

UNIVERSITY OF HELSINKI

REPORT SERIES IN PHYSICS

HU-P-D141

DYNAMIC NATURE OF PROTEINS
•
INTERPRETATION OF RESIDUAL DIPOLAR COUPLINGS

Martti Louhivuori

Division of X-ray Physics
Department of Physical Sciences
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 Auditorium B123 of Exactum
on June 1st 2007, at 12 o'clock.*

Helsinki 2007

Supervisor:

Professor Arto Annala
Department of Physical Sciences
University of Helsinki
Finland

Reviewers:

Professor Mark S. Johnson
Department of Biochemistry and Pharmacy
Åbo Akademi
Finland

Professor Ilpo Vattulainen
Institute of Physics
Tampere University of Technology
Finland

Opponent:

Doctor Martin Blackledge
Institut de Biologie Structurale Jean-Pierre Ebel
CEA–CNRS–Université Joseph Fourier
France

Report Series in Physics HU-P-D141
ISSN 0356-0961
ISBN 978-952-10-3247-9
ISBN 978-952-10-3248-6 (electronic)
<http://www.ethesis.helsinki.fi/>
Helsinki 2007
Yliopistopaino

M. Louhivuori: Dynamic nature of proteins—interpretation of residual dipolar couplings, University of Helsinki, 2007, 34 pages + appendices. University of Helsinki, Report Series in Physics, HU-P-D141.

ABSTRACT

Protein conformations and dynamics can be studied by nuclear magnetic resonance spectroscopy using dilute liquid crystalline samples. This work clarifies the interpretation of residual dipolar coupling data yielded by the experiments. It was discovered that unfolded proteins without any additional structure beyond that of a mere polypeptide chain exhibit residual dipolar couplings. Also, it was found that molecular dynamics induce fluctuations in the molecular alignment and doing so affect residual dipolar couplings. The finding clarified the origins of low order parameter values observed earlier.

The work required the development of new analytical and computational methods for the prediction of intrinsic residual dipolar coupling profiles for unfolded proteins. The presented characteristic chain model is able to reproduce the general trend of experimental residual dipolar couplings for denatured proteins. The details of experimental residual dipolar coupling profiles are beyond the analytical model, but improvements are proposed to achieve greater accuracy. A computational method for rapid prediction of unfolded protein residual dipolar couplings was also developed.

Protein dynamics were shown to modulate the effective molecular alignment in a dilute liquid crystalline medium. The effects were investigated from experimental and molecular dynamics generated conformational ensembles of folded proteins. It was noted that dynamics induced alignment is significant especially for the interpretation of molecular dynamics in small, globular proteins. A method of correction was presented.

Residual dipolar couplings offer an attractive possibility for the direct observation of protein conformational preferences and dynamics. The presented models and methods of analysis provide significant advances in the interpretation of residual dipolar coupling data from proteins.

Classification (INSPEC): A3325, A3520, A3620, A8715D, C7320

Keywords: residual dipolar coupling, protein dynamics, nuclear magnetic resonance spectroscopy, dilute liquid crystal

ACKNOWLEDGEMENTS

In a way this thesis you are now reading sums up the past four years of my life and yet in so many ways it does not. It certainly is a faithful account, a summarised version if you will, of almost all the work I have been involved in during that time. During these years I have found it almost impossible to convey to others what scholarly life is all about.

The thrill of discovery, the deep sensation of enlightenment when one finally understands, if only for a moment, something new, the satisfaction of seeing your program perform a simple task after a hard day of hacking—a task that could have been done by hand in a fraction of that time—not to mention the idle afternoons spent at the coffee table in deep discussion about things that really matter. They are all so intimately woven to the fabric of our daily routine that it seems nigh on impossible to give a coherent answer to the ever so persistent question: “What is it that you REALLY do?”

I have personally been blessed with a unique opportunity to work in a truly cross-disciplinary environment dividing my time and sharing the comradeship of colleagues in two outstanding laboratories: the X-ray and NMR labs. One could say, and I do fear some even might, that I have been a physicist using computer science to study chemistry in order to understand biology—surely, *philosophiæ naturalis* at its purest!

I want to express my gratitude to Juhani Keinonen and the Department of Physical Sciences for providing excellent facilities and adequate funding so that I have been free to concentrate on my work. I also want to thank Mårten Wikström and the Institute of Biotechnology for providing the facilities of the National Biological NMR Centre at my disposal. Without the collaboration in the NMR lab this work would not exist today. I am grateful to Mark Johnson and Ilpo Vattulainen for the kind comments and constructive suggestion you gave regarding this thesis. Credit is also due to Keijo, Seppo, Arto, Ritva and Pekka. The guidance, encouragement and support you offer us all in the X-ray lab is humbling. Thank you for everything.

I’d like to thank all my colleagues—past and present, near and far—for the collegiate, privileged atmosphere found in academia. I’m time after time gladly surprised by the easy, open-armed fashion one is always welcomed even by complete strangers. I’d like to thank the people in both the X-ray and NMR labs; Kaisa, Ulla, Heikki, Kim, Szabi, Merja, Marko, Tuomas, Liisa, Manu, Aleks (x2), Mikko, Kirsi, Arto, Kari (x2), Mika, Viljami, Teemu, Perttu, Kimmo, Peter, Ville, Maarit, Tapio and all the rest. You are the real reason I have dragged myself to work (almost) every day. Thanks for all the fish!

Especially I would like to acknowledge Kai and Renee. Thank you ever so much for the time we have spent together hitting our heads to metaphysical walls. Without you two I could not have achieved this. I’m also deeply indebted to my supervisor Arto for everything you have done for me. The attention and encouragement you

give every student, be they yours or not, is truly inspiring. Your scientific intuition and originality are the basis for this work. I would also like to note that it was your enthusiasm that finally lured me into the exciting realm of biophysics—something that I will always be grateful for.

But most of all I want to thank my wife Tanja for the support you have given me throughout the years and our children Ilta, Varpu and the lil'un for giving life a purpose and for filling our days with laughter! Ilman teitä en olisi mitään.

Helsinki, April 2007

A handwritten signature in black ink, appearing to read "Matt". The signature is written in a cursive, flowing style with a large initial 'M'.

Ex nihilo nihil fit

— René Descartes in
Principia philosophiæ

CONTENTS

List of original publications	ii
1 Introduction	1
2 Proteins	3
2.1 Ubiquitin	4
2.2 G protein	5
2.3 Troponin	7
3 Protein Motions	9
3.1 Protein dynamics	9
3.2 Folding and misfolding	10
4 Nuclear Magnetic Resonance	13
4.1 Residual dipolar couplings	13
4.2 Experimental methods	15
4.2.1 Sample preparation	15
4.2.2 Spectroscopy	17
5 Theoretical Models	18
5.1 Unfolded proteins	18
5.2 Dynamic molecular alignment	20
5.3 Computational methods	22
5.3.1 Polysim	22
5.3.2 Molecular dynamics simulations	24
6 Conclusions	26
Bibliography	27

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following four peer-reviewed articles and one manuscript submitted for peer-review referred to in the text as **Articles I–V**.

- Article I** **Louhivuori M**, Pääkkönen K, Fredriksson K, Permi P, Lounila J and Annila A (2003) On the origin of residual dipolar couplings from denatured proteins. *J Am Chem Soc* **125**: 15647–15650
- Article II** **Louhivuori M**, Fredriksson K, Pääkkönen K, Permi P and Annila A (2004) Alignment of chain-like molecules. *J Biomol NMR* **29**: 517–524
- Article III** Fredriksson K, **Louhivuori M**, Permi P and Annila A (2004) On the interpretation of residual dipolar couplings as reporters of molecular dynamics. *J Am Chem Soc* **126**: 12646–12650
- Article IV** **Louhivuori M**, Otten R, Lindorff-Larsen K and Annila A (2006) Conformational fluctuations affect protein alignment in dilute liquid crystal media. *J Am Chem Soc* **128**: 4371–4376
- Article V** **Louhivuori M**, Otten R, Salminen T and Annila A (2007) Evidence of molecular alignment fluctuations in aqueous dilute liquid crystalline media. *J Biomol NMR* (submitted)

Articles I, III & IV are re-printed by permission from the *Journal of American Chemical Society* (<http://pubs.acs.org/journals/jacsat/>) and **Article II** is re-printed by permission from the *Journal of Biomolecular NMR* (<http://www.springerlink.com/content/1573-5001/>).

AIM OF THE STUDY

The aim of the study was to address two perplexing observations by nuclear magnetic resonance spectroscopy concerning conformations of unfolded proteins and dynamics of folded proteins. First task was to clarify why unfolded proteins appear to have structural features when examined by residual dipolar couplings. Second objective was to understand why folded proteins appear to be highly mobile when studied with residual dipolar couplings. Both of these protein dynamics related questions have clear biological implications from protein function to amyloid formation.

Author's Contribution

The analytical models proposed in [Articles I–V](#) for the calculation of structural and dynamical nuclear magnetic resonance parameters from unfolded and folded proteins were developed by the author together with Arto Annala. The all-atom models presented in [Articles II–V](#) were devised by the author. *Polysim*, the software-package used in [Articles I–III](#), was written by the author based on a prototype algorithm by Kimmo Pääkkönen. The author wrote the software needed for data-analyses in the Articles. The author has been deeply involved in writing all of the Articles and has contributed to the preparation of the figures and tables in all of the Articles.

ABBREVIATIONS

DMA	dynamics modulated alignment
DNA	deoxyribonucleic acid
EXSY	exchange spectroscopy
GAF	Gaussian axial fluctuations
GDP	guanosine diphosphate
GTP	guanosine triphosphate
HSQC	hetero-nuclear single-quantum coherence
MD	molecular dynamics
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser enhancement
PDB	Protein Data Bank
RDC	residual dipolar coupling
RNA	ribonucleic acid
SVD	singular value decomposition
TnC	troponin C

INTRODUCTION

Proteins, those microscopic agents of life, are truly fascinating. Even though DNA and RNA have attracted the limelight and aroused the interest of most contemporaries, it is proteins that in our daily existence make or break the day. Proteins are large and complex biological macromolecules found in all living organisms. They usually adopt a protein-specific, semi-rigid spatial configuration known as a fold. The biological activity of a protein depends on its three-dimensional structure and dynamics. The interplay between activity and protein structures and dynamics can be studied using nuclear magnetic resonance (NMR) spectroscopy. It is a powerful method to acquire structural information at atomic resolution and motional information on many time scales. Functional assays are also easily performed during NMR measurements.

In this study I address two questions in protein NMR spectroscopy that have been puzzling the research community for the past few years. First, I will clarify why unfolded proteins appear to have structural features when examined by residual dipolar couplings (RDCs). Second, I will show why folded proteins appear to be highly mobile when studied using RDCs.

In 2001 [Shortle & Ackerman](#) published interesting data from denatured proteins that was said to infer the presence of native-like topology in denatured proteins. The interpretation was based on a common misconception to think that unfolded proteins—or to be more specific, fully unstructured proteins—would not exhibit residual dipolar couplings at all. As will be discussed in [section 5.1](#) and shown in [Article I](#), this is not true. Nonetheless, the misconception gave briefly momentum to claims of native-like residual structure in unfolded proteins. This misinterpretation of RDC data to support a potentially momentous idea about protein folding underlined the need for the first principles study ([Article I](#)) that resulted in intrinsic RDC profiles of unstructured proteins. In [section 5.3](#) and [Articles I & II](#) an analytical and a computational protocol are presented to predict the intrinsic RDC profiles.

In 2001 [Tolman *et al.*](#) and [Meiler *et al.*](#) published intriguing results about RDC-probed protein dynamics. They proposed that low RDC-based order parameter values report from substantial protein internal motions in the sub-microsecond time scale. As will be shown in [section 5.2](#) and [Article IV](#) there is an alternative explanation. All proteins undergo continually conformational changes. These conformational fluctuations affect the RDC signal measured from a protein solution in a dilute liquid crystalline medium. The low order parameters may arise from the changes in a molecule's alignment due to the conformational fluctuations. This dynamics modulated alignment (DMA) can be significant for small globular proteins in particular with an alignment dominated by electrostatic interactions. However, DMA has only a small effect on protein structure determination, as will be shown in [Article V](#).

