

Development of NMR Methods for the Study of Protein Folding

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim
Fachbereich Biochemie, Chemie und Pharmazie
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von
KAI SCHLEPCKOW
aus Karlsruhe

Frankfurt am Main
2008

(D30)

Dem Fachbereich Biochemie, Chemie und Pharmazie der
Johann Wolfgang Goethe-Universität als Dissertation vorgelegt.

Dekan: Prof. Dr. Harald Schwalbe

Gutachter: Prof. Dr. Harald Schwalbe und Prof. Dr. Clemens Glaubitz

Datum der Disputation:

Contents

List of Abbreviations	vii
Abstract	ix
1. Introduction	1
1.1. Protein Folding	1
1.2. NMR spectroscopy for the study of protein folding	5
1.2.1. Folding intermediates	5
1.2.2. Unfolded states of proteins	6
1.3. Thesis: projects and objectives	7
2. Theoretical Framework for the interpretation of NMR residual dipolar couplings of unfolded proteins	9
2.1. Introduction	9
2.1.1. The unfolded state and its study by NMR	9
2.1.2. Weak alignment and residual dipolar couplings	11
2.1.3. Prediction and simulation of residual dipolar couplings under weakly aligning conditions	19
2.2. Results and Discussion	26
2.2.1. Theoretical Framework	26
2.2.2. Predictions and comparison with experiment	37
2.3. Conclusion	51
3. Kinetic refolding of bovine α-lactalbumin (BLA)	53
3.1. Introduction	53

Contents

3.1.1.	Some general aspects	53
3.1.2.	Measurement of protein folding kinetics by stopped-flow techniques .	57
3.1.3.	Measurement of protein folding kinetics by NMR	58
3.1.4.	Bovine α -lactalbumin as a model system	60
3.2.	Materials and Methods	63
3.2.1.	CD spectroscopy	63
3.2.2.	Stopped-flow spectroscopy	63
3.2.3.	Heterologous expression and purification of ^{15}N -labeled bovine α - lactalbumin	64
3.2.4.	NMR resonance assignment of methyl and methylene groups in the native state	66
3.2.5.	Time-resolved NMR spectroscopy	67
3.3.	Results	68
3.3.1.	Tryptophan fluorescence and CD spectroscopic steady-state investiga- tions	68
3.3.2.	Folding kinetics by stopped-flow spectroscopy	71
3.3.3.	Heterologous expression and purification of ^{15}N -labeled bovine α - lactalbumin	77
3.3.4.	Assignment of backbone amide and side-chain aliphatic resonances .	79
3.3.5.	Folding kinetics by time-resolved NMR	87
3.4.	Discussion	94
3.4.1.	BLA refolding is (more) complex	94
3.4.2.	BLA folds <i>via</i> parallel pathways	97
3.4.3.	BLA and HEWL share a common folding mechanism	98
3.4.4.	Advances in NMR methodology	100
3.4.5.	Future prospects	101
4.	Photo-CIDNP studies on bovine α-lactalbumin in the presence of param- agnetic lanthanides	103
4.1.	Introduction	103
4.1.1.	Photochemically induced dynamic nuclear polarisation (Photo-CIDNP)	103

4.1.2. Paramagnetic lanthanides as shift and relaxation reagents and their effect on CIDNP	107
4.1.3. Lanthanide-induced refolding of bovine α -lactalbumin	108
4.2. Materials and Methods	110
4.2.1. Photo-CIDNP NMR spectroscopy	110
4.2.2. Organic synthesis of flavin I	111
4.3. Results	112
4.3.1. Photo-CIDNP of N-Acetylhistidine, N-Acetyltryptophan, and N-Acetyltirosine in the presence of ytterbium(III)	112
4.3.2. Organic synthesis of flavin I	118
4.3.3. Photo-CIDNP investigations on bovine α -lactalbumin using flavin I	120
4.4. Discussion	128
A. Appendix	131
A.1. Heterologous expression and purification of ^{15}N -labelled bovine α -lactalbumin	131
A.2. NMR resonance assignments	134
B. References	141
Zusammenfassung	159
Danksagung	165
Curriculum Vitae	167
List of Publications	171

List of Abbreviations

1D, 2D, 3D	one-dimensional, two-dimensional, three-dimensional
ACBP	acyl coenzyme binding protein
AEBSF	4-(2-Aminoethyl)benzenesulfonyl fluoride hydrochloride
apoMb	apomyoglobin
α S	α -synuclein
BLA	bovine α -lactalbumin
BPTI	bovine pancreatic trypsin inhibitor
C8E5	n-octyl-penta(ethylene glycol)
C12E5	n-dodecyl-penta(ethylene glycol)
<i>CmEn</i>	n-alkyl-poly(ethylene glycol)
CARA	computer aided resonance assignment
CD	circular dichroism
CIDNP	chemically induced dynamic nuclear polarization
CPBr	cetylpyridinium bromide
CTD	C-terminal domain
DHPC	dihexanoylphosphatidylcholine
DMN	dimethoxy-nitrophen
DMPC	dimyristoylphosphatidylcholine
DNA	desoxyribonucleic acid
DSS	2,2-dimethyl-2-silapentane-5-sulfonic acid
<i>E. coli</i>	<i>Escherichia coli</i>
fd	bacteriophage fd
FMN	riboflavin 5'-mononucleotide
GAF	Gaussian axial fluctuation

List of Abbreviations

GB	B1 domain of protein G
GdnCl	guanidine hydrochloride
HEWL	hen egg white lysozyme
HSQC	heteronuclear single quantum coherence
IMAC	immobilized metal ion affinity chromatography
IPTG	isopropyl β -D-1-thiogalactopyranoside
LB	Luria broth
MG	molten globule
mRNA	messenger ribonucleic acid
MS	mass spectrometry
NDBF	nitrodibenzofuran
Ni-NTA	Nickel nitrilotriacetic
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
OD ₆₀₀	optical density at 600 nm
PAG	polyacrylamide gel
PALES	Prediction of ALignmEnt from Structure
Pf1	bacteriophage Pf1
PP _{II}	polyproline II
ppm	parts per million
PRE	paramagnetic relaxation enhancement
RDC	residual dipolar coupling
RNase A	ribonuclease A
rpm	rounds per minute
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate - polyacrylamide gel electrophoresis
SNase	staphylococcal nuclease
States-TPPI	States time-proportional phase increment
UV	ultra-violet
WT	wild type

Abstract

To decipher the principles that govern a given amino acid sequence to attain a well-defined three-dimensional fold is one of the major goals in modern structural biology. The correct folding of proteins assures their functionality and, thereby, cell homeostasis. It is also well established that up to 30% of the proteins encoded in eukaryotic genomes exist in partially or fully unfolded states which are functional. Moreover, partially unfolded states of proteins are implicated in a number of diseases involving formation of fibrillar structures. The investigation of protein folding therefore requires techniques which not only allow to follow conformational transitions but to characterise the conformational states involved as well. In this respect, NMR spectroscopic techniques have proven essential in furthering the understanding of protein folding.

The focus of this thesis has been to further advance and develop existing NMR techniques for the study of protein folding. In order to do so, experimental as well as theoretical approaches have been pursued. From the theoretical side, a successful attempt to the development of a general theory for the treatment of residual dipolar couplings in the case of unfolded proteins has been undertaken. Information contained in residual dipolar couplings is especially valuable due to its long-range nature. The dynamic character of unfolded states of proteins, which may be composed of distinct subsets of conformations, renders reliable interpretation of data a non-trivial task. Statistical-coil-based approaches have been shown to be powerful in data interpretation. A consistent theory based on fundamental polymer physics, however, had not been presented so far.

The herein presented model addresses this problem building on the original work by Annala and co-workers. In this work, several shortcomings have been identified. These shortcomings have been corrected here leading to a general approach for the treatment of residual dipolar couplings of unfolded proteins. More specifically, it is shown that, in the case of fully unfolded

Abstract

proteins aligned by a steric mechanism, basic dependencies of dipolar couplings such as on chain length and location within the chain can be analysed in simple analytical terms. The main predictions of the model are compared to experimental data showing reasonable agreement. The presented mathematical framework is principally suited for various improvements which could include the treatment of long-range interactions and of the actual geometry of the given alignment medium.

From the experimental side, bovine α -lactalbumin has been chosen as a model system for the development of improved time-resolved 1D NMR methods aiming at the observation of conformational transitions by kinetic means. The presented results show that high-quality data can now be obtained at protein concentrations as low as $100\mu\text{M}$. Rate constants characterising distinct conformational transitions of up to 8 s^{-1} have been measured. These are the fastest rate constants which have been reported so far for protein folding events. The NMR data supplemented by complementary biophysical data furthermore demonstrate that the folding of bovine α -lactalbumin is more complex than has been anticipated. All data are consistent with a triangular folding mechanism involving parallel pathways of folding for formation of the native state of the protein. Interestingly, such a folding mechanism has also been found for the highly structurally homologous protein lysozyme from hen egg white. Evidence is presented that the guiding role of long-range interactions in the unfolded state of lysozyme for mediating intersubdomain interactions during folding is replaced in the case of bovine α -lactalbumin by the Ca^{2+} binding site.

In addition, initial studies for the extension of photochemically induced dynamic nuclear polarisation NMR techniques to the paramagnetic situation, namely when paramagnetic lanthanides such as ytterbium(III) are present in solution, have been conducted with the aim of further characterising a folding intermediate located on a parallel folding pathway of bovine α -lactalbumin (see above) in the future. It is shown, using the three polarisable amino acids tryptophan, tyrosine, and histidine, that photo-CIDNP enhancements can be obtained using equimolar amounts of Yb(III), albeit with less sensitivity. A CIDNP dye, namely flavin I, has been successfully synthesised which does not interfere destructively with Yb(III). Preliminary steady-state photo-CIDNP investigations of bovine α -lactalbumin using flavin I as the photosensitiser are presented which provide the basis for future real-time applications.

1. Introduction

1.1. Protein Folding

Protein folding, which constitutes the last step in the transformation of genetic information into well-defined biologically active three-dimensional structures of proteins [1], has been studied extensively for several decades. Groundbreaking work was done by Anfinsen and colleagues in the early 1960s. These scientists used bovine pancreatic ribonuclease denatured by 8M urea and by reduction of the disulfide bonds. The fact that the protein exhibited complete reformation of the disulfide bonds in their correct arrangement upon reoxidation with molecular oxygen unequivocally showed that the information for the 3D structure of the protein was contained in the amino acid sequence [2]. In the following years, further studies led to the conclusion that the 3D structure of a native protein in its physiological milieu is the one with lowest free energy (thermodynamic hypothesis) [3]. It became also evident that folding had to be a guided search for the native structure as simple calculations showed that a random search would take way too long to attain the structure on a biologically reasonable timescale. The possible number of conformations present in the denatured (or, unfolded) state, the starting point of the folding reaction, and the rapidity of folding observed experimentally called for a limited number of folding pathways and nucleation events [3, 4]. Until that time, protein folding data were shown to be interpretable by a two-state model implying the unfolded state and the native state of a protein as the only states being present at any time during folding.

Several reports in 1970/1971, however, provided evidence for the significant population of folding intermediates, defined conformational states with properties different from those of the unfolded state and the native state [5–8]. The following years saw a growing number of studies supporting the existence of folding intermediates. Clear evidence came from the studies by Creighton and co-workers working with bovine pancreatic trypsin inhibitor (BPTI)

1. Introduction

(for a review see [9]). These investigators used the disulfide bonds as probes and were able to trap and characterise folding intermediates which had only part of their disulfide network formed. Kinetic data seemed to be best interpretable in terms of a sequential folding model.



where U , I , and N indicate the unfolded state, intermediate state, and native state, respectively. This model suggests that folding is a sequence of well-defined steps where folding intermediates are transiently populated on the path to the native state. From a structural point of view, available data mainly supported a framework model for folding in which hydrogen-bonded secondary structure was formed first followed by formation of the tertiary structure [10].

One of the most well-studied folding intermediates is the molten globule. This conformational state is characterised by a high amount of native-like secondary structure but ill-defined tertiary interactions associated with a considerable extent of conformational dynamics on the micro- to millisecond timescale (for reviews see [11, 12]). The molten globule intermediate has been identified in both equilibrium and kinetic folding experiments and good agreement in terms of structural properties derived from combined equilibrium and kinetic studies has been demonstrated for α -lactalbumin [13] and apomyoglobin [14] underscoring the relevance of the molten globule in protein folding reactions.

It became already clear in the 1970s that kinetic complexity in folding could not only arise from the presence of folding intermediates but could also be due to conformational heterogeneity in the unfolded state. Slow isomerisation around prolyl peptide bonds was shown to be the reason for the presence of multiple unfolded forms [15–17].

Hydrogen exchange labelling in combination with 2D NMR subsequently appeared as a powerful approach to study protein folding reactions allowing atomic-level characterisation of both folding intermediates [18–20] and parallel pathways for folding [20].

The protein folding field experienced considerable advances when theorists introduced a new conceptual framework for understanding protein folding [21, 22]. This framework is commonly referred to as the 'new view' and describes protein folding in terms of energy landscapes. A schematic is given in Fig. 1.1.

According to this framework, the energy landscape of a foldable protein can be understood as the free energy of each conformation as a function of the degrees of freedom, such as the

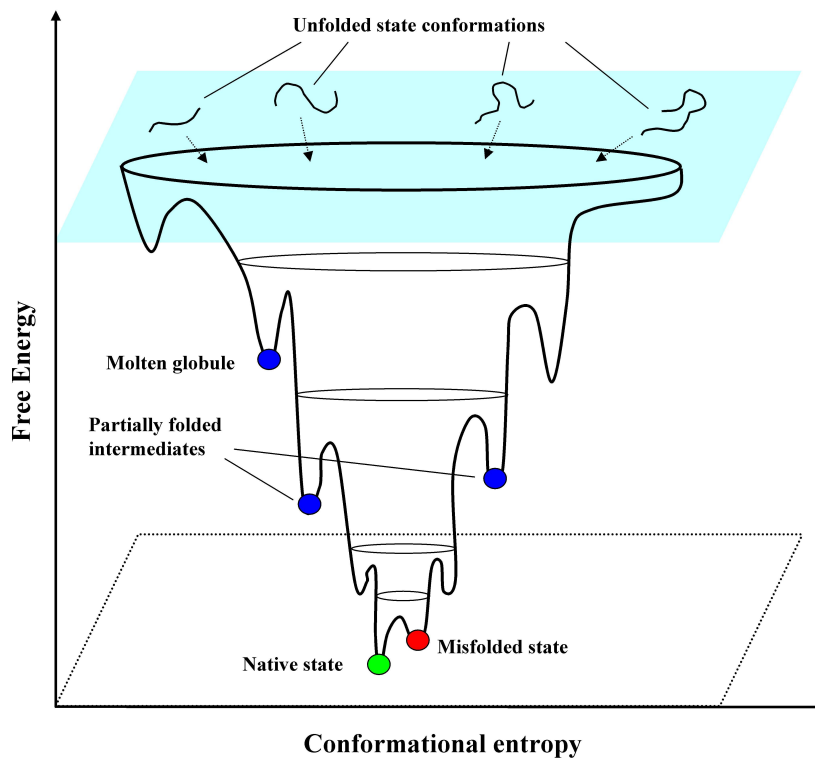


Figure 1.1.: Folding free energy landscape. Both free energy and conformational entropy decrease as the initially unfolded protein passes increasingly narrower ensembles of conformations on the way to the native state. Conformational states a protein may adopt during folding (e.g. molten globule) correspond to local minima on the free energy surface (native state: global minimum of free energy). The figure was adapted from Radford [23].

dihedral bond angles along the polypeptide backbone. The vertical axis corresponds to the 'internal free energy' of a given chain configuration while the many lateral axes represent the conformational coordinates. Generally, only one or two of the lateral axes are shown for illustrative reasons (cf. Fig. 1.1). According to the 'new view', protein folding is seen as a parallel flow process of an ensemble of chain molecules. Unfolded chain molecules may not follow a unique path but multiple paths which ultimately all lead to the same native structure. Some of these paths may involve folding intermediates while others may directly lead to the final structure. Specific time points during folding will consequently not be characterised by specific structures but rather by distributions of conformations about some mean. The 'new view' argues for replacing the pathway concept of sequential events with the funnel concept

1. Introduction

of parallel events [24,25].

The concept of an energy landscape also allows to incorporate the phenomenon of protein misfolding. Misfolding of proteins may occur when hydrophobic patches in the amino acid sequence becoming transiently exposed during protein synthesis in the cell interact with hydrophobic parts of nearby proteins [26]. Proteins may then become trapped in some incorrectly folded (misfolded) conformation. Under certain conditions, misfolding events can lead to the formation of large fibrillar structures associated with disorders such as Alzheimer's disease [27] or Parkinson's disease [28]. Living cells exhibit a number of chaperones and chaperonins, proteins and protein assemblies, respectively, assisting in folding newly synthesised proteins correctly, in order to minimise the probability of misfolding events and to assure cell homeostasis [1]. Misfolded states of proteins correspond to local free energy minima on the multidimensional free energy landscape for folding and may become relevant *in vivo* if populations increase due to, e.g. mutations in the amino acid sequence leading to destabilisation of the native state relative to the misfolded state [29,30].

The current view of how proteins fold to their native structures can be sketched as follows. Residual, transiently populated secondary and tertiary interactions in the unfolded state, which may be native-like or non-native in nature, can serve as initiation sites for folding (folding nuclei) [31]. Folding may then take place through either of three basic mechanisms. (i) Secondary structures form first followed by formation of the tertiary structure (*framework*). (ii) Hydrophobic collapse of the unfolded polypeptide chain based on tertiary interactions takes place first, which may lead to a molten globule folding intermediate, followed by secondary structure formation (*hydrophobic collapse*). (iii) Secondary and tertiary structure condense around a folding nucleus concomitantly (*nucleation – condensation*) [31]. It has been noted by Daggett and Fersht that the framework and hydrophobic collapse mechanisms indeed can be viewed as extreme cases of the nucleation-condensation mechanism [32]. The intrinsic stability of chain segments adopting secondary structural motifs determines whether secondary and tertiary interactions form in a stepwise manner (*framework* and *hydrophobic collapse*, respectively) or simultaneously (*nucleation – condensation*).

1.2. NMR spectroscopy for the study of protein folding

1.2.1. Folding intermediates

The investigation of protein folding using NMR spectroscopic techniques has been vital both to establish the existence of folding intermediates and to allow their atomic-level characterisation. In pioneering work, Schmid and Baldwin detected an early folding intermediate in the folding of ribonuclease A by protection of amide protons against exchange [33]. In this approach, the unfolded protein is first labelled with ^3H . Folding conditions are then chosen such that exchange of ^3H labels by solvent protons is fast compared to the final stages of folding. Any label being protected from exchange with solvent will be an indication of some kind of stable H-bonded structure formed early during folding, i.e. folding intermediates. As already anticipated by the authors in their original publication, these studies were subsequently extended to ^1H - ^2H exchange followed by analysis of NMR spectra [34]. By using this technique, several reports clearly demonstrated the significant population of folding intermediates such as for ribonuclease A [18], cytochrome c [19], barnase [35], and hen lysozyme [20]. The observation of heterogeneous kinetics on the single-residue level also provided evidence for multiple pathways in accord with the 'new view' of protein folding.

In the following years, real-time NMR emerged as a powerful approach to studying protein folding. Here, folding is initiated either outside or inside the NMR magnet and subsequently monitored by acquisition of 1D NMR spectra. This experimental scheme allows recording the disappearance and buildup of resonances indicative of the unfolded state and native state, respectively. These studies have underscored the existence and relevance of folding intermediates [36–39]. In addition, 2D real-time NMR has been used to investigate protein folding. 2D real-time NMR comprises experiments acquiring either a single (kinetic) ^1H , ^{15}N HSQC spectrum [40] or a set of ^1H , ^{15}N HSQC spectra [41] during folding. Provided that the folding reaction under study is sufficiently slow, these experiments give more detailed insight than 1D real-time experiments due to the additional spectral resolution in 2D.

Moreover, the application of photochemically induced dynamic nuclear polarisation (photo-CIDNP) techniques has given important insights into changes in hydrophobic clustering dur-

1. Introduction

ing protein folding [42–44]. In a specific application, polarisation is generated in an unfolded or partially folded state but detected in the well-resolved NMR spectrum of the native state [45] yielding site-resolved structural information on the level of hydrophobic amino acids. This has allowed a detailed structural characterisation of the molten globule intermediate of α -lactalbumin [46]. Other studies including relaxation dispersion NMR have even succeeded in the 3D structure determination of folding intermediates [47–49].

Recent work has also concentrated on the improvement of already existing NMR techniques. Optimised rapid mixing devices [50] as well as photo-induced folding initiation [51] result in shorter experimental dead-times enabling measurement of faster folding events with rate constants of up to 8 s^{-1} [52]. Improved NMR pulse schemes may turn out to be useful in making 2D real-time NMR a routine approach to study protein folding processes on a second timescale [53]. In this regard, the application of single-scan multidimensional NMR [54, 55] or Hadamard NMR [56] in combination with improvements in sensitivity [57] might as well be important for future developments.

1.2.2. Unfolded states of proteins

The potential of NMR spectroscopy for providing atomic-level structural information on unfolded proteins was first recognised in the early 1980s when NMR studies on proteins such as RNase A and hen lysozyme showed that non-random structural propensities were persistent in their denatured states [58, 59]. Strong support came from investigations using the N-terminal domain of 434 repressor where NOE data collected under highly denaturing conditions were consistent with the existence of hydrophobic clusters in the unfolded state [60]. These elements of residual structure were soon regarded as important for the initiation of protein refolding. Secondary chemical shifts [61] and $^3J_{\text{HNH}\alpha}$ coupling constants [62–64] were shown in the following years to provide valuable complementary information on residual secondary structural propensities in unfolded proteins. The analysis of ^{15}N R_2 relaxation rates using a model of segmental motion in combination with site-directed mutagenesis proved to be vital in furthering the understanding of structural properties of unfolded states of proteins involving the detection of long-range contacts between local hydrophobic clusters [64–68]. PREs [69–71] and RDCs [70, 72, 73] have both underscored the occurrence of long-range interactions and provided more quantitative measures of conformational distributions in en-

sembles of unfolded proteins. A consistent description of the unfolded state is, however, still lacking since, due to the limited number of experimental data, the calculation of complete conformational distributions is not possible at present [74].

1.3. Thesis: projects and objectives

The first of three chapters presenting and summarising results obtained within the course of this PhD thesis is devoted to the use of RDCs in characterising structural and dynamical properties of unfolded proteins. Within the past seven years, RDCs have been measured for a variety of unfolded proteins as well as for intrinsically unstructured proteins and short peptides. The statistical coil model [75, 76] has proven to be powerful in interpreting these data in terms of conformational distributions. The development of solid theoretical grounds on the basis of polymer physics is nonetheless highly desirable in order to be able to fundamentally understand the structural and dynamical information contained in RDCs. The herein presented work builds on the approach by Annala and co-workers who model the unfolded protein and the alignment medium by a random flight chain and a two-dimensionally infinite plane obstructing unhindered chain motion, respectively [77]. The aims of this project are (i) to further advance the model and (ii) to validate predictions made by the model using available experimental data.

The second project aims at the development of improved time-resolved NMR methods for studying protein folding allowing both measurements at lower protein concentrations using cryoprobe technology and routine detection of folding events on the millisecond timescale. Bovine α -lactalbumin is used as the model protein since its folding has been studied in great detail [13, 78–80]. Further aims are (i) the interpretation of the kinetic data supplemented by complementary data from stopped-flow fluorescence and CD spectroscopy in terms of a folding mechanism and (ii) to gain further insight into the properties of the non-*MG* folding intermediate which has been detected on the Ca^{2+} -induced refolding pathway monitored by real-time photo-CIDNP experiments [44].

The third chapter deals with steady-state photo-CIDNP measurements which shall provide a basis for future real-time photo-CIDNP investigations of the Yb^{3+} -induced refolding of bovine α -lactalbumin. Yb^{3+} has been shown to successfully substitute for Ca^{2+} showing

1. Introduction

both similar refolding characteristics as compared to Ca^{2+} and an increased chemical shift dispersion in 1D ^1H NMR experiments suitable for site-resolved kinetic studies [81]. The aims of this project are as follows. First, the investigation of the basic effects of Yb^{3+} on CIDNP of the three polarisable amino acids, namely tryptophan, tyrosine, and histidine. Second, to find and synthesise a suitable CIDNP dye not interfering with Yb^{3+} and, third, to conduct preliminary photo-CIDNP experiments using bovine α -lactalbumin.

2. Theoretical Framework for the interpretation of NMR residual dipolar couplings of unfolded proteins

2.1. Introduction

2.1.1. The unfolded state and its study by NMR

Proteins comprise the most important constituent of living cells. It is well established that many of them require a well-defined three-dimensional structure to be functional in the cellular context. Proteins which are unstructured, however, play an equally important role. This important role can be inferred from different viewpoints. The unfolded state, for example, displays an unstructured state being encountered during mRNA translation and marking the starting point of protein folding which is essentially the last step in the transformation of genetic information into biologically active protein molecules [1] (for a classification of unstructured states of proteins refer to Fig. 2.1).

Understanding protein folding is of great importance in order to decipher how structural and thus functional integrity are being maintained in living systems. It requires the detailed characterisation of all the states being populated during folding including the unfolded state. In addition, the unfolded state is constantly populated given the small energetic barriers between the folded and unfolded states of a protein and the relatively small free activation energies associated with the structural transition.

2. Theoretical Framework for the interpretation of NMR residual dipolar couplings of unfolded proteins

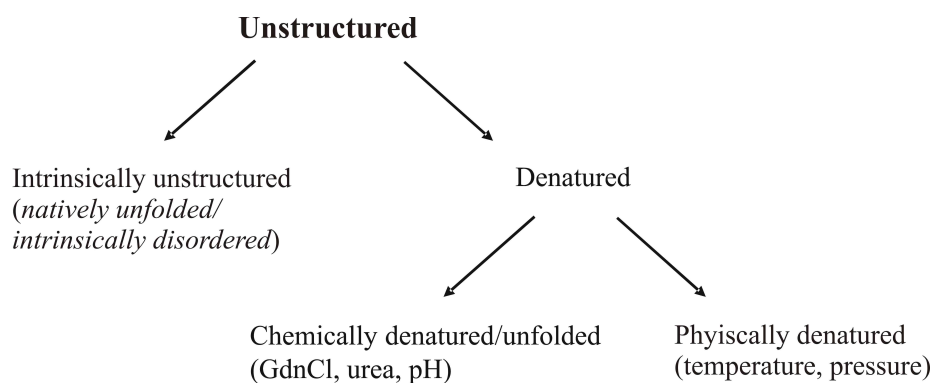


Figure 2.1.: Classification of unstructured states of proteins

Interestingly, a significant part of proteins is functional whilst being partially or fully unstructured comprising up to 30% of the proteins being encoded in eukaryotic genomes [82]. A characteristic feature of intrinsically unstructured proteins is their malleability. This malleability enables intrinsically unstructured proteins to bind to different target molecules as intermolecular interfaces are larger as compared to folded proteins. This class of proteins is therefore ideally suited for the regulation of diverse cellular activities. Their involvement in cell cycle control and transcriptional and translational regulation has indeed been shown. Often, nucleic acids act as binding partners and binding is commonly associated with the induction of folding of at least part of the protein [82, 83].

