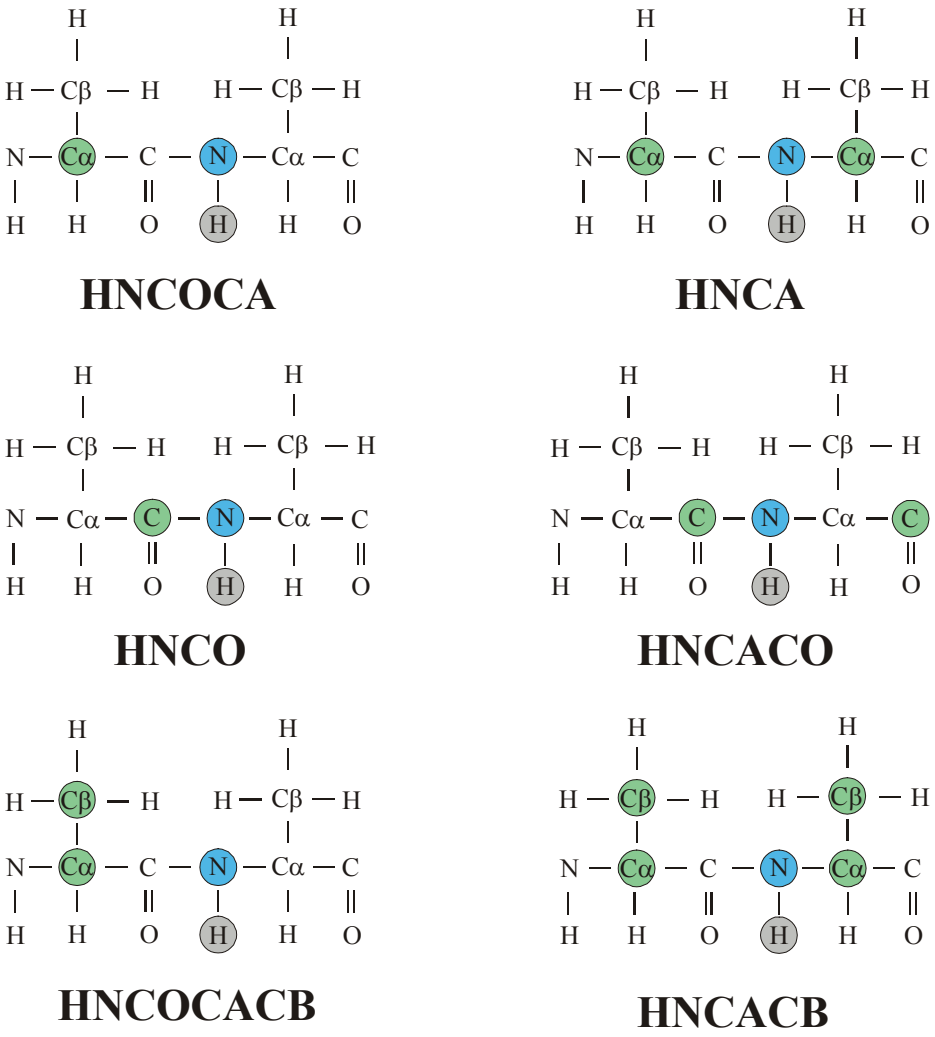

Figure 3. Schematic presentation of the typical NMR experiments that are used for protein backbone assignment. In favorable cases only a subset of the experiments is needed.

There are several software packages that do backbone assignment automatically [17 - 21]. For well behaving proteins these methods usually work, but if the spectral quality deteriorates so does the completeness and accuracy of the sequential assignment. This is also true for manual analysis, so the problem is intrinsically related to the sample preparation and spectroscopy, not to the data analysis itself.

One major advantage with sequential assignments based on ^{15}N, ^{13}C labeled proteins is that it relies exclusively on through bond connectivities. This is in contrast to proton

only assignments where also through space (NOESY) connectivities are needed. The backbone assignment with labeled samples is relatively straightforward and mechanical in nature. The time required to make backbone assignment manually varies from a week to couple of months, depending on the complexity of the spectra and the experience of the spectroscopist. Using automatic assignment is considerably faster, as the algorithms usually run in minutes. Most of the packages are semi-automatic, as they require some manual interference, at least in difficult cases. Many require sensible input data with not too much noise, which often means that peak picking needs to be done manually. Automatic peak pickers have also become better, so they may soon replace manual peak picking.

Side chain assignments and NOESY assignments are considerably harder to obtain than the main chain assignment. Especially automating NOESY spectrum assignment has received lots of attention [22 – 28]. Often the success in NOESY assignment is critically dependent on the correctness and completeness of amino acid typing and sequential assignment. There are recent attempts that try to bypass the assignment of signals altogether by using method similar to X-ray crystallography. In these, a proton density cloud analogous to the electron density is calculated and used directly to fit the covalent structure [29, 30].

Automation of data acquisition and analysis are in general a very important area of NMR structural studies, especially when considering the structural biology program. Traditional manual structure determination is too slow to meet the demands of high-throughput structure determination. A full manual structure determination takes months or even years. In principle this time could be shortened to days if all the steps from peak picking to structure calculations would be automated. Measurements of NMR spectra would then be the most time consuming step. Automation would allow NMR to become comparable in speed with x-ray crystallography and an even stronger contributor to the research in structural biology than it is currently.

2.2.2 Residual dipolar couplings

2.2.2.1 Theory

Partial alignment as the source of residual dipolar couplings has a long history. Studies on the magnetic-field-aligned liquid crystals were performed a long time ago [31]. The observation of magnetic alignment, where molecules with sufficiently large magnetic susceptibility anisotropies align in magnetic field, was made with d_6-benzene in 1978 [32]. Paramagnetic systems were the first in which residual dipolar couplings were applied to macromolecules [33, 34] and proteins [35]. However, most biological macromolecules are diamagnetic, hence the intrinsic alignment in the magnetic field is very small. Not many proteins have large magnetic anisotropies which limits the use of intrinsic magnetic alignment. Also the amount of alignment is often so small that the magnitude of the measured dipoles is only fractions of hertz.

The amount of orientation can be enhanced by liquid crystals, however to this end the ordinary liquid crystals are too dense. Many interactions are seen in the spectrum and cannot be analyzed for large biological macromolecules. It was only after the use of isotope labels became widespread and by the advent of dilute aqueous liquid crystals that applications to proteins became feasible.

A dipolar coupling contains information of the angle between the direction of the magnetic field and the vector between dipolarly coupled nuclei (Figure 4). For the residual dipolar couplings to become observable the molecule has to be partially aligned, otherwise the rotational motion in solution will very effectively average out dipolar couplings. On the other hand if the molecule was strongly aligned, as for example when bound to a membrane, the spectral complexity would increase substantially due to large dipolar couplings, making the spectrum intractable. In dilute liquid crystals proteins align only slightly, typically the order of alignment is only few parts in thousand.