The thesis is organised as follows. I begin by giving a brief overview of the particular proteins that were used in this study. Then I will describe the dynamic nature of proteins and its implications to protein NMR spectroscopy and especially to RDC studies of unfolded and conformationally labile proteins. I will proceed with a brief glance at NMR spectroscopy and a longer one at residual dipolar couplings. After this the prediction of RDCs for unfolded proteins and the effects of DMA will be discussed. The underlying principles of the proposed theoretical and computational models will also be described. I will conclude with remarks on the presented results. The rigorous and detailed work is presented in the attached four peer-reviewed articles and one manuscript submitted for peer-review.

PROTEINS

Proteins play a crucial role in every living organism from *E. coli*¹ to humans. They are involved in all biological functions making an organism alive just like cog wheels make a clockwork tick. There exists hordes of unique proteins with multitudes of roles and functions e.g. from cell signaling to molecular recycling. Since proteins have been extensively studied, we have nowadays an excellent understanding of protein structure and a fair picture of their dynamics. A lot of work has been done and will be done to relate biological activity to molecular structure and dynamics. (Creighton, 1984 | Voet & Voet, 2004 | Whitford, 2005)

Structure Proteins are single-branch polymer chains built from amino acids linked together. In nature there are 20 different proteinogenic² amino acids. Each amino acid contains a nitrogen, an alpha carbon and a carbonyl that make up the main chain, or backbone, of a polymer. The nitrogen and hydroxyl carbon of adjacent amino acids form a semirigid O=C-N-H peptide plane. A unique side-chain starts from the alpha carbon. The side-chain can be as small as a single hydrogen (glycine) or as big as eleven heavy atoms in a double aromatic ring weighing ~150 Da (tryptophan).

Many natural polymer chains fold into distinct three-dimensional structures. Strictly speaking, it is these functional structures that are defined as proteins. The fold of a protein is the structural framework that supports its functional activity. The structure of a protein is usually described using a four-tier hierarchy. Starting from the amino acid sequence—or the *primary* structure—and proceeding to secondary, tertiary and finally quaternary structure of the protein.

The fold of a protein contains common structural elements referred to as the *secondary* structure of the polypeptide. Typical elements are the helical backbone (alpha helix) and multiple backbone strands running in parallel or anti-parallel (beta sheet). The *tertiary* structure, i.e. the overall three-dimensional structure, of a polypeptide contains several secondary structures. It may also include several separate structural units, called domains, linked together loosely or tightly, that are nevertheless part of the same polymer chain. Two or more proteins may also form larger complexes by binding with each other. A protein complex may be built either from only one type of proteins, i.e. a homomer, or from different proteins, i.e. a heteromer. The structure of a protein complex is referred to as the *quaternary* structure of the constituent proteins.

¹*E. coli* (lat. abbr.) *Escherichia coli* are a common intestinal bacteria that are usually harmless to humans.

²*proteinogenic* (lat.) Literally *protein building*, i.e. an amino acid found in proteins and encoded for in the DNA.

In this work the following proteins have been used to develop and to verify the models proposed in [Articles I–V](#) that will be discussed in more detail later on in [chapter 5](#). These particular proteins were chosen mainly because there are abundant experimental and computational data available for them. Furthermore, troponin C was chosen as an example of large-scale conformational changes.

2.1 Ubiquitin

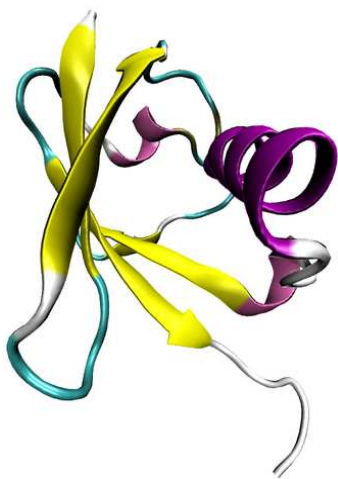


Figure 1. Ubiquitin.
(PDB: 1UBQ)

Ubiquitin is a small ~ 8.5 kD protein found universally in eukaryotes ([Goldstein *et al.*, 1975](#)). Its amino acid sequence is highly conserved in evolution. Only a few variations are known besides the common form found also in human cells ([Vijay-Kumar *et al.*, 1987](#)). Ubiquitin consists of 76 amino acids in a single polypeptide chain with three α -helices, four β -strands forming a single β -sheet and a flexible coil-like C-terminus ([Figure 1](#)).

Function Ubiquitin is primarily associated with protein degradation in numerous processes from endocytosis³ to apoptosis⁴ ([Hershko & Ciechanover, 1998](#) | [Roos-Mattjus & Sistonen, 2004](#)). A misfolded protein is initially tagged with ubiquitin and then reduced to its constituent amino acids by a proteasome while ubiquitin is released for new use ([Figure 2](#)) ([Ardley & Robinson, 2004](#)). A malfunction in or an overloading of the ubiquitin-proteasome system may cause pathological conditions such as cancers ([Nakayama & Nakayama, 2006](#)) or neuro-degenerative disorders ([Ardley & Robinson, 2004](#)).

Data available A wealth of experimental data, both by NMR spectroscopy and X-ray crystallography, is available on ubiquitin. Ubiquitin has been widely used as a model protein in NMR method development since it gives excellent spectra. Structural models with resolutions as high as 1.3 Å are also available from the Protein Data Bank (PDB).

³*endocytosis* (lat.) Absorption of external material by a cell.

⁴*apoptosis* (lat.) Programmed cell death.

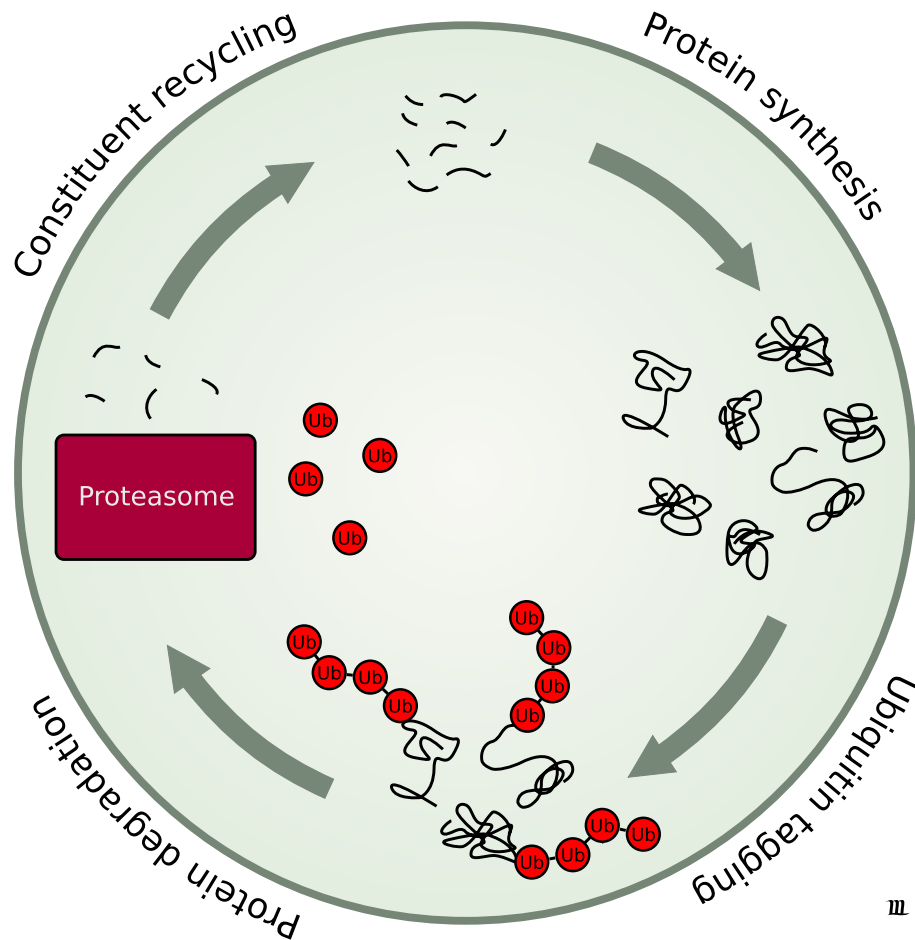


Figure 2. Ubiquitin facilitated protein degradation. Misfolded proteins are tagged with ubiquitin, disassembled by a proteasome and re-used in protein synthesis.

2.2 G protein

Heterotrimeric guanosine triphosphate (GTP) binding protein, or *G protein*⁵, consists of three subunits named alpha, beta and gamma. The alpha subunit (G_α) has two domains, a GTPase domain and an α -helical domain, and is loosely bound to the two other subunits. The subunits beta and gamma on the other hand are tightly bound together and are often referred to as the beta-gamma complex ($G_{\beta\gamma}$).

There exists more than 16 mammalian variants of the alpha subunit, 5 variants of the beta subunit (G_β) and 14 variants of the gamma (G_γ) subunit. Many, but not all, possible combinations of these are found in nature. In our studies two variants of the beta subunit were used. The G_β has 56 amino acids in a single polypeptide chain with four β -strands, that form a single β -sheet, and a solitary α -helix (Figure 3).

⁵Confusingly, *G proteins* is also the name of the super-family of regulatory GTP hydrolases that the heterotrimeric G protein belongs to.

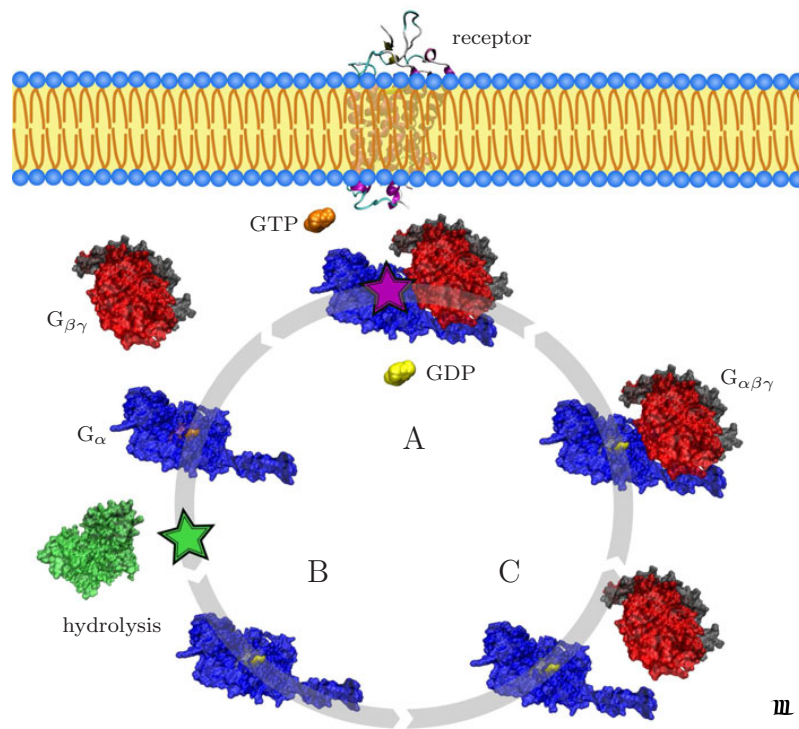


Figure 4. G protein cycle. (A) A GDP to GTP exchange in a receptor-bound G protein dissociates alpha (blue) and beta–gamma (red/black) subunits. (B) GTP is hydrolysed back to GDP. (C) Alpha subunit re-associates with the beta–gamma complex to form once again an intact G protein.