Unstructured conformations of proteins also occur in the cell in the course of protein trafficking across membranes [84] and under stress conditions [85].

The different roles of unstructured proteins in the cell emphasise the relevance of their detailed study. Since early studies dating back about 40 years were primarily concerned with questions of protein stability and whether the fold of a protein is solely governed by its amino acid sequence [2] and proteins consequently had to be actively denatured to access their unfolded states, the majority of investigations has so far dealt with denatured proteins although the number of studies on unstructured proteins under physiological conditions has significantly increased meanwhile.

The focus in this chapter will be on chemically denatured (unfolded, cf. Fig. 2.1) proteins. Chemical denaturation of proteins is commonly accomplished by the use of GdnCl or urea [86].

Early as well as recent investigations reported *random coil* behaviour of unfolded proteins (the term *random coil* refers to an ensemble of conformations with random distribution of dihedral angles). These conclusions were drawn from intrinsic viscosity [87] and small angle scattering experiments [88]. While these experiments enable accessing global properties of the denatured protein, NMR gives residue- and atom-resolved information on the distribution of conformations [89]. Interestingly, several NMR studies reported hydrophobic clustering [60, 65] and residual secondary [58] and even tertiary structure [65, 72, 90] being present in unfolded proteins. In the meantime, successful attempts to reconcile these seemingly contradictory observations have been published [76, 91]. A detailed and consistent description of the unfolded state is, however, still lacking. Questions as to the relevance of the polyproline II (PP_{II}) conformation in unfolded states of proteins are still a matter of debate [92, 93]. NMR will likely contribute significantly to a better understanding of the unfolded state since a variety of experimental parameters can be exploited to gain insight into the structure as well as the dynamics of these states in an atomic-resolved manner [89]. One especially valuable NMR observable is the direct dipole-dipole interaction between nuclear spins (residual dipolar couplings, RDCs), e.g., between the spins of a ¹⁵N nucleus and a ¹H nucleus. RDCs directly report on the average orientation of internuclear vectors with respect to the magnetic field. As they are correlated to each other via a common frame, they in principle provide unique long-range structural information.

2.1.2. Weak alignment and residual dipolar couplings

The observation of anisotropic interactions in solution NMR spectroscopy is generally not possible since they are averaged out to zero which is due to rapid molecular tumbling in solution. However, it was recognised 45 years ago that these may become observable if some net ordering is imposed onto the molecule under investigation. Saupe and Englert showed that nematic liquids become homogeneously ordered by the application of an external magnetic field leading to their more or less parallel orientation relative to the field [94]. This alignment created an anisotropic environment for benzene being dissolved in *p*-azoxyanisole which was used as the ordering medium. Due to the net ordering of benzene, residual resonance splittings (commonly referred to as *residual dipolar couplings*) became observable in the ¹H NMR spectrum. Although it was clear that partial ordering of molecules in solution would

2. Theoretical Framework for the interpretation of NMR residual dipolar couplings of unfolded proteins

give access to valuable structural information, this methodology could only be applied to small molecules since spectra of larger molecules such as biopolymers became very difficult to analyse due to the large number of spin-spin interactions. In principle, a molecule can be aligned in the presence of a magnetic field solely because of its magnetic susceptibility anisotropy. But it was only with the advent of high-field NMR spectrometers that this effect led to a measurable alignment of molecules. In pioneering studies, Bothner-By and co-workers were able to observe dipolar couplings in a paramagnetic molecule by using at 14.1 T magnet (corresponding to 600MHz ^1H resonance frequency) [95]. The occurrence of dipolar couplings in diamagnetic species, which manifested as a magnetic-field-dependent change in scalar J coupling constants, was subsequently reported [96]. It took, however, more than 10 years until 1995 the Prestegard group published the first report on the observation of dipolar couplings in a field-aligned protein which was paramagnetic cyanometmyoglobin [97]. At the same time, field-alignment of duplex and quadruplex DNA molecules was also demonstrated [98]. It was evident from these studies that the angular dependence of the dipolar coupling would - with respect to a fixed molecular frame - provide long-range structural information which was otherwise inaccessible from measurements of NMR parameters such as scalar J couplings or nuclear Overhauser enhancements (NOEs) which are strictly local. The Prestegard group also emphasised that residual dipolar couplings may allow for the characterisation of protein dynamics [99].

In 1997, Tjandra and Bax showed that, by using bicelles, large disk-shaped bodies consisting of a mixture of dihexanoyl phosphatidylcholine (DHPC) and dimyristoyl phosphatidylcholine (DMPC), proteins could be aligned in a magnetic field such that residual dipolar couplings could accurately be measured while retaining the high resolution of NMR spectra as interactions to only nearby nuclei were detectable (Fig. 2.2) [101]. By using human ubiquitin, the authors also showed that the addition of lipids to the protein sample did not lead to significant changes in chemical shifts and ^{15}N R_2 relaxation rates. Moreover, the axial component of the alignment tensor, which describes the orientation of the molecule relative to the field, almost agreed with that of the nearly axially symmetric rotational diffusion tensor [102]. These results indicated that the alignment of the protein was primarily induced by its shape and not by weak interactions with the ordering medium. The groundbreaking studies by the Prestegard and Bax groups were the basis for many applications of residual dipolar couplings in the

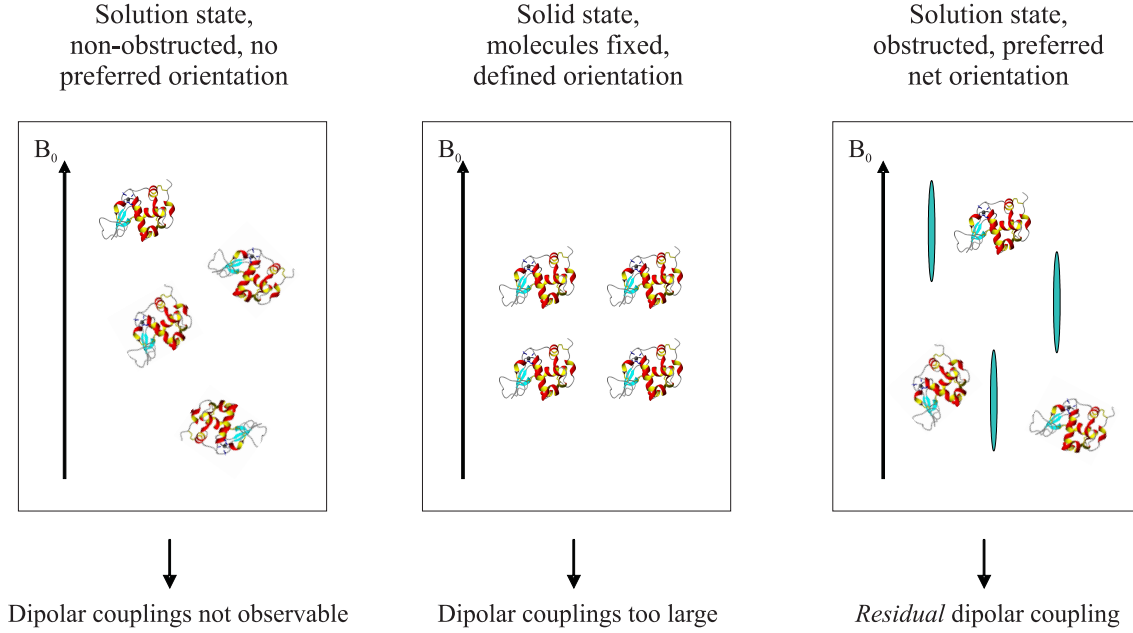


Figure 2.2.: Left, non-obstructed solution state: due to motional averaging, $\langle P_2(\cos \theta) \rangle = 0$ and dipolar couplings are not observable (in case of sufficiently long DNA molecules, however, magnetic susceptibility anisotropy might cause weak alignment). Middle, solid state: $\langle P_2(\cos \theta) \rangle = 1$, dipolar couplings are large and lead to complicated spectra which cannot be analysed. Right, obstructed solution state: $\langle P_2(\cos \theta) \rangle = 10^{-4} - 10^{-3}$, residual dipolar couplings become observable due to weak alignment of the protein relative to the magnetic field. The protein cartoon represents the crystal structure of bovine α -lactalbumin [100].

investigation of biomolecular structure and dynamics. In the following, the basic equations will be derived [103].

In the case of a rigid molecule where all nuclei have fixed positions, the residual dipolar coupling between two interacting nuclei is given by

$$D_{PQ} = D_{PQ}^{max} \langle P_2(\cos \theta) \rangle \quad (2.1)$$

where

$$D_{PQ}^{max} = \frac{\mu_0 \hbar \gamma_P \gamma_Q}{4\pi^2 R_{PQ}^3} \quad (2.2)$$

and

2. Theoretical Framework for the interpretation of NMR residual dipolar couplings of unfolded proteins

$$\langle P_2(\cos \theta) \rangle = \left\langle \frac{3 \cos^2 \theta - 1}{2} \right\rangle \quad (2.3)$$

are the dipolar coupling in the case of full alignment (static dipolar coupling) and the second Legendre polynomial of $\cos \theta$, respectively. The angular brackets refer to the time or ensemble average of the angle θ giving the orientation of the internuclear vector with respect to the magnetic field. This average is equivalent for the isotropic and anisotropic (dilute liquid crystal) cases. Typically, the $\langle P_2(\cos \theta) \rangle$ term is on the order of 10^{-4} to 10^{-3} . The constants in eq. (2.2) are the magnetic permittivity of vacuum, μ_0 , Planck's constant, \hbar , the magnetogyric ratio of spin X, γ_X , and the internuclear distance, R_{PQ} .

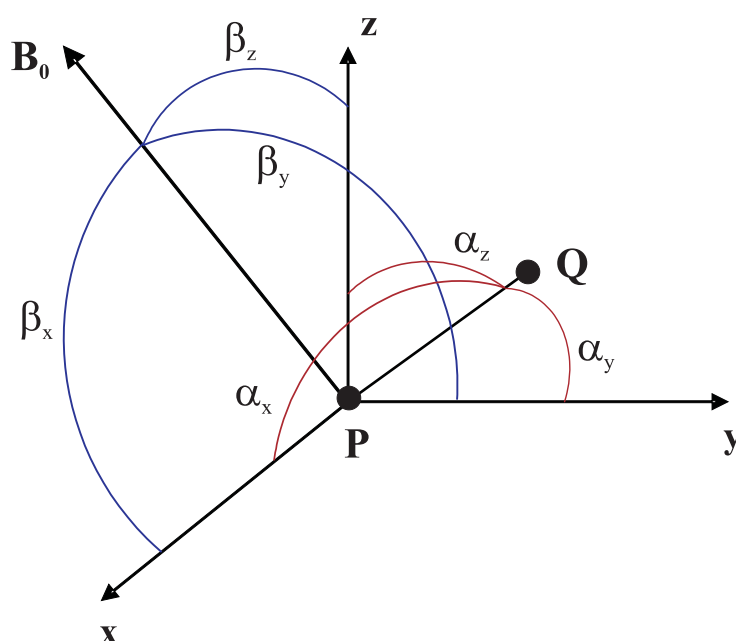


Figure 2.3.: Orientation of an internuclear vector PQ with respect to the magnetic field vector B_0 . Angles $(\alpha_x, \alpha_y, \alpha_z)$ refer to the orientation of the internuclear vector PQ with respect to the axes of the molecule-fixed coordinate system. Angles $(\beta_x, \beta_y, \beta_z)$ refer to the time-dependent angles of the axes of the molecular frame relative to the magnetic field vector B_0 .

Fig. 2.3 shows that the angle θ may be expressed in terms of angles $(\alpha_x, \alpha_y, \alpha_z)$ giving the orientation of the internuclear vector PQ within the molecular frame and angles $(\beta_x, \beta_y, \beta_z)$ describing the time-dependent orientation of the molecular frame relative to the magnetic field. Eq.(2.3) may then be rewritten as

$$\langle P_2(\cos \theta) \rangle = \frac{3}{2} \langle (\cos \beta_x \cos \alpha_x + \cos \beta_y \cos \alpha_y + \cos \beta_z \cos \alpha_z)^2 \rangle - \frac{1}{2} \quad (2.4)$$

Writing $C_i = \cos \beta_i$ and $c_i = \cos \alpha_i$ this may be recast as

$$\langle P_2(\cos \theta) \rangle = \frac{3}{2} [\langle C_x \rangle^2 c_x^2 + \langle C_y \rangle^2 c_y^2 + \langle C_z \rangle^2 c_z^2 + 2 \langle C_x C_y \rangle c_x c_y + 2 \langle C_x C_z \rangle c_x c_z + 2 \langle C_y C_z \rangle c_y c_z] - \frac{1}{2} \quad (2.5)$$

Using $S_{ij} = \frac{3}{2} \langle C_i C_j \rangle - \frac{1}{2} \delta_{ij}$, where δ_{ij} is the Kronecker delta function, one obtains

$$\langle P_2(\cos \theta) \rangle = \sum_{ij=(x,y,z)} S_{ij} \cos \alpha_i \cos \alpha_j \quad (2.6)$$

S is a 3×3 matrix and commonly referred to as the Saupe order matrix. $\langle C_x \rangle^2 + \langle C_y \rangle^2 + \langle C_z \rangle^2 = 1$ and since $\langle C_i C_j \rangle = \langle C_j C_i \rangle$, S is symmetric and therefore contains only 5 independent elements. If the structure of the molecule is known (i.e. angles $(\alpha_x, \alpha_y, \alpha_z)$ are known) S can be solved using singular value decomposition (SVD) [104]. Usually, much more than the minimum number of 5 dipolar couplings are measured so that S is overdetermined. The Saupe order matrix is subsequently diagonalised by an Euler rotation and the principal axis frame is generated. If we write $\langle C_i \rangle^2 = \frac{1}{3} + A_{ii}$, the dipolar coupling may be expressed in polar coordinates $\theta = \alpha_z$; $c_z = \cos \theta$; $c_x = \sin \theta \cos \phi$; $c_y = \sin \theta \sin \phi$:

$$D_{PQ}(\theta, \phi) = \frac{3}{2} D_{PQ}^{max} [\cos^2 \theta A_{zz} + \sin^2 \theta \cos^2 \phi A_{xx} + \sin^2 \theta \sin^2 \phi A_{yy}] \quad (2.7)$$

where A_{xx} , A_{yy} and A_{zz} define the alignment tensor A . By definition, $|A_{zz}| > |A_{yy}| > |A_{xx}|$ and $A_{yy} + A_{xx} = -A_{zz}$. Using the trigonometric relationships $2 \sin^2 \phi = 1 - \cos 2\phi$ and $2 \cos^2 \phi = 1 + \cos 2\phi$, eq. (2.7) may be recast as

$$D_{PQ}(\theta, \phi) = \frac{3}{2} D_{PQ}^{max} [P_2(\cos \theta) A_{zz} + \frac{1}{2} \sin^2 \theta \cos 2\phi (A_{xx} - A_{yy})] \quad (2.8)$$

If we define an axial component of the alignment tensor $A_a = \frac{3}{2} A_{zz}$, and a rhombic component $A_r = (A_{xx} - A_{yy})$ we obtain

$$D_{PQ}(\theta, \phi) = D_{PQ}^{max} [P_2(\cos \theta) A_a + \frac{3}{4} A_r \sin^2 \theta \cos 2\phi] \quad (2.9)$$

2. Theoretical Framework for the interpretation of NMR residual dipolar couplings of unfolded proteins

The alignment tensor A describes the orientation of the molecule relative to the field. The maximum value for A_a is 1 which corresponds to full alignment of the z axis of the tensor with the static field. In typical experiments, A_a is on the order of 10^{-3} . Eq. (2.9) is also sometimes rewritten as

$$D_{PQ}(\theta, \phi) = D_{PQ}^a \left[(3 \cos^2 \theta - 1) + \frac{3}{2} R \sin^2 \theta \cos 2\phi \right] \quad (2.10)$$

where D_{PQ}^a and $R = \frac{A_r}{A_a}$ are the magnitude of residual dipolar coupling tensor and the rhombicity, respectively.

RDCs and structure

From eqs. (2.9, 2.10) it is evident that after determination of the alignment tensor A , a measured dipolar coupling will not be defined by a unique (θ, ϕ) pair. This orientational degeneracy consequently limits the use of RDCs in biomolecular structure determination. Many applications, however, have underlined the importance of RDCs as an additional source of structural information. Early studies reported on the use of RDCs in validating structures which had been determined using conventional x-ray crystallographic or NMR spectroscopic techniques. The basic concept is to predict dipolar couplings from the structure and to compare them to the measured couplings which, of course, have to be excluded from the structure calculation process. For the prediction of RDCs the alignment tensor has to be determined. This can be done by using either the measured couplings or by simulation [105]. Once the alignment tensor has been obtained and dipolar couplings have been calculated the Q value can be computed which is a measure of the agreement of the measured couplings with the initial structure [106, 107]. Lower Q values indicate better agreement. Apparently, the unique long-range structural information content in RDCs makes them ideally suited to refine initial structures. This is done by driving an existing structure into a conformation which is in agreement with the measured RDCs and accounted for in the structure calculation process by an additional dipolar energy term [108, 109].

The orientational degeneracy inferred from eq.(2.10) seems to make *de novo* protein structure determination from only RDCs impossible. However, the determination of mutually independent alignment tensors, generally accomplished by measuring RDCs in different alignment media, can at least partially raise the inherent degeneracy in a single RDC data set [110]. In

fact, Blackledge and co-workers have shown that the backbone structure of ubiquitin can accurately be calculated using two different RDC data sets [111]. First, the degeneracy in the orientation of the peptide planes could be reduced to the correct orientation and its mirror image using four different types of peptide-plane RDCs. Second, the remaining ambiguity could be raised by the tetrahedral geometry requirements at the junctions connecting subsequent peptide planes using two types of dipolar couplings involving the C^α atom (for a chiral motif the degeneracy is completely raised if two independent tensors are available). It is clear that this approach requires many types of couplings to be measured throughout the protein backbone. In case that data sets are less complete, RDCs may still be sufficient to create structural models. In one example, this is done in combination with database folds. A primary-sequence-based homology model can be validated using measured RDCs as soon as the assignment of the backbone resonances is complete. This has been shown for the FMN-binding domain of sulfite reductase flavoprotein [112]. Experimental dipolar couplings may then additionally be used to refine the initial structural model. Another approach termed *molecular fragment replacement* was introduced by Delaglio *et al.*. Here, experimental dipolar couplings of a 7-residue fragment are compared to fragments of equal length in a database with a given structural resolution. By shifting a 7-residue window along the protein sequence one by one the database is searched $N-6$ times (N is the number of residues in the sequence) for the fragments best-fitting the measured RDCs. This allows to finally construct a model of the protein [113].

Moreover, RDCs have successfully been used in the determination of the quaternary geometry of protein-protein complexes an early example of which was the structure of the phosphoryl transfer complex between the N-terminal domain of enzyme I and HPr [114]. The long-range orientational character of RDCs is especially helpful in such cases as NOE data are generally sparse for intermolecular interfaces and thus insufficient to deduce the relative orientation of complex partners. RDCs have therefore also been used in the determination of structures of proteins and their small-molecule binding partners. Early examples include the FKBP/FK506 complex [115] and the interaction of α -methyl mannose with mannose-binding protein [116].

RDCs and dynamics

Shortly after the first demonstration of residual dipolar couplings in a field-aligned protein, cyanometmyoglobin [97], Tolman *et al.* in a follow-up study showed that dipolar couplings

2. Theoretical Framework for the interpretation of NMR residual dipolar couplings of unfolded proteins

may also allow for the characterisation of protein dynamics on timescales which had so far been inaccessible to study [99]. This was again demonstrated for the paramagnetic protein cyanometmyoglobin. Based on largely systematic differences between RDC measurements and available x-ray [117] and solution [118] structures it was concluded that slow collective motions of helices within the protein had to exist. It was also Prestegard and co-workers showing that not only the relative orientation of domains in a protein but also differential domain dynamics may be inferred from RDC data. In a study involving the protein barley lectin, alignment tensors were determined for each domain revealing differences in the axial and rhombic components (A_a and A_r , respectively). These differences were interpreted in terms of differential reorientational dynamics of the two domains [119].

The same group then introduced the terminology of the generalised degree of order (GDO) as an indicator of internal motion. After the order matrices for each ring in a trimannoside had been solved, order parameters were determined showing that motion for the linkage between rings I and III is limited while there is significant motion for the linkage between rings II and III [120]. Application of the GDO approach to human ubiquitin resulted in good agreement with the x-ray structure [121] but it was additionally noted by the authors that the data were also consistent with a dynamic model for the protein [122].

Presumable dynamics in ubiquitin slower than the rotational correlation time were then the subject of a couple of further investigations. It was the Griesinger group developing a theoretical basis for a model-free interpretation of backbone amide ^1H , ^{15}N RDCs in terms of local internal dynamics. RDCs were collected for ubiquitin from 11 different alignment media and order parameters were subsequently determined for all residues of the protein. These were on average smaller than the average Lipari-Szabo order parameter commonly extracted from spin relaxation measurements [123, 124] and accordingly interpreted as an additional disorder present in NH vector orientations [125, 126]. At about the same time, Tolman suggested a different theoretical approach for the interpretation of dipolar couplings in terms of internal dynamics. As in the approach by Griesinger and co-workers, RDCs from multiple alignment media were needed for the determination of order parameters. Again, differences from spin-relaxation-derived order parameters were observed indicating larger amplitudes of internal motion [127].

The studies by the Griesinger group [125, 126] were subsequently taken on by Clore and

Schwieters who reexamined the dipolar coupling data [128]. The result was that all dipolar coupling data could be accounted for by a single structure representation. Interestingly, improvement in dipolar R factors was observed when an ensemble of two conformations was considered. It was concluded that the dipolar coupling data could be accounted for by motions of much smaller amplitude than suggested in the study by Peti *et al.* [126]. The authors also noted, however, that large-scale motions were not generally excluded by the data [128].

A more recent approach to the extraction of local dynamic information from RDC data involves the investigation of the influence of small-amplitude anisotropic peptide plane motions on RDC averaging [129, 130]. These motions are described by a Gaussian axial fluctuation (GAF) model [131]. The authors are able to provide an improved description of the alignment tensor when an average peptide plane reorientation amplitude (σ_{av}) is included into alignment tensor analysis. Secondary structural elements of investigated proteins exhibited average motional amplitudes comparable to those identified from spin relaxation studies. It was stated that this result does not exclude the possibility of significant local fluctuations from the average behaviour. The approach of Bernadó and Blackledge [129, 130] is advantageous over the model-free approaches in that it allows for the detection of internal protein dynamics using fewer alignment media.

Summing up, it has to be stated, however, that a solid theoretical basis for reliable interpretation of RDC data in dynamical terms is still being awaited.

2.1.3. Prediction and simulation of residual dipolar couplings under weakly aligning conditions

Folded proteins

At the time when the measurement of RDCs of biomolecular systems was introduced it was immediately recognised that molecular alignment in nearly neutral dilute liquid crystal media such as phospholipid bicelles was primarily determined by the solute's shape [102]. This was taken on by Zweckstetter and Bax when they introduced the program PALES (Prediction of ALignmEnt from Structure) [105]. The basic principle in this approach is to exclude the fraction of molecules, generally assumed to be randomly oriented and uniformly distributed in the sample, which clashes with the phospholipid bilayers. In practice, the molecule's orientational

2. Theoretical Framework for the interpretation of NMR residual dipolar couplings of unfolded proteins

distribution is computationally evaluated excluding every position and orientation causing obstruction of the molecule by an infinite wall which represents the alignment medium. The individual alignment matrices of each non-obstructed position and orientation are then averaged yielding the alignment tensor A . Thus, the (X-ray or NMR) structure is the sole input needed. By using this approach, the authors demonstrated good agreement between predicted and measured dipolar couplings as long as uncharged alignment media were considered.

In contrast to the simulation approach by Zweckstetter and Bax, García de la Torre and co-workers developed a model for the prediction of RDCs [132]. In their approach, the protein and the aligning medium were approximated by an axially symmetric ellipsoid and a planar wall, respectively. An orientational probability distribution function was derived for the ellipsoid in the presence of the planar wall from geometrical considerations which finally yielded an analytical expression allowing for the calculation of dipolar couplings. The expression for the case of bicelles with their normals oriented perpendicular to the direction of the magnetic field is given by

$$D_{PQ} = -D_{PQ}^{max} S_{LS} S_{corr} \frac{3 \cos^2 \alpha_{PQ} - 1}{2} \frac{[3I_2(p, 1) - I_0(p, 1)]}{4[(\frac{\delta}{2b})(\frac{1}{v} - 1) - I_0(p, 1)]} \quad (2.11)$$

where S_{LS} is the Lipari-Szabo generalised order parameter [123, 124], S_{corr} accounts for incomplete bicelle alignment relative to the magnetic field (usually S_{corr} is about 0.8 [105]), α_{PQ} is the angle of the internuclear vector with the main axis of the ellipsoidal particle, δ is the bicelle thickness, and v is the volume fraction of the bicelles. $I_n(p, 1)$ is given by

$$I_n(p, 1) = \int_0^1 c^n \sqrt{\frac{a^2}{b^2} + (1 - \frac{a^2}{b^2})c^2} dc \quad (2.12)$$

where a and b are the semiaxes of the axially symmetric ellipsoid with $b > a$; c is the shorthand notation for $\cos \theta$.

In order to compute RDCs for a real protein, its shape must be represented by a respective ellipsoid. For this purpose, the authors make use of the radius of gyration tensor G which describes the second moments of position of the atoms in the molecule. G is calculated from the structure and the eigenvalues are then used to determine a , b and α_{PQ} . Dipolar couplings are finally calculated using eq.(2.11). Therefore, the inputs needed are δ , v and the structure.

In another approach, the fact that the asymmetries in the shape governing molecular align-

ment are already contained in the moments of inertia tensor I is exploited for the calculation of dipolar couplings [133]. Azurmendi and Bush first compute I from the structure (I differs from G only in that the atom positions are weighted by their mass). Assuming that the eigenvectors of A and I are parallel and that the eigenvalues of A are proportional to the asymmetries of the molecule in the planes defined by pairs of eigenvectors of I , A can be built up using the following relationship:

$$A_{ii} = I_{ii} - \frac{I_{jj} + I_{kk}}{2} (i, j, k = 1, 2, 3) \quad (2.13)$$

with I_{ii} , I_{jj} , and I_{kk} being the eigenvalues of the inertial tensor. If the structure is available, the alignment tensor will be completely defined except for a scale factor which depends on experimental conditions meaning that only one variable remains for prediction of dipolar couplings.