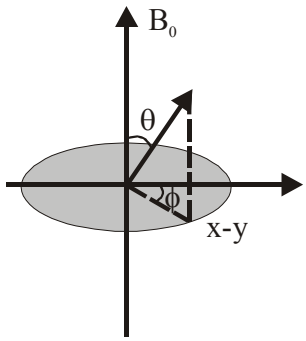

Figure 4. B_0 points the direction of the magnetic field. Angle θ is the angle between dipolar coupling vector and the magnetic field, ϕ is the projection angle of the dipolar coupling vector.

Residual dipolar couplings can be calculated with following formula:

$$D^{AB}(\theta,\phi) = D_a^{AB}\left[(3\cos^2\theta - 1) + \frac{3}{2}R\sin^2\phi\ \cos 2\phi\right] \quad (1)$$

where θ is the angle between dipolar coupling vector and the magnetic field, and ϕ is the projection angle of the bond vector to the x-y plane (Figure 4) [36]. Rhombicity R expresses the amount of departure from the cylindrical symmetry. D_a is the axial component of the molecular alignment tensor that can be calculated from the equation 2:

$$D_a^{AB} = -\mu_0(h/2\pi)\gamma_A\gamma_B/\left(4\pi^2 r_{AB}^3\right) \quad (2)$$

D_a depends on gyromagnetic ratios of the nuclei γ_A and γ_B, magnetic permittivity of vacuum μ_0 and distance between the nuclei r_{AB}. h is Plank's constant. Importantly the magnitude of D_a is proportional to the amount of alignment S governed by the density of the liquid crystal and the molecular shape.

The information contained in residual dipolar couplings is complementary to the NOE data, especially when examined from the structure determination point of view. An NOE is a short-range constraint, and often not very precise. For example in the DNA

and RNA studies, a slight bending of the double stranded DNA or RNA can be much more easily detected via residual dipolar couplings than by using NOEs. The detection of long range effects would require greater precision from NOE measurements than can reasonably be expected.

There is an abundance of measurable dipolar couplings, both in the main chain and in the side chains [37, 38]. The most commonly measured residual dipolar coupling is the NH coupling. It can be measured simply from only ^{15}N labeled sample, it is large and generally the easiest dipole to measure owing to the large chemical shift dispersion of the amides. Other dipolar couplings include NCO, NCA, COCA, HNCA, HNCO and HACA (Figure 5). HACA is a large coupling but the others are smaller, implying that the precision will be smaller as well. Nevertheless all couplings provide valuable restraints. Therefore the use of several dipolar couplings allows a more precise structure determination.

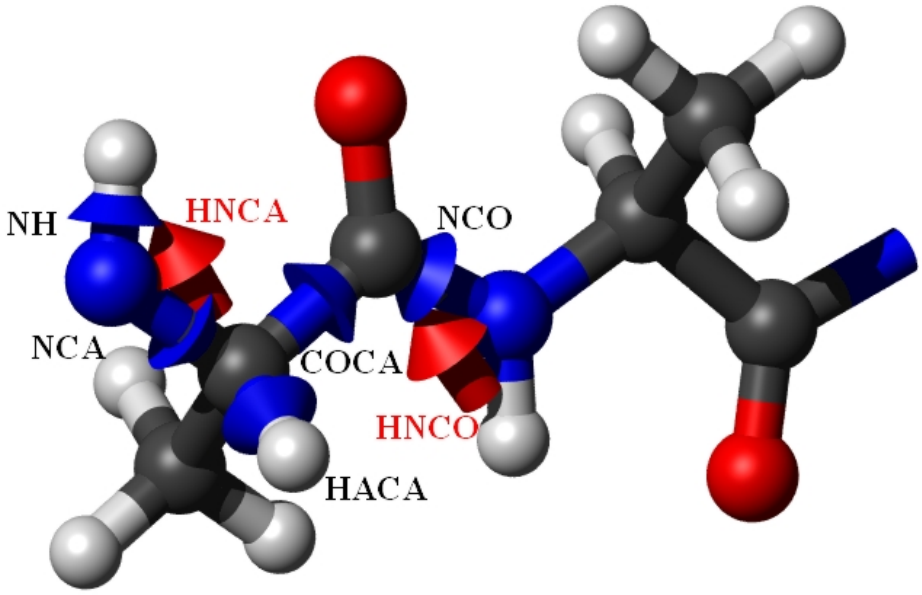

Figure 5. Arrows show some residual dipolar couplings that can be measured from the main chain.

One complication in the analysis of dipolar couplings is that they do not have an one-to-one correspondence between dipolar coupling value and polar angle θ (Figure 6).

Dipolar couplings cannot differentiate the sign of the alignment vector i.e. the functional form is even. This will result in mirror image solutions for the alignment tensor. Differentiating between these can be accomplished by recording the dipolar couplings in several alignments. This assumes that the conformation remains invariant.

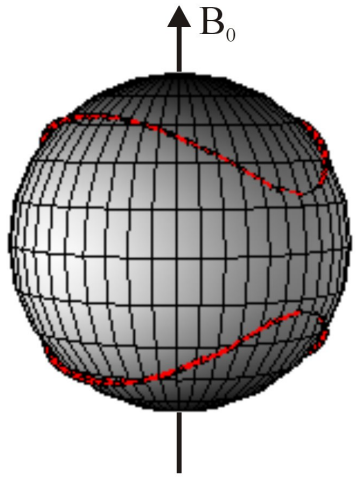

Figure 6. Dipolar couplings cannot differentiate the sign of the alignment vector. Single measurement restricts the orientation of the internuclear vector to two, possibly distorted, cones.

2.2.2.2 Liquid Crystals

The mechanism of enhanced alignment may be complex, including steric, electrostatic and specific surface-to-surface interactions. If the steric hindrances dominate, it is possible to predict the dipolar couplings from the structure. This can be done either by simulation [39] or by analytical method [40].

A liquid crystal is a mixture between solvent and ordered particles in it. The particles orient themselves in the magnetic fields. As the particles in liquid crystals or the filaments of a matrix are oriented, they induce a weak alignment of proteins and solutes in general. This happens primarily due to a steric obstruction that will affect the concentrations for the various rotational states. Also interactions, for example orientation due to opposing surface charges between the protein and the liquid crystal particle, or weak binding of the protein to the particles can influence the molecular

alignment. If the chemical shifts change significantly in the anisotropic medium there is the possibility that the structure has changed, which can make the interpretation of dipolar couplings unreliable.