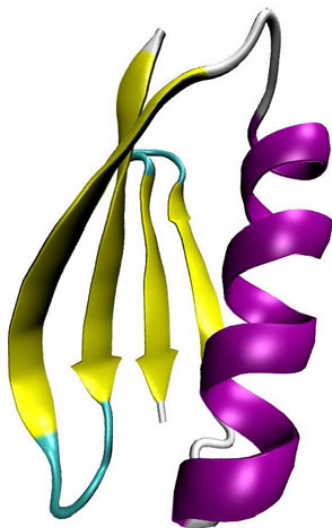


Figure 3. G protein subunit β_3 . (PDB: 1P7E)

Function Membrane-associated G proteins function as molecular switches in many intracellular signaling cascades (Feldman & Hegele, 2000). Extracellular factors cause conformational changes in G protein coupled receptors that enable the binding of G proteins to the receptors. A receptor-bound G protein exchanges a guanosine diphosphate (GDP) to a guanosine triphosphate after which it is released from the receptor following the dissociation of the alpha subunit and the beta–gamma complex (Figure 4) (Offermanns, 2003). The G_{α} -GTP and $G_{\beta\gamma}$ complexes may then independently activate downstream effectors initiating signaling cascades (Daaka, 2004 | Offermanns, 2003). Eventually they will re-associate after the alpha subunit has hydrolysed the GTP back to a GDP. G proteins mediate many diverse functions such as vision, smell and taste and G protein abnormalities are involved in many diseases and disorders ranging from cholera and diabetes to cardiomyopathy and night blindness (Malbon, 2005 | Offermanns, 2003 | Raju, 2000).

Data available G proteins have been extensively studied using various methods. High resolution (1.5 Å and better) structures are available for many of the different subunit variants.

2.3 Troponin

Troponin is a calcium-sensitive protein complex involved in muscle contraction. It is found in slightly different variants in both skeletal and cardiac muscles. Troponin consists of three subunits labeled C, I and T. Troponin C (TnC) binds calcium ions, troponin I (TnI) inhibits actomyosin ATPase and troponin T (TnT) anchors the complex to tropomyosin—another component of the thin filaments in myofibrils.

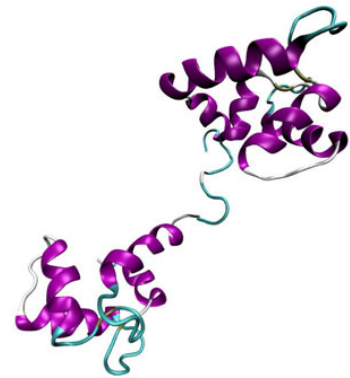


Figure 5. Calcium saturated cardiac troponin C. (PDB: 1LA0)

TnC Troponin C is a 18 kDa protein with globular N- and C-terminal domains connected by a helical linker (Figure 5). It has four calcium-binding sites with two at each of the terminal domains. C-terminal binding sites III and IV are occupied by either a Ca^{2+} or a Mg^{2+} under normal physiological conditions (Robertson *et al.*, 1981). In cardiac troponin C (cTnC) binding site I is inactive leaving only one active N-terminal calcium-binding site compared to the two present in skeletal troponin C (sTnC). The N-terminus of TnC undergoes a distinct conformational state change on calcium binding from a closed state to an open state, even though in cTnC this transition is complete only after TnI binding (Sorsa *et al.*, 2004).

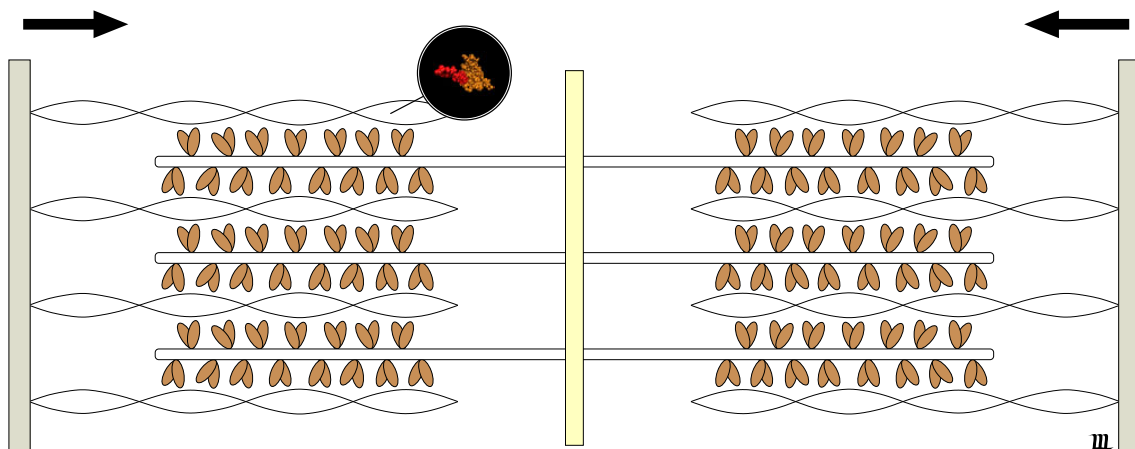


Figure 6. In muscle contraction thin and thick filaments of myofibrils slide past each other shortening the muscle cell. Troponin binds to the thin filament (depicted as helical strands) in regular intervals.

Function Troponin has a crucial role in the calcium-ion regulated muscle contraction system (Gagné *et al.*, 1995). The binding of Ca^{2+} to TnC exposes a large hydrophobic region facilitating TnI binding to TnC. This disrupts ATPase inhibition (Weeks & Perry, 1978) and is also the impetus for a cascade of events that move tropomyosin along the thin filament exposing myosin binding sites (Lehman *et al.*, 2001 | Xu *et al.*, 1999). This in turn leads to muscle contraction (Figure 6).

Data available High resolution X-ray and NMR structures of TnC are readily available in both calcium-saturated and calcium-free states for TnC–TnI complexes.

PROTEIN MOTIONS

Proteins are dynamic molecules. Each and every one of the thousands of atoms in a protein are constantly buffeted by forces exerted by other atoms in the protein as well as in the medium and jostled into transiently better positions in a never-ending “atomic dance”.

3.1 Protein dynamics

Even though most native proteins are in a stable folded state, all proteins exhibit internal motions on various time scales and amplitudes. Local fluctuations at picosecond–nanosecond time scale and global fluctuations at microsecond–millisecond time scale (Figure 7). This coarse classification is of course only indicative as there are many superimposed motional modes in proteins.

Local fluctuations Fast internal motions include bond vibrations and librations as well as peptide plane motions. The infinitesimal local atomic vibrations take place as fast as picoseconds or faster whereas the larger reorientations associated with peptide plane movements take place in nanoseconds (Bouvignies *et al.*, 2005 | Kay *et al.*, 1989 | Lakomek *et al.*, 2006). All these local, fast motions can be assessed in laboratory frame NMR measurements through relaxation time (T1, T2) (Kay *et al.*, 1989 | Nirmala & Wagner, 1988 | Palmer, 2001 | Skelton *et al.*, 1993) and hetero-nuclear nuclear Overhauser enhancement (NOE) (Campbell & Sykes, 1993 | Clore *et al.*, 1990 | Kay *et al.*, 1989) measurements. Anything beyond the rotational correlation time τ_c of a protein (Kay *et al.*, 1989 | Lipari & Szabo, 1982), on the

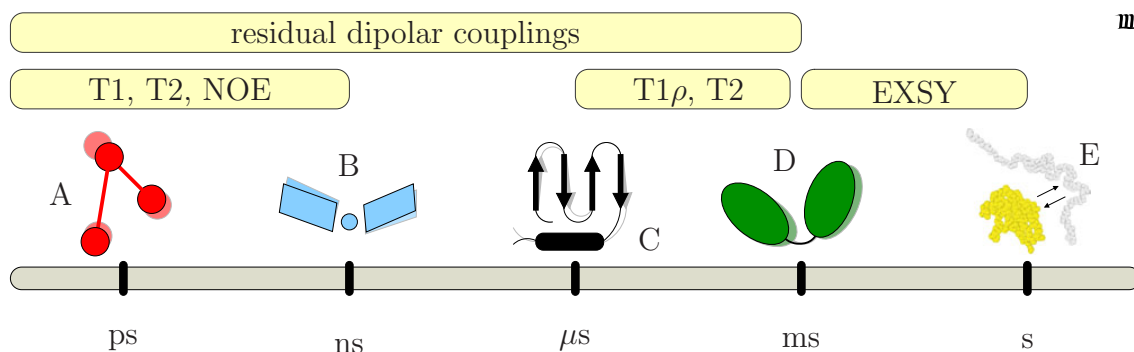


Figure 7. Protein motions range from picoseconds to seconds: (A) bond vibrations and librations, (B) peptide-plane motions, (C) secondary structure fluctuations, (D) domain motions, and (E) protein folding. RDCs bridge the methodological gap from τ_c to μs .

order of 10 ns (Tolman *et al.*, 1997), is too slow to be observed by these techniques (Lakomek *et al.*, 2006). Local fluctuations are reflected in residual dipolar coupling data.

Global fluctuations Slow internal motions range from secondary structure motions in the microseconds to protein domain motions in the milliseconds. For example hydrogen bond formation and enzymatic reactions occur at these time scales (James, 2003 | Massi *et al.*, 2005). In NMR spectroscopy these motions can be obtained by rotating frame measurements in particular from relaxation dispersion studies (Akke & Palmer, 1996 | Kay *et al.*, 1989 | Kopple *et al.*, 1986 | Palmer, 2004 | Palmer *et al.*, 2001 | Tjandra *et al.*, 1995).

Supra- τ_c backbone motions The amount of motion in folded proteins between rotational correlation time and the microsecond–millisecond time range has been a subject of debate for some years now (cf. Bax & Tjandra, 1997a | Clore & Schwieters, 2004 | Meiler *et al.*, 2001, 2003 | Peti *et al.*, 2002 | Tolman *et al.*, 1997). The discussion has revolved around two main issues. First, the reported RDC-based order parameter values (Meiler *et al.*, 2001 | Tolman *et al.*, 2001) are perplexingly low. Second, recent evidence (Lakomek *et al.*, 2005) of increased mobility in solvent exposed residues relative to core residues has implied *novel* correlated supra- τ_c backbone motions.

The dynamic molecular alignment, discussed later on in [chapter 5](#) and in [Articles IV & V](#), suggests an explanation for the low order parameter values with no need for additional dynamical modes. This interpretation does not question the presence of correlated motions. As Lakomek *et al.* (2005) have pointed out:

... as NH vectors are parallel in β -sheets, [alignment tensor] fluctuations do not affect our conclusions.

Alignment fluctuations are anisotropic and hence residues are affected unevenly. There is no saying *a priori* how DMA affects the reported results. Nonetheless, as we show in [Articles IV & V](#) these effects appear small in the backbone of a folded protein, which supports the validity of the correlated supra- τ_c motions.

3.2 Folding and misfolding

The fold of a protein is essential for its ability to interact with other molecules, be they receptors, regulators, epitopes or ligands. The idea of complementary epitopes, i.e. that the binding of two molecules necessitates complementary structures, has been known (Pauling, 1940) for decades. Recently it has been understood (Dyson & Wright, 2005 | Wright & Dyson, 2002) that not all proteins, and certainly not every part of a protein, belong to a stable fold. There is a great diversity in the stability of protein structures presumably contributing to diverse biological activities. Thus research interests have been shifting from structure determination towards folding, transient structures and internal-dynamics facilitated binding processes.

Folding Proteins start to fold as soon as they are formed in ribosomes, even before they are completely synthesised (Dobson, 2003 | Gay *et al.*, 1995). The whole process is over in less than a second (Zwanzig *et al.*, 1992). Considering the large number of degrees of freedom in proteins, folding happens astonishingly fast. The quickness of this process is often associated with the so called Levinthal’s paradox. It states that not even the age of the universe would be long enough for a protein to explore all its possible conformations and find the correct fold unbiased.

Even though it was shown 15 years ago by Zwanzig *et al.* (1992) that it is indeed only a virtual paradox if some elementary physical restrictions are applied to the system—not to mention that Levinthal himself concluded already in his original speech (Levinthal, 1969) that a random search is not an efficient method of finding a correct fold and therefore is not how it happens in nature—Levinthal’s paradox is frequently picked up as a genuine problem. Nonetheless, even without unheeded exaggerations the characterisation of a folding process is a genuine problem (Dobson, 2003 | Finkelstein & Galzitskaya, 2004).