Almond and Axelsen again used the radius of gyration tensor G as a suitable descriptor of molecular shape in their study [134]. They first constructed ellipsoids characterised by the square roots of the eigenvalues of G , $(\lambda_1, \lambda_2, \lambda_3)$. Order matrices were then calculated numerically using an obstructing planar wall [105]. Positions and orientations leading to intersection of the surface of the ellipsoid with the wall were excluded. The diagonal components of the order matrix (S_{xx}, S_{yy}, S_{zz}) were analysed for different distances from the wall and different ellipsoidal shapes characterised by the dimensionless quantity

$$\sigma = \frac{\lambda_2 - \lambda_3}{\lambda_1 - \lambda_3} \quad (2.14)$$

with $\lambda_1 > \lambda_2 > \lambda_3$. This finally yielded the following simple expression for the calculation of the alignment tensor components:

$$S_{ii} \propto \lambda_i - \frac{1}{2}(\lambda_j + \lambda_k) \quad (2.15)$$

This expression is subject to a scaling depending on the concentration of the aligning medium. It is evident that dipolar couplings for a real protein may therefore be computed in an easy and very fast fashion. After calculation of G from the structure of the protein, $(\lambda_1, \lambda_2, \lambda_3)$ may be used to determine the diagonal components of A via eq.(2.15).

Comparison with the studies by Fernandes *et al.* [132] and Azurmendi and Bush [133] shows that the approach by Almond and Axelsen gives the best agreement with results from simulation using PALES [105]. Problems in the earlier studies arise as the eigenvalues of

2. Theoretical Framework for the interpretation of NMR residual dipolar couplings of unfolded proteins

G [132] and I [133] enter directly in the respective equations. This leads to non-linear dependencies whereas an almost linear dependence for the diagonal components of A is expected from simulation [105] and from the approach by Almond and Axelsen [134].

Meanwhile, a few studies addressing prediction of macromolecular alignment in charged liquid crystal media have appeared. In two similar but independent studies it was shown that in the case of dilute suspensions of filamentous phages such as Pf1 [135] and fd [136] weak alignment is dominated by short-range steric and long-range electrostatic interactions meaning that residual dipolar couplings may be predicted from the molecular shape and charge distribution [137, 138]. The report of Zweckstetter *et al.* [138] essentially builds on the earlier introduced steric obstruction model [105] and has been further extended in the meantime to low pH conditions and surfactant liquid crystalline systems [139–141]. The approach of Wu *et al.* in turn deals with the alignment of nucleic acids in Pf1 medium and, interestingly, shows that for usual salt strengths and nucleic acid sizes weak alignment is of completely electrostatic nature [142].

Unfolded proteins

Recently, RDCs have been measured on a variety of unstructured proteins and small peptides [70, 72, 73, 90, 143–160]. The interpretation of the RDC data is non-trivial due to the highly heterogeneous character of the ensemble of conformations. The earliest reports provided evidence for a native-like topology within the unfolded ensemble [72, 143, 144]. In a more recent study, however, it was argued in favour of simple local conformational propensities [149]. Data interpretation would clearly be aided if RDCs could be predicted for such systems. Models for steric [105, 132–134] and/or electrostatic [137–139] alignment have been successfully used to predict dipolar couplings in cases where the biomolecule is characterised by a single predominating conformation. Given the heterogeneity in ensembles of unstructured conformations alternative approaches are necessary which explicitly take this dynamic behaviour into account.

Addressing this problem, Sosnick and co-workers have proposed an approach to predicting RDCs of unfolded proteins using the concept of the statistical coil [62–64]. A statistical coil model of the unfolded state is built based on (ϕ, ψ) backbone dihedral angle occurrences in regions outside of, and not adjacent to, helices, sheets, and turns in high-resolution x-ray

structures [76]. Probabilities of the form $P(a_i, b_i)$ are accordingly defined where a_i and b_i denote the amino acid identity and the Ramachandran basin ($\alpha_R, \alpha_L, \beta, \gamma$, and polyproline II), respectively. The corresponding energy is then defined as

$$U(a_i, b_i) = -RT \ln P(a_i, b_i) \quad (2.16)$$

A statistical potential can be generated which, upon inclusion of nearest neighbour effects [161, 162], adopts the following form:

$$U(\text{total}) = \sum_i U(a_i, b_i) + \sum_i \delta U(a_i, b_i, a_{i+1}, b_{i+1}) \quad (2.17)$$

where the second term corresponds to the nearest neighbour contribution. An ensemble of conformations is constructed according to the statistical potential by performing a Monte Carlo [163, 164] simulation. Steric overlap between residues is then removed by an excluded volume energy function. Each member of the conformational ensemble is individually aligned and the alignment tensor is subsequently calculated *via* the moments of inertia tensor I . RDCs are finally averaged over the whole ensemble of conformations. The authors show reasonable agreement with experiment which is significantly improved when correlations between adjacent residues are accounted for underscoring the importance of nearest neighbor effects. Importantly, agreement with experiment requires a preponderance of extended backbone conformers in the coil library.

Almost at the same time a conceptually similar approach was introduced by Blackledge and co-workers [75]. Here, the conformational ensemble is generated using the *flexible – meccano* algorithm by random extraction of amino acid-specific (ϕ, ψ) combinations from a database of loop structures, found in high-resolution x-ray structures, with all residues in helices and sheets being removed. Residues preceding proline are considered as an additional amino acid type because of restricted conformational sampling. Residue-specific volume exclusion is used to avoid clashes between different residues of the same conformer. Each member of the ensemble is again individually aligned and A is being calculated by virtue of its similarity to the radius of gyration tensor G [134]. Calculated RDCs are scaled with respect to experimental RDCs. The approach by Bernado *et al.* [75] seems to be advantageous over the approach by Jha *et al.* [76] for two reasons. First, the asymmetry in the shape of the conformations is accounted for by making use of G in calculating A , and second, the relative

2. Theoretical Framework for the interpretation of NMR residual dipolar couplings of unfolded proteins

extent of alignment can be predicted for each conformer. The latter fact is important since conformations with higher aspect ratios make the largest contribution to the RDC signal.

In the meantime, further developments to the prediction of RDCs of unstructured proteins have been accomplished. By applying distance constraints between specific C^β atoms it has been possible to reproduce experimental RDCs of the intrinsically unstructured protein α -synuclein showing that contacts between residues farther apart in the sequence (here, between the N- and C-termini) may have non-negligible effects on sequential RDC profiles [165]. Moreover, by a combination of the *flexible – meccano* algorithm [75] and simulation of electrostatic alignment [138] RDCs have been predicted for α -synuclein dissolved in charged ordering media [166].

While the statistical-coil-based approaches are powerful in reproducing experimentally measured dipolar couplings, they do not provide a basis for analysing fundamental RDC dependencies such as on chain length or liquid crystal concentration. An alternative therefore constitutes the prediction of RDCs of unstructured proteins from fundamental polymer theory. The simplest model of an unstructured polypeptide chain in this context is the random flight chain. The random flight chain is a chain of infinitely thin segments which are equal in length and oriented randomly with respect to each other (Fig. 2.4).

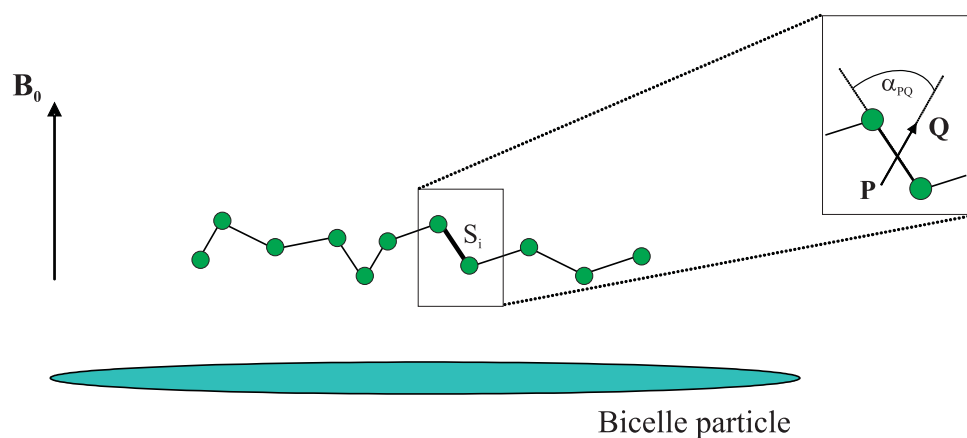


Figure 2.4.: Random flight chain model of an unfolded polypeptide chain. The protein backbone is simply represented by segments of equal length connecting subsequent C^α atoms. Weak alignment is conferred onto the protein by a bicelle particle which is oriented perpendicular to the magnetic field. The zoom in shows the orientation of an internuclear vector PQ relative to the segment S_i .

This model has first been used by Annala and co-workers to provide a theoretical basis for the interpretation of dipolar couplings exhibited by unfolded proteins [77]. The equation to be solved in order to derive dipolar couplings is

$$D_{PQ} = D_{PQ}^{max} \left\langle \frac{3 \cos^2 \alpha_{PQ} - 1}{2} \right\rangle \langle P_2(\cos \theta) \rangle \quad (2.18)$$

$\langle P_2(\cos \theta) \rangle$ has to be evaluated for each chain segment and α_{PQ} is the angle of the internuclear vector relative to the main axis of the axially symmetric segment. The barrier oriented perpendicular to B_0 in Fig. 2.4 is the source of the non-vanishing $\langle P_2(\cos \theta) \rangle$. For the calculation of $\langle P_2(\cos \theta) \rangle$ of a segment S_i , the spatial distribution functions for the half-chains preceding and succeeding S_i have to be considered (Fig. 2.4). The distributions are binomial in nature but may safely be approximated by exponential distributions. The distribution function W for a half-chain is then given by

$$W(z, h_z, c, n) = \sqrt{\frac{2}{\pi n}} \left[\exp\left(-\frac{(z + 0.5c)^2}{2n}\right) - \exp\left(-\frac{(2h_z - z + 0.5c)^2}{2n}\right) \right] \quad (2.19)$$

where the first term corresponds to the free distribution and the second term denotes the modification needed due to the obstruction in the z direction (Fig. 2.4) [167]. c is $\cos \theta$ and h_z and n are the distance of S_i from the obstructing barrier and the number of segments of the respective half-chain, respectively.

Louhivuori *et al.* showed reasonable agreement with data from simulation using PALES [105] and extended their approach to valence chains to account for steric hindrance between chain residues later on [168].

However, the formalism presented in the initial study [77] contains three shortcomings. The most serious one is to add the (weighed) probabilities of statistically independent events rather than multiplying them (see eq. (1) in [77]). Secondly, only one obstacle has been accounted for, while, according to the model, the random flight chain is confined between two obstacles. And finally, the one-dimensional random walk formalism has been employed for describing the probabilities of the possible chain conformations while, in general, the three-dimensional formalism must be used. These shortcomings will in the following be addressed in the *Results and Discussion* section. A mathematical framework will be elaborated allowing for the analysis of fundamental dependencies of RDCs of random flight chains on chain length, chain locus, and obstacle density. Predictions will be compared with experiment.

2.2. Results and Discussion

2.2.1. Theoretical Framework

General approach

A random flight chain is constructed as a set of line segments connecting subsequent steps of a three-dimensional random walk. The unfolded polypeptide is, therefore, represented in the model by a sequence of infinitely thin rods of equal length attached one to another at the tips. Each rod/segment represents a structural subunit of the polypeptide. The segments are randomly oriented, there is no interaction, including steric hindrance, with the other segments.

The dipolar coupling between nuclei P and Q depends on the angle β between the internuclear vector and the magnetic field [169]:

$$D_{PQ} = \frac{\mu_0 \hbar \gamma_P \gamma_Q}{4\pi^2 R_{PQ}^3} \left\langle \frac{3 \cos^2 \beta - 1}{2} \right\rangle. \quad (2.20)$$

Here, γ_P and γ_Q are the magnetogyric ratios of nuclei P and Q, and R_{PQ} is the internuclear distance.

In the case of axially symmetric segments, the dipolar couplings can be expressed via the angle θ between the axis of the segment to which the nuclei belong and the magnetic field (see Fig. 2.5) :

$$D_{PQ} = \frac{\mu_0 \hbar \gamma_P \gamma_Q}{4\pi^2 R_{PQ}^3} \frac{3 \cos^2 \alpha_{PQ} - 1}{2} \left\langle \frac{3 \cos^2 \theta - 1}{2} \right\rangle, \quad (2.21)$$

where α_{PQ} is the angle between the internuclear vector and the axis of the segment.

Since the angle $\theta = \theta(t)$ changes with time due to fluctuations of the chain, one has to average the coupling over the time of the measurements:

$$\langle \cos^2 \theta \rangle = \frac{1}{\tau} \int_0^\tau \cos^2 \theta \, dt. \quad (2.22)$$

For residual dipolar couplings, the experimental parameter is averaged over a much longer time than the characteristic times of thermal motion of segments of the polypeptide chain. If no special aligning conditions are imposed (alignment due to the presence of the magnetic field is negligible in the case of diamagnetic proteins), the chain will freely fluctuate in solution

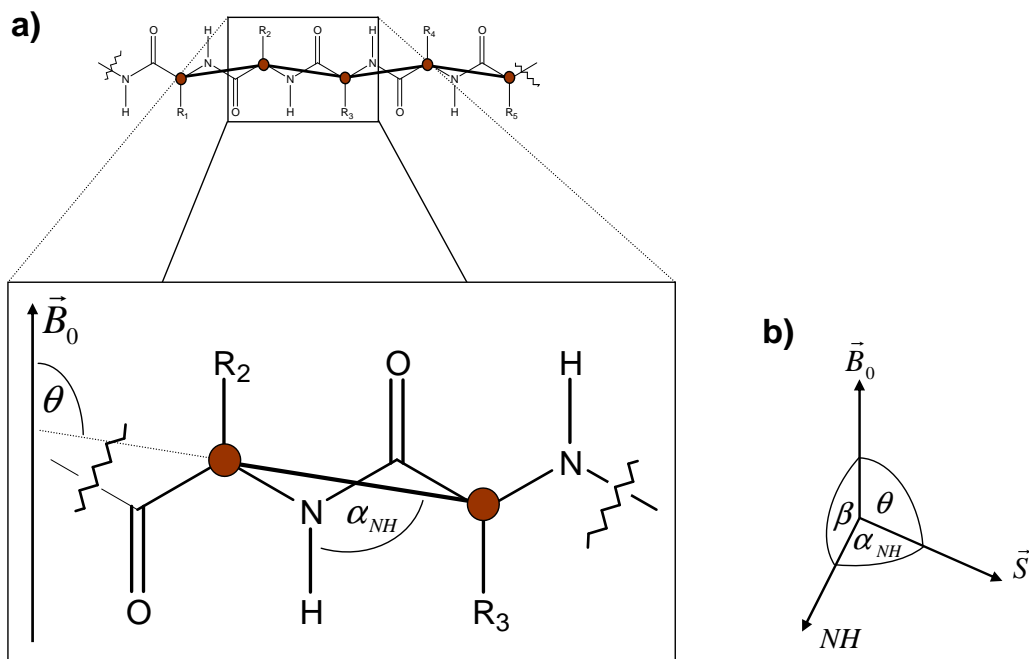


Figure 2.5.: a) Representation of an unfolded polypeptide chain by a random flight chain. The segments of the random flight chain are being built by connecting subsequent C^α atoms. Additional information about the molecular structure of the polypeptide chain is being neglected. H^α atoms are not shown. Zoom in: A given segment is at angle θ with respect to the magnetic field. The vector connecting nuclei N and H belonging to the segment is at angle α_{NH} (corresponding to α_{PQ} in eq. (2.21)) with respect to the unit vector along the segment axis. The R_n indicate residue-specific side-chains. b) The angles between the magnetic field vector B_0 , the unit vector S along the segment of interest, and the internuclear vector connecting nuclei N and H.

2. Theoretical Framework for the interpretation of NMR residual dipolar couplings of unfolded proteins

and the segment S_i will spend equal times at each possible value of θ and averaging over a long time τ will result in zero coupling, $\langle \cos^2 \theta \rangle = 1/3$, $D_{PQ} = 0$. If, however, segments of the chain have a preferred orientation in space, dipolar couplings become an observable quantity and can serve as an important source of information about the structural properties of the polypeptide. The optimal experimental conditions correspond to a very small degree of alignment, $\langle \cos^2 \theta \rangle - 1/3 \sim 10^{-3} - 10^{-4}$. The random flight chain model is suitable for the NMR experiments in which alignment is caused by a spatial obstruction of chain motions due to various kinds of oriented media (bicelles in an external magnetic field, strained gels, etc.).

The statistical mechanics postulates [170] that averaging over an infinite time is equivalent to averaging over the phase space of the system with a probability distribution function describing how frequently the system visits each point of its phase space. Since τ is much longer than the characteristic times of fluctuations and, additionally, since large ensembles of the polypeptide chains contribute to the measured signal, one can replace the averaging over time in eq. (2.21) by averaging over the configurational space of the polypeptide.

In this section we show how the results of the theory of random walks allow one to construct a probability distribution function which depends on the angle θ only. To obtain such a function $P(\theta)$ one has to integrate over all other variables characterising the configurational space of a polypeptide. The distributions $P(\theta)$ are different for each chain segment i , and they also depend on the length of the chain N , $P(\theta) = P_i^{(N)}(\theta)$; these indices will be omitted as their values are always clear from the context. In the next section this general recipe of finding $P(\theta)$ will be applied in calculating the RDCs for a particular experimental setup.

The desired distribution $P(\theta)$ gives the probability $d\mathcal{P}$ of finding a given segment S_i of the chain oriented at an angle in an infinitesimally small vicinity of θ , $d\mathcal{P} = P(\theta)d\theta$. Then

$$\langle \cos^2 \theta \rangle = \int_0^\pi P(\theta) \cos^2 \theta \sin \theta d\theta = \int_{-1}^1 P(\theta) \cos^2 \theta d(\cos \theta). \quad (2.23)$$

The sine of θ in eq. (2.23) appears as a remnant of the integration over the complete phase space of the chain in which the integration over the polar angle $d\Omega = \sin \theta d\theta d\phi$ is performed. The probability distribution must be properly normalized,

$$\int_{-1}^1 P(\theta) d(\cos \theta) = 1. \quad (2.24)$$

If no alignment is imposed, all values of θ are equally probable, $P(\theta) = 1/2$, $\langle \cos^2 \theta \rangle = 1/3$, and $D_{PQ} = 0$.

The function $P(\theta)$ can generally be thought of as the ratio of the number of polypeptide conformations in which the given segment S_i is at angle θ with respect to the magnetic field to the total number of polypeptide conformations (cf. generic definition of probability, the number of successful events divided by the total number of events). In continuous space, it is impossible to explicitly enumerate all the conformations of the polypeptide chain. Therefore, one has to operate with continuous distributions characterising the density of conformations in a space of continuous variables.

In order to find the density of conformations $P(\theta)$ the results of the theory of random walks are used which connect the starting point of a walk and the number of random steps with a probability to arrive at some ending point. The segment S_i , for which $P(\theta)$ is sought, divides the chain into two parts (before and after S_i) of lengths $N_1 = i - 1$ and $N_2 = N - i$. Each subchain corresponds to a random walk characterised by the starting point (defined by the position and orientation of the segment S_i), number of steps (N_1 and N_2 , respectively), and by the ending point (ending points of the chain). Therefore, the function $P(\theta)$ can be expressed *via* the random walk distribution functions integrated over the positions of the end points of the chain and over the position and azimuthal orientation of the given segment S_i . In this probability distribution all other variables of the configurational space of the polypeptide chain (i.e., the coordinates of the rest of the segments) have already been integrated over.

A function $W(\mathbf{r}_0, \mathbf{r}_1, \mathbf{r}_2, \mathbf{s}_i)$ proportional to the density of polypeptide conformations is introduced such that the given segment S_i is oriented along the unit vector $\mathbf{s}_i = \{\cos \phi \sin \theta, \sin \phi \sin \theta, \cos \theta\}$ (the magnetic field is taken to be along the z -axis, θ and ϕ are the polar angles) while the ends of the polypeptide chain are located at the points \mathbf{r}_1 and \mathbf{r}_2 . Vector \mathbf{r}_0 points to the middle of S_i (see Fig. 2.6). The expression for $P(\theta)$ can then be written as

$$P(\theta) = \frac{\mathcal{N}(\theta)}{\int_{-1}^1 \mathcal{N}(\theta) d(\cos \theta)}, \quad (2.25)$$

where

$$\mathcal{N}(\theta) = \int d\mathbf{r}_0 \int_0^{2\pi} d\phi \int d\mathbf{r}_1 \int d\mathbf{r}_2 W(\mathbf{r}_0, \mathbf{r}_1, \mathbf{r}_2, \mathbf{s}_i). \quad (2.26)$$

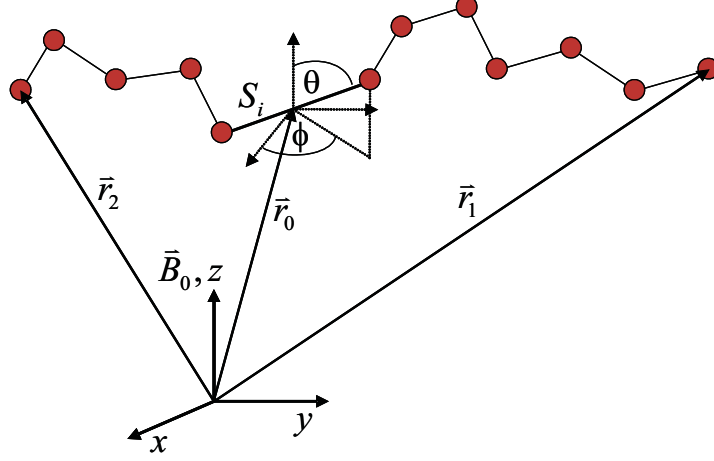


Figure 2.6.: Definition of the angle ϕ and vectors \mathbf{r}_0 , \mathbf{r}_1 , \mathbf{r}_2 . Segment S_i is enlarged for illustrative purposes.

Roughly speaking, eq. (2.26) gives the number of conformations of the chain for a given position \mathbf{r}_0 and orientation \mathbf{s}_i of the segment S_i by integrating over the number of conformations with positions of the end points \mathbf{r}_1 and \mathbf{r}_2 . Then, an integration over the position (\mathbf{r}_0) and orientation (ϕ) of S_i is performed. Integration over \mathbf{r}_0 , \mathbf{r}_1 , and \mathbf{r}_2 is carried out over the whole accessible space. This results in the total number of conformations in which S_i is at angle θ with respect to the z -axis. The total number of conformations is obtained by integrating $\mathcal{N}(\theta)$ over the interval $(0, \pi)$.

The densities of conformations for the subchains before and after the segment S_i are given by a probability distribution of a random walk with the corresponding length, starting and ending points. This distribution is denoted by $G(\mathbf{a}, \mathbf{r})$. For the first subchain the ending point of the random walk is \mathbf{r}_1 , the starting point is one of the tips of segment S_i , $\mathbf{a}_1 = \mathbf{r}_0 + \mathbf{s}_i/2$. The segment length is taken to be 1, defining therefore the scale for all length-related quantities. For the second subchain, the random walk ends at \mathbf{r}_2 and it starts from the other tip of the segment S_i , $\mathbf{a}_2 = \mathbf{r}_0 - \mathbf{s}_i/2$. The subchains are independent, hence the probability distributions are being multiplied, that is, the number of conformations of the whole chain is a product of numbers of conformations of the subchains:

$$W(\mathbf{r}_0, \mathbf{r}_1, \mathbf{r}_2, \mathbf{s}_i) = G(\mathbf{r}_0 + \mathbf{s}_i/2, \mathbf{r}_1)G(\mathbf{r}_0 - \mathbf{s}_i/2, \mathbf{r}_2). \quad (2.27)$$

The random walk probability distribution $G(\mathbf{a}, \mathbf{r})$ which is to be substituted into eq. (2.27) should be determined based on the particular geometry of obstructing media, but it can also accommodate further improvements in the model (self-avoiding random walk, non-Markovian processes, etc.).

In the simplest case – no alignment, no interactions between the segments of the chain – $G(\mathbf{a}, \mathbf{r})$ was first obtained by Lord Rayleigh [171]. It reads as [167]

$$G(\mathbf{a}, \mathbf{r}) = \frac{1}{(2\pi n/3)^{3/2}} \exp\left(-\frac{3(\mathbf{r}-\mathbf{a})^2}{2n}\right). \quad (2.28)$$

Here, n is the number of the steps in the random walk. Substitution of eq. (2.28) into eqs. (2.27), (2.26) and (2.25) gives a constant (1/2), which does not depend on θ , as it is expected.

It is noted that, in principle, the binominal distributions should be used in eq. (2.28) instead of the exponential distributions. In the limit of the infinitely long chain, $n \rightarrow \infty$, the binominal distributions coincide with the exponential distributions. Therefore, strictly speaking, the analysis based on the use of the exponential functions is applicable for sufficiently long subchains, i.e. for segments which are far from the ends of a long chain. However, already for $n > 2$, the exponential distribution is a rather good approximation.

Alignment media in the random flight chain model are represented by a set of barriers obstructing the random walk, and the function $G(\mathbf{a}, \mathbf{r})$ has to be modified correspondingly. The role of the obstructing media is to exclude all the conformations of the chain from the configurational space which cross any of the barriers. This is achieved by modeling the obstructing media as a set of absorbing barriers. Indeed, the probability distribution of a random walk in the presence of absorbing barriers disregards the pathways connecting the starting and the ending points, which cross any of the barriers. Representing the obstructing media as a set of reflecting barriers would contradict the physical picture behind the random flight chain model, since in this case some random trajectories (=conformations) are counted twice, see, e.g. [167].

It is instructive to make the following general remark. The integral over the probability

2. Theoretical Framework for the interpretation of NMR residual dipolar couplings of unfolded proteins

distribution $G(\mathbf{a}, \mathbf{r})$ in the presence of obstructing media is less than 1,

$$\mathcal{J}(\mathbf{a}) = \int G(\mathbf{a}, \mathbf{r}) d\mathbf{r} < 1. \quad (2.29)$$

In the language of the random walk, the difference $1 - \mathcal{J}$ reflects the probability of absorption. In the language of chain conformations, the integral \mathcal{J} corresponds to the fraction of possible polypeptide conformations in the obstructed space as compared to the number of conformations of a non-obstructed chain.

After integration over \mathbf{r}_1 and \mathbf{r}_2 in eq. (2.26) one arrives at

$$\mathcal{N}(\theta) = \int d\mathbf{r}_0 \int_0^{2\pi} d\phi \mathcal{J}(\mathbf{r}_0 + \mathbf{s}_i/2) \mathcal{J}(\mathbf{r}_0 - \mathbf{s}_i/2). \quad (2.30)$$

Steric alignment in the absence of long-range interactions

Now, a particular type of experiment for the measurement of RDCs is considered in which alignment of the polypeptide chains is achieved by placing bicelles into the solution. Bicelles are large, disk-shaped bodies which weakly interact with polypeptides and orient in the presence of an externally applied magnetic field [172]. It is assumed that the bicelles are aligned perpendicular to the magnetic field. Neglecting the finite size of the bicelles, they are approximated by a set of parallel infinite planes.