The first used liquid crystal medium was bicelles [41]. These discoidal particles will form from a mixture of lipids. Depending on the temperature and the concentration, lipids form discs or stacked lamellas [42]. Commonly dimyristoylphosphatidylcholine (DMPC) and dihexanoylphosphatidylcholine (DHPC) are used, although many other lipids form bicelles. Typical concentrations of bicellar media are 3 – 10 % (w / v) [41]. There are several modifications to bicelles which alter the pH and temperature ranges where the bicelles are stable, or change their surface charge [43].

Rod-shaped viruses form also liquid crystal [44 – 46]. Bacteriophage Pf1 is extensively studied and used as liquid crystal medium. It contains circular single stranded DNA which is packed inside a protein coat. The length of the rod is approximately 20000 Å and the diameter is about 60 Å [44]. In magnetic field the DNA, and thus the whole virus, orients itself along the magnetic field. Pf1 is very stable and tolerates temperatures at least from 5 – 60 °C.

Other liquid crystal media that have been used with proteins include purple membranes [47], cellulose crystallites [48] and mixture of cetylpyridinium halide, n-hexanol and sodium halide [49]. Also uniaxial matrices obtained by stretching or compression e.g. polyacrylamide gels can also be used to align proteins [50 – 52]. Polyacrylamide gels are inert and stable material over a wide temperature range.

Obtaining dipolar couplings in several different alignments is desirable. This, however, can be problematic, as not all liquid crystal media are suitable for all proteins. One way of obtaining different orientations with same media is to change the charge distribution on the surface of the orienting particles [53] to balance between steric obstruction and Coulombic interactions.

2.2.2.3 Measuring residual dipolar couplings

The basic assumption when using the dipolar couplings is that the interaction between a liquid crystal and a protein does not change the structure or dynamics of the protein. Indicators of this potential problem are the chemical shifts; they should be close to the solution values. Small changes (< 0.1 ppm for ^{15}N and ^{13}C, < 0.01 ppm for ^{1}H) are to be expected due to chemical shift anisotropy effect [54, 55]. In the case of enzymes, one could also try to measure the activity to confirm the function of the protein.

Residual dipolar couplings should preferably have magnitudes that allow enough precision in the measurements. Too high degree of alignment may cause proton-proton couplings to show in the spectrum, broadening the signals and degrading the spectral quality. On the other hand, if the concentration is too small, the line widths will be sharp but the magnitude of the dipolar couplings will be small. By carefully balancing between the magnitude and the line broadening it should be possible to find a compromise that gives good precision. This is especially important for measuring dipolar couplings which are small in magnitude.

The common method of measuring dipolar coupling is to obtain the frequency difference of the doublet peaks. To facilitate the analysis it is possible to record multiplets into two sub spectra instead of in the same spectrum [56]. This reduces number of the peaks to half, thus reducing the potential overlap and allowing more precise measurements of the frequencies.

To measure one residual dipolar coupling, in fact four peak centers have to be extracted. Two of those are from the reference peaks in the water to obtain the scalar coupling, which has to be subtracted from the dipolar coupling. Reference spectra give usually sharp and well-defined peaks. In the liquid crystal peaks may broaden due to proton-proton dipolar couplings if the concentration of the liquid crystalline medium is too high.

2.2.2.4 Applications of residual dipolar couplings

Residual dipolar couplings are very useful parameters. In structure calculations residual dipolar couplings can be used as additional restraints [57]. They serve to improve the precision and the accuracy of the structures. This can be especially important for RNA and DNA studies, where NOE data is sparse, as well as for loop regions in proteins [58 – 60]. Residual dipolar couplings can also be used to calculate rough structure in the absence of NOE information. Although the precision will not be high, overall fold can be calculated [61 – 63]. This can be achieved by backbone assignment alone, which is much faster than doing the side chain assignment and the NOESY spectrum analysis. It is even possible to use residual dipolar couplings directly to find the homologous structure by comparing measured dipolar couplings to those calculated from structures [64 – 66]. Results from these fast methods can be used to judge if more thorough studies are worthwhile.

In protein-ligand or protein-protein interaction studies residual dipolar couplings can be used to assess structural changes [II]. By fitting residual dipolar couplings to a structure a good fit implies no large structural changes. If there are regions where the fit is poor, it is indicative of structural or dynamic changes. It is important to fit the dipolar couplings to the structure and not simply compare them, as the values of the residual dipolar couplings themselves change due to a reorientation of the molecule in the liquid crystal lattice even if the structure had remained relatively unaltered.

Large complexes can be constructed from known subunits by using residual dipolar couplings to determine the orientations of the subunits [67 – 70]. This can be useful for domains separated by a flexible linker, which do not have NOEs between the domains. Residual dipolar couplings are measured for both domains, and the domains are then oriented independently of each other. The relative orientation is not immediately clear, as dipolar couplings cannot differentiate the sign of the dipole, and thus four symmetrical solutions are acquired per subunit. If there is a linker, some of the combinations may be excluded because of the linker length. Measuring dipolar couplings in several orientations can resolve the symmetry problem [71, 72].

Residual dipolar couplings can be used to study dynamics. RDCs are sensitive to a wide range of time scales, making them good probes for biologically interesting motions [73]. Studies of dynamics are still few and not well established [74 – 76]. Recently a formalism which allow for simultaneous extraction of structural and dynamic parameters from dipolar couplings was suggested [77].

3. Aim of the study

The aim of this study was to study protein structures with modern NMR techniques, in particular using residual dipolar couplings. The first two papers focus on studies of troponin C by methods of protein structure determination and use of dipolar couplings to detect structural changes in TnC. In the third paper dipolar couplings together with SAXS are used to determine the structure of a calmodulin-trifluoroperazine complex. In the fourth paper highly denatured ubiquitin is examined by residual dipolar couplings.

4. Applications

4.1 Troponin C: assignment and structure calculation

Structure determination of the N-terminal domain of cardiac troponin C was initiated using unlabeled samples. Due to the severe overlap in the unlabeled spectra several temperatures were used to aid the unambiguous assignment of the signals. NOESY, COSY and a set of TOCSY spectra were recorded in 30, 40 and 50 ºC. These were used to obtain amino acid type identification, sequential assignment and initial structures. The assignments were later confirmed with the aid of ^{15}N labeled sample, which was also used to facilitate the structural refinement. The lack of a ^{13}C label did however limit the precision of the resulting structures.

The structure determination revealed that the calcium-loaded form of N-terminal domain of the cardiac troponin C was in its closed conformation, contrary to that of the skeletal form. This was also confirmed by structural studies done by another group [7, 78]. Helix angles, which typically change when the structure opens, are closer to the skeletal apo NTnC angles than to skeletal Ca^{2+} NTnC angles.