Misfolding Protein misfolding has potentially devastating effects. Such grave pathological disorders as Creutzfeldt–Jakob disease, scrapie, bovine spongiform encephalopathy⁶ and Alzheimer’s disease are caused by protein misfolding (Prusiner, 1997, 1998 | Thomas *et al.*, 1995). It shows how crucial it is for proteins to adopt their correct form.

A misfolded protein is usually tagged with ubiquitin to be disassembled in the ubiquitin-proteasome. Thus our cells do not usually accumulate misfolded proteins in quantities to develop any adverse effects. But in some cases the ubiquitin-proteasome may e.g. become blocked disrupting the “waste reprocessing” system.

Folding kinetics The principle stages of protein folding are understood (Dobson, 2003 | Finkelstein & Galzitskaya, 2004). The formation of local structure prior to global structure. / Hydrophobicity as a driving force in global structure formation. / The emergence of stable folding cores as a vital intermediate that stabilises the overall structure. / The importance of hydrogen bonds as the “glue” that holds the pieces together. These are all well known factors in protein folding and yet the details of any particular folding pathway are still not easily derived from these general principles (Finkelstein & Galzitskaya, 2004) but rather requires laborious case-by-case studies.

The folding of a protein is often pictured as a stochastic process of energy-minimisation whereby the molecule adopts a stable configuration through internal interactions as well as interactions with the surrounding medium. The energy landscape associated with this process is usually depicted as a funnel that guides the configurations to converge at a focal point marking the folded structure (Figure 8). Even though it is often implicitly assumed that this energy minimum is a global one, in reality there is no need for it to be one. The only biological necessity is that the folded structure is stable enough under biological conditions for the protein

⁶*Bovine spongiform encephalopathy* (BSE) is most commonly known by its pseudonym *mad cow disease*.

to perform its function for a reasonable length of time, or in other words, that the energy minimum, be it local or global, is deep and steep enough to fulfill this goal.



Figure 8. A folding funnel?

Unfolded proteins It is easy, and common enough, to view unfolded proteins as being devoid of any structural features, flexible “spaghetties” taking whichever form fortune fancies regardless of the inherent properties of the amino acid sequence. Nonetheless, it is evident that the atoms of any molecule prefer certain configurations over others and that the existence of different kinds of side-chains in different sequences will create a conformational bias that may or may not be insignificant ([Article II](#)).

What is not intuitive is the form of this bias. Is the bias towards native-like configurations or some other residue-specific natural configurations? Surprisingly, recent evidence ([Lindorff-Larsen *et al.*, 2004](#) | [Ohnishi *et al.*, 2004](#) | [Shortle & Ackerman, 2001](#) | [Zagrovic *et al.*, 2002](#)) points out that not only is the bias towards native-like configurations but that it is also readily observable in the RDCs of denatured proteins. Even though it is shown in [section 5.1](#) and [Article I](#) that the mere presence of RDC signals does not imply residual structure in denatured proteins, it should be noted that not all claims of native-like residual structure are backed by RDC data interpretations. The evidence suggests that a folded protein is not just some bag full of compressed strings ready to burst but more a favourable ordering of relaxed residues—the natural form of the protein in the environment in question, in fact.

Transition states Metastable transition states seem to have a key role in protein folding ([Duan & Kollman, 1998](#) | [Finkelstein & Galzitskaya, 2004](#) | [Lindorff-Larsen *et al.*, 2004](#) | [Pandit *et al.*, 2006](#)). Folding catalysts work e.g. by stabilising the transition states ([Fischer & Schmid, 1990](#)). By doing so they direct the folding process in the right direction. Transition states can be viewed as logistical-hubs collecting molecules into the vicinity of the desired target location, i.e. native fold. If there is a fast, clear path from the transition state to the final state, the transition state acts as a gateway drawing the molecule from a larger subset of states towards the correct fold. Studying transition states can be difficult as they are transient, but recent computer studies ([Duan & Kollman, 1998](#) | [Rhee *et al.*, 2004](#)) give hope that precise knowledge could be gained through computational studies as well. RDC measurements may reveal early conformational preferences that initiate protein folding.

NUCLEAR MAGNETIC RESONANCE

Modern multi-dimensional protein NMR spectroscopy is a powerful spectroscopic technique (Cavanagh *et al.*, 1996). Every magnetic nucleus precesses in a magnetic field with a certain, nucleus-specific frequency. In nuclear magnetic resonance spectroscopy the magnetic field experienced by a nucleus is composed of the stationary spectrometer-field and local time-dependent fields. The local fields contain information about protein structure and dynamics practically at atomic resolution.

Spectroscopy Multi-dimensional hetero-nuclear spectra give detailed information of the target molecule enabling the assignment, i.e. identification of signal sources, and structure determination of proteins as large as several kD. Much of the information is contained in interactions, i.e. couplings, between the nuclei.

Magnetic nuclei may be coupled indirectly by through-bond scalar couplings and directly by through-space dipolar couplings. Both of these couplings may transfer magnetisation, in general coherence, from one nucleus to another. The transfer can be controlled by sequences of radio-frequency pulses⁷ and delays. For the scope of this work it is important to understand the basis of residual dipolar couplings.

4.1 Residual dipolar couplings

Dipolar coupling is a direct magnetic interaction between two nuclear spins. It depends inversely on the 3rd power of distance and on the orientation of the inter-nuclear vector relative to the magnetic field. Owing to the orientational dependence, dipolar couplings average to zero when a molecule tumbles isotropically in a solution. In a dilute liquid crystal the rotational diffusion of a solute is slightly anisotropic due to steric and electrostatic interactions with the liquid crystal particles. Thus a *residual* of the dipolar coupling remains for observation.

Residual dipolar couplings are highly versatile parameters. They contain information of the relative orientation and the distance between two nuclei as well as the slight alignment of the protein. RDCs have been used e.g. in structure determination and refinement (Bouvignies *et al.*, 2006b), for protein model assembly from peptide fragments (Kontaxis *et al.*, 2005) and domains (Mattinen *et al.*, 2002), for protein fold recognition (Annala *et al.*, 1999) and to identify conformational exchange (Pääkkönen *et al.*, 2000). RDCs have also been the basis for protein dynamics studies (Meiler *et al.*, 2001 | Tolman *et al.*, 2001), which is the topic of [Article II](#).

⁷Radio-frequency pulses are short bursts, typically several microseconds in length, of electromagnetic waves oscillating at frequencies on the order of GHz.

Alignment media In a solution protein-molecules are constantly in Brownian motion. They reorient and diffuse all the time. The free, isotropic rotation of the molecules causes dipolar couplings to average to zero. By introducing some alignment with an anisotropic medium, such as bicelles, bacteriophages or polyacrylamid gels, directional information is recovered (Bax & Tjandra, 1997b | Tjandra & Bax, 1997).

Dipolar couplings are on the order of kHz, but residual dipolar couplings introduced by these dilute alignment media are preferably on the order of Hz to allow high-resolution NMR spectroscopy. They are scalable in the limits of the particular alignment medium. The residual dipolar coupling between two nuclei A and B depends on the orientation of the inter-nuclear vector between the two nuclei. It is usually written in the principal axis system of an aligned molecule as

$$D^{AB}(\theta, \phi) = D_{\max}^{AB} \left[\frac{1}{2} (3 \cos^2 \theta - 1) + \frac{3}{4} R \sin^2 \theta \cos 2\phi \right], \quad (1)$$

where rhombicity R is a measure of asymmetry in the molecular shape and D_{\max}^{AB} is the normalised maximum residual dipolar coupling between nuclei A and B. The angles θ and ϕ define the orientation of the connecting vector in respect to the alignment frame.

Alignment frame The effective alignment of a molecule is described by a 3×3 tensor called the alignment tensor. Rhombicity is defined as the ratio between the non-axial and axial components of this tensor, or $R = \frac{2}{3} (A_{xx} - A_{yy}) / A_{zz}$, where A_{ii} are the diagonal elements of the symmetric, traceless alignment tensor \mathbf{A} . The elements of \mathbf{A} are defined in the laboratory frame as

$$A_{ij} = \frac{1}{2} \langle 3 \cos \theta_i \cos \theta_j - \delta_{ij} \rangle \quad \forall i, j \text{ in } x, y, z, \quad (2)$$

where θ_i denotes the angle between the i^{th} molecular axis and the external magnetic field. The orientation of an internally rigid molecule in a liquid sample could be defined by a single alignment tensor whereas to define the orientation of a structurally fluctuating protein may require the use of multiple alignment tensors. The use of only one alignment tensor appears to be a good approximation for most folded proteins when determining their structure. Only when one is interested in some highly mobile or structurally labile regions of a protein—say certain side-chains—it is of paramount importance to use more than one alignment tensor to map out the conformational state space (Article IV). Then the only common frame of reference is the laboratory frame.

The alignment tensor of a molecule with a known structure may be determined from measured RDCs through singular value decomposition (SVD) (Losonczi *et al.*, 1999). Non-redundant alignment tensor elements $\mathbf{a} = [A_{yy}, A_{zz}, A_{xy}, A_{xz}, A_{yz}]$ can be solved from the linear equation $\mathbf{C}\mathbf{a} = \mathbf{d}$, defined as

$$\begin{pmatrix} c_{yy}^1 - c_{xx}^1 & c_{zz}^1 - c_{xx}^1 & 2c_{xy}^1 & 2c_{xz}^1 & 2c_{yz}^1 \\ c_{yy}^2 - c_{xx}^2 & c_{zz}^2 - c_{xx}^2 & 2c_{xy}^2 & 2c_{xz}^2 & 2c_{yz}^2 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ c_{yy}^n - c_{xx}^n & c_{zz}^n - c_{xx}^n & 2c_{xy}^n & 2c_{xz}^n & 2c_{yz}^n \end{pmatrix} \begin{pmatrix} A_{yy} \\ A_{zz} \\ A_{xy} \\ A_{xz} \\ A_{yz} \end{pmatrix} = \begin{pmatrix} D_1 \\ D_2 \\ \cdot \\ \cdot \\ D_n \end{pmatrix} \quad (3)$$

with $c_{ij}^k = \cos \theta_i^k \cos \theta_j^k$, by using the decomposition $\mathbf{C}^{-1} = \mathbf{V}\mathbf{W}\mathbf{U}^T$, where \mathbf{V} is an $N \times N$ orthogonal matrix, \mathbf{U} is an $M \times N$ column-orthogonal matrix and \mathbf{W} is an $N \times N$ diagonal matrix with non-negative elements (Losonczi *et al.*, 1999). SVD is both a fast and robust method making it an efficient tool for RDC data analysis.

It is a common practice to represent the quality of a fit between a structural model and experimental RDCs as a normalised, squared distance called the quality factor. The quality factor Q is defined as

$$Q = \sum_i (D_i - D'_i)^2 / \sum_i D_i^2, \quad (4)$$

where D_i is the experimental and D'_i is the back-calculated residual dipolar coupling of residue i . Two sets of dipolar couplings are also often compared using the correlation coefficient as a measure of fit. Pearson correlation coefficient R between two vectors x and y is defined as

$$R = \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2] [n \sum y^2 - (\sum y)^2]}}, \quad (5)$$

where n is the length of the vectors. The quality factors of high-resolution NMR structural models are usually about 0.2.

4.2 Experimental methods

Experiments were carried out to validate the theoretical models for RDC interpretation. Primarily two synthetic model polymers were studied using hetero-nuclear correlation NMR spectroscopy. Residual dipolar couplings of both E₂₁- and E_{21X}-polymers were measured using two-dimensional ¹³C-¹H correlation spectroscopy in Varian 600 MHz and 800 MHz NMR spectrometers.