Typical interplanar distances encountered in experiments are 400-600 Å in the case of phospholipid bicelles [102] and a few nanometers in the case of *n*-alkyl-poly(ethylene glycol)/*n*-alkyl alcohol bicelles [173], respectively. The length of one amino acid residue ($C^\alpha - C^\alpha$ distance ≈ 3.8 Å) is taken to be the unit of length in the herein elaborated model, hence, an L value of 150 corresponding to an interplanar distance of about 570 Å seems to be a reasonable choice for the calculations.

We note that the model – random walk between two infinite planes oriented perpendicular to the z -axis while there are no obstructions in the x - and y -directions – implies the 1D character of the random walk problem. However, the random walk probability distribution has to be taken from the 3D random walk formalism rather than from the 1D formalism. This is due to the fact that, in the 3D case, the characteristic width of the random walk probability distribution (which translates into the radius of gyration, end-to-end distance, etc.) is a factor of $\sqrt{3}$

smaller than in the 1D case. When applying the 1D formalism in this model, one effectively replaces the steps of *equal length* with steps of *equal projection* onto the z -axis.

The probability distribution of the 3D random walk between two parallel infinite absorbing barriers is easily obtained by generalization (based on symmetry considerations) of a 1D walk in the presence of one barrier, [167]

$$G(\mathbf{a}, \mathbf{r}) = \frac{1}{(2\pi n/3)^{3/2}} \left[\exp\left(-\frac{3(\mathbf{r}-\mathbf{a})^2}{2n}\right) - \exp\left(-\frac{3(\mathbf{r}'-\mathbf{a})^2}{2n}\right) - \exp\left(-\frac{3(\mathbf{r}''-\mathbf{a})^2}{2n}\right) \right]. \quad (2.31)$$

Here, \mathbf{a} is the starting point of an n -steps walk, \mathbf{r} is the ending point, \mathbf{r}' is the point symmetric to \mathbf{r} with respect to mirror reflection in the plane of the first barrier, $\mathbf{r}' = \mathbf{r} - 2D'\mathbf{n}'$ (where D' is the distance between \mathbf{r} and the first barrier, and \mathbf{n}' is the unit normal to the plane of the first barrier). Analogously, $\mathbf{r}'' = \mathbf{r} - 2D''\mathbf{n}''$ is the image of the point \mathbf{r} in the plane of the second barrier (Fig. 2.7).

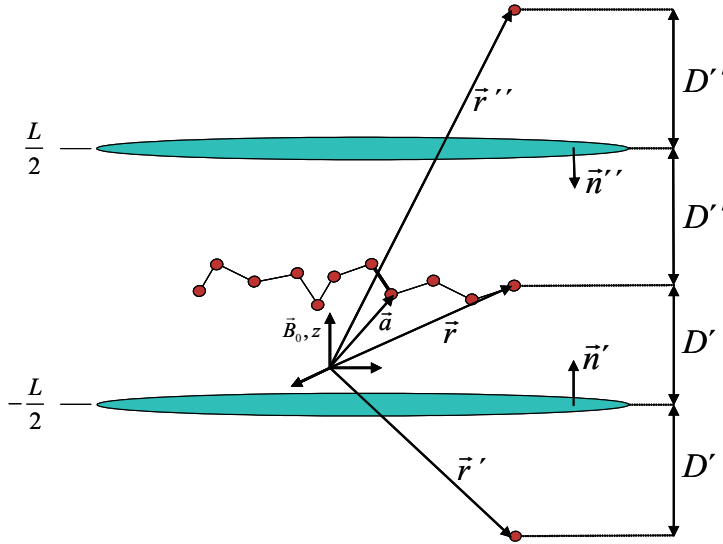


Figure 2.7.: Definition of the vectors \mathbf{r}' and \mathbf{r}'' . The two ellipsoids represent two bicelle particles oriented perpendicular with respect to B_0 . The bicelles are approximated by infinite planes in the calculations. The segment which divides the chain in two sub-chains is thickened for illustrative purposes.

Eq. (2.31) is valid for $n < L$. This is the situation where the chain cannot "feel" two barriers simultaneously. However, completely or almost completely extended conformations for which the size of the chain approximately equals N are extremely unlikely. Rather, the typical size

2. Theoretical Framework for the interpretation of NMR residual dipolar couplings of unfolded proteins

of the chain is of the order of \sqrt{N} . Therefore, in fact, eq. (2.31) works well until $N \sim L^2$, i.e. it is satisfied for most of the experimental conditions.

The second and the third terms in the square brackets in eq. (2.31) correspond to exclusion of the trajectories (i.e. conformations) which cross the first and the second barriers, respectively. These terms are responsible for the integrals being less than 1 (eq. (2.29)).

This integral is now being calculated. For convenience, it is assumed that the barriers are perpendicular to the z -axis, crossing it at points $-L/2$ and $L/2$ (see Fig. 2.7). Then, integrations over x and y result in the factor $(2\pi n)/3$. The integration over z from $-L/2$ to $L/2$ results in a set of error functions $\text{Erf}(x)$:

$$\mathcal{J}(\mathbf{a}) = \int G(\mathbf{a}, \mathbf{r}) d\mathbf{r} = \text{Erf} \left[\sqrt{\frac{3}{2n}} \left(\frac{L}{2} - a_z \right) \right] + \text{Erf} \left[\sqrt{\frac{3}{2n}} \left(\frac{L}{2} + a_z \right) \right] - 1. \quad (2.32)$$

Substituting eq. (2.32) into eq. (2.30) gives

$$\begin{aligned} \mathcal{N}(\theta) = \int d\mathbf{r}_0 \int_0^{2\pi} d\phi \\ \left\{ \text{Erf} \left[\sqrt{\frac{3}{2N_1}} \left(\frac{L}{2} - z_0 - c/2 \right) \right] + \text{Erf} \left[\sqrt{\frac{3}{2N_1}} \left(\frac{L}{2} + z_0 + c/2 \right) \right] - 1 \right\} \times \\ \left\{ \text{Erf} \left[\sqrt{\frac{3}{2N_2}} \left(\frac{L}{2} - z_0 + c/2 \right) \right] + \text{Erf} \left[\sqrt{\frac{3}{2N_2}} \left(\frac{L}{2} + z_0 - c/2 \right) \right] - 1 \right\}, \end{aligned} \quad (2.33)$$

where $N_1 = i - 1$ and $N_2 = N - i$ are the lengths of the subchains before and after the given segment S_i , ($N_1 + N_2 + 1 = N$), z_0 is the z -component of the vector \mathbf{r}_0 , and

$$c \equiv \cos \theta \quad (2.34)$$

is the z -component of the vector \mathbf{s}_i .

In eq. (2.33), due to the symmetry of the problem, no dependencies on x_0 , y_0 , and the azimuthal angle ϕ are present. The integrals over these variables produce the infinitely large factor $2\pi\sigma$, where σ is the (infinite) surface of the barriers. However, exactly the same factor is produced in the denominator of eq. (2.25), so the value of $P(\theta)$ remains finite. Therefore, the probability distribution is expressed *via* the one-dimensional integral over the position of the middle of the given segment S_i , z_0 :

$$P(c) = \frac{\mathcal{E}(c)}{\int_{-1}^1 \mathcal{E}(c) d(c)}, \quad (2.35)$$

where

$$\begin{aligned} \mathcal{E}(c) = & \int_{-L/2+|c|/2}^{L/2-|c|/2} \left\{ \text{Erf} \left[\sqrt{\frac{3}{2N_1}} \left(\frac{L}{2} - z_0 - c/2 \right) \right] + \text{Erf} \left[\sqrt{\frac{3}{2N_1}} \left(\frac{L}{2} + z_0 + c/2 \right) \right] - 1 \right\} \times \\ & \left\{ \text{Erf} \left[\sqrt{\frac{3}{2N_2}} \left(\frac{L}{2} - z_0 + c/2 \right) \right] + \text{Erf} \left[\sqrt{\frac{3}{2N_2}} \left(\frac{L}{2} + z_0 - c/2 \right) \right] - 1 \right\} dz_0. \end{aligned} \quad (2.36)$$

At this stage of derivation, the reason for non-zero RDCs can be clearly seen. Indeed, $\mathcal{E}(c)$ is just a renormalised number of chain conformations for which the segment of interest is at particular angle θ ($c \equiv \cos \theta$). In eq. (2.36), the $|c|/2$ terms in the limits of the integration are due to the fact that the center of the given segment can not approach the barriers closer than $|c|/2$. The conformations in which the segment of interest is parallel to the barriers ($c = 0$) have more available space and, therefore, are more populated.

Eq. (2.36) can be further simplified by expanding the error functions in a series over $|c|/2$. Using the standard Taylor series

$$\begin{aligned} \text{Erf}(x + \delta) = & \text{Erf}(x) + \frac{2}{\sqrt{\pi}} \exp(-x^2) \delta - \frac{2}{\sqrt{\pi}} x \exp(-x^2) \delta^2 \\ & + \frac{2}{\sqrt{\pi}} \frac{2x^2 - 1}{3} \exp(-x^2) \delta^3 - \frac{2}{\sqrt{\pi}} \frac{2x^2 - 3}{6} x \exp(-x^2) \delta^4 + \dots \end{aligned} \quad (2.37)$$

one obtains after some algebra

$$\begin{aligned} \mathcal{E}(c) = & L - \frac{2}{\sqrt{\pi}} \sqrt{\frac{2(N_1 + N_2)}{3}} - \frac{2}{\sqrt{\pi}} \sqrt{\frac{3}{2(N_1 + N_2)}} c^2 + \frac{1}{\pi} \sqrt{\frac{1}{N_1 N_2}} |c|^3 + \\ & \frac{1}{2\sqrt{\pi}} \sqrt{\frac{3}{2(N_1 + N_2)^3}} c^4 - \frac{3}{20\pi} (N_1 + N_2) \sqrt{\frac{1}{N_1^3 N_2^3}} |c|^5 + \dots \end{aligned} \quad (2.38)$$

As shown in Fig. 2.8, one has to include terms up to $|c|^5$ in order to achieve accuracy (10^{-5}) needed for the calculations of the RDCs.

The denominator in eq. (2.35) is calculated then as follows.

2. Theoretical Framework for the interpretation of NMR residual dipolar couplings of unfolded proteins

$$\int_{-1}^1 \mathcal{E}(c) \mathbf{d}c = 2L - \frac{4}{\sqrt{\pi}} \sqrt{\frac{2(N_1 + N_2)}{3}} - \frac{4}{3\sqrt{\pi}} \sqrt{\frac{3}{2(N_1 + N_2)}} + \frac{1}{2\pi} \sqrt{\frac{1}{N_1 N_2}} + \frac{1}{5\sqrt{\pi}} \sqrt{\frac{3}{2(N_1 + N_2)^3}} - \frac{1}{20\pi} (N_1 + N_2) \sqrt{\frac{1}{N_1^3 N_2^3}} + \dots \quad (2.39)$$

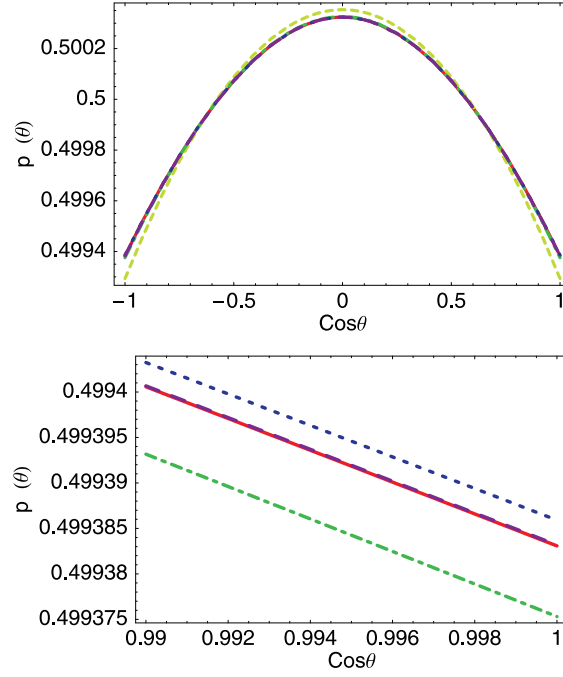


Figure 2.8.: Comparison of $P(\theta)$ for the center segment of a 21mer chain calculated with the series representation (eq. (2.38)) (dashed-dotted green, (including terms up to $|c^3|$), dotted blue (including terms up to c^4), and dashed purple (including terms up to $|c^5|$) curves) and with the exact expression (eq. (2.36)) (red). The bottom graph is a zoom in of the bottom right part of the top graph.

Eqs. (2.35), (2.38), (2.39) define the probability distribution for a given segment to be found at a particular angle θ . With the given probability distribution, the average angular distribution in eq. (2.21) is now given by:

$$D_{PQ} \approx \frac{\mu_0 \hbar \gamma_P \gamma_Q}{4\pi^2 R_{PQ}^3} \frac{3 \cos^2 \alpha_{PQ} - 1}{2} \times$$

$$\frac{-\frac{8}{15\sqrt{\pi}}\sqrt{\frac{3}{2(N_1+N_2)}} + \frac{1}{4\pi}\sqrt{\frac{1}{N_1N_2}} + \frac{4}{35\sqrt{\pi}}\sqrt{\frac{3}{2(N_1+N_2)^3}} - \frac{N_1+N_2}{32\pi}\sqrt{\frac{1}{N_1^3N_2^3}}}{2L - \frac{4}{\sqrt{\pi}}\sqrt{\frac{2(N_1+N_2)}{3}} - \frac{4}{3\sqrt{\pi}}\sqrt{\frac{3}{2(N_1+N_2)}} + \frac{1}{2\pi}\sqrt{\frac{1}{N_1N_2}} + \frac{1}{5\sqrt{\pi}}\sqrt{\frac{3}{2(N_1+N_2)^3}} - \frac{N_1+N_2}{20\pi}\sqrt{\frac{1}{N_1^3N_2^3}}}. \quad (2.40)$$

Eq. (2.40) is the final result of the derivation.

2.2.2. Predictions and comparison with experiment

In order to reveal the two general patterns in the behaviour of RDCs which are predicted by the random flight chain model, eq. (2.40) is further simplified by expanding over $1/L$ and $1/\sqrt{N}$ provided that $L \gg \sqrt{N} \gg 1$. This condition is fulfilled for typical ranges of experimental parameters.

Only the first terms are kept in the expansion:

$$D_{PQ} \approx -\frac{\mu_0 \hbar \gamma_P \gamma_Q}{4\pi^2 R_{PQ}^3} \frac{3 \cos^2 \alpha_{PQ} - 1}{2} \times \frac{1}{L} \left[\frac{4}{15\sqrt{\pi}} \sqrt{\frac{3}{2(N_1 + N_2)}} - \frac{1}{8\pi} \sqrt{\frac{1}{N_1 N_2}} \right], \quad (2.41)$$

$$N = N_1 + N_2 + 1.$$

From eq. (2.41) it is evident that shorter chains have larger (in absolute terms) RDCs at fixed distance L (see also Figs. 2.9, 2.10).

From the first glance, this is a counter-intuitive result. Indeed, one might expect that longer chains should be more affected by the presence of the obstacles as compared to shorter chains. However, within the same theoretical framework, one can show that the chain clews almost never approach the obstacles closer than their characteristic radius $\sim \sqrt{N}$; the conformations in which the chain is thinly spread along the obstacle are statistically insignificant. In the vicinity of the obstacle, the average shape of the chain resembles a sphere flattened on one side, in contrast to the unperturbed spherical shape characteristic of non-obstructed chains. The deformation is most pronounced at the surface of the chain clew while the inner region is less affected. From this point of view, it is evident that the influence of the obstacle is more significant for clews with higher surface-to-volume ratios. That is, shorter chains are more affected by the obstacles and therefore have larger RDCs.

2. Theoretical Framework for the interpretation of NMR residual dipolar couplings of unfolded proteins

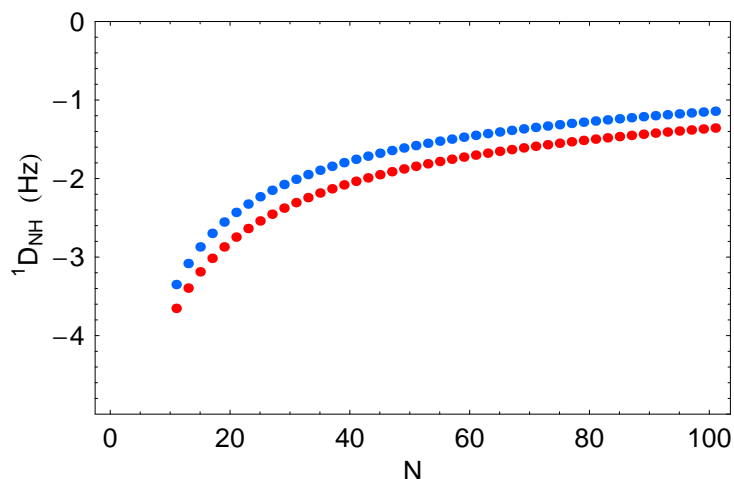


Figure 2.9.: NH residual dipolar couplings as a function of the chain length N for the segment next to the end segment (upper curve) and the center segment (lower curve).

The second general feature predicted by the random flight chain model is that RDCs are larger for the segments in the middle of the chain as compared to the end segments,^a leading to the bell-like shape of the RDC profile, as illustrated in Fig. 2.10. This feature can also be explained in terms of average shapes of the chain clews. As it has been noted, the chain clews resemble flattened spheres with maximum deformation on one side of the clew, at the contact with the obstacle. In this situation, the end segments (located mostly on the surface of the clew) are affected most, while the middle segments are affected less. However, when the clew rotates (keeping the distance to the obstacle constant) other end segments become affected by the obstacle while the middle segments are still constrained. This rotational degree of freedom leads to the fact that on average the middle segments are more strongly aligned and RDCs in such segments are larger.

The presented data in Figs. 2.9 and 2.10 correspond to $^1D_{NH}$ couplings using a value of -21700 Hz for the constant term in eq.(2.21). The angle α_{NH} (α_{PQ} , see eq.(2.21)) was taken to be 90° , a value close to $\approx 85^\circ$ dictated by the rigid part of the geometry of the chain. It may easily be checked that deviations from 90° do not exceed $10\text{-}15^\circ$ (0.26 radian) in all relevant conformations (α -helix, β -sheet, polyproline II) being populated in unfolded polypeptides. Such deviations are insignificant for the RDC values, since $\cos^2(90^\circ \pm 15^\circ) \approx 0.07 \ll 1$ (see

^aIt is assumed that the internuclear vectors throughout the chain are on average oriented at the same angle α_{PQ} with respect to the symmetry axis of the corresponding segment

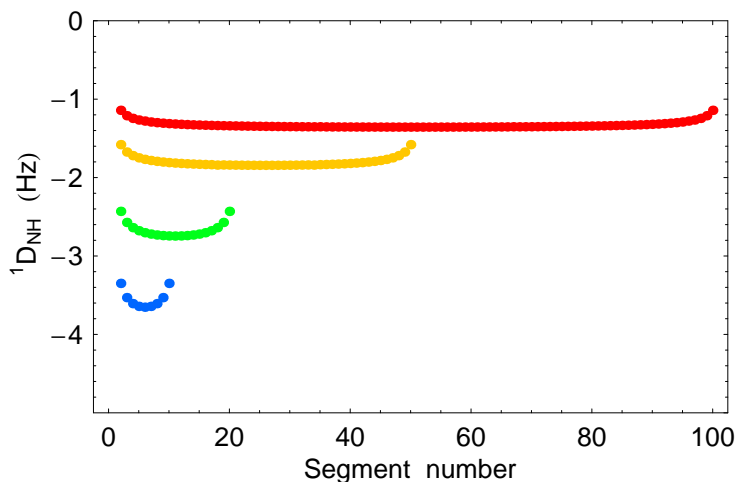


Figure 2.10.: NH residual dipolar couplings as a function of the segment number for random flight chains of different lengths: 11 segments (blue), 21 segments (green), 51 segments (yellow), and 101 segments (red).

eqs. (2.40) and (2.41)).

Fig. 2.11 shows RDC profiles calculated with the expressions from ref. [77] (two upper curves) and with the use of the present formalism (two lower curves). It is seen, that the shortcomings mentioned in the introduction lead to the three-times smaller RDCs (in absolute terms), and also to a less pronounced bell-shape.

Two important remarks need to be made when comparing the theoretical predictions with experimental data. First, the random flight chain model provides a model for an unfolded protein in which long-range interactions are not taken into account. Second, the derived expressions only hold true in the case of a steric alignment mechanism since electrostatic contributions as they occur, e.g., when using the filamentous bacteriophages Pf1 [135] and fd [136] are not accounted for. The comparison is therefore limited to RDCs which have been experimentally measured in bicelle and polyacrylamide alignment media [174, 175]. Moreover, the derivation of section 2.2.1 is carried out for obstruction in one dimension, while for polyacrylamide motion of the polypeptide chain is obstructed in all three dimensions. Therefore, certain deviations from the predicted behaviour are possible for the polyacrylamide media.

For the complete set of experimental conditions and details about residual structure concerning all investigated proteins and peptides, please, refer to Table 2.1.

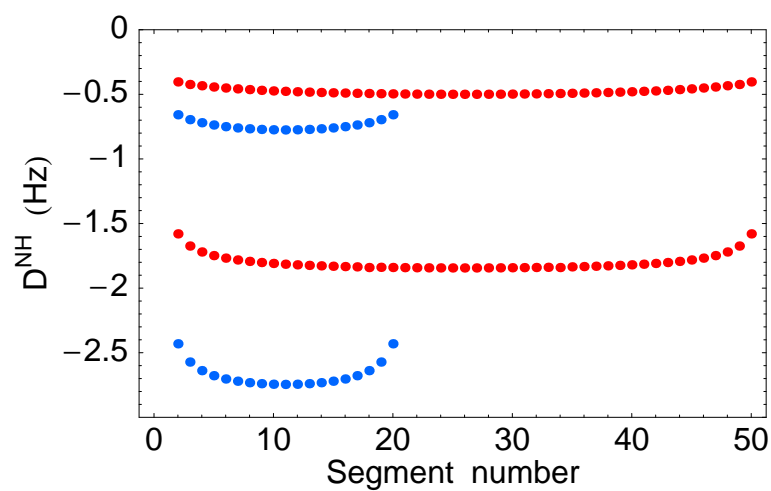


Figure 2.11.: Comparison of NH residual dipolar couplings as a function of the segment number for random flight chains consisting of 21 and 51 segments (blue and red, correspondingly). The upper profiles are obtained with the expression from ref. [77], the lower profiles are the present results reproduced from Fig. 2.10.

Polypeptide (chain length)	Origin of unfolding	Alignment media	Residual structure	¹ D _{PQ}	Model agreement
SNase ^a Δ131Δ (131) [72]	Truncation; truncation + urea	12% PAG (compr.); 5% CmEr ^d /alcohol	Long-range; native-like topology	¹ D _{NH}	- ^c
SNase ^a (149) and SNase ^a Δ131Δ (131) [143]	Urea; acid; truncation; truncation + urea; truncation + detergent	6-14% PAG (compr.) ^b ; 8-12% PAG (str.) ^b ; 3-7% C8E5 ^d /1-octanol; 6.5% CPBr ^e /1-hexanol	Long-range; native-like topology	¹ D _{NH}	- ^c
Mutants (131) and fragments (101) of SNase ^a Δ131Δ [144]	Truncation; truncation + urea	12% PAG (compr.) ^b	Long-range; native-like topology; local propensities	¹ D _{NH}	- ^c
Peptides (2, 4, 10, and 15) [145]	Intrinsically unstructured	18% PAG (compr.) ^b ; 8-10% C8E5 ^d /1-octanol	local propensities	¹ D _{NH}	- ^f
S-peptide of RNaseA ^g (22) [146]	Decreased salt concentration; increased temperature	5.2% C8E5 ^d /1-octanol	α-helical	¹ D _{NH}	No bell shape

^aSNase: Staphylococcal nuclease

^bPAG: Polyacrylamide gel (compressed/stretched)

^cNH RDCs only available for part of the sequence

^dCmEr/C8E5/C12E5: n-Alkyl-poly(ethylene glycol)/n-Octyl-penta(ethylene glycol)/n-Dodecyl-penta(ethylene glycol)

^eCPBr: Cetylpyridinium bromide

^fResidue assignment not given

^gRNaseA: Ribonuclease A

2. Theoretical Framework for the interpretation of NMR residual dipolar couplings of unfolded proteins

Polypeptide (chain length)	Origin of unfolding	Alignment media	Residual structure	$^1D_{DQ}$	Model agreement
GB1 ^h (56) [147]	Acid + increased temperature	6% PAG (str.) ^b ; fd ⁱ (~30mg/mL)	Not native-like	$^1D_{NH}$	$_{-f}$
Eglin C (70) [90]	Acid + urea	12% PAG (compr.) ^b	Native-like global structure	$^1D_{NH}$, $^1D_{C\alpha H\alpha}$	No bell shape
ACBP ^j (86) [148]	GdnCl; acid	7% PAG (str.) ^b	α -helical propensity; native-like long-range interactions in acid solution; suggested loss of long-range interaction in acid solution upon I27A mutation	$^1D_{NH}$	Approximate bell shape for GdnHCl; no bell shape for acid
apoMb ^k (153) [149]	Acid + urea; acid	10% PAG (compr. + str.) ^b	α -helical propensity for acid; neither secondary nor tertiary interactions for acid + urea	$^1D_{NH}$	Bell shape for acid + urea; no bell shape for acid
Polyglutamic acid (21) [150]	Intrinsically unstructured	6.5% DMPC/DHPC ^l	No res. structure reported	RDC values not given explicitly	RDCs only measured for E11 and E21

^hGB1: B1 domain of protein G

ⁱfd: fd bacteriophage

^jACBP: Acyl-coenzyme A binding protein

^kapoMb: apo-Myoglobin

^lDMPC/DHPC: Dimyristoyl-/Dihexanoyl-phosphatidylcholine

Polypeptide (chain length)	Origin of unfolding	Alignment media	Residual structure	$^1D_{\text{PQ}}$	Model agreement
CTD ^m of T4 fibrin (29) [151]	Increased temperature	7% PAG (compr. + str.) ^b	Local propensities	$^1D_{\text{NH}}$, $^1D_{\text{C}^{\alpha}\text{H}^{\alpha}}$	No bell shape
αS^n (140) and fragment thereof (32) [70]	Intrinsically unstructured; urea; increased temperature; polyamine binding	PF1 ^o (10mg/mL); 5% C8E5 ^d /octanol	Long-range interactions, weakened to different extents by polyamine ligands and disrupted by elevated temperature; disruption of hydrophobic clusters by urea	$^1D_{\text{NH}}$	Approximate bell shape for full-length αS in C8E5 ^d /octanol in presence of 8M urea and upon spermine binding, respectively (in the latter case PRE ^p data still show existent long-range interaction
SNase ^a (149) and fragment thereof (102) [152]	Acid + urea	7.5% PAG (str.) ^b	No long-range interactions	$^1D_{\text{NH}}$	Approximate bell shape in both cases
Mutants of αS^n (140) [153]	Intrinsically unstructured; urea	PF1 (10mg/mL) ^o ; 5% C8E5 ^d /octanol	For A18C basically as for WT ^q ; A30P and A53T induce local effects and affect long-range interactions as compared to WT	$^1D_{\text{NH}}$	Approximate bell shape for A30P and A53T in C8E5 ^d /octanol and for A30P in C8E5 ^d /octanol/8M urea, respectively

^mCTD; C-terminal domainⁿ αS ; α -Synuclein^oPF1; PF1 bacteriophage^pPRE; Paramagnetic relaxation enhancement^qWT; Wild type

2. Theoretical Framework for the interpretation of NMR residual dipolar couplings of unfolded proteins

Polypeptide (chain length)	Origin of unfolding	Alignment media	Residual structure	$^1D_{DQ}$	Model agreement
SNase ^a Δ131Δ (131) [154]	Truncation + urea	6% PAG (str.) ^b ; composite of 6% PAG (str.) ^b and Pf1 ^c (3 and 10 mg/mL) at different relative angles	Tertiary interactions	$^1D_{NH}$	Approximate bell shape in 6% PAG (str.) ^b
α S (140) [155]	Intrinsically unstructured	5% C8E5 ^d /octanol	Long-range interaction	$^1D_{NH}$	No bell shape
Tau (441) [156]	Intrinsically unstructured	5% C12E5 ^d /1-hexanol	Local propensities	$^1D_{NH}$	No bell shape
Peptides (9 and 12) [157]	Intrinsically unstructured	10% PAG (str.) ^b ; Pf1 ^c (20 mg/mL)	Local propensities; pronounced effects induced by G, P, W, and Y	$^1D_{NH}$, $^1D_{C\alpha H^e}$	Approximate bell shape in PAG ^b for longer peptide; slight deviation from bell shape in PAG ^b for shorter peptides
Peptides of varying length and amino acid composition/sequence (6, 7, 8, 12, 15) [158]	Intrinsically unstructured	18% PAG (compr.)	Preference for extended conformation	$^1D_{NH}$, $^1D_{C\alpha H^e}$	Approximate bell shape for $^1D_{NH}$ of Ac-YGEGSGAGTGDG; other sequences either too short (6-8 res.) or RDCs only available for part of the sequence (12/15 res.)