Distance difference matrices were calculated to compare the structure against skeletal apo- and Ca^{2+} NTnC structures. The maps showed close similarity between our structure and the skeletal apo form, while our structure and skeletal calcium form were different from each other. Differences between our structure and skeletal calcium form are similar to those between skeletal apo- and calcium form, which further assures that skeletal apo NTnC is very similar to cardiac Ca^{2+} NTnC.

The closed form of cardiac NTnC does not explain why TnI would bind only to the Ca^{2+} form of NTnC and not to the apo form. Some changes in the structure or dynamics must contribute to the specificity. Keeping that in mind, dynamics of the NTnC were studied on the basis of ^{15}N line width analysis and relaxation times measurements. The defunctional calcium binding loop, along with a few other

residues, emerged as highly mobile in both measurements. Thus it was reasoned that while calcium binding does not open cardiac NTnC, it does change its dynamics.

An alternative explanation for broad line widths and altered relaxation rates could be monomer-dimer exchange. This is indistinguishable from slow exchange. 9 % TFE was used in the sample preparation to prevent dimerization. However, as the residues with the broadest lines did not form a contact surface, it seems that the conformational change is more probable than dimerization.

4.2 Dipolar couplings

4.2.1 Conformational changes

Cardiac troponin C undergoes structural change when TnI binds to the calcium saturated form [8]. Troponin C was therefore used as a model system to see if this structural change could be detected by using only dipolar couplings.

NH residual dipolar couplings were measured from NTnC in its apo- and calcium-saturated forms as well as in complex with a TnI peptide. They were used to follow the structural changes in the different states of troponin C. Each set of dipoles was compared against all three structures. To compare dipoles against the structures, the structure must be oriented correctly and the theoretical dipolar couplings from the oriented molecule must be calculated. This can be done most conveniently by using singular value decomposition (SVD) [79]. A C program was written to this end.

For the apo and TnI forms the measured dipoles fit best with the corresponding structures, and clearly worse to the opposite structures. These measurements showed that the quality of the fit decreases clearly when the structure changes.

The calcium form was more problematic, its dipoles did not fit particularly well to any structure. Indeed the worst fit was against the calcium form, which was expected to give the best fit. One explanation to this could be a conformational exchange between two (or more) forms, of which evidence was observed also in our earlier study [I].

This study of conformational changes in cardiac NTnC shows that it is possible to detect large conformational changes by using dipolar couplings. Better sensitivity for detection could be achieved when using several different dipolar couplings. It would also give more reliable results due to redundancy and allow measurements of some amino acids that could not be measured due to overlap.

4.2.2 Building large structures from subunits

Dipolar couplings can be used to determine the relative orientations of protein domains [79 – 81]. The data will give several symmetry related solutions. Finding the correct solution among the symmetry alternatives may be possible by considering steric limitations imposed by the linker region, or possibly by chemical shift changes in the case of separate molecules. It is possible that these methods don't leave only one possibility. A long linker may allow several solutions, and in binding studies the possible conformational changes may cause chemical shift changes in other places in the molecules than in the binding interface, making identifying the binding site nontrivial.

SAXS is a method that can be used together with residual dipolar couplings for building large structures from subunits. SAXS gives the distribution of distances between the atoms. The measured data can also be compared with values calculated from the structure by using a program called CRYSOL [82]. SAXS is not very sensitive to the rotations of the domains, although it can help determine the mirror solution problem of the dipolar couplings. The SAXS data is, however, sensitive to the distance between the subunits. SAXS measurement is easy to carry out, and the conditions are the same as in the NMR experiment, so measuring it does not involve much additional work.

The combination of these two methods takes advantage of each method in measuring the parameter for which it is most suitable. Dipolar couplings are sensitive to orientation, but cannot identify symmetrical solutions or tell anything about the distance between the molecules. SAXS suites well for measuring distances, and while

its sensitivity for the orientation is poor, it can still separate symmetry alternatives for a reasonably asymmetric molecule.

Calmodulin was used as a test system for building a molecular complex using dipolar couplings and SAXS. Calmodulin has been extensively studied and thus it serves as an example case. In solution, calmodulin domains are free to rotate and move with respect to each other, within the limits of the short linker. Calmodulin can bind trifluoroperazine (TFP) to form a complex of four TFP molecules per calmodulin. The structure of the complex is known [83]. In the complex the domains are fixed, forming a compact globular complex. The knowledge of the crystal structure allowed for a comparison with the model derived from NMR and SAXS measurements and thus the analysis of the accuracy and robustness of the developed method.

The NH residual dipolar couplings were measured from the calmodulin-TFP complex and fitted against the crystal structure of the complex (1LIN). In this fitting the domains of the structure had been separated from each other to allow each domain to rotate independently. A program for the data fitting was written in Python, and also jackknife and Monte Carlo methods were implemented for error analysis. The angles of the domains were reasonably tolerant to errors, with 1 Hz measurement error or 10 % missing data not severely misaligning the molecules.

The earlier C program was converted to Python because of better scalability and maintenance. Although Python is an interpreted language and not compiled, as C is, it was possible to maintain the speed by first extending Python with a small C module that contained the SVD algorithm, and later by just using standard python extension called Numerical Python. Python programs can be run without recompiling in several different environments, e.g. Unix, PC and Mac, which was also seen as an advantage. Due to the change in programming environment the development of the program was fast with little need for debugging. The resulting program was better structured due to the use of object oriented programming, and easy to extend.

Molecules oriented by dipoles served as an input to program CRYSOL for SAXS calculations. A Python program was made for translating the coordinates of the domains and controlling CRYSOL in order to perform a grid search. These

calculations resolved the distance between the domains and enabled the selection of the correct symmetry solution and allowed the building of a model of the system.

The resulting model was very similar to the crystal structure of the complex. In the crystal structure, the domains were slightly closer together than in the model. This might be a true difference between the solution and crystal structure or due to incompleteness of the model. The calmodulin-TFP complex presented a challenge for the method owing to the TFP ligands. There was no RDC data from the TFP molecules. Thus they were not in the model although they were present in the measurement of course. This causes uncertainty for the comparison of the calculated and measured SAXS data, as the model and the actual molecules in the measurements differed. It was possible to obtain an improved fit by correcting for the lacking scattering power by either assuming a spherical scatterer placed in between the domains or by placing the TFP molecules, one at a time, in the center region.