4.2.1 Sample preparation

A E₂₁-polymer-bicelle solution with E₂₁ concentration of ~0.5 mM and bicelle concentrations from ~6.3% (w/v) to ~38.3% (w/v) were prepared as described below. Similar E_{21X}-polymer-bicelle samples were also prepared.

Synthetic E₂₁-polymers In order to understand the RDCs of an unstructured polymer it is useful to have a molecule devoid of any structural characteristics and yet detectable. A monotonous polymer containing only a single type of amino acid is the closest one gets to complete lack of structural features, especially if one does not choose amino acids like proline that include inherent structural features. Alas, it is virtually impossible to distinguish between the residues of a homopolymer longer than a couple of residues. The chemical environments of all non-terminal residues are nearly identical making it difficult to discern between their signals. To overcome this problem, we decided to use 21 amino acid polyglutamates labeled with magnetic isotopes only at E11 and E21.

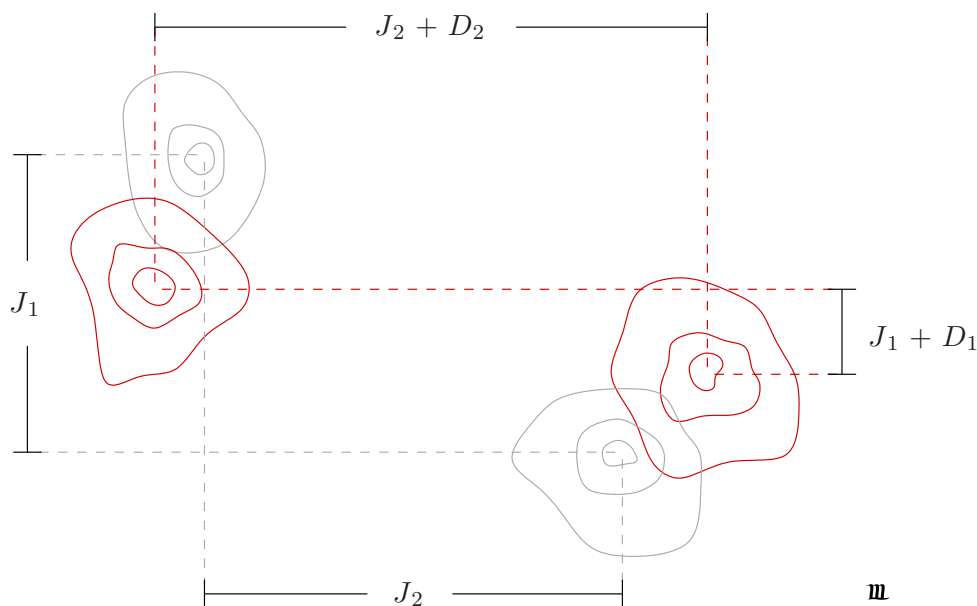


Figure 9. A residual dipolar coupling (D -coupling) is measured from a doublet splitting ($J+D$ -coupling) by subtracting the effect of the scalar coupling (J -coupling) measured independently. With spin-state selective pulse sequence one can measure from a two-dimensional correlation spectra, like the one depicted, two types of residual dipolar couplings (1 and 2) simultaneously.

The required amino acids were bought from *Cambridge Isotope Laboratories Inc.* and the E_{21} -polymers were synthesised by *Larodan Fine Chemicals Ab.* The alpha carbon and nitrogen of E11 and E21 were labeled with isotopes ^{13}C and ^{15}N .

Synthetic E_{21X} -polymers To investigate conformational preferences and the effects of structural formation in RDCs we designed a synthetic E_{21X} -polymer with an amino acid sequence of EEECEEE SEEEEEEE DEECEEE. This polymer could potentially form a disulfide bridge between the two cysteines, which would serve as a model of nascent structure. The slight variation in the sequence would allow us to identify the labeled residues C4, S8, E11, D15 and E21. The E_{21X} -polymers were produced in a similar manner to E_{21} -polymers.

Bicelles A DMPC–DHPC bicelle medium was prepared according to the recipe of [Ottiger & Bax \(1999\)](#). First, (40.4 ± 0.1) mg of DMPC⁸ and (9.6 ± 0.1) mg of DHPC⁹ from *Avanti Polar Lipids* were mixed in a molar ratio of 3:1. Then (1.1 ± 0.1) mg of positively charged CTAB¹⁰ from *Aldrich-Chemie* were added to promote the formation of bicelles as well as to stabilise them ([Losonczi & Prestegard, 1998](#)). The mixture was dissolved in D_2O to a solution with an appropriate bicelle concentration. After adding D_2O the solution was repeatedly subjected to vigorous vortexing (*Scientific Industries, Vortex-genie 2, G-560*) and ice baths under ultrasonic micro-

⁸DMPC (abbr.) 1,2-Dimyristoyl-*sn*-Glycero-3-Phosphocholine.

⁹DHPC (abbr.) 1,2-Dihexanoyl-*sn*-Glycero-3-Phosphocholine.

¹⁰CTAB (abbr.) cetyltrimethylammoniumbromid, $\text{CH}^3(\text{CH}_2)^{15}\text{N}(\text{Br})(\text{CH}_3)^3$.

vibrations (*Elma, Transsonic 310/H*) to help the solvation.

4.2.2 Spectroscopy

Residual dipolar couplings are obtained from measurements in two phases. At first scalar, or J -, couplings are measured from the target molecules diluted in D_2O or water and then the cumulative effect of both dipolar, or D -, and scalar couplings are measured from the target molecules diluted in an alignment medium. The RDC values are then obtained by a straightforward subtraction of the J -couplings from the combined ($J+D$)-couplings (Figure 9).

NC^α , $H^\alpha N$, $H^\alpha C$, $C^\alpha C$, $C^\alpha C^\beta$, and $H^\alpha C^\alpha$ couplings were measured from two-dimensional proton-carbon correlation spectra acquired at 30 °C. Pulse sequences employed included standard ^{13}C -HSQC, constant time ^{13}C -HSQC and (S_3 -IS)-HSQC (Cavanagh *et al.*, 1996). Measurement of coupling constants from unfolded proteins results easily in crowded spectra with narrow peak-dispersion. To counter this, spin-state selection (Meissner *et al.*, 1997 | Ottiger *et al.*, 1998 | Würtz & Permi, 2007) was used. With spin-state selection the in-phase and anti-phase components of a spectra are obtained separately to reduce spectral overlap. This method is useful in particular for unfolded proteins that tend to have a narrow signal-dispersion. Also the measurement of small couplings requires dense media that may compromise spectral quality.

THEORETICAL MODELS

The aim of this study was to clarify the interpretation of protein conformational preferences and dynamics from RDC data. In order to establish the innate RDC profile of polypeptide chains—against which all recorded RDC data should be contrasted prior to any interpretation attempts—we computed both analytically and through simulations the RDC profiles of simple, idealistic chain models ([Articles I–III](#)). Conformational transiency and the limits of its influence on measured RDC data was investigated using molecular dynamics simulated conformational ensembles to seek for evidence of dynamic molecular alignment ([Articles IV & V](#)).

5.1 Unfolded proteins

Unfolded proteins have been traditionally pictured as being devoid of any structural features beyond the intrinsic molecular structure of a polypeptide chain. However, this view was challenged ([Ohnishi *et al.*, 2004](#) | [Shortle & Ackerman, 2001](#) | [Zagrovic *et al.*, 2002](#)) by showing that at least some proteins retain structural features similar to the folded state even under extreme denaturing conditions.

The simple chain models presented in [Articles I–III](#) were a pioneering effort to produce the RDC profiles of unfolded proteins. These intrinsic RDC profiles serve as the base-lines, the no-additional-structure-present-profiles, to which any data should be compared to. Any statistically meaningful deviation from such a RDC profile will then be a tell-tale signal of additional or nascent structure beyond that of a mere random polypeptide chain.

Characteristic chain model An unstructured protein, i.e. a polypeptide chain without secondary structure, may be modeled as a sequence of residue-sized segments in what is called a valence chain. When specific side-chain characteristics are omitted, we have a model rather similar to the classical random-walk in three dimensions. The only difference is that the segments of a valence chain are not freely-jointed but rather restricted to rotate on a valence-cone as depicted in [Figure 10](#).

The average end-to-end distance $\langle r_{ee}^2 \rangle$ of an infinite valence chain scales proportionally to the number of segments n and to the length of a segment l squared. The constant scaling factor called the characteristic ratio C_∞ is defined as

$$C_\infty = \lim_{n \rightarrow \infty} \langle r_{ee}^2 \rangle / nl^2 . \quad (6)$$

The characteristic ratio can be used to reduce a valence chain to a random-walk in three dimensions. This is done by redefining a valence chain as a sequence of freely-jointed characteristic segments that have a length defined by the characteristic ratios.

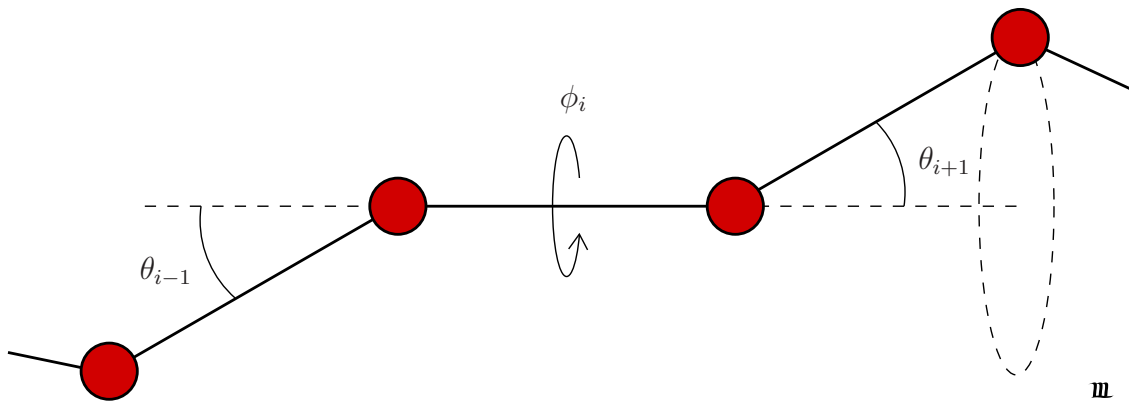


Figure 10. Valence chain. The relative orientations of the segments are governed by segment-specific angles θ and ϕ .

In other words, the rigidity imposed by the valence angles is replaced by effective segment lengths ([Article II](#)).

Residual dipolar couplings The probability distribution of an ensemble of characteristic chains in the vicinity of an obstructing surface can be calculated using the classical results of Smoluchowski and Markoff ([Chandrasekhar, 1943](#)). An unobstructed characteristic chain has the probability

$$P'(n, x) = \sqrt{3/2\pi C_\infty n l^2} e^{-3x^2/2C_\infty n l^2} \quad (7)$$

to be at a distance of x from the starting position after n steps. The addition of an obstructing surface, i.e. the alignment medium, at distance m from the starting position modifies [Equation 7](#) to

$$P(n, x, m) = P'(n, x) - P'(n, 2m - x). \quad (8)$$

[Equation 8](#) serves as the basis for the proposed ([Article I](#)) approximation of the density function $W(m, \theta, n)$ for segment n at distance m and angle θ from the obstructing surface.

$$W(m, \theta, n) = \int_{-L}^m \left[\frac{n-1}{N} P(n-1, x - C_\infty^n l, m - C_\infty^n l) + \frac{N-n}{N} P(N-n, x + C_\infty^{N-n} l, m + C_\infty^{N-n} l) \right] dx, \quad (9)$$

where L is the distance between two obstructing surfaces and N is the number of segments in the chain. The characteristic ratios C_∞^n and C_∞^{N-n} are calculated cumulatively from the characteristic ratios of the individual amino acids in each sub-chain as shown in [Article II](#).