Polypeptide (chain length)	Origin of unfolding	Alignment media	Residual structure	$^1D_{PQ}$	Model agreement
Ubiquitin (76) [73]	Acid + urea	10% PAG	Native-like long-range structure (turn structure in N terminus)	$^1D_{NH}$, D_{H^N, H^N}	Approximate bell shape
Lysozyme (all-Ala) (129) [160]	Intrinsically unstructured + acid	7% PAG	Long-range interactions	$^1D_{NH}$	No bell shape

Table 2.1.: Summary of experimental conditions and residual structure exhibited by unfolded polypeptides on which RDCs have been measured. The most right column reports on whether the sequential distribution of RDCs agrees with the first prediction of the model.

$^1D_{\text{NH}}$ dependence on segment number

The $^1D_{\text{NH}}$ plotted as a function of the segment number follow a symmetric bell shape (see Fig. 2.10). Inspection of experimentally determined $^1D_{\text{NH}}$ reveals bell-like shapes in a few cases [73, 148, 149, 152, 157, 158]. Saying bell-like we talk of profiles in which NH RDCs are (i) on average larger in the middle of the chain as compared to the chain ends, and (ii) exhibit a uniform sign. The first report of such a profile was the paper by Poulsen and co-workers [148]. They measured $^1D_{\text{NH}}$ of the GdnCl-denatured state of acyl-coenzyme A binding protein (ACBP). Except for 2 amino acid residues (change in sign) an overall bell shape is observed. However, this is not the case for the acid-denatured state of ACBP as the sign of the dipolar coupling changes along the sequence multiple times. Since it is known that denaturants such as GdnCl lead to more extended chain conformations whereas acid denaturation promotes more compact states of unfolded proteins, the GdnCl-denatured state of ACBP might be a good model case for a fully unfolded protein following the prediction made by the random flight chain model. A higher degree of compactness in acid-denatured ACBP, however, favors elements of residual structure and therefore leads to deviations from the bell-shaped distribution of RDCs exhibited by the random flight chain. Indeed, the occurrence of hydrophobic long-range interactions has been reported [148].

The conclusion drawn from the RDC data collected on the GdnCl-denatured and acid-denatured states of ACBP is supported by the report by Mohana-Borges *et al.* [149]. Corresponding NH RDCs were measured on urea-denatured and acid-denatured apoMb. The urea-denatured state of apoMb exhibits a bell-shaped RDC profile whereas in the case of acid-denatured apoMb the dipolar coupling sign changes along the amino acid sequence several times. This is illustrated in Fig. 2.12 where NH RDCs measured for the two different unfolded states of apoMb and predicted NH RDCs for a chain of equal length are shown. It is noted that the simple model being used in this work, of course, does not capture any amino acid specific features that are clearly observed in experiments.

Bell-shaped RDC distributions were observed in two further studies. In the first study, NH RDCs for the urea-denatured state of full-length staphylococcal nuclease (SN) reported by Sallum *et al.* follow the prediction made by the random flight chain model [152]. This is as well the case for the second study where NH RDCs were measured on model peptides of 9 and 12 residues in length, respectively [157]. Although no denaturant was present in the

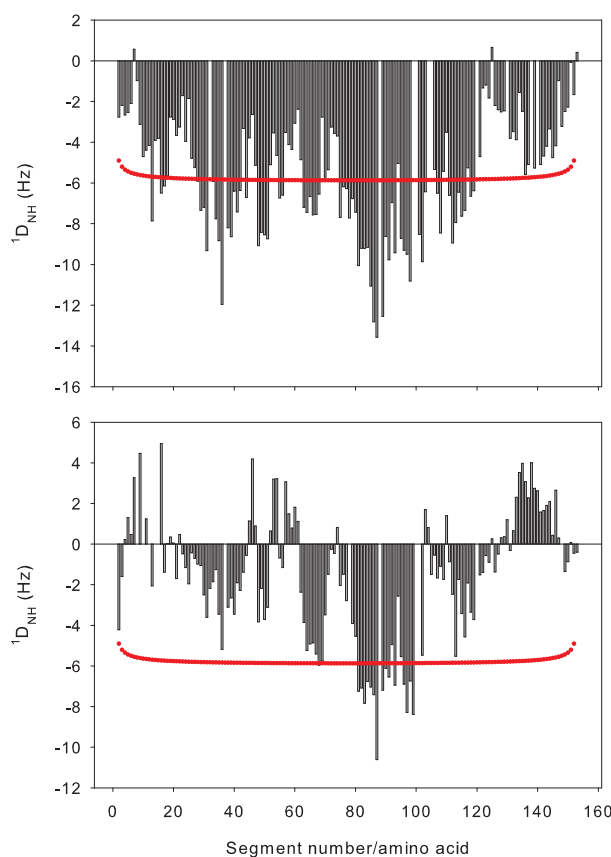


Figure 2.12.: Experimental $^1D_{NH}$ for the urea- and acid-denatured states of apoMb (black bars; top and bottom graphs, respectively) and predicted $^1D_{NH}$ for a chain of equal length (red). RDC profiles of apoMb were taken from Mohana-Borges *et al.* [149], digitized and replotted. For the prediction an interplanar distance of $L=38$ was taken.

experiments [157], the lengths of the peptides chosen and their amino acid composition preclude the occurrence of long-range interactions and (in most cases) local order which makes them an ideal model for a fully unfolded polypeptide chain. Therefore, not surprisingly, for almost all peptides studied in 10% PAG (stretched) an approximately bell-shaped RDC profile is observed [157]. Likewise, the same reasoning applies to the 12-residue peptide Ac-YGEGSGAGTGDG for which $^1D_{NH}$ were measured in 18% PAG (compressed) in the absence of denaturant [158]. In a further study, RDCs were measured for the central and C-terminal residues of a 21-residue polyglutamate chain [150]. The axial alignments of the respective C^α fragments, which were derived from $C^\alpha H^\alpha$, $C^\alpha CO$, and $C^\alpha C^\beta$ RDCs, show a sixfold larger value for the central fragment as compared to the C-terminal fragment. While a sequential

2. Theoretical Framework for the interpretation of NMR residual dipolar couplings of unfolded proteins

RDC profile is not available, this observation is in qualitative agreement with the prediction that the chain center shows larger alignment than the chain ends.

As mentioned above, the presence of residual intramolecular interactions within ensembles of unfolded proteins might lead to deviations from the characteristic distribution of $^1D_{\text{NH}}$ observed along a random flight chain which is likely the case for the acid-denatured states of ACBP [148] and apoMb [149]. Therefore, $^1D_{\text{NH}}$ profiles of proteins which are reported to exhibit long-range interactions in their unfolded state ensembles were analysed in the following. Deviations from the predicted bell-shaped profile are to be expected since long-range interactions are not explicitly taken into account in the random flight chain model.

The pioneering work of Shortle and co-workers on the $\Delta 131\Delta$ construct of SN is an especially attractive case in this context as the persistence of a native-like topology under even highly denaturing conditions (8M urea) was reported in 3 subsequent studies [72, 143, 144]. However, NH RDCs are available only for part of the sequence as the complete backbone resonance assignment could not be undertaken in an earlier study which was attributed to exchange broadening [61]. Hence, an analysis of the profile's shape was not conducted.

The most recent RDC study on $\Delta 131\Delta$ is consistent with the earlier reports [72, 143, 144] stating a 'specific tertiary structural arrangement, not inconsistent with a natively like topology' [154]. Interestingly, the $^1D_{\text{NH}}$ profile in 6% PAG (stretched) alone only slightly deviates from the bell shape predicted by the random flight chain model suggesting that the assumed tertiary structural arrangement is not very much reflected in the RDCs measured in PAG alone.

For another protein, eglin C, a native-like global structure was also reported [90]. A bell-shaped profile cannot be inferred from sequential NH RDCs of the urea-denatured state supporting the authors' conclusion that residual structure is being populated. The finding that native-like hydrophobic long-range interactions persist in acid-denatured ACBP is in line with the above reasoning [148]. Recent work on α -synuclein (αS) also showed that long-range contacts greatly affect the sequential $^1D_{\text{NH}}$ profile which is reflected in large couplings in specific parts of the amino acid sequence [70, 153]. Perturbation of these long-range contacts by the addition of urea [70] and by single point mutations [153] resulted in significant decreases in respective RDCs. The modulation of long-range contacts using site-directed mutagenesis has also been demonstrated for hen lysozyme [160].

Overall, it is concluded that long-range interactions in unfolded states of proteins are likely

reflected in $^1D_{\text{NH}}$ profiles which significantly deviate from the bell shape exhibited by random flight chains. Of course, it cannot be ruled out that intramolecular interactions present within unfolded states of proteins may compensate each other such that NH RDCs still follow a symmetric bell shape. An example may be the case of the acid- and urea-denatured state of ubiquitin. In this report, native-like long-range interactions were shown to exist for this state by the measurement of long-range $D_{\text{HN},\text{HN}}$ while the $^1D_{\text{NH}}$ profile is still approximately bell-shaped [73].

It is furthermore noted that there are further reports on NH residual dipolar couplings in short peptides and unfolded proteins. In some cases residue assignments were not given which did not allow to include these data into the comparison [145, 147]. In other cases local structural propensities might also lead to deviations from the predicted bell-shaped profile [146, 151, 156].

$^1D_{\text{NH}}$ dependence on chain length

The random flight chain model predicts stronger alignment for shorter chains as compared to longer chains (Figs. 2.10 and 2.9).

As the random flight chain represents a model for a fully unfolded polypeptide chain this prediction will be solely compared to experimental data from peptides and unfolded states of proteins which are reported to lack long-range contacts in their conformational ensembles [148, 149, 152, 157]. Based on the model, corresponding RDCs should in principle be a direct measure of the extent of alignment. Chain lengths of the polypeptides which were considered are 9 and 12 (model peptides, [157]), 86 (ACBP, [148]), 102 (truncation fragment of SN, [152]), 149 (full-length SN, [152]) and 153 residues (apoMb, [149]), respectively. $^1D_{\text{NH}}$ were all measured in PAG with acrylamide concentrations ranging from 7% [148] to 10% [149, 157]. It is evident that these differences in the acrylamide concentration and further differences in experimental parameters such as temperature impair a direct comparison of most of the data sets.

In two of the above mentioned studies, however, NH RDCs were measured for chains of different length under identical experimental conditions. This is the case for the model peptides investigated by Grzesiek and co-workers [157] and full-length SN and the truncation fragment thereof studied by Sallum *et al.* [152]. From pure inspection of the reported dipolar

2. Theoretical Framework for the interpretation of NMR residual dipolar couplings of unfolded proteins

couplings it is not possible to conclude whether the shorter chain exhibits stronger alignment in comparison to the longer chain. Reported dipolar couplings were therefore digitised for all relevant polypeptides and their sequence averages were determined. This should be a reasonable approach as the amino acid sequences of the polypeptides studied in each report are very similar. The sequence-averaged dipolar coupling should thus be a good measure of the overall chain alignment.

The sequenced-averaged NH dipolar coupling of the 9-residue peptide is -4.79 Hz whereas in the case of the longer 12-residue peptide it equals -4.10 Hz [157]. Similarly, for the 102-residue truncation fragment of SN an average NH dipolar coupling of -4.96 Hz was determined while for the 149-residue full-length protein an average NH dipolar coupling of -4.36 Hz was computed [152]. Therefore, in both studies the longer chain shows a (in absolute terms) smaller sequenced-averaged NH dipolar coupling. This result is in agreement with the prediction made by the random flight chain model. Even the $\sim 1/\sqrt{N}$ dependence inferred from eq. (2.41) is remarkably precisely represented in the calculated average dipolar couplings as is illustrated in Table 2.2.2.

Table 2.2.: Scaling of the sequence-averaged ${}^1D_{\text{NH}}$ with the chain length for two available independent studies [152, 157]. In each report, polypeptides exhibit similar amino acid sequences and experimental conditions did not vary. The average ${}^1D_{\text{NH}}$ for the 9-residue peptide is the average of the ${}^1D_{\text{NH}}$ of 9 different 9-residue peptides differing only in the amino acid at position 5 (peptides having I, N, Q, T, D, E, K, V, and L at position 5, respectively, were analysed).

Ref	Length $N^{(1)}$	$\langle {}^1D_{\text{NH}} \rangle (N^{(1)})$	Length $N^{(2)}$	$\langle {}^1D_{\text{NH}} \rangle (N^{(2)})$	$\frac{\langle {}^1D_{\text{NH}} \rangle (N^{(2)})}{\langle {}^1D_{\text{NH}} \rangle (N^{(1)})}$	$\sqrt{\frac{N^{(1)}}{N^{(2)}}}$
[152]	102	-4.96	149	-4.36	0.88	0.83
[157]	9	-4.79	12	-4.10	0.86	0.85

It is noted that ${}^1D_{\text{NH}}$ were measured for peptides of varying length under identical experimental conditions in another study [158]. However, the fact that the studied sequences are either too short (6-8 residues) or RDCs are only available for part of the sequence (12 and 15 residues) did not allow to conduct an analysis as it was undertaken for the two other studies.

Furthermore, it is important to state that in the studies investigated here [152, 157] weak alignment was accomplished by using PAG whereas the formalism was derived for the situa-

tion in which weak alignment is achieved by the presence of regularly arranged infinite planes resembling the bicelle case. The good agreement of the analysed experimental data with the prediction made by the model might therefore indicate that the situation in polyacrylamide gels is well described by the herein presented formalism. Systematic studies of the effect of chain length on the extent of alignment under constant experimental conditions are nevertheless needed in order to evaluate the prediction.

2.3. Conclusion

In this chapter, a general mathematical framework is presented which is suitable for describing various characteristics of random flight chains, including RDCs in various alignment media. This framework is used to derive values of RDCs within the model suggested by Annala *et al.*. It is shown that RDCs can be represented in the form of an expansion over a small parameter $1/\sqrt{N}$ (where N is the number of chain segments). Only few first terms of this expansion are necessary for an accurate approximation of the exact formula. The series representation makes it possible to obtain analytical dependencies of the RDCs on obstacle concentration, chain length, and on position of a segment within the chain.

Two general features of RDCs in unfolded polypeptide chains are predicted by the random flight chain model. First, RDCs are larger for segments in the middle of the chain as compared to the end segments. Second, shorter chains exhibit larger RDCs for a given obstacle concentration.

An extensive search of the available literature has been performed in order to find experimental data supporting or disproving the predictions of the model. Enough evidence has been collected to argue that the first feature is an important characteristic of RDCs of unfolded polypeptides which can serve to recognise fully unfolded states in further experiments. The evidence supporting the second predicted feature is less solid, as there are only two experimental studies providing possibilities to compare theory and experiment. The data from both studies appear to agree with the theoretically predicted behaviour, but more systematic investigations are required.

The presented framework is an important step towards a solid theoretical foundation for the analysis of experimentally measured RDCs of unfolded proteins in the case of alignment me-

2. Theoretical Framework for the interpretation of NMR residual dipolar couplings of unfolded proteins

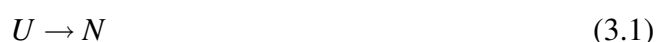
dia such as polyacrylamide gels and neutral bicelle systems which align biomacromolecules by a steric mechanism. Various improvements and generalisations are possible within the suggested approach.

3. Kinetic refolding of bovine α -lactalbumin (BLA)

3.1. Introduction

3.1.1. Some general aspects

The study of the kinetics characterising the transitions between different conformational states during protein folding is an important aspect in exploring the multidimensional potential energy surface of folding (cf. section 1.1). Kinetic measurements can, in principle, yield valuable information both on the number of conformational states involved in folding and on how they are mechanistically linked [176]. The power of kinetic investigations of protein folding will be illustrated in the following using three examples of increasing complexity. Let us first consider a simple two-state protein folding transition which only involves the unfolded (U) and folded (or native, N) states.



This will be an irreversible transition under conditions strongly favouring the native state. The folding reaction will therefore be characterised by a single rate constant k . Rate laws describing the time-dependent changes in the concentrations of states U and N are easily formulated in the framework of classical reaction kinetics.

$$\frac{dU}{dt} = -kU \quad (3.2)$$

$$\frac{dN}{dt} = kU \quad (3.3)$$

3. Kinetic refolding of bovine α -lactalbumin (BLA)

Solving this system of differential equations is straightforward and yields single exponential functions for the disappearance and buildup of the U and N conformational states, respectively (Fig. 3.1a).

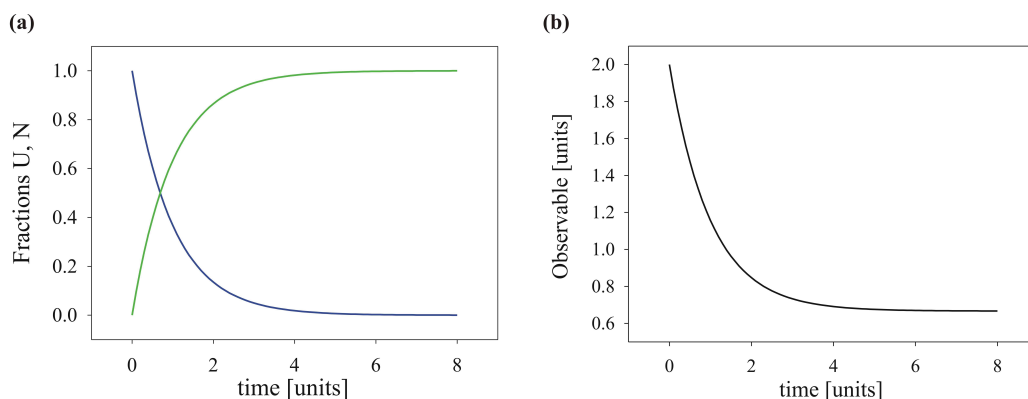


Figure 3.1.: Two-state folding. (a) Time-dependent changes in the concentrations of the unfolded (U , blue) and native (N , green) states. (b) Hypothetical change in an arbitrary observable over time which is assumed to be three-fold lower in the N state as compared to the U state. k was chosen to be equal to 1 units^{-1} in the simulation.

Measurements of protein folding kinetics are often ensemble measurements. The signal being detected will thus be the sum of the contributions of the different conformational states. Fig. 3.1b shows the hypothetical change in an arbitrary observable over time which is assumed to be three-fold lower in the N state.

The synthetic data shown in Fig. 3.1b can be fitted using a single exponential function and the associated rate constant accordingly corresponds to the transition from the U to the N state.

Let us now move to a more complex situation in which an obligatory folding intermediate (I) is populated in the folding transition.



Again, folding steps are assumed to be irreversible. Two rate constants k_1 and k_2 will thus describe the formation of the I and N states, respectively. The system of differential equations corresponding to the time-dependent changes in the concentrations of U , I , and N may be solved analytically. Hypothetical time courses are depicted in Fig. 3.2a.

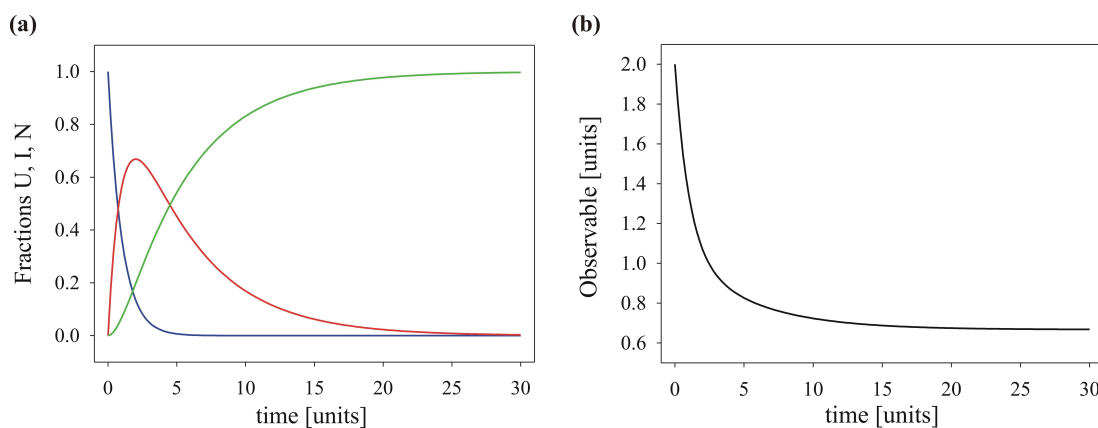


Figure 3.2.: Sequential three-state folding. (a) Time-dependent changes in the concentrations of the unfolded (U , blue), intermediate (I , green), and native (N , red) states. (b) Hypothetical change in an arbitrary observable over time. The I and N states are assumed to attain $1/2$ and $1/3$, respectively, of the value of the observable in the U state. In the simulation, k_1 and k_2 were chosen to be equal to 1 units^{-1} and 0.2 units^{-1} , respectively.

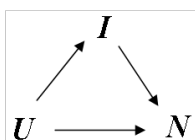


Figure 3.3.: Triangular folding mechanism involving parallel pathways for formation of the N state.

In the case of ensemble measurements the signal will again be the sum of the contributions of the different conformational states involved. Fig. 3.2b shows the theoretical change in an arbitrary observable which is assumed to gradually decrease during the folding transition; for details see the legend of Fig. 3.2.

The analysis of this curve shows that a double exponential function provides the best fit. Corresponding rate constants k_1 and k_2 identify the formation of the intermediate and native states, respectively.

In the third folding scenario, formation of a folding intermediate I is non-obligatory. Fig. 3.3 describes folding as the formation of N through parallel folding channels.

It is assumed that folding steps are irreversible and formation of the I state and the N state

3. Kinetic refolding of bovine α -lactalbumin (BLA)

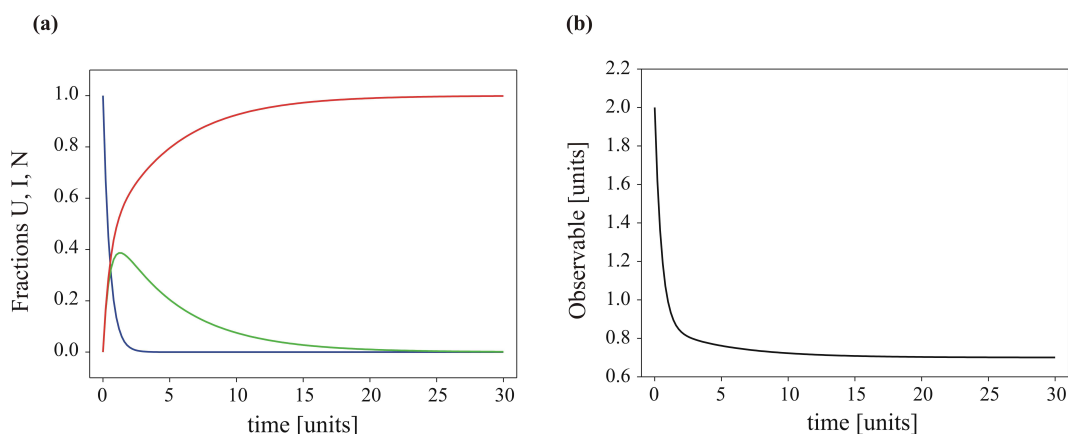


Figure 3.4.: Parallel three-state folding. (a) Time-dependent changes in the concentrations of the unfolded (U , blue), intermediate (I , green), and native (N , red) states. (b) Hypothetical change in an arbitrary observable over time. The I and N states are assumed to attain $1/2$ and $1/3$, respectively, of the value of the observable in the U state. In the simulation, k_1 and k_2 were chosen to be equal to 1 units^{-1} and 0.2 units^{-1} , respectively.

from the U state occur equally fast (k_1). Conversion of the intermediate state into the native state will be described by rate constant k_2 . Figs. 3.4a and 3.4b show hypothetical time courses for the conformational states involved in the folding reaction and an arbitrary observable, respectively.

Again, as in sequential two-state folding (eq. 3.4, Fig. 3.2), a double exponential function provides the best fit to the synthetic data. Importantly, rate constant k_1 will now be the sum of the (identical) rate constants characterising formation of the intermediate and the native state from the unfolded state.