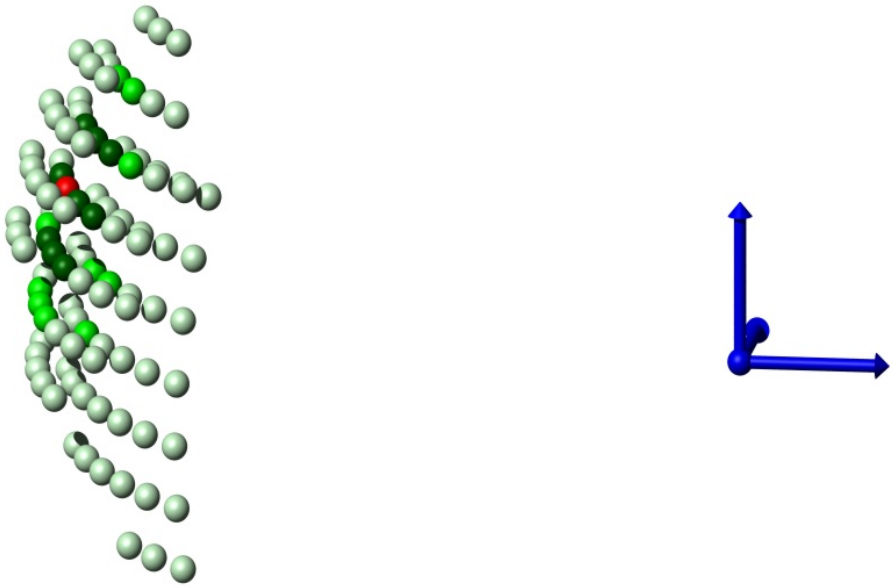

Figure 7. Spatial precision of the combined NMR-SAXS for the domains. The center of the axis system marks center of the N-domain. Red ball marks the minimum error sum from the fit, i.e. the center of the C-domain. The grid points were separated by 1 Å. Color shading mark the change in the error sum, with smaller values in darker shades. Figure was prepared with the program MOLMOL [9].

Despite the shortcomings in the model, the results were reasonable (Figure 7). The distance between the domains is obtained with precision of ± 1 Å, while perpendicular precision is ± 4 Å.

Building complexes from smaller domains can be used when the whole system cannot be studied, but the structures of the parts are known. In such a case this method is quite straightforward, requiring only backbone assignment as a prerequisite. Measuring SAXS adds little to the workload, but allows easy characterization of the spatial positions of the domains.

4.2.3 Denatured conformers

Recently it was noticed that denatured proteins in dilute liquid crystals give dipolar couplings [84]. This was interpreted as a sign of long-range order in the denatured ensemble. This conjecture was based on the thought that due to conformational isomerism and dynamic nature of the denatured protein dipolar couplings should average out to zero. There is, however, an alternative explanation. Steric hindrance in main chain torsion angles is present even in a random chain. This gives raise to a persistence length λ_p which describes the distance that is required until the initial direction of the chain is lost i.e. residues behave independently of each other. Persistence length varies depending on the amino acid type, but typically extends over a few residues. It means that at least short peptides are more likely to be elongated, explaining the anisotropic tumbling in dilute liquid crystals, which gives rise to the observed dipolar couplings.

A computer simulation for valence chains was created for testing the hypothesis. The persistent length was simulated by constraining the valence angle between the segments. Also calculations with Gaussian chains where the angles between the segments were unconstrained were performed. Resulting PDB files were fed to the program PALES for the calculation of dipoles. From the output files of PALES the average dipoles were calculated.

Results from the computer simulations agreed with theoretical calculations, and reasonably well with measurements. There remained, however, some differences to the measured dipoles. Variation in the dipolar coupling values in the measurements were larger than those found in the simulations. The difference might at least partly stem from the difference between a homopolymer used in the simulation and the heteropolymer used in the measurements. The degrees of freedom of the chain depend also on amino acid type, so in real case larger variations are to be expected. It may also be that the experimental conditions did not result into true valence chain, as discussed in work [IV].

The work presents a reasonable explanation for the origin of the dipolar couplings in denatured states of proteins. There are possibilities to use these results in the analysis of denatured states. The emergence of secondary structures is expected to stiffen the chain and to change the N-H vector orientations, to promote an easy detection via RDCs. Indeed, in simulation, where random ubiqitin structures were calculated, with a few fixed secondary structures in place, a clear change in residual dipolar couplings was seen [Figure 8]. Conditions in this simulation were unrealistic as strict angle restraints were used. In emerging secondary structures angular restraints would probably be weak. However, it illustrates the directions of the change that should be seen in dipolar couplings.

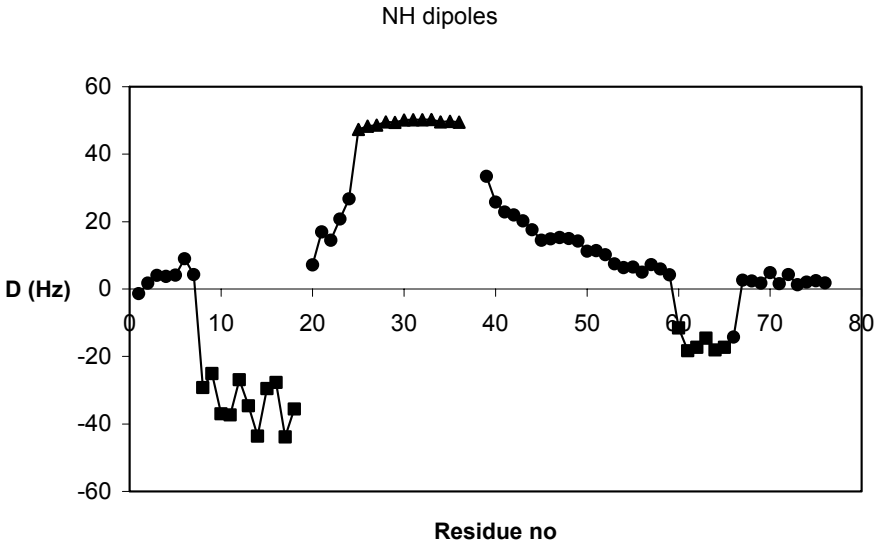

Figure 8. Residual dipolar couplings from a simulation where parts of the sequence where restrained to secondary structures. Circles represent unrestrained residues, squares alpha helical residues and triangles beta sheet residues.

5. Conclusions

Residual dipolar couplings are new, valuable parameters to further complement the rich set of NMR observables. RDCs can be utilized once the backbone assignment is available, which is obtained relatively fast by standard procedures, at least when compared to the complete side chain assignment or structure determination. The applications of RDCs include improved accuracy in structure determination, fast screening of backbone folds, building bigger complexes from smaller parts, studies of conformational changes and characterizing folding intermediates and dynamics.

The number of determined structures has increased rapidly during last years, and structural biology program will further accelerate structure determination. For understanding protein function also its dynamics and interactions with other molecules need to be studied. RDCs can serve as a screening method for detecting structural changes due to ligand binding.