In an axially symmetric system [Equation 1](#) reduces to

$$\langle D^{AB} \rangle = D_{\max}^{AB} (3 \langle \cos^2 \theta \rangle - 1) / 2 \quad (10)$$

and since $\langle \cos^2 \theta \rangle$ can be expressed using [Equation 9](#) as

$$\langle \cos^2 \theta \rangle = \frac{\int_0^{2\pi} \cos^2 \theta |\sin \theta| \int_0^L W(m, \theta, n) dm d\theta}{\int_0^{2\pi} |\sin \theta| \int_0^L W(m, \theta, n) dm d\theta} \quad (11)$$

one can calculate numerically the RDC profiles of valence chains in alignment media as is shown in [Articles I–III](#).

The RDC profiles predicted by this model are monotonous and bell-shaped for polypeptides that have only one kind of amino acids (Figure 4 in [Article I](#)). Variation in the amino acid sequence translates directly to variation in the RDC profile (Figure 3 in [Article II](#)). “Stiffer” residues like proline increase the signal whereas “nimble” residues like glycine decrease the signal. Complex amino acid sequences result in oscillating RDC profiles that are similar to experimental RDC profiles measured from denatured proteins.

Alas, the intricate details of experimental RDC profiles are beyond the capabilities of the proposed analytical model. This is in part due to the limited supply of characteristic ratios and effective valence angles reported for polymers. Accurate valence angles would enable an accurate calculation of cumulative characteristic ratios for complex amino acid sequences. Another clear limitation in the proposed model is the lack of side-chains and other local topological details. This could be naturally taken into account by expanding the current model to e.g. incorporate multiple branches, i.e. a detailed structure, in the calculation of the characteristic ratios. One such approach would be the recently proposed amino acid bulkiness approximation ([Cho *et al.*, 2007](#)). Hydration and other secondary effects could likewise be taken into account with suitable modifications to the chain geometry based e.g. on the recently assessed ([Bernadó *et al.*, 2006](#)) residue-specific solvent accessibilities in unfolded proteins.

The limitations in the theoretical model led us to consider simulated all-atom models as a more meaningful approach. Nevertheless, the analytical models give a clear picture of what is going on at the face of the liquid crystalline particle.

5.2 Dynamic molecular alignment

The dynamic nature of protein structures manifests in RDC measurements in many ways. Here we focus on what we call dynamic molecular alignment. The alignment frame of a protein is in a continuous state of change due to the never ceasing internal atom position shifts. The fluctuation of the alignment frame are presumed to be small and had been dismissed as insignificant. In [Article IV](#) we show that at least for small and globular proteins DMA can be significant and should be taken into account.

Cause Dynamics modulated alignment implies that the individual contributions to the RDC signal are uneven among the conformations of the representative ensemble. A slight change in the shape or the orientation of the magnetic moment translates into a change in the probability, or relative population, of the conformation. A conformation in a less favored shape or with a magnetic moment that is more in

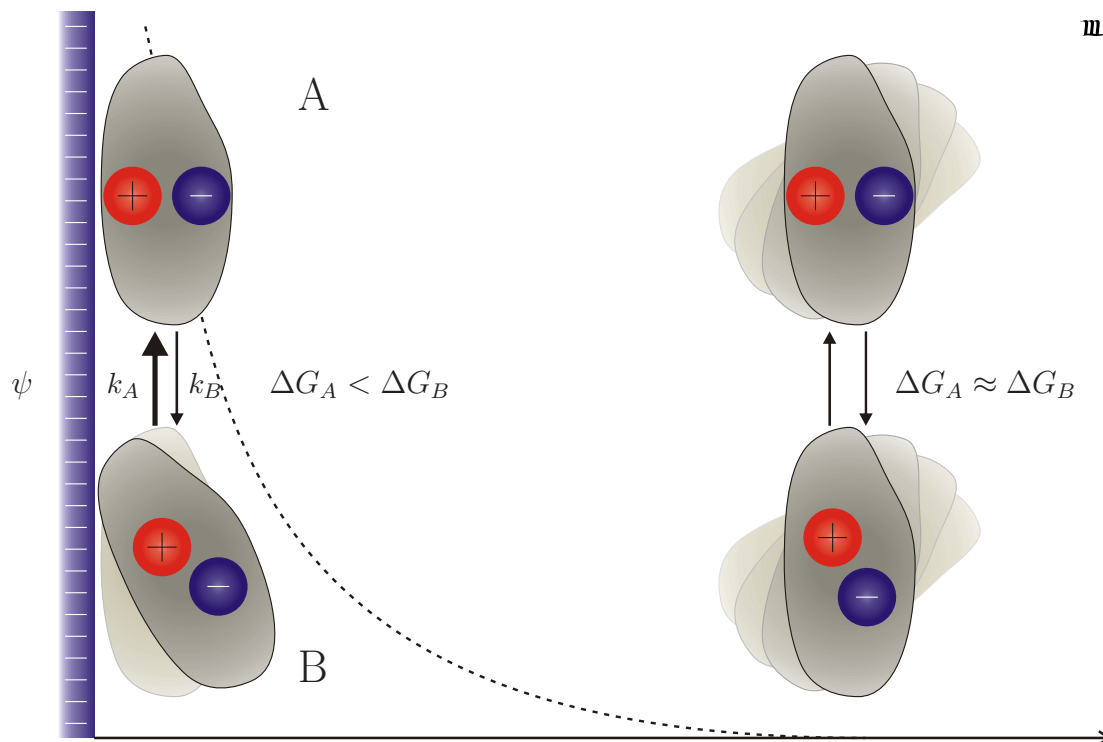


Figure 11. Schematic representation of two nearly identical conformations in a dilute liquid crystalline sample. Conformation B’s magnetic moment is slightly askew making it energetically less favourable for it to orient exactly parallel to the nematic crystal. Consequently, conformation A is slightly more probable than conformation B. (Adapted by permission from *J Am Chem Soc* **2006**, 128: 4371–4376. Copyright 2006 American Chemical Society.)

discord with the steric alignment frame leads to a lower probability as a consequence of a decrease in the accessible state space (Figure 11).

Each conformation’s relative contribution to the RDC signal can be calculated as the average contribution over all residues n or

$$C = \frac{1}{n} \sum_{i=1}^n D_i / \langle D_i \rangle, \quad (12)$$

where D_i is the RDC signal of residue i (Bax *et al.*, 2001). As we have shown in Article IV this is roughly equal to the relative population, or probability, of the conformation. In order to take DMA into account the RDC signals of every conformation in an ensemble should be weighted according to their relative contributions, i.e.

$$\mathbf{d} = \mathbf{D}\mathbf{c}^T/N, \quad (13)$$

where \mathbf{D} is a $N \times n$ matrix of RDC signals for the whole ensemble, \mathbf{c} is a N sized vector $[C_1, C_2, \dots, C_N]$ of relative contributions and \mathbf{d} is a n sized vector $[D_1, D_2, \dots, D_n]$ of true, unbiased RDC signals.

Effect The effects of DMA are not uniform, universal nor even unique. DMA-bias is not the same for each residue in the molecule, it can be significantly different

even for two molecules with an almost identical structure and it is not possible, and hardly sensible, to deconvolute the measured RDC signal *a priori* without the conformational ensemble. To counter the effect of DMA in refinement protocols, each conformation's DMA-bias should be removed after a tentative conformational ensemble is obtained but before comparison is done with measurement data. Alas, it is hard to envision a feasible implementation of this scheme in the current architectures even with modern hardware.

In a simulation we obviously have full knowledge of all conformations and their alignments. The problem of obtaining RDCs from a dynamic ensemble can be solved directly. In practice one attempts to solve the indirect problem of obtaining dynamics from RDCs. This is done from experimental RDCs by using [Equation 3](#). It takes at least 5 non-redundant measurements to determine the alignment tensor. Since the scale of the obtained order parameters is relative, they are usually scaled to match the Lipari-Szabo model-free values. All in all this deconvolution is an impressive achievement. It is very difficult to deconvolute all conceivable motional modes from the data as is shown in [Article V](#).

Gaussian axial fluctuations A conceptually different, complementary explanation for the observed lower-than-expected order parameters has also been devised in the form of the Gaussian axial fluctuations (GAF) model ([Bremi & Brüschweiler, 1997](#) | [Brüschweiler & Wright, 1994](#)). In the GAF-model the internal motions of a protein are implicitly assumed to be decoupled from the molecular alignment ([Bouvignies *et al.*, 2006a](#)). The dynamic information in RDCs is then attributed to Gaussian fluctuations of individual peptide planes. The GAF-model has been used to show correlated hydrogen-bond mediated motions ([Bouvignies *et al.*, 2005](#)) as well as to determine simultaneously protein backbone structure and dynamics ([Bouvignies *et al.*, 2006a](#)). Removal of any effects due to DMA before applying the GAF-model would yield information on the dynamic properties of individual peptide planes without artefacts from molecular alignment fluctuations.

5.3 Computational methods

In addition to liquid-state NMR experiments the proposed theoretical models were checked against computational models. A model for estimating the RDC-profile of unfolded proteins presented in [Articles I–III](#) was devised using torsion-angle dynamics generated conformational ensembles. In [Articles IV & V](#) the presence and significance of DMA was analysed from molecular dynamics trajectories as well as from experimental conformational ensembles.

5.3.1 Polysim

A complete from-sequence-to-RDCs analysis-chain called `polysim` was written in the programming language Python. It employed two external programs: CYANA ([Güntert *et al.*, 1997](#)) and PALES ([Zweckstetter & Bax, 2000](#) | [Zweckstetter *et al.*, 2004](#)). CYANA was used to generate random conformations of a polypeptide with a given

amino acid sequence. PALES was used to predict the alignment of the conformations in a dilute liquid crystal medium. From the PALES predictions the average RDCs were calculated with an error estimate. Polysim makes it possible to get a rough estimate of the RDC-profile of any unfolded protein with minimal effort. The procedure is very simple and relatively fast on modern desktop computers.

Torsion-angle dynamics In torsion-angle dynamics (Jain *et al.*, 1993 | Stein *et al.*, 1997) the representation of a molecule is reduced to its “free” angles and energy functions and forces are expressed in the reduced coordinates akin to classical Lagrangian dynamics. CYANA and its predecessors are implementations of torsion-angle dynamics for NMR structure determination. Protein conformations are sampled in vacuum against a fixed set of experimental constraints using e.g. simulated annealing. Solvation effects are included in the form of implicit water models and a pervasive heat bath. The reduced coordinate space used in torsion-angle dynamics eases the computational burden involved in the calculations making it an attractive candidate for the generator of the large conformational ensembles needed to represent unfolded proteins. Nonetheless, the limitations of the implicit water model make CYANA susceptible to conformational artefacts that may be significant.

Statistical coil model A couple of years after the publication of polysim Sosnick and coworkers presented (Jha *et al.*, 2005) another protocol for the generation of more realistic conformational ensembles of unfolded proteins. At the same time a similar approach was devised independently by Bernadó *et al.* (2005). Their ingenious approaches utilise the vast amount of structural information stored in the Protein Data Bank. First, they collect a fragment database from the unstructured regions of the stored protein structures. This coil/loop-library is then used to build random conformations by statistically sampling the extracted fragments. In this way the conformational preferences are adopted without knowing specifically what they are. The results of the statistical coil models have been promising but some problems still remain (Dames *et al.*, 2006). Recently it was shown (Cho *et al.*, 2007) that an even simpler model based on amino acid bulkiness can also explain many aspects of unfolded protein RDC profiles.

Alignment prediction The generation of a realistic, representative conformational ensemble is only the first step in the prediction of residual dipolar couplings. The second, just as crucial, step is the prediction of a conformation’s effective alignment in dilute liquid crystal media. PALES accomplishes this with great success in uncharged alignment media (Zweckstetter & Bax, 2000) and with modest success in charged alignment media (Zweckstetter *et al.*, 2004). Attractive potentials seem to be especially challenging for PALES and the results in such media are considered unrealistic. Nonetheless, it was shown recently by Skora *et al.* (2006) that for unfolded proteins electrostatic interactions are crucial in the alignment process and can be reliably estimated using PALES.