Comparison of the hypothetical data obtained for the three folding scenarios illustrates that the number of exponentials needed to best fit the data gives an indication of the number of states involved and how these states might be interconnected but does not provide a clear proof for some folding mechanism. It is noted that a double exponential refolding behaviour may even result from slow isomerisation around Xaa-Pro peptide bonds in the unfolded state [17]. Therefore, further experiments are generally required to verify (or falsify) a proposed mechanism of folding.

3.1.2. Measurement of protein folding kinetics by stopped-flow techniques

A common approach to the study of chemical reactions in solution is (i) to introduce a sudden change, i.e. a change in solution conditions or in temperature, making the system unstable and (ii) to detect how the system returns to its stable equilibrium state. Both these processes apparently have to be fast compared to the time taken for regaining the stable equilibrium state. In general, an abrupt change in solution conditions may be accomplished by mixing two solutions of differing composition. When mixing the solutions manually, the time needed for mixing (usually on the order of a few seconds) sets an upper limit for the maximum speed of a chemical reaction which would still be detectable. Hartridge and Roughton already noted in 1923 that there is a need for the development of an alternative method as opposed to manual mixing to be able to observe even faster reactions [177]. They designed an apparatus which allowed rapid mixing of two solutions. After mixing, the solutions enter and flow down a tube which serves observation purposes (continuous flow method). It became thus possible to observe progressing chemical reactions in the millisecond time regime. This method was later on extended to less rapid chemical reactions. Here, the flow of the mixture is stopped shortly after full flow velocity has been reached (stopped flow) [178]. Further developments of the stopped-flow technique focussed on the improvement of deadtimes and mixing efficiencies [179–181]. A schematic representation of a stopped-flow apparatus as it is used in nowadays' applications is shown in Fig. 3.5.

The stopped-flow method has long been recognised as a powerful approach to study protein folding events taking place on a millisecond-to-second timescale. A solution of protein unfolded by high concentrations of chemical denaturants such as GdnCl or urea is commonly rapidly mixed with a solution containing refolding buffer. In this way, conditions are created which strongly favour the native state. Folding of the protein is thus initiated and may be detected by a range of spectroscopic techniques such as tryptophan (or tyrosine) fluorescence, far-UV CD, near-UV CD, visible absorption, IR absorption, and small angle x-ray scattering. Information on changes in tertiary interactions, backbone structure, and overall shape and compactness during folding becomes thus available. Most of these spectroscopic techniques yield a signal which is not only a time-average but also averages over all conformations being populated. To discern the contributions of distinct conformational states sequential mixing

3. Kinetic refolding of bovine α -lactalbumin (BLA)

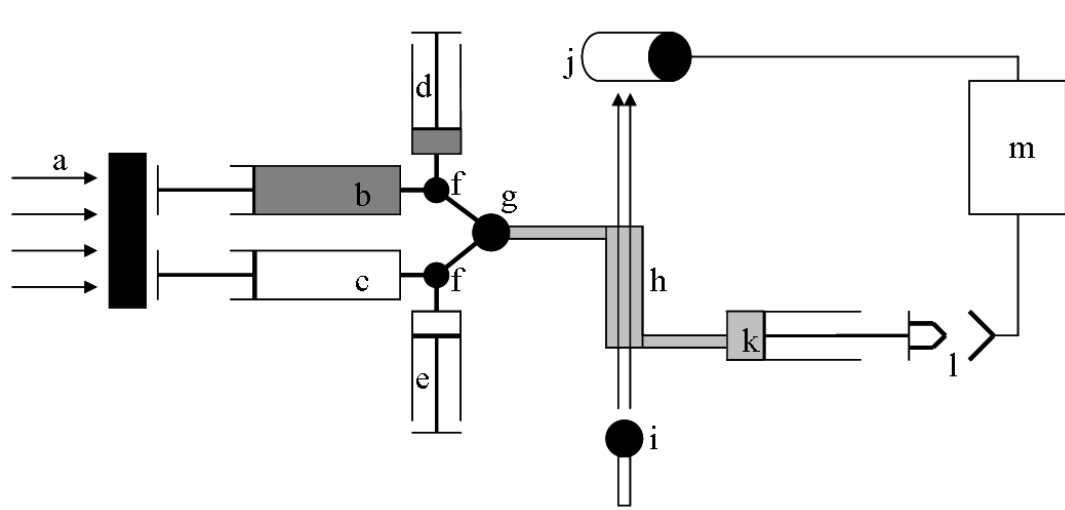


Figure 3.5.: Stopped-flow setup. a trigger 1, b solution 1, c solution 2, d resevoir syringe 1, e resevoir syringe 2, f three-way valve, g mixing chamber, h cuvette/optical cell, i lamp, j photomultiplier, k stop syringe, l trigger 2, m detector.

experiments have been devised. The so-called interrupted refolding assay allows, e.g. to specifically monitor the time-evolution of the native (N) state [182, 183]. In a first mixing step, two solutions containing denatured protein and refolding buffer, respectively, are mixed as in a conventional single mixing stopped-flow experiment. After a variable time t_i , refolding is interrupted by mixing the refolding protein with unfolding buffer. The principal setup of such a sequential mixing experiment is depicted in Fig. 3.6.

If experimental conditions in the second mixing step are chosen such that unfolding of a possible intermediate state I will be too fast to be monitored, the unfolding amplitude will directly correspond to the amount of native protein which was present when refolding was interrupted. By varying delay time t_i , it is effectively possible to determine the time-evolution of the native state. Fitting of the data with an appropriate exponential function then yields valuable information such as the number of pathways involved to form the native state.

3.1.3. Measurement of protein folding kinetics by NMR

NMR spectroscopy is well-suited for kinetic studies of protein folding since it provides site-resolved kinetic information and allows accessing folding timescales ranging from microsec-

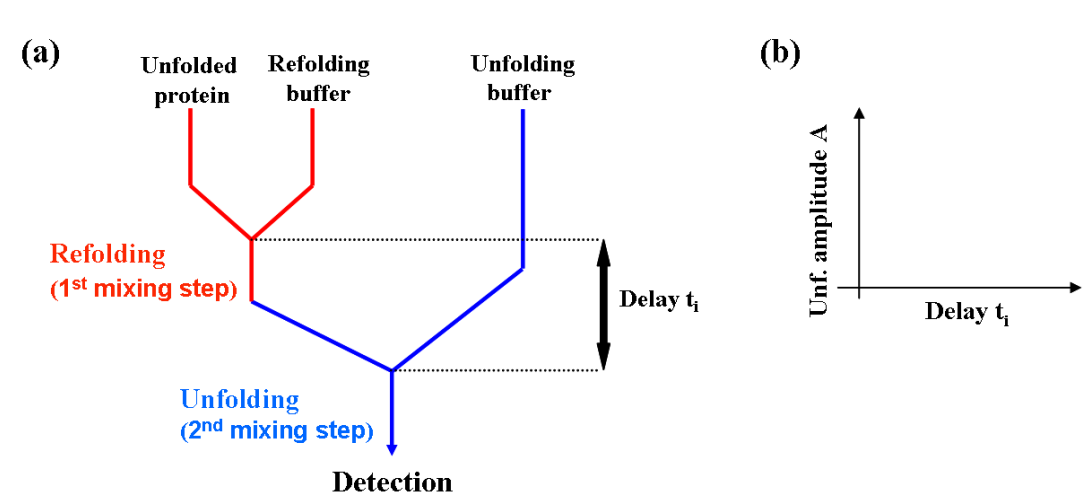


Figure 3.6.: General setup for an interrupted refolding experiment. (a) Refolding of the protein is initiated in a first mixing step. After a variable delay time t_i , refolding is interrupted using unfolding buffer in a second mixing step. Given highly denaturing conditions, the unfolding amplitude is a direct measure of the amount of native protein present at time t_i . (b) Unfolding amplitudes as a function of t_i . Data are analysed using an appropriate exponential function.

onds to days. For this purpose, many different NMR techniques have been developed. Native-state hydrogen exchange [184], line-shape analysis [185, 186], and relaxation dispersion [47] experiments provide kinetic information under equilibrium conditions. The relaxation dispersion NMR approach has been shown to be especially useful since it provides rate constants for transitions between distinct conformational states along with information on populations and interconformational chemical shift differences. These equilibrium methods are restricted to the study of fast-to-very-fast folding transitions occurring on the microsecond-millisecond timescale. For monitoring folding transitions occurring on slower timescales, additional experimental approaches are desired. Such approaches constitute, e.g. a rapid change in solution conditions to create an unstable non-equilibrium situation and the concomitant detection of the relaxation processes re-establishing equilibrium. Not surprisingly, the use of the stopped-flow technique in conjunction with NMR spectroscopy was already introduced in 1971 [187] and several advancements of the stopped-flow NMR technique have been reported since then covering a wide range of chemical and biochemical applications [188–192]. An alternative approach to the stopped-flow NMR technique is rapid injection NMR. Here, the experimental

3. Kinetic refolding of bovine α -lactalbumin (BLA)

setup is devised such that small volumes (tens of microliters) can be rapidly injected into the NMR tube from above the NMR active volume [79,193]. Rapid mixing techniques in conjunction with NMR have proven to be powerful in attaining detailed insight into folding processes including proteins such as bovine α -lactalbumin, *E.coli* dihydrofolate reductase, and CDK inhibitor p19^{INK4d} [18, 19, 38, 42, 79, 194, 195]. These studies include indirect approaches such as hydrogen exchange pulse labelling which combines labelling of specific amide hydrogens and their subsequent detection using 2D NMR experiments [18, 19]. Direct NMR approaches, on the other hand, encompass real-time 1D [36, 38, 79, 183], 2D [41, 195], 3D [196], diffusion [194], and photo-CIDNP [42, 44] experiments. The dimensionality of the experiment to be chosen essentially depends on the speed of the folding process being studied. Since measurement time increases with dimensionality, fast and slow-to-very-slow processes will mainly be studied by 1D and 2D/3D NMR, respectively.

A further alternative to induce a sudden change in solution conditions for initiation of folding has been introduced by Schwalbe and co-workers [44, 51, 197, 198]. In this approach, a small molecule is released from its cage by a short laser pulse within the NMR active volume by means of laser coupling (Fig. 3.7).

The small molecule subsequently binds to and thereby initiates folding of its target partner. This approach has been successfully applied to both proteins [44, 51] and RNA [197, 198]. Formation of the folded state is routinely detected by application of time-resolved 1D NMR techniques. Photo-induced folding initiation is advantageous over rapid mixing techniques in that the deadtime of the experiment is only limited by the time needed for sufficient illumination of the sample. The approach furthermore enables to acquire highly reproducible data by circumventing the problem of mixing artefacts.

3.1.4. Bovine α -lactalbumin as a model system

Bovine α -lactalbumin (BLA) is the regulatory subunit of lactose synthase which catalyses the last step in the biosynthesis of lactose in the mammary gland [199]. BLA consists of 123 amino acids and the polypeptide chain is folded into an α - and a β -subdomain connected by a Ca²⁺ binding loop at the junction of the two subdomains [200]. Moreover, the polypeptide chain is interconnected by four disulfide bonds (6-120, 61-77, 73-91, and 28-111) making the protein highly stable against exposure to heat [201,202] or chemical denaturants [203]. Figure

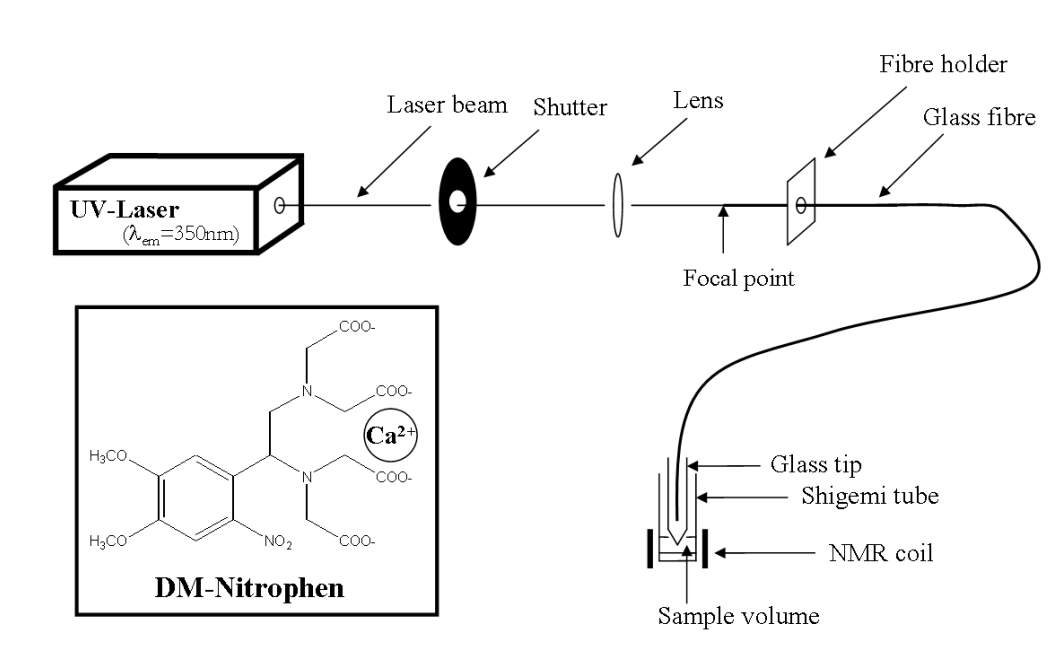


Figure 3.7.: Laser-coupled NMR spectroscopy. Experimental setup for the measurement of photo-induced folding kinetics. The small-molecule ligand to be released may be Ca²⁺ which can be caged by the organic compound DM-nitrophen.

3.8 shows the x-ray structure of Ca²⁺-loaded BLA (holo-BLA).

The native state of BLA binds Ca²⁺ strongly in the loop region indicated in Fig. 3.8 with a dissociation constant of $K_D=0.3\text{nM}$ at pH 7.2 and 25°C [204]. BLA is a popular model system for studies of protein folding [205, 206]. About 30 years ago, it was Kuwajima who showed that chemical denaturation of BLA proceeds via a folding intermediate [207, 208] showing properties similar to the acid-denatured state (A state) of BLA [208]. Subsequent folding studies by Sugai and co-workers revealed that in the dead-time of a stopped-flow refolding experiment a molten globule (MG) folding intermediate is formed which is equivalent to the intermediate state observed during equilibrium denaturation [13, 209]. Interestingly, the same authors also showed that the kinetic folding of the highly homologous protein lysozyme from hen egg white (HEWL) proceeds via a similar structural intermediate although the biological functions of the two proteins are very different (HEWL catalyzes the cleavage of glycosidic bonds) [13, 209]. Folding of the BLA MG intermediate to the native state has been documented to be highly cooperative and to be highly dependent on the Ca²⁺ concentration. Folding in

3. Kinetic refolding of bovine α -lactalbumin (BLA)

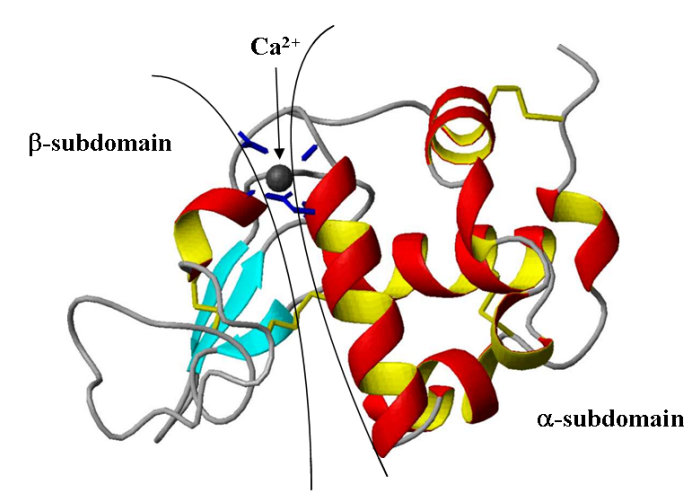


Figure 3.8.: 2.3 Å resolution crystal structure of holo-BLA (M90V) [100]. The Ca^{2+} ion at the junction of the two subdomains is depicted as a black sphere. Ion-liganding moieties (side-chains of D82, D87, D88, and backbone carbonyls of K79 and D84) are shown in blue. Disulfide bonds are coloured yellow.

the absence of Ca^{2+} is about two orders of magnitude slower as compared to folding in the presence of a molar excess of Ca^{2+} [78, 80]. Formation of the native state from the *MG* under strongly native conditions has been observed to occur mono- and biexponentially for the apo and holo proteins, respectively [13, 78, 209]. On the basis of classical double-jump experiments, the second slow phase detected for holo-BLA has been attributed to slow isomerisation around Xaa-Pro peptide bonds in the unfolded state [78]. This phase is likely to be present in the folding of the apo protein as well but probably difficult to detect as it seemingly almost coincides with the major folding phase [209]. Based on these findings, BLA folding can be sketched as follows



where *U*, *MG*, and *N* correspond to the unfolded, molten globule, and native states, respectively. The first step occurs on a micro- to millisecond timescale and is not detectable in stopped-flow experiments. The second step, however, is observable and takes place on a timescale of (milli)seconds to minutes depending on the specific conditions chosen. The accelerating effect of Ca^{2+} on the folding of BLA has been demonstrated to be due to a significant

reduction of the free activation energy in the formation of the transition state during folding of the *MG* to the native state [78,80,210]. BLA has also been extensively used as a model system in the development of NMR methodology for the study of protein folding [40,53,79,211].

3.2. Materials and Methods

3.2.1. CD spectroscopy

CD steady-state spectra in the far-UV region were recorded using a 0.2 mm quartz cuvette, a band width of 1 nm, a data pitch of 0.2 nm, and a response time of 2 s. The scanning speed was 50 nm/min. Each spectrum was the result of 5 accumulations and protein concentrations were typically 60-70 μM . Spectra were taken from 10mM Tris/HCl pH 7 and 4M urea at 35°C.

3.2.2. Stopped-flow spectroscopy

Stopped-flow CD and fluorescence measurements were performed using an π^* -180 CDF spectrometer (Applied Photophysics). Protein refolding experiments were initiated by mixing protein and Ca^{2+} solutions containing 4M urea and 50mM Tris/HCl at pH 7 in a 1:1 volume ratio (35°C). Concentrations after mixing were 100 μM (BLA) and 200 μM (Ca^{2+}), respectively. Detection started after a dead time of 1.8ms. Tryptophan fluorescence was excited at 289nm and monitored above 320nm using a cut-off filter. Excitation at longer wavelengths (298nm) to correct for possible protein concentration effects was shown to have no effect in subsequent kinetic analyses. The CD signal was monitored at 235nm. A cell path length of 2 mm and slit widths of 4 nm were used in the CD measurements. Kinetic traces which were analysed are averages of at least 15 single measurements except for those which were analysed with respect to the urea and Ca^{2+} concentration dependencies of the kinetics. Here, typically 3-5 traces were averaged. Kinetic analysis was performed using the software package SigmaPlot 10.0 (Systat).

In the interrupted refolding assay, BLA and Ca^{2+} solutions were firstly mixed in a 1:6 ratio leading to exactly the same concentrations as after mixing in the conventional stopped-flow experiments. After delay time t_i , the refolding protein was mixed in a 1:6 ratio with 8.8M urea solution at pH 7 leading to 8M urea final concentration. The Ca^{2+} concentration in the

3. Kinetic refolding of bovine α -lactalbumin (BLA)

second mixing step was adjusted such that it equalled $200\mu\text{M}$ after mixing. The data were normalized between 0 (unfolded) and 1 (native) by dividing each unfolding amplitude by the unfolding amplitude of the completely refolded protein. Experiments were carried out using an SX.18MV instrument (Applied Photophysics).

3.2.3. Heterologous expression and purification of ^{15}N -labeled bovine α -lactalbumin

Transformation of *E.coli* BL21 (DE3) cells

E.coli BL21 (DE3) cells (Invitrogen) were transformed with the plasmid pHisBLA-RI carrying the gene for a His₆-tagged version of bovine α -lactalbumin. The procedure was started with thawing an aliquot of $50\mu\text{l}$ cells on ice. In the following, $0.2\mu\text{l}$ of undiluted plasmid solution was added and the cells were then incubated on ice for 5 min. After an incubation at 42°C for 30 seconds (heat shock), the cells were again transferred onto ice (2 min). Cells were supplemented with $250\mu\text{l}$ LB media and shaken for 15 min at 37°C . $5\mu\text{l}$ were then transferred on a predried LB agar plate containing the antibiotic ampicillin ($200\mu\text{g/ml}$). Cells were spread on the plate using a sterile bent glass rod. The plate was incubated over night at 37°C .

Protein production

Several colonies were scraped off the plate on the next day and resuspended in $200\mu\text{l}$ of ^{15}N -M9 minimal medium supplemented with ampicillin ($200\mu\text{g/ml}$); OD_{600} was measured. At 5.30 pm, 100ml of ^{15}N -M9-ampicillin medium were inoculated to an initial OD_{600} of 0.00001 (corresponding to $5\mu\text{l}$ cell suspension of $\text{OD}_{600}=0.2$). This pre-culture was shaken over night at 37°C and transferred into 1l of ^{15}N -M9-ampicillin medium on the next morning followed by continued shaking at 37°C . At an OD_{600} of 0.7-0.9, protein production was induced by adding 0.5g of isopropyl- β -D-thiogalactopyranoside (IPTG). The protein was stored in the form of inclusion bodies in the cell interior. After shaking of 3.5-4 hours at 37°C , the cells were harvested by centrifugation (GSA rotor, 6500rpm, 30 min, 4°C). Cells were resuspended in 20ml sonication buffer, lysed by four sonication steps at 50 % for 1min (1min pause in between) and centrifuged again (JA25.50 rotor, 15000rpm, 30 min, 4°C). The cell pellet was subsequently washed using 20ml washing buffer and centrifuged as before. The course of

protein expression upon induction with IPTG and the washing procedure after cell harvest were documented by SDS-PAGE [212].

Protein purification

Cells were resuspended in 2ml H₂O, 1.7g urea and 12ml solubilisation buffer were added, and the suspension was supplemented to 20mM β -mercaptoethanol (1.4 μ l/ml). The resulting mixture was centrifuged (JA25.50 rotor, 15000rpm, 30min, 4°C). The supernatant was diluted 1:1 with IMAC I buffer in preparation of the first purification step. The protein solution was loaded onto 5ml Ni-NTA column (Superflow, Qiagen) which had been equilibrated before with 40ml IMAC I. The column is washed with 10ml IMAC I and 15ml IMAC II. IMAC II has a higher salt (NaCl) concentration than IMAC I and therefore removes protein components which might still stick to the column due to electrostatic interactions. His₆-tagged α -lactalbumin was then eluted stepwise with 20mM, 50mM, 100mM, 250mM, and 500mM imidazole (mixture of IMAC I and IMAC III, 15ml each step, 5ml/min flow rate).

Table 3.1.: Elution of His₆-tagged α -lactalbumin from a Ni-NTA column using imidazole. Protocol for mixing buffers IMAC I and IMAC III to achieve respective imidazole concentrations.

[Imidazole] [mM]	IMAC III [ml]	IMAC I [ml]
20	0.6	14.4
50	1.5	13.5
100	3	12
250	7.5	7.5
500	15	0

Column fractions were checked by SDS-PAGE [212]. Fractions containing α -lactalbumin were pooled and dialysed (Zellutrans ROTH, Cat E659.1) against 7l of refolding buffer supplemented with urea and oxidised/reduced glutathione as a redox shuffling system (for correct disulfide bond formation). This was followed by two further dialysis steps against 7l of refolding buffer supplemented with glutathione and 7l of 'pure' refolding buffer, respectively. Every step of dialysis took 12 hours. After the third step, the concentration of His₆-tagged

3. Kinetic refolding of bovine α -lactalbumin (BLA)

α -lactalbumin was determined by absorption (extinction coefficient of His₆- α -lactalbumin (280nm) = 28840M⁻¹cm⁻¹). In the following, the His₆-tag was cleaved off the protein using 420U trypsin (Sigma bovine pancreas Cat T-8802, dissolve to 7600 U/ml in 25mM acetic acid) per mg of protein. Incubation proceeded over night and was stopped with 0.5mM 4-(2-Aminoethyl)benzenesulfonyl fluoride hydrochloride (AEBSF, fresh stock solution of 50mM in H₂O) acting as an inhibitor of trypsin. The protein was further purified by ion exchange chromatography. After the sample had been diluted 1:2 with H₂O and 1:2 with 20mM sodium phosphate, 10mM NaCl, pH 6.5, it was loaded onto a HiPrep Q FF (Amersham) column equilibrated in the same sodium phosphate buffer. α -lactalbumin was eluted using a NaCl gradient (0-300mM in 150ml, flow rate 4ml/min). Fractions were checked by SDS-PAGE [212] and those containing α -lactalbumin were pooled and subsequently concentrated using an Amicon chamber and a centricon having molecular weight cutoffs of 10kDa. The concentration of α -lactalbumin was determined by the absorption at 280nm. The protein was finally lyophilised and could be stored at -20°C.

The expression and purification procedure which has been followed here was established and optimised in earlier work [68,213].

3.2.4. NMR resonance assignment of methyl and methylene groups in the native state

Assignment of backbone amide groups in BLA under the refolding conditions (4M urea, 50mM Tris pH 7) was started from the previously reported backbone amide assignment of BLA at pH 6.3 and 20°C [80]. The acquisition of ¹H,¹⁵N HSQC [214] spectra of a 1.2mM BLA sample in 4M urea, 50mM Tris pH 7 and in presence of a 5-fold excess of Ca²⁺ at six different temperatures (20°C, 22°C, 24°C, 28°C, 32°C, and 35°C) using an 800MHz NMR spectrometer equipped with a 5mm ¹H(¹³C/¹⁵N) z-axis gradient cryogenic probe was performed in order to assign as many backbone amide groups as possible by transfer of the above mentioned assignment [80] to the conditions applied here. 4 scans were acquired for each spectrum using 256 increments in the indirect dimension (spectral width: 2351 Hz). 2048 complex points were collected in the proton dimension (spectral width: 12821 Hz). A qsrine window function was used for apodization.

Further backbone amide assignment was based on a 3D ¹H,¹⁵N NOESY-HSQC [215] data

set acquired at 35°C. Data acquisition was done at a 600MHz NMR spectrometer equipped with a 5mm $^1\text{H}(^{13}\text{C}/^{15}\text{N})$ z-axis gradient cryogenic probe by recording 8 scans with 1024 complex points in ω_3 (spectral width: 8389 Hz), 96 increments in ω_2 (Echo-Antiecho, spectral width: 1702 Hz) and 180 increments in ω_1 (States-TPPI, spectral width: 8389 Hz), respectively. A q-sine window function was used for apodization. ^1H chemical shifts were referenced relative to 2,2-Dimethyl-2-silapentane-5-sulfonate, sodium salt (DSS). ^{15}N chemical shift referencing was done indirectly [216]. Assignment of valine γ methyl, leucine δ methyl, isoleucine γ methylene, γ methyl, and δ methyl groups was based on amide-proton-methyl(methylene)-proton NOEs. Backbone amide and methyl/methylene group assignments were carried out using CARA (computer aided resonance assignment, www.nmr.ch). Standard Bruker pulse programs were used throughout the NMR measurements.