Many of the proteins, especially in eukaryotes, are part of bigger molecular assemblies. It is not always possible to determine the structure of the whole, intact complex. Alternative approach is to solve the structures of modules of the complex, proteins or domains. These individual parts can be assembled together for example by using RDCs and SAXS. This method does have its limitations, at least on NMR part, as for very large complexes the fast relaxation may prevent the measurement.

Denaturated proteins give observable RDCs on the contrary to the first expectations. With the appropriate interpretation RDCs can be used to study properties of denatured proteins, possibly even to detect the emergence of secondary structures. Indeed, measurements of RDCs from denaturated proteins could develop into a new method for the challenging research field of protein folding.

References

1. Wuthrich, K. *NMR of proteins and nucleic acids.* 1986, Wiley: New York

2. Pervushin, K., Riek, R., Wider, G., Wuthrich, K. 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. *Proc. Nat. Acad. Sci. USA* 1997, 94, 14147 - 14151

3. Kawasaki, H., Kretsinger, R. H. Calcium-binding proteins. 1: EF-hands. *Protein Profile* 1994, 1, 343-390 (343 – 517)

4. Satyshur, K. A., Rao, S. T., Pyzalska, D., Drendel, W., Greaser, M., Sundaralingam, M. Refined structure of chicken skeletal muscle troponin C in the two-calcium state at 2-Å resolution. *J. Biol. Chem.* 1988, 263, 1628 - 1647

5. Herzberg, O., James, M. N. G. Refined crystal structure of troponin C from turkey skeletal muscle at 2.0 Å resolution. *J. Biol. Chem.* 1988, 203, 761 - 779

6. Potter, J. D., Gergley, J. The calcium and magnesium binding sites on troponin and their role in the regulation of myofibrillar adenosine triphosphatase. *J. Biol. Chem.* 1975, 250, 4628 - 4633

7. Sia, S. K., Li, M. X., Spyracopoulos, L., Gagne, S. M., Liu, W., Putkey, J. A., Sykes, B. D. Structure of cardiac muscle troponin C unexpectedly reveals a closed regulatory domain. *J. Biol. Chem.* 1997, 272, 18216 - 18221

8. Li, M. X., Spyracopoulos, L., Sykes, B. D. Binding of cardiac troponin I147-163 induces a structural opening in human cardiac troponin C. *Biochemistry* 1999, 38, 8289 - 8298

9. Koradi, R., Billeter, M., Wüthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. J. Mol. Graphics 1996, 14, 51-55

10. Leavis, P. C., Gergely, J. Thin filament proteins and thin filament-linked regulation of vertebrate muscle contraction. *CRC Crit. Rev. Biochem.* 1984, 16, 235-305

11. Krudy, G. A., Kleerekoper, Q., Guo, X., Howarth, J. M., Solaro, R. J., Rosevear, P. R. NMR studies delineating spatial relationships within the cardiac troponin I-troponin C complex. *J. Biol. Chem.* 1994, 269, 23731 - 23735

12. Farah, C. S., Miyamoto, C. A., Ramos, C. H. I., da Silva, A. C. R., Quaggio, R. B., Fujimori, K., Smillie, L. B., Reinach, F. C. Structural and regulatory functions of

the NH2- and COOH-terminal regions of skeletal muscle troponin I. *J. Biol. Chem.* 1994, 269, 5230 - 5240

13. Tripet, B., Van Eyk, J. E., Hodges, R. S. Mapping of a second actin-tropomyosin and a second troponin C binding site within the C terminus of troponin I, and their importance in the Ca2+-dependent regulation of muscle contraction. *J. Mol. Biol.*1997, 271, 728 - 750

14. Pearlstone, J. R., Sykes, B. D., Smillie, L. B. Interactions of structural C and regulatory N domains of troponin C with repeated sequence motifs in troponin I. *Biochemistry* 1997, 36, 7601 - 7606

15. Syska, H., Wilkinson, J. M., Grand, R. J., Perry, S. V. The relationship between biological activity and primary structure of troponin I from white skeletal muscle of the rabbit. *Biochem. J.* 1976, 153, 375 - 387

16. Robertson, S. P., Johnson, D. J., Holroyde, M. J., Kranias, E. G., Potter, J. D., Solaro, R. J. The effect of troponin I phosphorylation on the Ca^{2+}-binding properties of the Ca^{2+}-regulatory site of bovine cardiac troponin. *J. Biol. Chem.* 1982, 257, 260 - 263

17. Zimmerman, D. E., Kulikowski, C. A., Huang, Y., Feng, W., Tashiro, M., Shimotakahara, S., Chien, C., Powers, R., Montelione, G. T. Automated analysis of protein NMR assignment using methods from artificial intelligence. *J. Mol. Biol.* 1997, 269, 592 - 610

18. Leutner, M., Gschwind, R. M., Liermann, J., Schwarz, C., Gemmecker, G., Kessler, H. Automated backbone assignment of labeled proteins using the threshold accepting algorithm. *J. Biomol. NMR* 1998, 11, 31 - 43

19. Atreya, H. S., Sahu, S. C., Chary, K. V., Govil, G. A tracked approach for automated NMR assignments in proteins (TATAPRO). *J. Biomol. NMR* 2000, 17, 125 - 136

20. Guntert, P., Salzmann, M., Braun, D., Wuthrich, K. Sequence-specific NMR assignment of proteins by global fragment mapping with the program MAPPER. J. Biomol. NMR 2000, 129 - 137

21. Moseley, H. N. B., Monleon, D., Montelione, G. T. Automatic determination of protein backbone resonance assignments from triple resonance nuclear magnetic resonance data. *Meth. Enzymol.* 2001, 339, 91 - 108

22. Nilges, M., Macias, M. J., O'Donoghue, S., Oschkinat, H. Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from β-spectrin. *J. Mol. Biol.* 1997, 269, 408 - 422

23. Mumenthaler, C., Guntert, P., Braun, W., Wuthrich, K. Automated combined assignment of NOESY spectra and three-dimensional protein structure determination. *J. Biomol. NMR* 1997, 10, 351 - 362

24. Görler, A., Gronwald, W., Neidig, K-P., Kalbitzer, H. R. Computer assisted assignment of ^{13}C or ^{15}N edited 3D-NOESY-HSQC spectra using back calculated and experimental spectra. *J. Magn. Reson.* 1999, 137, 39 - 45

25. Duggan, B. M., Legge, G. B., Dyson, H. J., Wright, P. E. SANE (Structure Assisted NOE Evaluation): an automated model-based approach for NOE assignment. *J. Biomol. NMR* 2001, 19, 321 - 329