The rapidity of the whole conformation generation–alignment prediction chain is of utmost importance with unfolded proteins. The conformational ensembles need

to contain easily tens of thousands of conformations to be representative for a small protein. This means a much heavier computational burden rendering e.g. full MD simulations unfeasible in all but the most rigorous, time-consuming studies.

Molecular dynamics simulations are routinely used to generate dynamic ensembles of folded proteins (cf. DePristo *et al.*, 2003 | Lindorff-Larsen *et al.*, 2005 | Richter *et al.*, 2007 | Schwieters *et al.*, 2006), but the enormity of an unfolded protein's state space would require orders of magnitude longer trajectories. Maybe with modern grid computing techniques this would still be feasible, but hardly practical. Then again, MD simulations could be quite fruitful in conjunction with alignment prediction in charged alignment media. The mapping of protein–nematogen¹¹ interactions in the vicinity of the nematogen could serve as a basis for an accurate alignment prediction.

5.3.2 Molecular dynamics simulations

The dynamic conformational ensembles of folded ubiquitin, G protein β_3 domain and troponin C used in [Articles IV & V](#) were generated with MD simulations. The initial structures were first heated and then left to equilibrate after which conformational snap-shots were collected at regular intervals. The simulations were done with the MD simulation software GROMACS (Berendsen *et al.*, 1995 | Lindahl *et al.*, 2001 | van der Spoel *et al.*, 2005) using implicit water in the form of a generalised Born model (Im *et al.*, 2003).

Generalised Born model The use of explicit water molecules in MD simulations adds a considerable computational load in the form of hundreds of additional independent molecules. This has been traditionally avoided by using implicit solvent models (Roux & Simonson, 1999), i.e. by replacing the water molecules with pseudo-potentials that mimic the effects of the solvent. This is of course done at the expense of realism. The generalised Born model (Still *et al.*, 1990) is a highly successful implicit solvent model (Bashford & Case, 2000 | Im *et al.*, 2003 | Lee *et al.*, 2002).

The solvation potential and the solvation energy of an arbitrary solvent molecule can be defined by solving the Poisson-Boltzmann equation (Bertonati *et al.*, 2007 | Sharp & Honig, 1990). The generalised Born model is a continuum dielectric model that gives a reasonable approximation of the solvation energy (Im *et al.*, 2003). It makes two approximations to the Poisson-Boltzmann equation (Onufriev *et al.*, 2002). First, self-interactions are expressed as *effective Born radii* that give the distances between specific atoms and effective spherical dielectric boundaries (Im *et al.*, 2003). Second, pair-wise interactions are estimated by an analytic function depending on the effective Born radius and the inter-atomic distance of the two atoms. The importance of determining the effective Born radii accurately is therefore paramount (Onufriev *et al.*, 2002).

The generalised Born model is computationally fast and yet accurate enough to be of good use in studies of biological systems.

¹¹*nematogen* (lat.) Nematic liquid crystal, i.e. translucent crystalline material of semi-ordered molecules in liquid phase.

Molecular dynamics or torsion-angle dynamics simulations of unfolded or denatured proteins are challenging. Not only is there a large amount of motional freedom in the molecule but also the force fields are developed to be used with folded proteins. For example the presence of denaturants is a dramatic change to the solution parameters. Residual dipolar couplings could provide data for the development of specific force fields for denatured and unfolded proteins.

CONCLUSIONS

In this study I have clarified how to interpret residual dipolar coupling data for protein structure and dynamics. Unfolded proteins were shown to display residual dipolar couplings without any additional structure beyond that of a polypeptide chain (section 5.1, Articles I–III). Also, molecular dynamics were shown to induce alignment modulation and to affect RDC signals (section 5.2, Articles IV & V). The results suggest the following procedures.

Before attempting to interpret RDC data from an unfolded protein, one needs to establish the expected “background noise” from the mere amino acid sequence. Any deviation from this baseline is then a sign of additional structural features. The prediction of such a profile is of course intricate since RDCs are fundamentally sensitive to even the smallest structural changes of which there is no short of in unfolded proteins. The poor signal–noise ratio forces one to use large structural ensembles making one prone to any inbuilt statistical failure of the methods in use. The simplicity offered by the presented theoretical model is therefore appealing. With this model it is possible to learn quickly whether an unfolded protein has any plausible additional structural features remaining.

Molecular dynamics induced modulation of alignment can be countered by using multiple alignment frames and by taking conformational probabilities into account in the averaging process. The principle of one alignment frame per conformation guarantees that dynamics induced alignment fluctuations do not introduce artefacts in RDC interpretations. The correlation between the general degree of order and the probability of a conformation offers a convenient, readily available route for calculating the weighted average RDC signal of an ensemble of conformations. Applying these two corrections to data analysis is straightforward, but in the case of experimental analysis one does not necessarily have all that is needed readily available. Nonetheless, applying the corrections to a tentative structural ensemble prior to further analysis would reduce the artefacts in question.

The results and models presented in this work address important open questions in studies of protein structure and dynamics by NMR spectroscopy. To understand the dynamic properties of proteins is to understand a fundamental biological mechanism with far-reaching consequences. I hope this work will benefit our efforts to understand the plentiful wonders of nature.

BIBLIOGRAPHY

- Akke M & Palmer A G (1996) Monitoring macromolecular motions on microsecond to millisecond time scales by $R_{1\rho}$ - R_1 constant relaxation time NMR spectroscopy. *J Am Chem Soc*, **118**: 911–912 | [10](#)
- Annala A *et al.* (1999) Recognition of protein folds via dipolar couplings. *J Biomol NMR*, **14**: 223–230 | [13](#)
- Ardley H C & Robinson P A (2004) The role of ubiquitin-protein ligases in neurodegenerative disease. *Neurodegenerative Dis*, **1**: 71–87 | [4](#)
- Bashford D & Case D A (2000) Generalized Born models of macromolecular solvation effects. *Annu Rev Phys Chem*, **51**: 129–152 | [24](#)
- Bax A & Tjandra N (1997a) Are proteins even floppier than we thought? *Nature Struct Biol*, **4**: 254–256 | [10](#)
- Bax A & Tjandra N (1997b) High-resolution heteronuclear NMR of human ubiquitin in an aqueous liquid crystalline medium. *J Biomol NMR*, **10**: 289–292 | [14](#)
- Bax A *et al.* (2001) Dipolar couplings in macromolecular structure determination. *Methods Enzymol*, **339**: 127–174 | [21](#)
- Berendsen H J C *et al.* (1995) GROMACS: A message-passing parallel molecular dynamics implementation. *Comput Phys Commun*, **91**: 43–56 | [24](#)
- Bernadó P *et al.* (2005) A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proc Nat Acad Sci USA*, **102**: 17002–17007 | [23](#)
- Bernadó P *et al.* (2006) Sequence-specific solvent accessibilities of protein residues in unfolded protein ensembles. *Biophys J*, **91**: 4536–4543 | [20](#)
- Bertonati C *et al.* (2007) Poisson-Boltzmann calculations of non-specific salt effects on protein-protein binding free energies. *Biophys J*, **92**: 1891–1899 | [24](#)
- Bouvignies G *et al.* (2005) Identification of slow correlated motions in proteins using residual dipolar and hydrogen-bond scalar couplings. *Proc Nat Acad Sci USA*, **102**: 13885–13890 | [9](#), [22](#)
- Bouvignies G *et al.* (2006a) Simultaneous determination of protein backbone structure and dynamics from residual dipolar couplings. *J Am Chem Soc*, **128**: 15100–15101 | [22](#)

- Bouvignies G *et al.* (2006b) Ultrahigh-resolution backbone structure of perdeuterated protein GB1 using residual dipolar couplings from two alignment media. *Angew Chem Int Ed*, **45**: 8166–8169 | [13](#)
- Bremi T & Brüschweiler R (1997) Locally anisotropic internal polypeptide backbone dynamics by NMR relaxation. *J Am Chem Soc*, **119**: 6672–6673 | [22](#)
- Brüschweiler R & Wright P E (1994) NMR order parameters of biomolecules: A new analytical representation and application to the gaussian axial fluctuation model. *J Am Chem Soc*, **116**: 8426–8427 | [22](#)
- Campbell A P & Sykes B D (1993) The two-dimensional transferred nuclear Overhauser effect: Theory and practice. *Annu Rev Biophys Biomol Struct*, **22**: 99–122 | [9](#)
- Cavanagh J *et al.*, *Protein NMR Spectroscopy* (Academic Press, San Diego, 1996) | [13](#), [17](#)
- Chandrasekhar S (1943) Stochastic problems in physics and astronomy. *Rev Mod Phys*, **15**: 1–89 | [19](#)
- Cho M K *et al.* (2007) Amino acid bulkiness defines the local conformations and dynamics of natively unfolded α -synuclein and tau. *J Am Chem Soc*, **129**: 3032–3033 | [20](#), [23](#)
- Clore G M & Schwieters C D (2004) How much backbone motion in ubiquitin is required to account for dipolar coupling data measured in multiple alignment media as assessed by independent cross-validation? *J Am Chem Soc*, **126**: 2923–2938 | [10](#)
- Clore G M *et al.* (1990) Analysis of the backbone dynamics of interleukin-1 β using two-dimensional inverse detected heteronuclear ^{15}N - ^1H NMR spectroscopy. *Biochemistry*, **29**: 7387–7401 | [9](#)
- Creighton T E, *Proteins—Structures and Molecular Principles* (W. H. Freeman and Company, New York, 1984) | [3](#)
- Daaka Y (2004) G proteins in cancer: The prostate cancer paradigm. *Sci STKE*, **216**: re2 | [6](#)
- Dames S A *et al.* (2006) Residual dipolar couplings in short peptides reveal systematic conformational preferences of individual amino acids. *J Am Chem Soc*, **128**: 13508–13514 | [23](#)
- DePristo M A *et al.* (2003) Ab initio construction of polypeptide fragments: Efficient generation of accurate, representative ensembles. *Proteins: Struct Funct Genet*, **51**: 41–55 | [24](#)
- Dobson C M (2003) Protein folding and misfolding. *Nature*, **426**: 884–890 | [11](#)

- Duan Y & Kollman P A (1998) Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science*, **282**: 740–744 | [12](#)
- Dyson H J & Wright P E (2005) Intrinsically unstructured proteins and their functions. *Nature Rev Mol Cell Biol*, **6**: 197–208 | [10](#)
- Feldman R D & Hegele R A (2000) G-protein polymorphisms and maternal/neonatal metabolism: still a weight for the answer. *Lancet*, **355**: 1201–1202 | [6](#)
- Finkelstein A V & Galzitskaya O V (2004) Physics of protein folding. *Phys Life Rev*, **1**: 23–56 | [11](#), [12](#)
- Fischer G & Schmid F X (1990) The mechanism of protein folding. Implications of in vitro refolding models for de novo protein folding and translocation in the cell. *Biochemistry*, **29**: 2205–2212 | [12](#)
- Gagné S M *et al.* (1995) Structures of the troponin C regulatory domains in the apo and calcium-saturated states. *Nature Struct Biol*, **2**: 784–789 | [8](#)
- Gay G d P *et al.* (1995) Folding of a nascent polypeptide chain in vitro: Cooperative formation of structure in a protein module. *Proc Nat Acad Sci USA*, **92**: 3683–3686 | [11](#)
- Goldstein G *et al.* (1975) Isolation of a polypeptide that has lymphocyte-differentiating properties and is probably represented universally in living cells. *Proc Nat Acad Sci USA*, **72**: 11–15 | [4](#)
- Güntert P *et al.* (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol*, **273**: 283–298 | [22](#)
- Hershko A & Ciechanover A (1998) The ubiquitin system. *Annu Rev Biochem*, **67**: 425–479 | [4](#)
- Im W *et al.* (2003) Generalized Born model with a simple smoothing function. *J Comput Chem*, **24**: 1691–1702 | [24](#)
- Jain A *et al.* (1993) A fast recursive algorithm for molecular dynamics simulation. *J Comput Phys*, **106**: 258–268 | [23](#)
- James L C (2003) Antibody multispecificity mediated by conformational diversity. *Science*, **299**: 1362–1367 | [10](#)
- Jha A K *et al.* (2005) Statistical coil model of the unfolded state: Resolving the reconciliation problem. *Proc Nat Acad Sci USA*, **102**: 13099–13104 | [23](#)
- Kay L E *et al.* (1989) Backbone dynamics of proteins as studied by ¹⁵N inverse detected heteronuclear NMR spectroscopy: Application to Staphylococcal nuclease. *Biochemistry*, **28**: 8972–8979 | [9](#), [10](#)
- Kontaxis G *et al.* (2005) Molecular fragment replacement approach to protein structure determination by chemical shift and dipolar homology database mining. *Methods Enzymol*, **394**: 42–78 | [13](#)