3.2.5. Time-resolved NMR spectroscopy

Methyl/methylene group refolding kinetics were recorded under the experimental conditions employed in the stopped-flow experiments and followed in real-time by recording consecutive single-scan 1D ^1H NMR spectra using the jump-return-echo excitation scheme optimised for the methyl/methylene group region ($\tau_{\text{JR}}=80\mu\text{s}$) [217]. Experiments were carried out using a 700 MHz NMR spectrometer equipped with a 5mm $^1\text{H}(^{13}\text{C}/^{15}\text{N})$ z-axis gradient cryogenic probe. In order to improve the time resolution in the real-time experiment, an acquisition time of 105ms (2048 points) and a recycle delay of 500ms were chosen. Urea signals were suppressed by selective low power irradiation. The delay between acquisitions of two subsequent scans was 638ms. 64 single-scan spectra were recorded before two equivalents of Ca^{2+} ions were released from the caged compound DM-nitrophen by a short laser pulse ($\tau_{\text{Laser}} = 200\text{ms}$). The application of the laser pulse was accomplished by a UV argon ion laser ($\lambda_{\text{max}} = 350\text{nm}$, 3.7). Maximum output power at the end of the optical fiber was 4.5W. Refolding was monitored by a second set of 64 single-scan spectra in an interleaved manner by inserting a variable delay τ_{var} between the end of the laser irradiation (τ_{Laser}) and the start of the acquisition of the first spectrum after release of Ca^{2+} . Experiments with values of $\tau_{\text{var}}=0\text{ms}$, 100ms, 200ms, and 300ms were undertaken. Each experiment was repeated three times. Temperature increases due to laser irradiation were less than 0.5°C. Data processing was performed with Felix 2004 (Accelrys). Data of corresponding delays τ_{var} were summed up in order to improve

3. Kinetic refolding of bovine α -lactalbumin (BLA)

signal-to-noise. Time-dependent peak volumes were extracted, normalised, and finally plotted vs. time in order to obtain kinetic parameters using single, double, and triple exponential functions. In order to elucidate whether a more complex exponential function fits the data significantly better than a simpler exponential function an F test analysis for each peak analysed was undertaken [218,219]. In general, the corresponding F value is calculated as follows:

$$F = \frac{\nu_2(\chi_1 - \chi_2)}{\chi_2(\nu_1 - \nu_2)} \quad (3.6)$$

with $\nu = N - n$ (N : number of data points, n : number of model parameters with $n_2 > n_1$) and χ being the degrees of freedom for a certain model and the sum of squared errors between the experimental values and the model-based predicted values, respectively. In other words, the F value depicts the ratio of the relative difference in sum-of-squares and the relative difference in degrees of freedom. The computed F values are then evaluated using an exact F distribution [220]. A critical F value can be defined as $F_{0.01, N_1, N_2}$ with $N_1 = n_2 - n_1$ and $N_2 = N - n_2$ indicating that, given a number of data points N and the numbers of model parameters n_1, n_2 , there is only a 1% probability that the improvement in the fit by the more complex model is achieved by chance, provided that the computed F value is at least as large as the critical F value. The F test was performed using the software package SigmaPlot 10.0 (Systat).

3.3. Results

3.3.1. Tryptophan fluorescence and CD spectroscopic steady-state investigations

As described below, Ca^{2+} -induced refolding experiments were conducted by means of stopped-flow fluorescence spectroscopy at an intermediate urea concentration (4M). The initial and final fluorescence values at various urea concentrations allowed to analyse the urea equilibrium unfolding behaviour of the apo and holo states of BLA. As can be seen from Fig. 3.9, the two forms of the protein exhibit very different unfolding transitions pointing to a strong dependence of stability on the concentration of Ca^{2+} ions.

From previous studies it is known that, at intermediate concentrations of denaturant at 25°C and neutral pH, apo-BLA populates the unfolded (U), MG (A), and native (N) states [221].

It has also been demonstrated that, in 60mM Tris/HCl pH 8 at 37°C (conditions similar to those employed in the present work), the tertiary structure of the *N* state is completely disrupted [222] yielding the apo-*MG* conformation (α -lactalbumin is known to populate a molten globular conformation in the absence of Ca^{2+} and at low ionic strength (apo-*MG*) which is similar to the low pH-induced *A* state molten globule [223]). The *U*, *A*, and apo-*MG* states are therefore expected to be significantly populated under the conditions chosen in the present study (4M urea, 50mM Tris/HCl pH 7, 35°C). As these states have no persistent tertiary structure, the unfolding transition shown in Fig. 3.9 displays the disruption of the secondary structure present in the variously unfolded conformers of apo-BLA. The unfolding occurs in a two-stepped manner which is in accordance with earlier results [203].

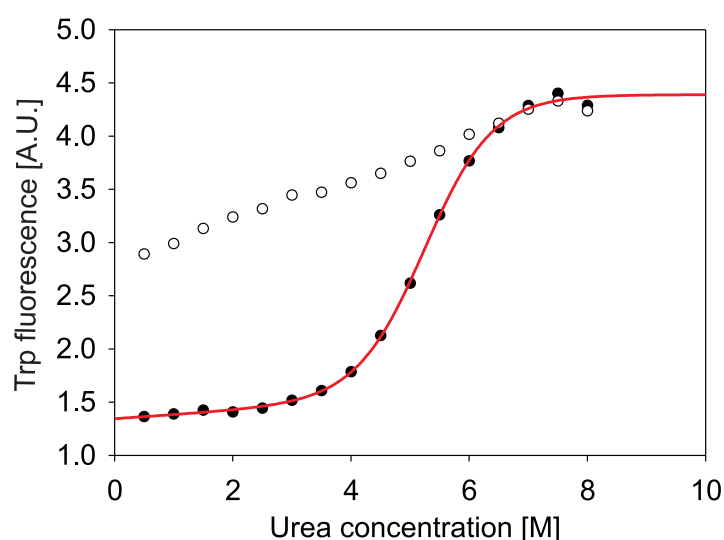


Figure 3.9.: Urea equilibrium unfolding of apo- (empty circles) and holo-BLA (filled circles). Shown is the tryptophan fluorescence above 320nm in dependence of the urea concentration (data taken from stopped-flow measurements). The red line represents the fit to the unfolding data in the presence of two eq. Ca^{2+} assuming two-state unfolding (see text for details).

The situation in the presence of two eq. Ca^{2+} is different. Here, at intermediate concentrations of denaturant, the *U*, *A*, and *N* states will be significantly populated [221]. Compared with the situation for the apo protein, the addition of Ca^{2+} at 4M urea concentration, therefore, leads to a shift from the molten globule states to the *N* state. Since Ca^{2+} does not bind to the *U* state [221] and the *A* state binds Ca^{2+} only very weakly (about 10^3 M^{-1}) [78], it will mainly be the apo-*MG* population which binds Ca^{2+} and subsequently folds to the *N*

3. Kinetic refolding of bovine α -lactalbumin (BLA)

state. The apo-*MG* has previously been shown to undergo a folding transition upon binding of Ca^{2+} [224, 225].

The urea-induced unfolding of the holo protein shows a sigmoidal transition (Fig. 3.9). Based on the assumption of simple two-state unfolding, the data were analyzed according to the approach by Santoro and Bolen [226] yielding $\Delta G_0^{N-U} = 5.6 \text{ kcal mol}^{-1}$ and $m = 1.06 \text{ kcal mol}^{-1} \text{ M}^{-1}$ for the free energy of unfolding in the absence of urea and for the m value reporting on the amount of surface area exposed to solvent upon unfolding, respectively. The midpoint of the unfolding transition is defined as the $\Delta G_0^{N-U}/m$ ratio and was accordingly calculated to be 5.3M. A sigmoidal urea-induced unfolding transition of Ca^{2+} -loaded BLA with a midpoint of transition of about 5M was also observed in an earlier study in which the experimental conditions were slightly different (20 μM instead of 100 μM protein, 1.5 instead of two equivalents of Ca^{2+} ions, pH 8 instead of pH 7, and 20 $^\circ\text{C}$ instead of 35 $^\circ\text{C}$) [203].

It has been previously shown that the addition of Ca^{2+} to 4M urea-denatured BLA is accompanied by a significant decrease in α -helicity [51]. These results were confirmed by the measurement of far-UV CD spectra of BLA in the absence of Ca^{2+} and in the presence of two equivalents of Ca^{2+} under the refolding conditions applied in the present work. The spectra are shown in Fig. 3.10. The shift of the equilibrium between the apo-*MG* and *N* states is evidenced by a decrease in the mean residue ellipticity at 222 nm (a measure of the α -helical content) by 3600 $\text{deg cm}^2 \text{ dmol}^{-1}$ upon addition of Ca^{2+} .

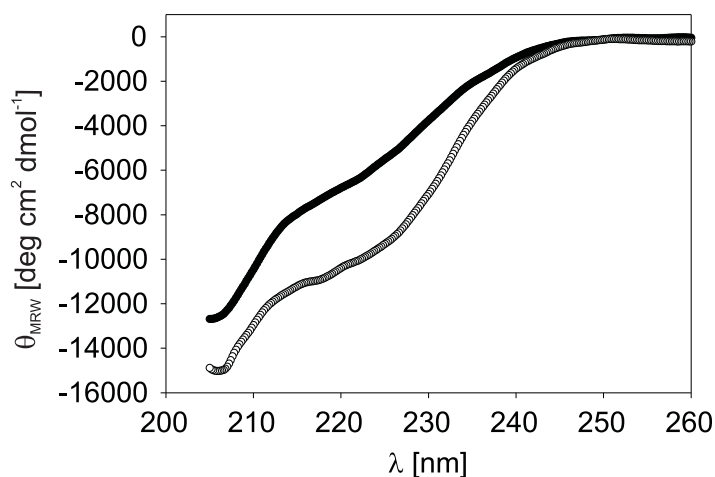


Figure 3.10.: Far-UV CD spectra of BLA in the absence of Ca^{2+} (filled circles) and in the presence of two eq. Ca^{2+} (empty circles). Shown is the mean residue ellipticity.

3.3.2. Folding kinetics by stopped-flow spectroscopy

Refolding as seen by tryptophan fluorescence, far-UV CD, and near-UV CD

The Ca^{2+} -induced refolding of BLA was investigated under non-equilibrium conditions by stopped-flow methods using far-UV CD and tryptophan fluorescence spectroscopies. Refolding was initiated by mixing protein and Ca^{2+} solutions in a 1:1 ratio at a constant urea concentration of 4M. The kinetic data (Fig. 3.11) were fitted using a triple exponential function which, on the basis of an F statistical analysis [218, 219], was shown to significantly better fit the data than a double exponential function. Refolding rate constants are given in Table 3.2.

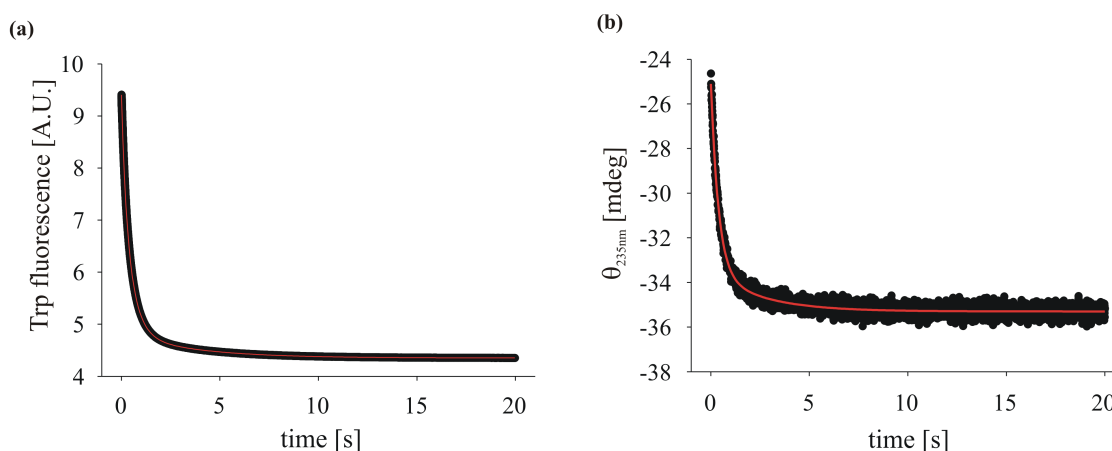


Figure 3.11.: Stopped-flow spectroscopic detection of the Ca^{2+} -induced refolding of BLA using tryptophan fluorescence (a) and far-UV CD (235nm) (b). Red lines represent triple exponential fits to the data.

Table 3.2.: Results from triple exponential fits to the data shown in Fig. 3.11

Probe	k_1 [s^{-1}]	A_1 [%]	k_2 [s^{-1}]	A_2 [%]	k_3 [s^{-1}]	A_3 [%]
Trp fluorescence	4.3 ± 0.04	30.8 ± 0.8	2.0 ± 0.01	60.0 ± 0.8	0.29 ± 0.001	9.8 ± 0.1
Far-UV CD (235nm)	9.5 ± 4.3	8.9 ± 3.4	2.4 ± 0.1	74.9 ± 3.4	0.38 ± 0.01	15.7 ± 0.7

The existence of three kinetic phases in the stopped-flow-detected kinetics shows that the Ca^{2+} -induced refolding of BLA cannot be described by a simple two-state transition only involving the apo-*MG* and *N* states but rather involves the population of additional states which may be due to either folding intermediates [227] or conformational heterogeneity in the initial

3. Kinetic refolding of bovine α -lactalbumin (BLA)

state [15, 228] or final state [229, 230]. Comparison of the results from kinetics monitored by far-UV CD and tryptophan fluorescence reveals that the same processes are being observed by both methods (Table 3.2). Refolding was also monitored using near-UV CD spectroscopy (277 nm). Data analysis shows that a double exponential provides the best fit to the data (Fig. 3.12) which is in contrast to the result obtained for tryptophan fluorescence and far-UV CD. Double exponential fits to the data of the latter probes show, however, negligible differences compared to near-UV CD (Table 3.3). It is expected that the near-UV CD data could, in principle, also be best described by a triple exponential function if signal-to-noise would be improved by averaging over a larger amount of separate measurements.

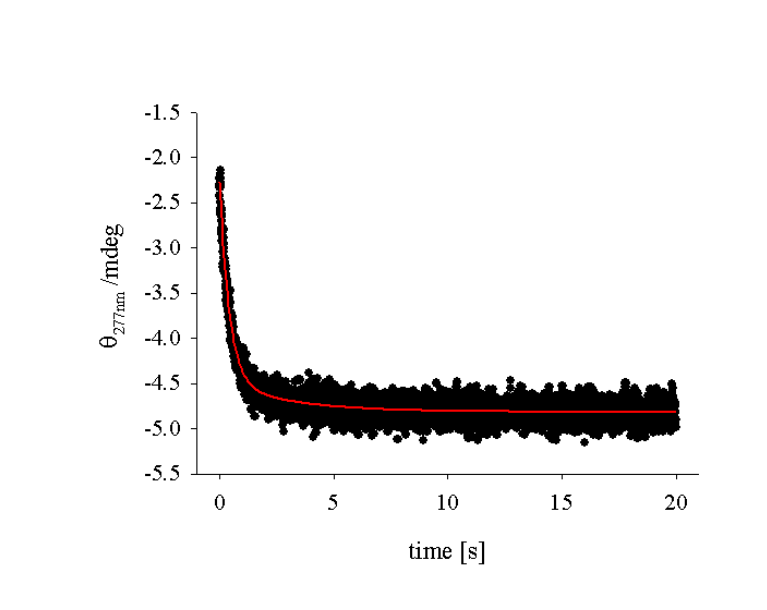


Figure 3.12.: Kinetic refolding of BLA detected by stopped-flow near-UV CD (277 nm). The red line represents a double exponential fit to the data.

Table 3.3.: Results from double exponential fits to the data shown in Figs. 3.11, 3.12

Probe	k_1 [s^{-1}]	A_1 [%]	k_2 [s^{-1}]	A_2 [%]
Trp fluorescence	2.86 ± 0.01	81.7 ± 0.1	0.52 ± 0.004	18.3 ± 0.1
Far-UV CD (235nm)	2.61 ± 0.04	82.3 ± 0.8	0.40 ± 0.01	17.7 ± 3.5
Near-UV CD (277nm)	2.38 ± 0.07	86.5 ± 1.4	0.34 ± 0.03	13.5 ± 8.3

Urea concentration dependence

The influence of the urea concentration on the refolding kinetics was investigated in analogy to a chevron type of kinetic analysis. In such an analysis, the refolding and unfolding kinetics are measured at varying concentrations of denaturant. The plot of the natural logarithm of the corresponding rate constant against the denaturant concentration may, in principle, give access to valuable information such as to the occurrence of folding intermediates. Since folding of BLA is not induced by a change in the denaturant concentration but by the addition of Ca^{2+} , only folding transitions can be monitored so that the 'unfolding limb' of the chevron plot can not be determined.

The urea concentration dependence of the refolding kinetics of BLA was determined using tryptophan fluorescence since it provides a significantly better signal-to-noise in comparison to far- and near-UV CD (cf. Figs. 3.11, 3.12). Refolding was measured at constant urea concentrations in 0.5M steps between 0.5M and 8M. The results from triple exponential fits to the kinetics between 0.5M and 5M are shown in Fig. 3.13.

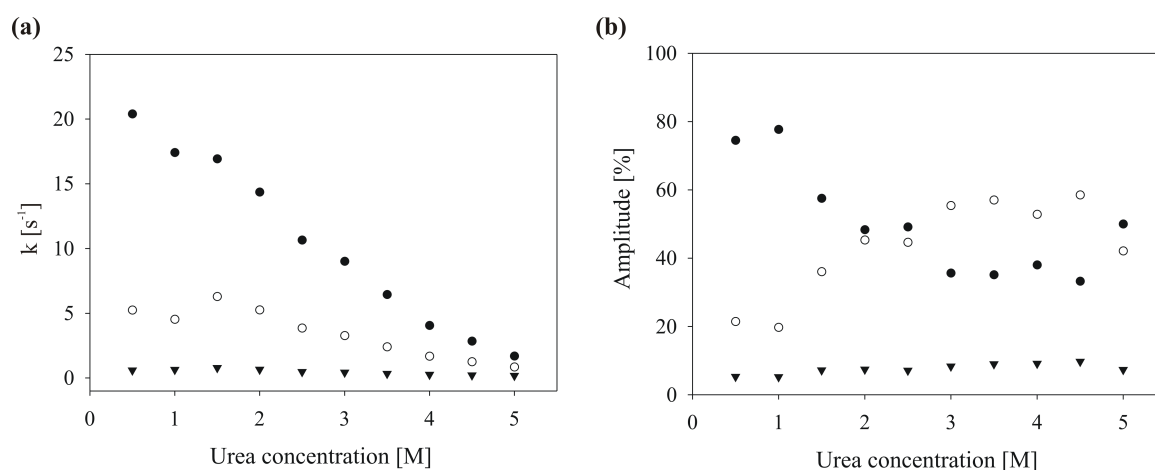


Figure 3.13.: Urea concentration dependence of stopped-flow-fluorescence-detected kinetic refolding of BLA. Filled circles, empty circles, and inverted triangles indicate the very fast, fast, and slow refolding phases, respectively. (a) Refolding rate constants. (b) Refolding phase amplitudes.

It can be seen that refolding is fastest at 0.5M urea and progressively decreases in speed as the urea concentration increases. Refolding rate constants at 0.5M urea are $k_1 = 20.4 \pm 0.1 \text{ s}^{-1}$, $k_2 = 5.3 \pm 0.1 \text{ s}^{-1}$, and $k_3 = 0.6 \pm 0.02 \text{ s}^{-1}$. k_1 is most affected by a change in urea

3. Kinetic refolding of bovine α -lactalbumin (BLA)

concentration whereas k_2 is less affected. k_3 is least affected when increasing the urea concentration. It can, therefore, be concluded that k_1 and k_2 correspond to folding events while the behaviour of k_3 might indicate that this phase corresponds to slow isomerisation around Pro-Xaa peptide bonds in the unfolded state, a process known to be independent of the denaturant concentration [231,232]. The values of the amplitudes associated with the three rate constants have also been determined in dependence of the urea concentration (Fig. 3.13b). These indicate the different contributions of the folding phases to the total change in fluorescence upon mixing protein and Ca^{2+} solutions. While A_1 and A_2 decrease and increase significantly, respectively, when going to higher urea concentrations, A_3 remains largely unaffected at values below 10 % in the studied concentration range.

At urea concentrations above 5M, the folding phases seem to merge as double and single exponentials give the best fits to the kinetic data at concentrations of 5.5M and 6M, respectively (data not shown). Interestingly, a fast decrease in tryptophan fluorescence is followed by a slow increase at 6.5M urea (data not shown). Moreover, stopped-flow experiments at concentrations between 7M and 8M detect merely slow increases in tryptophan fluorescence (data not shown). A more detailed analysis as to the origin of the apparent change in behaviour for urea concentrations above 6M has not been undertaken so far. For urea concentrations concerned, changes in fluorescence were much less pronounced as seen at concentrations below 6M.

Ca^{2+} concentration dependence

Refolding kinetics were further studied at varying Ca^{2+} concentrations to investigate the effect of Ca^{2+} on the three folding phases (Fig. 3.14a). Again, tryptophan fluorescence was the probe of choice in stopped-flow experiments which is due to signal-to-noise considerations.

It is seen that refolding is markedly accelerated when the Ca^{2+} concentration is increased which is in agreement with previous studies [80,210]. The largest effect is seen for k_1 which increases by an order of magnitude between 0.1mM and 10mM Ca^{2+} . k_2 increases as well (by a factor of 3) but the effect is much less pronounced. No effect was detected for k_3 . Refolding rate constants at 10mM Ca^{2+} were determined to be $k_1 = 33.2 \pm 0.8 \text{ s}^{-1}$, $k_2 = 4.0 \pm 1.0 \text{ s}^{-1}$, and $k_3 = 0.4 \pm 0.6 \text{ s}^{-1}$.

The differences between the initial and final fluorescence values for each Ca^{2+} concentra-

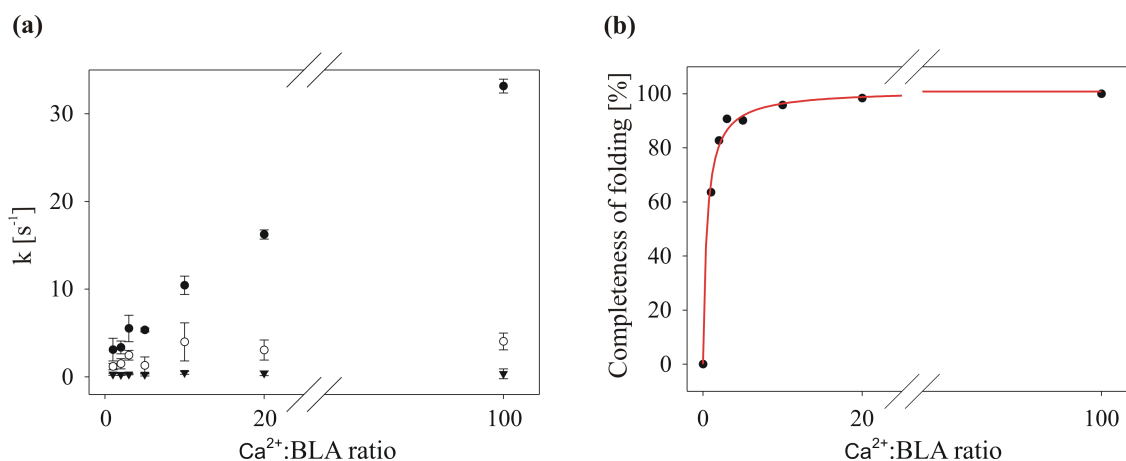


Figure 3.14.: Ca^{2+} concentration dependence of stopped-flow-fluorescence-detected kinetic refolding of BLA. (a) Filled circles, empty circles, and inverted triangles indicate the very fast, fast, and slow refolding phases, respectively. Error bars denote fitting errors. (b) Apparent binding constant of Ca^{2+} to BLA ($K_{bind} = 1.9 \cdot 10^4 \text{ M}^{-1}$). The amplitude of the total change in fluorescence upon addition of 100 eq. Ca^{2+} was defined as 100% 'folding completeness'. Changes in amplitude at lower Ca^{2+} :BLA ratios were scaled accordingly. The red line shows the fit to a one-site saturation binding model.

tion also allowed to determine the apparent binding constant of Ca^{2+} to the apo-*MG* under the refolding conditions which was calculated to be $1.9 \cdot 10^4 \text{ M}^{-1}$ (Fig. 3.14b). At low urea concentrations, the binding constant for goat α -lactalbumin has been shown to be independent of the urea concentration while a significant drop is seen in the *apparent* binding constant when increasing the urea concentration [204]. The binding constant for BLA under the present refolding conditions is therefore expected to be significantly higher than $1.9 \cdot 10^4 \text{ M}^{-1}$. Binding of Ca^{2+} should, therefore, be strong and have no effect on the analysis of the conformational changes detected in the stopped-flow kinetics.

Interrupted refolding

To discriminate between parallel folding pathways and a strictly sequential formation of the N state through a folding intermediate the time evolution of the N state was monitored in an interrupted refolding experiment [183]. In this experiment, refolding of BLA is induced in a first mixing step. After a variable delay time t_i , a second mixing step is performed leading

3. Kinetic refolding of bovine α -lactalbumin (BLA)

to complete unfolding of molecules that had already reached the native state after t_i . The unfolding amplitude observed after the second mixing step is therefore a measure for the amount of native protein that was present when refolding was interrupted. Unfolding was measured after refolding times t_i ranging from 33ms to 15 seconds and the resulting unfolding amplitudes derived from a fit to the data were plotted vs. the corresponding delay times t_i (Fig. 3.15).

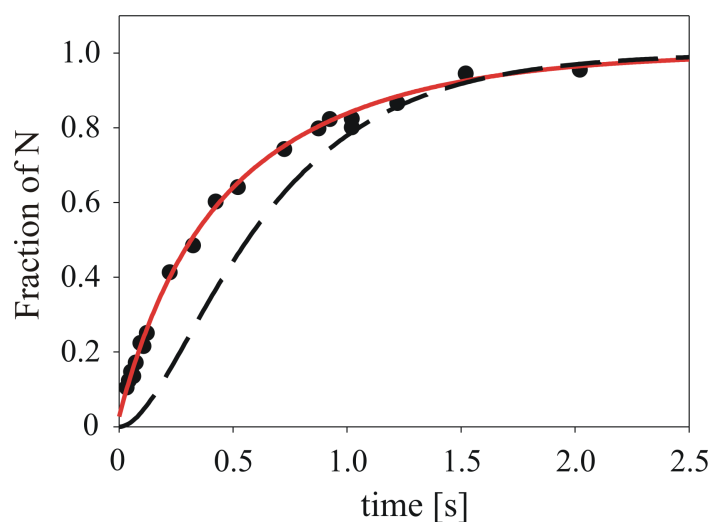


Figure 3.15.: Time course of formation of native BLA measured by interrupted refolding experiments. The solid line (red) represents a double exponential fit of the data with the parameters given in Table 3.7. The dashed line is a simulation of folding according to the sequential model (eq. 3.4) with $k_1 = 5 \text{ s}^{-1}$ and $k_2 = 2 \text{ s}^{-1}$. These values are in accordance with the rate constants observed with other methods (see Tables 3.2, 3.7).