26. Herrmann, T., Guntert, P., Wuthrich, K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* 2002, 319, 209 - 227

27. Gronwald, F., Moussa, S., Elsner, R., Jung, A., Ganslmeier, B., Trenner, J., Kremer, W., Neidig, K-P., Kalbitzer, H. R. Automated assignment of NOESY NMR spectra using a knowledge based method (KNOWNOE). *J. Biomol. NMR* 2002, 23, 271 - 287

28. Herrmann, T., Guntert, P., Wuthrich, K. Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J. Biomol. NMR* 2002, 24, 171 - 189

29. Grishaev, A., Llinas, M. CLOUDS, a protocol for deriving a molecular proton density via NMR. *Proc. Nat. Acad. Sci.* 2002, 99, 6707 - 6712

30. Grishaev, A., Llinas, M. Protein structure elucidation from NMR proton densities. *Proc. Nat. Acad. Sci.* 2002, 99, 6713 - 6718

31. Emsley, J. W., Lindon, J. C. NMR Spectroscopy Using Liquid Crystal Solvents, 1975, Oxford, Pergamon Press

32. Lohman, J. A. B., MacLean, C. Alignment effects on high resolution NMR spectra, induced by the magnetic field. *Chem. Phys.* 1978, 35, 269 - 274

33. Bastiaan, E. W., MacLean, C., Van Zilj, P. C. M., Bothner-By, A. A. High-resolution NMR of liquids and gases: Effects of magnetic-field-induced molecular alignment. *Ann. Rep. NMR Spec.* 1987, 19, 35 - 77

34. Bastiaan, E. W., MacLean, C. Molecular orientation in high-field high-resolution NMR. *NMR-Basic Princ. Prog.* 1990, 25, 17 - 43

35. Tolman, J. R., Flanagan, J. M., Kennedy, M. A., Prestegard, J. H. Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. *Proc. Natl. Acad. Sci. USA* 1995, 92, 9279 - 9283

36. Clore, G. M., Gronenborn, A. M., Bax, A. A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. *J. Magn. Res.* 1998, 133, 216 - 221

37. Bax. A., Kontaxis, G., Tjandra, N. Dipolar couplings in macromolecular structure determination, *Methods Enzymol.*, 2001, 339, 127 - 174

38. Prestegard, J. H., Al-Hashimi, H. M., Tolman, J. R. NMR structures of biomolecules using field oriented media and residual dipolar couplings *Q. Rev. Biophys.* 2000, 33, 371 - 424

39. Zwecksetter, M., Bax, A. Prediction of sterically induced alignment in a dilute liquid crystalline phase: aid to protein structure determination by NMR. *J. Am. Chem. Soc.* 2000, 122, 3791-3792

40. Fernandes, M. X., Bernado, P., Pons, M., Torre, J. G. An analytical solution to the problem of the orientation of rigid particles by planar obstacles. Application to Membrane systems and to the calculation of dipolar couplings in protein NMR spectroscopy. *J. Am. Chem. Soc.* 2001, 123, 12037 - 12047

41. Tjandra, N., Bax, A. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* 1997, 278, 1111 - 1114

42. Nieh, M-P., Glinka, C. J., Krueger, S., Prosser, R. S., Katsaras, J. SANS study on the effect of lanthanide ions and charged lipids on the morphology of phospholipid mixtures. *Biophys. J.* 2002, 82, 2487 - 2498

43. Prestegard, J. H., Kishore, A. I. Partial alignment of biomolecules: an aid to NMR characterization. [Review] *Curr. Opin. Chem. Biol.* 2001, 5, 584 - 590

44. Hansen, M. R., Mueller, L., Pardi, A. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat. Struct. Biol.* 1998, 5, 1065 - 1074

45. Clore, G. M., Starich, M. R., Gronenborn, A. M. Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses. *J. Am. Chem. Soc.* 1998, 120, 10571 - 10572

46. Hansen, M. R., Hanson, P., Pardi, A. Filamentous bacteriophage for aligning RNA, DNA, and proteins for measurement of nuclear magnetic resonance dipolar coupling interactions. *Methods Enzymol.* 2000, 317, 220 - 240

47. Sass, J., Cordier, F., Hoffmann, A., Rogowski, M., Cousin, A., Omichinski, J. G., Lowen, H., Grzesiek, S. Purple membrane induced alignment of biological macromolecules in the magnetic field. *J. Am. Chem. Soc.* 1999, 121, 2047 - 2055

48. Prosser, R. S., Losonczi, J. A., Shiyanovskaya, I. V. Use of a novel aqueous liquid crystalline medium for high-resolution NMR of macromolecules in solution. *J. Am. Chem. Soc.* 1998, 120, 11010 - 11011

49. Barrientos, L. G., Dolan, C., Gronenborn, A. M. Characterization of surfactant liquid crystal phases suitable for molecular alignment and measurement of dipolar couplings. *J. Biomol. NMR* 2000, 16, 329 - 337

50. Tycko, R., Blanco, F. J., Ishii, Y., Alignment of Biopolymers in Strained Gels: A New Way To Create Detectable Dipole-Dipole Couplings in High-Resolution Biomolecular NMR *J. Am. Chem. Soc.* 2000, 122, 9340 - 9341

51. Sass, H. J., Musco, G., Stahl, S. J., Wingfield. P. T., Grzesiek, S. Solution NMR of proteins within polyacrylamide gels: diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes. *J. Biomol. NMR* 2000, 18, 303 - 309

52. Ruckert, M., Otting, G. Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *J. Am. Chem. Soc.* 2000, 122, 7793 - 7797

53. Losonczi J. A., Prestegard, J. H. Improved dilute bicelle solutions for high-resolution NMR of biological macromolecules. *J. Biomol. NMR* 1998, 12, 447 - 451

54. Bax, A., Tjandra, N. High-resolution heteronuclear NMR of human ubiquitin in an aqueous liquid crystalline medium. *J. Biomol. NMR* 1997, 10, 289 - 292

55. Ottiger, M., Bax, A. Characterization of magnetically oriented phospholipid micelles for measurement of dipolar couplings in macromolecules. *J. Biomol. NMR* 1998, 12, 361 - 372

56. Tolman, J. R., Prestegard, J. H. Measurement of amide 15N-1H one-bond couplings in proteins using accordion heteronuclear-shift-correlation experiments. *J. Magn. Reson. Ser. B* 1996, 112, 245 - 252

57. Tjandra, N., Omichinski, J. G., Gronenborn, A. M., Clore, G. M., Bax, A. Use of dipolar H-1-N-15 and H-1-C-13 couplings in the structure determination of

magnetically oriented macromolecules in solution. *Nat. Struct. Biol.* 1997, 4, 9279 - 9283