- Kopple K D *et al.* (1986) Conformations of cyclic octapeptides. 3. cyclo-(D-Ala-Gly-Pro-Phe)₂. Conformations in crystals and a T_{1ρ} examination of internal mobility in solution. *J Am Chem Soc*, **108**: 4637–4642 | [10](#)
- Lakomek N A *et al.* (2005) Side-chain orientation and hydrogen-bonding imprint supra-τ_c motion on the protein backbone of ubiquitin. *Angew Chem Int Ed*, **44**: 7776–7778 | [10](#)
- Lakomek N A *et al.* (2006) A thorough dynamic interpretation of residual dipolar couplings in ubiquitin. *J Biomol NMR*, **34**: 101–115 | [9](#), [10](#)
- Lee M S *et al.* (2002) Novel generalized Born methods. *J Chem Phys*, **116**: 10606–10614 | [24](#)
- Lehman W *et al.* (2001) Troponin organization on relaxed and activated thin filaments revealed by electron microscopy and three-dimensional reconstruction. *J Mol Biol*, **307**: 739–744 | [8](#)
- Levinthal C, *Mossbauer Spectroscopy in Biological Systems, Proceedings of a meeting held at Allerton House, Monticello, Illinois* (University of Illinois Press, Urbana, 1969), pp. 22–24 | [11](#)
- Lindahl E *et al.* (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J Mol Model*, **7**: 306–317 | [24](#)
- Lindorff-Larsen K *et al.* (2004) Transition states for protein folding have native topologies despite high structural variability. *Nature Struct Mol Biol*, **11**: 443–449 | [12](#)
- Lindorff-Larsen K *et al.* (2005) Simultaneous determination of protein structure and dynamics. *Nature*, **433**: 128–132 | [24](#)
- Lipari G & Szabo A (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J Am Chem Soc*, **104**: 4546–4559 | [9](#)
- Losonczi J A & Prestegard J H (1998) Improved dilute bicelle solutions for high-resolution NMR of biological macromolecules. *J Biomol NMR*, **12**: 447–451 | [16](#)
- Losonczi J A *et al.* (1999) Order matrix analysis of residual dipolar couplings using singular value decomposition. *J Magn Reson*, **138**: 334–342 | [14](#), [15](#)
- Malbon C C (2005) β-catenin, cancer, and G proteins: Not just for Frizzleds anymore. *Sci STKE*, **292**: pe35 | [6](#)
- Massi F *et al.* (2005) Microsecond timescale backbone conformational dynamics in ubiquitin studied with NMR r_{1ρ} relaxation experiments. *Protein Sci*, **14**: 735–742 | [10](#)

- Mattinen M L *et al.* (2002) Quaternary structure built from subunits combining NMR and small-angle X-ray scattering data. *Biophys J*, **83**: 1177–1183 | [13](#)
- Meiler J *et al.* (2001) Model-free approach to the dynamic interpretation of residual dipolar couplings in globular proteins. *J Am Chem Soc*, **123**: 6098–6107 | [1](#), [10](#), [13](#)
- Meiler J *et al.* (2003) Dipolar couplings in multiple alignments suggest α helical motion in ubiquitin. *J Am Chem Soc*, **125**: 8072–8073 | [10](#)
- Meissner A *et al.* (1997) Integration of spin-state-selective excitation into 2D NMR correlation experiments with heteronuclear ZQ/2Q π rotations for $^1J_{\text{XH}}$ -resolved E.COSY-type measurement of heteronuclear coupling constants in proteins. *J Biomol NMR*, **10**: 89–94 | [17](#)
- Nakayama K I & Nakayama K (2006) Ubiquitin ligases: cell-cycle control and cancer. *Nature Rev Cancer*, **6**: 369–381 | [4](#)
- Nirmala N R & Wagner G (1988) Measurement of ^{13}C relaxation times in proteins by two-dimensional heteronuclear ^1H - ^{13}C correlation spectroscopy. *J Am Chem Soc*, **110**: 7557–7558 | [9](#)
- Offermanns S (2003) G-proteins as transducers in transmembrane signalling. *Prog Biophys Mol Biol*, **83**: 101–130 | [6](#)
- Ohnishi S *et al.* (2004) Direct demonstration of structural similarity between native and denatured eglin C. *Biochemistry*, **43**: 4064–4070 | [12](#), [18](#)
- Onufriev A *et al.* (2002) Effective Born radii in the generalized Born approximation: The importance of being perfect. *J Comput Chem*, **23**: 1297–1304 | [24](#)
- Ottiger M & Bax A (1999) Bicelle-based liquid crystals for NMR-measurement of dipolar couplings at acidic and basic pH values. *J Biomol NMR*, **13**: 187–191 | [16](#)
- Ottiger M *et al.* (1998) Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J Magn Reson*, **131**: 373–378 | [17](#)
- Palmer A G (2001) NMR probes of molecular dynamics: Overview and comparison with other techniques. *Annu Rev Biophys Biomol Struct*, **30**: 129–155 | [9](#)
- Palmer A G (2004) NMR characterization of the dynamics of biomacromolecules. *Chem Rev*, **104**: 3623–3640 | [10](#)
- Palmer A G *et al.* (2001) Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol*, **339**: 204–238 | [10](#)
- Pandit A D *et al.* (2006) Small proteins fold through transition states with native-like topologies. *J Mol Biol*: 755–770 | [12](#)

- Pauling L (1940) A theory of the structure and process of formation of antibodies. *J Am Chem Soc*, **62**: 2643–2657 | 10
- Peti W *et al.* (2002) Model-free analysis of protein backbone motion from residual dipolar couplings. *J Am Chem Soc*, **124**: 5822–5833 | 10
- Prusiner S B (1997) Prion diseases and the BSE crisis. *Science*, **278**: 245–251 | 11
- Prusiner S B (1998) Prions. *Proc Nat Acad Sci USA*, **95**: 13363–13383 | 11
- Pääkkönen K *et al.* (2000) Conformations of the regulatory domain of cardiac troponin C examined by residual dipolar couplings. *Eur J Biochem*, **267**: 6665–6672 | 13
- Raju T N K (2000) The nobel chronicles. *Lancet*, **355**: 2259 | 6
- Rhee Y M *et al.* (2004) Simulations of the role of water in the protein-folding mechanism. *Proc Nat Acad Sci USA*, **101**: 6456–6461 | 12
- Richter B *et al.* (2007) The MUMO (minimal under-restraining minimal over-restraining) method for the determination of native state ensembles of proteins. *J Biomol NMR*, **37**: 117–135 | 24
- Robertson S P *et al.* (1981) The time-course of Ca^{2+} exchange with calmodulin, troponin, parvalbumin, and myosin in response to transient increases in Ca^{2+} . *Biophys J*, **34**: 559–569 | 7
- Roos-Mattjus P & Sistonen L (2004) The ubiquitin-proteasome pathway. *Annals Med*, **36**: 285–295 | 4
- Roux B & Simonson T (1999) Implicit solvent models. *Biophys Chem*, **78**: 1–20 | 24
- Schwieters C D *et al.* (2006) Using Xplor–NIH for NMR molecular structure determination. *Prog Nucl Magn Reson Spectrosc*, **48**: 47–62 | 24
- Sharp K A & Honig B (1990) Calculating total electrostatic energies with the non-linear Poisson-Boltzmann equation. *J Phys Chem*, **94**: 7684–7692 | 24
- Shortle D & Ackerman M S (2001) Persistence of native-like topology in a denatured protein in 8 M urea. *Science*, **293**: 487–489 | 1, 12, 18
- Skelton N J *et al.* (1993) Practical aspects of two-dimensional proton-detected ^{15}N spin relaxation measurements. *J Magn Reson*, **102**: 253–264 | 9
- Skora L *et al.* (2006) Charge-induced molecular alignment of intrinsically disordered proteins. *Angew Chem Int Ed*, **45**: 7012–7015 | 23
- Sorsa T *et al.* (2004) The contractile apparatus as a target for drugs against heart failure: Interaction of levosimendan, a calcium sensitiser, with cardiac troponin C. *Mol and Cell Biochem*, **266**: 87–107 | 7

- van der Spoel D *et al.* (2005) GROMACS: Fast, flexible and free. *J Comput Chem*, **26**: 1701–1718 | [24](#)
- Stein E G *et al.* (1997) Torsion-angle molecular dynamics as a new efficient tool for NMR structure calculation. *J Magn Reson*, **124**: 154–164 | [23](#)
- Still W C *et al.* (1990) Semianalytical treatment of solvation for molecular mechanics and dynamics. *J Am Chem Soc*, **112**: 6127–6129 | [24](#)
- Thomas P J *et al.* (1995) Defective protein folding as a basis of human disease. *Trends Biochem Sci*, **20**: 456–459 | [11](#)
- Tjandra N & Bax A (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science*, **278**: 1111–1114 | [14](#)
- Tjandra N *et al.* (1995) Rotational dynamics of calcium-free calmodulin studied by ¹⁵N-NMR relaxation measurements. *Eur J Biochem*, **230**: 1014–1024 | [10](#)
- Tolman J R *et al.* (1997) NMR evidence for slow collective motions in cyanometmyoglobin. *Nature Struct Biol*, **4**: 292–297 | [10](#)
- Tolman J R *et al.* (2001) Structural and dynamic analysis of residual dipolar coupling data for proteins. *J Am Chem Soc*, **123**: 1416–1424 | [1](#), [10](#), [13](#)
- Vijay-Kumar S *et al.* (1987) Comparison of the three-dimensional structures of human, yeast, and oat ubiquitin. *J Biol Chem*, **262**: 6396–6399 | [4](#)
- Voet D & Voet J G, *Biochemistry* (John Wiley & Sons, New Jersey, 2004), 3rd edn. | [3](#)
- Weeks R A & Perry S V (1978) Characterization of a region of the primary sequence of troponin C involved in calcium ion-dependent interaction with troponin I. *Biochem J*, **173**: 449–457 | [8](#)
- Whitford D, *Proteins—structure and function* (John Wiley & Sons, Chichester, 2005) | [3](#)
- Wright P E & Dyson H J (2002) Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *J Mol Biol*, **293**: 321–331 | [10](#)
- Würtz P & Permi P (2007) SESAME-HSQC for simultaneous measurement of NH and CH scalar and residual dipolar couplings. *Magn Reson Chem*, **45**: 289–295 | [17](#)
- Xu C *et al.* (1999) Tropomyosin positions in regulated thin filaments revealed by cryoelectron microscopy. *Biophys J*, **77**: 985–992 | [8](#)
- Zagrovic B *et al.* (2002) Native-like mean structure in the unfolded ensemble of small proteins. *J Mol Biol*, **323**: 153–164 | [12](#), [18](#)

Zwanzig R *et al.* (1992) Levinthal's paradox. *Proc Nat Acad Sci USA*, **89**: 20–22 | [11](#)

Zweckstetter M & Bax A (2000) Prediction of sterically induced alignment in a dilute liquid crystalline phase: Aid to protein structure determination by NMR. *J Am Chem Soc*, **122**: 3791–3792 | [22](#), [23](#)

Zweckstetter M *et al.* (2004) Prediction of charge-induced molecular alignment of biomolecules dissolved in dilute liquid-crystalline phases. *Biophys J*, **86**: 3444–3460 | [22](#), [23](#)