58. Bayer, P., Varani, L., Varani, G. Refinement of the structure of protein-RNA complexes by residual dipolar coupling analysis *J. Biomol. NMR* 1999, 14, 149 - 155

59. Ramos, A., Grunert, S., Adams, J., Micklem, D. R., Proctor, M. R., Freund, S., Bycroft, M., St Johnston, D., Varani, G. RNA recognition by a Staufen double-stranded RNA-binding domain *EMBO J.* 2000, 19, 997 - 1009

60. Lynch, S. R., Puglisi, J. D. Application of residual dipolar coupling measurements to identify conformational changes in RNA induced by antibiotics *J. Am. Chem. Soc.* 2000, 122, 7853 - 7854

61. Fowler, C. A., Tian, F., Al-Hashimi, H. M., Prestegard, J. H. Rapid determination of protein folds using residual dipolar couplings. *J. Mol. Biol.* 2000, 304, 447 – 460

62. Hus, J. C., Marion, D., Blackledge, M. De novo determination of protein structure by NMR using orientational and long-range order restraints. *J. Mol. Biol.* 2000, 298, 927 - 936

63. Hus, J. C., Marion, D., Blackledge, M. Determination of protein backbone structure using only residual dipolar couplings. *J. Am. Chem. Soc.* 2001, 123, 1541 - 1542

64. Annila, A., Aitio, H., Thulin, E., Drakenberg, T. Recognition of protein folds via dipolar couplings. *J. Biomol. NMR* 1999, 14, 223-230

65. Delaglio, F., Kontaxis, G., Bax, A. protein structure determination using molecular fragment replacement and NMR dipolar couplings. *J. Am. Chem. Soc.* 2000, 122, 2142 - 2143

66. Andrec, M., Du, P. C., Levy, R. M. Protein structural motif recognition via NMR residual dipolar couplings. *J. Am. Chem. Soc.* 2001, 123, 1222 - 1229

67. Clore, G. M. Accurate and rapid docking of protein –protein complexes on the basis of intermolecular nuclear Overhauser enhancement data and dipolar couplings by rigid body minimization. *Proc. Natl. Acad. Sci. USA* 2000, 97, 9021 - 9025

68. Skrynnikov, N. R., Goto, N. K., Yang, D. W., Choy, W. Y., Tolman, J. R., Mueller, G. A., Kay, L. E. Orienting domains in proteins using dipolar couplings measured by liquid-state NMR: differences in solution and crystal forms of maltodextrin binding protein loaded with beta-cyclodextrin. *J. Mol. Biol.* 2000, 295, 1265 - 1273

69. Bewley, C. A., Clore, G. M. Determination of the relative orientation of the two halves of the domain-swapped dimer of cyanovirin-N in solution using dipolar couplings and rigid body minimization. *J. Am. Chem. Soc.* 2000, 122, 6009 - 6016

70. Goto, N. K., Skrynnikov, N. R., Dahlquist, F. W., Kay, L. E. What is the average conformation of bacteriophage T4 lysozyme in solution? A domain orientation study using dipolar couplings measured by solution NMR. *J. Mol. Biol.* 2001, 308, 745 - 764

71. Ramirez, B. E., Bax, A. Modulation of the alignment tensor of macromolecules dissolved in a dilute liquid crystalline medium. *J. Am. Chem. Soc.* 1998, 120, 9106 - 9107

72. Al-Hashimi, H. M., Valafar, H., Terrell, M., Zartler, E. R., Eidsness, M. K., Prestegard, J. H. Variation of molecular alignment as a means of resolving orientational ambiguities in protein structures from dipolar couplings. *J. Magn. Reson.* 2000, 143, 402 - 406

73. Tolman, J. R., Flanagan, J. M., Kennedy, M. A., Prestegard, J. H. NMR evidence for slow collective motions in cyanometmyoglobin. *Nat. Struct. Biol.* 1997, 4, 292 - 297

74. Fischer, M. W. F., Losonczi, J. A., Weaver, J. L., Prestegard, J. H. Domain orientation and dynamics in multidomain proteins from residual dipolar couplings. *Biochemistry* 1999, 38, 9013 - 9022

75. Skrynnikov, N. R., Goto, N. K., Yang, D. W., Choy, W. Y., Tolman, J. R., Mueller, G. A., Kay, L. E. Orienting domains in proteins using dipolar couplings measured by liquid-state NMR: differences in solution and crystal forms of maltodextrin binding protein loaded with β-cyclodextrin'. *J. Mol. Biol.* 2000, 295, 1265 - 1273

76. Meiler, J., Prompers, J. J., Peti, W., Griesinger, C., Bruschweiler, R. Model-free approach to the dynamic interpretation of residual dipolar couplings in globular proteins. *J. Am. Chem. Soc.* 2001, 123, 6098 - 6107

77. Tolman, J. R., Al-Hashimi, H. M., Kay, L. E., Prestegard, J. H. Structural and dynamic analysis of residual dipolar coupling data for proteins. *J. Am. Chem. Soc.* 2001, 123, 1416 - 1424

78. Spyracopoulos, L., Li, M. X., Sia, S. K., Gagne, S. M., Chandra, M., Solaro, R. J., Sykes, B. D. Calcium-induced structural transition in the regulatory domain of human cardiac troponin C. *Biochemistry* 1997, 36, 12138 - 12146

79. Losonczi, J.A., Andrec, M., Fischer, M. W. F., Prestegard, J. H. Order matrix analysis of residual dipolar couplings using singular value decomposition. *J. Magn. Reson.* 1999, 138, 334 - 342

80. Al-Hashimi, H. M., Valafar, H., Terrell, M., Zartler, E. R., Eidsness, M. K., Prestegard, J.H. Variation of molecular alignment as a means of resolving orientational ambiguities in protein structure from dipolar couplings. *J. Magn. Reson.* 2000, 143, 402 - 406

81. Skrynnikov, N. R., Goto, N. K., Yang, D., Choy, W. Y., Tolman, J. R., Mueller, G. A., Kay, L. E. Orienting domains in proteins using dipolar couplings measured by liquid-state NMR: differences in solution and crystal forms of maltodextrin binding protein loaded with beta-cyclodextrin. *J. Mol. Biol.* 2000, 295, 1265 - 1273

82. Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystalogr.* 1992, 25, 495 - 503

83. Vandonselaar, M., Hickie, R. A., Quail, J. W., Delbaere, L. T. Trifluoroperazine-induced conformational change in Ca(2+)-calmodulin. *Nat. Struct. Biol.* 1994, 1, 795 - 801

84. Shortle, D., Ackerman, M. S. Persistence of native-like topology in a denatured protein in 8 M urea. *Science* 2001, 293, 487 – 489