The results show that the N state of BLA is formed in two reactions with rate constants of $k_1 = 4.7 \pm 3.4 \text{ s}^{-1}$ ($27 \pm 26 \%$ amplitude) and $k_2 = 1.5 \pm 0.3 \text{ s}^{-1}$ ($70 \pm 27 \%$ amplitude) which agree well with the two fastest time constants detected in fluorescence and CD experiments and with the time constants observed in NMR (see Tables 3.2, 3.7). The fit extrapolates to about 2.7 % native protein at $t=0$. This residual may be due to experimental error or might indicate that a small amount of native BLA is formed during the rapid burst phase reaction within the first millisecond of folding. A single exponential fit cannot describe the data equally well and leads to an extrapolated 8 % of native protein present at time $t=0$. The slowest reaction observed with optical probes and NMR is not observed in interrupted refolding experiments.

This might be due to the small amplitude of this reaction. Comparison of the measured time course of formation of native BLA to the expected time course for folding according to a sequential model assuming an obligatory folding intermediate on a strictly linear pathway clearly shows the absence of the expected lag phase (Fig. 3.15).

3.3.3. Heterologous expression and purification of ^{15}N -labeled bovine α -lactalbumin

Transformation of BL21 (DE3) *E. coli* cells and expression of His₆-tagged α -lactalbumin were performed as described in section 3.2.3. Fig. 3.16 shows the SDS-PAGE [212] result of the expression and affinity chromatography (Ni-NTA) procedures.

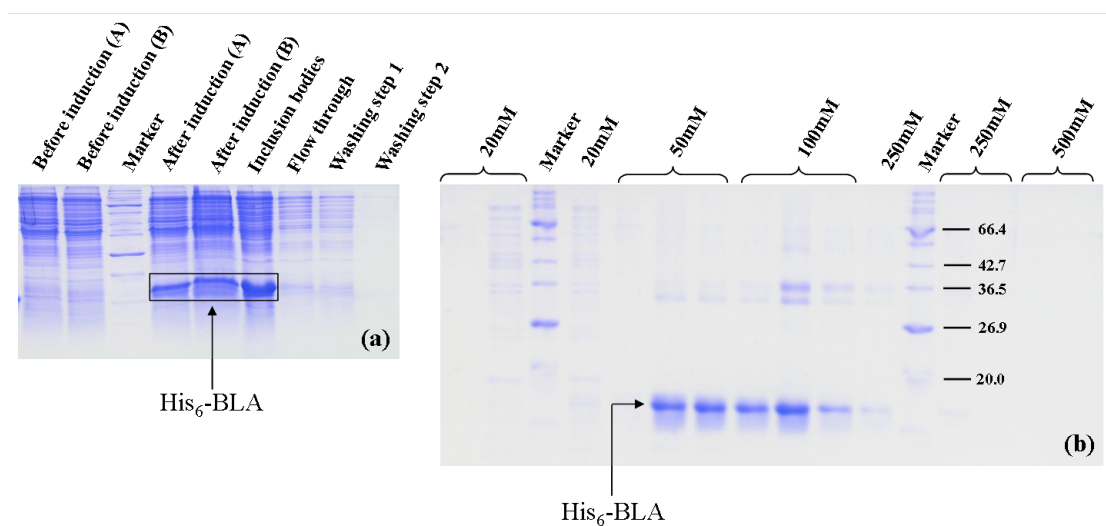


Figure 3.16.: SDS-PAGE [212] analysis of expression and purification of His₆-tagged α -lactalbumin. (a) (A) and (B) indicate different cultures. (b) His₆-BLA elutes primarily at imidazole concentrations between 50mM and 100mM. Numbers on the right side of the gel refer to molecular weights of proteins (in kDa) contained in a broad range marker.

It is evident that expression of His₆- α -lactalbumin only occurs after induction with IPTG. The lanes referring to the inclusion bodies, the flow through, and the washing steps further indicate that the protein is almost exclusively found in the inclusion bodies. The purification of the resolubilised inclusion bodies via a Ni-NTA column results in absolute protein yields of 12-40 mg/l depending on the specific batch. The lanes referring to imidazole concentrations

3. Kinetic refolding of bovine α -lactalbumin (BLA)

of 50mM to 100mM in Fig. 3.16 indicate that the contribution of higher-molecular-weight proteins (MWs of about 30-55 kDa) to the protein yield is negligible. It turns out that the addition of urea and solubilisation buffer to the inclusion bodies leads to efficient resolubilisation if it is allowed to proceed over night at room temperature.

Purification of α -lactalbumin after His tag cleavage by ion exchange chromatography is straightforward. Fig. 3.17 displays the profile of an exemplary chromatographic run.

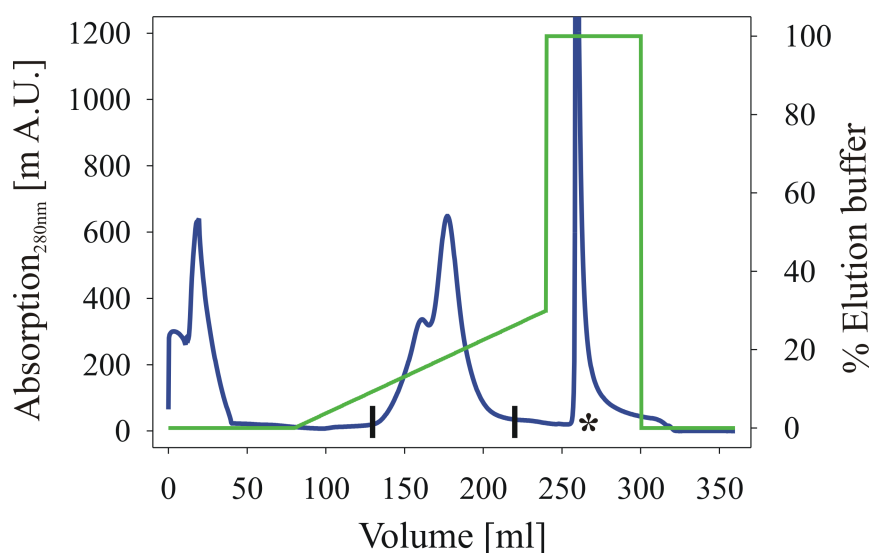


Figure 3.17.: Ion exchange chromatography of α -lactalbumin after His tag cleavage. Blue and green lines indicate the absorption at 280nm and the NaCl gradient used for protein elution, respectively. Cleaved BLA elutes as a double peak at NaCl concentrations between 100mM and 250mM. The sharp peak at about 260ml volume has been cut off for illustrative purposes. For the vertical bars and the asterisk refer to Fig. 3.18.

Several rounds of expression and purification showed that the double peak elution as seen in Fig. 3.17 is reproducible. The origin of the heterogeneous elution of the protein could, however, not be clarified. It is also noted that the order of the dilution steps of the cleaved α -lactalbumin before loading onto the ion exchange column is important to prevent precipitation of the protein. Water and sodium phosphate buffer must not mixed before added to the protein solution. The cleaved protein is first to be diluted with water and subsequently with the buffer solution. α -lactalbumin otherwise precipitates due to an abrupt change in ionic strength (the refolding buffer contains 50mM NaCl whereas the sodium phosphate buffer contains only

10mM NaCl). It is also recommended to concentrate the protein solution before the dilution steps to reduce the total volume which is to be loaded onto the ion exchange column.

Fig. 3.18 shows that α -lactalbumin elutes from the column in high purity as additional bands in the SDS-PAGE [212] analysis are absent.

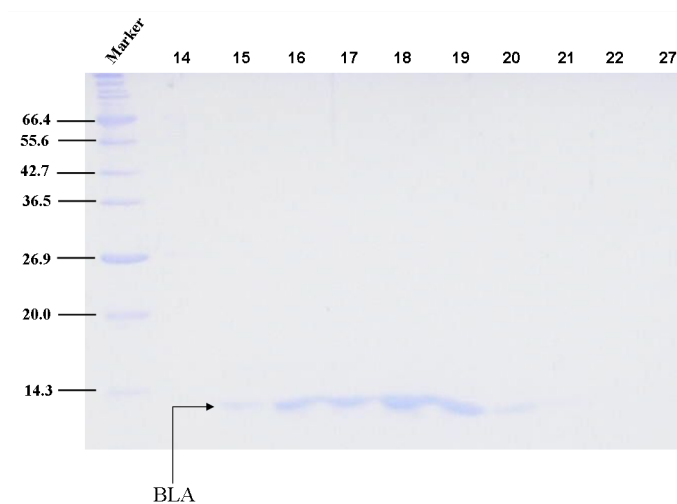


Figure 3.18.: SDS-PAGE [212] analysis of the fractions of an ion exchange chromatographic run after His tag cleavage. Lanes 14-22 refer to the fractions within the region of the double peak in Fig. 3.17 (indicated by vertical bars). Lane 27 refers to the sharp peak in Fig. 3.17 (indicated by an asterisk). Numbers on the left side of the gel refer to molecular weights of proteins (in kDa) contained in a broad range marker.

Protein yields after His tag cleavage and purification are very low (on the order of 2-5 mg/l). It seems as if digestion of His₆- α -lactalbumin with trypsin results in a significant loss of protein. SDS-PAGE [212] analysis also reveals that the fractions corresponding to the peaks at the beginning and the end of the chromatographic run, respectively, do not contain any protein (data not shown).

3.3.4. Assignment of backbone amide and side-chain aliphatic resonances

For the assignment of methyl and methylene resonances as probes of native state formation during folding, the following strategy was applied. First, ¹⁵N-labelled bovine α -lactalbumin

3. Kinetic refolding of bovine α -lactalbumin (BLA)

was expressed and purified as outlined in the previous section. Second, a ^{15}N , ^1H HSQC [214] spectrum was recorded under the conditions of an earlier study (50mM HEPES pH 6.3 at 20°C) where an almost complete backbone amide resonance assignment had been reported [80]. This spectrum was compared to a spectrum under conditions resembling the situation when refolding has taken place (4M urea, 50mM Tris/HCl pH 7 at 20°C). At this stage, as many resonances as possible were assigned on the basis of the previously published assignment [80]. Third, a temperature series of ^{15}N , ^1H HSQC spectra was conducted to transfer the assignment from 20°C to 35°C , the temperature at which refolding experiments were performed. Fourth, a ^{15}N NOESY HSQC [215] data set at 35°C was acquired and analysed to complete the backbone amide resonance assignment. Assignment of methyl and methylene resonances was finally undertaken on the basis of amide-methyl(methylene) NOEs.

Fig. 3.19 displays the ^{15}N , ^1H HSQC spectrum of BLA in 50mM HEPES pH 6.3 at 20°C .

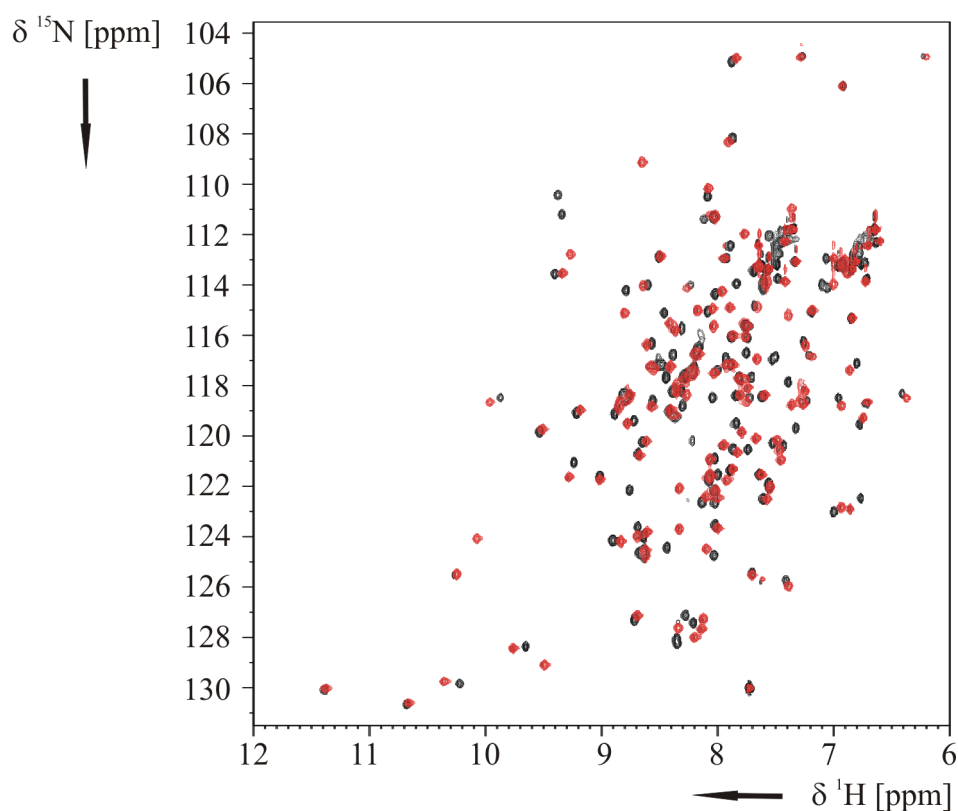


Figure 3.19.: BLA in 50mM HEPES pH 6.3 at 20°C in the absence of Ca^{2+} (black) and presence of two eq. Ca^{2+} (red). Protein concentration was 1.23mM. Spectra shown are not referenced.

The spectrum in black was recorded in the absence of Ca^{2+} . The dispersion of chemical shifts and the sharpness of the peaks clearly show that BLA behaves as a well-folded protein under these conditions. Only at higher temperatures, the tertiary structure disrupts resulting in the apo-*MG* conformation [46, 222]. The addition of Ca^{2+} (red spectrum) therefore does not change the overall spectral characteristics although a significant number of peaks exhibits considerable shifts. In the following, ^{15}N , ^1H HSQC spectra of BLA in 4M urea, 50mM Tris/HCl pH 7 were acquired. Fig. 3.20 shows the protein at 35°C in the absence of Ca^{2+} representing the initial state of the refolding reaction.

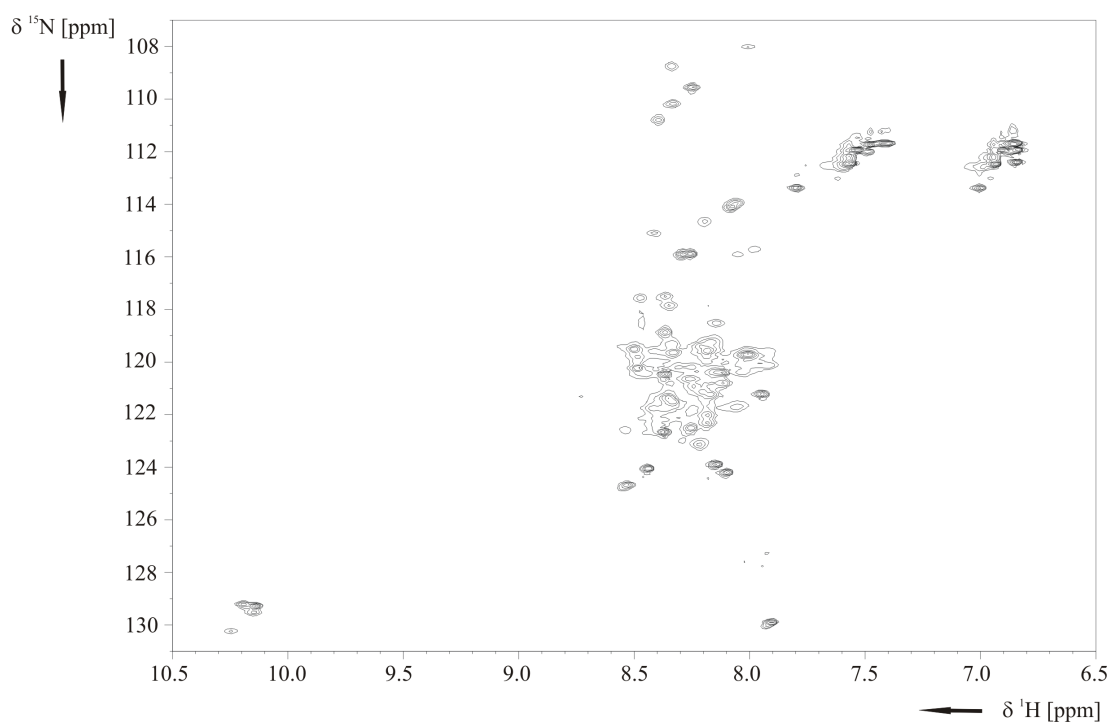


Figure 3.20.: BLA in 4M urea, 50mM Tris/HCl pH 7 at 35°C in the absence of Ca^{2+} (spectrum not referenced). Protein concentration was 0.49mM.

The poor chemical shift dispersion indicates that the protein is unfolded. The limited number of observable peaks and severe line broadening as evident from the middle part of the spectrum further show that the majority of the ensemble of conformers is in a partially folded, *MG*-like conformation. Very pronounced spectral changes are visible upon addition of two eq. Ca^{2+} (Fig. 3.21, blue spectrum).

BLA apparently undergoes a conformational transition to a well-folded state as indicated

3. Kinetic refolding of bovine α -lactalbumin (BLA)

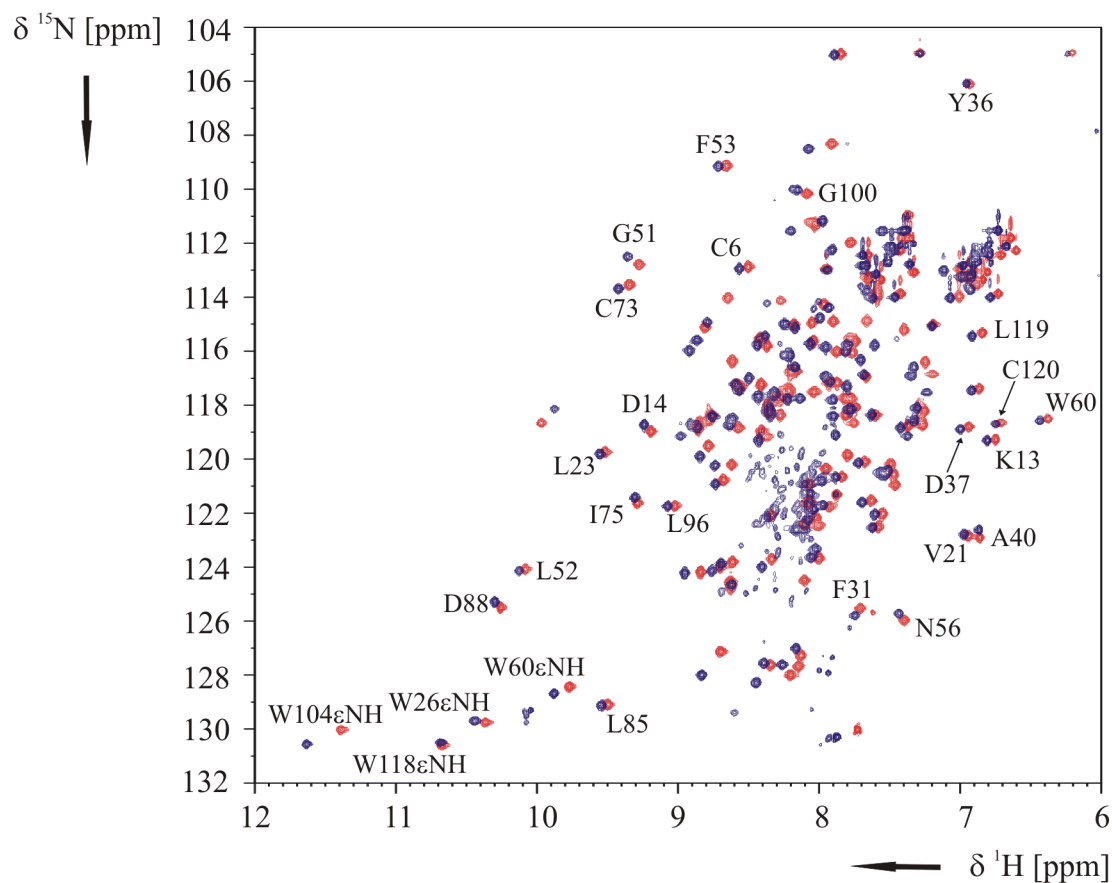


Figure 3.21.: Comparison of ^{15}N , ^1H HSQC spectra of BLA at 20°C in the presence of 2 eq. Ca^{2+} . (a) 50mM HEPES pH 6.3 (red). (b) 4M urea, 50mM Tris/HCl pH 7 (blue). Residues which could be assigned are indicated by their respective labels. Protein concentrations were 1.23mM (a) and 0.49mM (b), respectively. Spectra shown are not referenced.

by the characteristic dispersion in chemical shifts and the appearance of sharp peaks. A few weak peaks in the range between 8.0 and 8.5 ppm ^1H chemical shift are indicative of a minor population of unfolded protein showing that two eq. Ca^{2+} are not sufficient to fully shift the conformational equilibrium to the folded state. The red spectrum in Fig. 3.21 displays BLA in HEPES buffer in the presence of Ca^{2+} (cf. Fig. 3.19). By using the previously reported backbone amide assignment, performed at 50 mM HEPES pH 6.3 at 20°C [80], it was possible to transfer the assignment of 26 peaks (see labels in Fig. 3.21). In the following, a temperature series of ^{15}N , ^1H HSQC spectra was undertaken. Spectra were measured at 20°C, 22°C, 24°C, 28°C, 32°C, and 35°C. Fig. 3.22 shows an overlay of four spectra.

Spectra analysis resulted in transfer of 23 assignments from 20°C to 35°C. The combined use of a ^{15}N NOESY HSQC spectrum measured at 35°C, database chemical shift distributions, and the available assignment of HN, HA and N resonances by Balbach *et al.* [80] made it possible to assign 97% (116 out of 120) of the backbone amide resonances. Moreover, 88 HA resonances were assigned. Finally, methyl and methylene resonances in the range of -0.60 to 0.62 ppm (cf. Fig. 3.26b) were assigned on the basis of amide-methyl(methylene) NOEs. Figs. 3.23 and 3.24 illustrate the assignment process using the strip scope tool in CARA (computer aided resonance assignment, www.nmr.ch). A complete list of assignments can be found in the appendix.

3. Kinetic refolding of bovine α -lactalbumin (BLA)

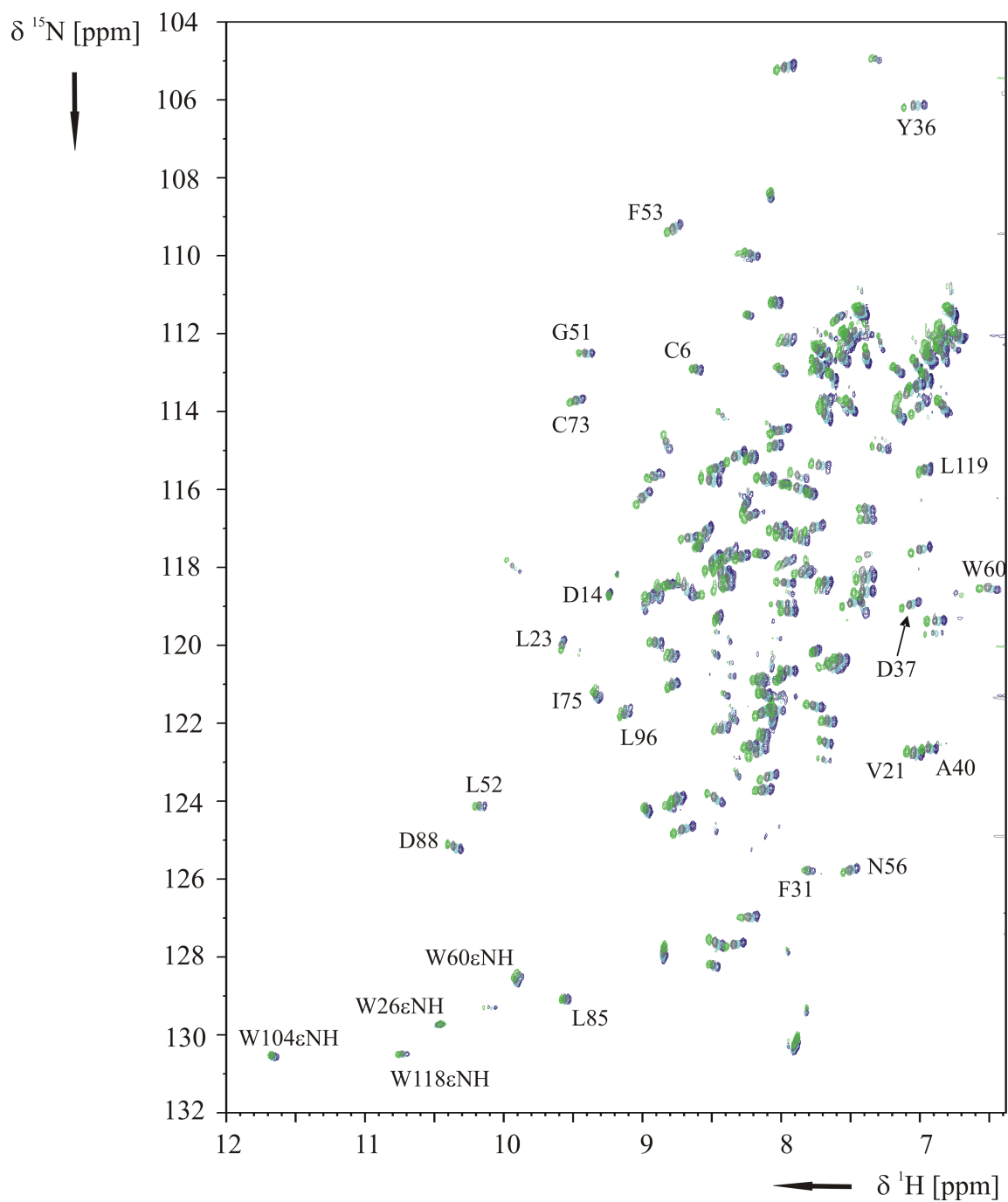


Figure 3.22. ^{15}N , ^1H HSQC spectra of BLA in 4M urea, 50mM Tris/HCl pH 7 in the presence of 5 eq. Ca^{2+} at different temperatures (20°C (dark blue), 24°C (light blue), 28°C (gray), 35°C (green)). Residues which could be assigned are indicated by their respective labels.

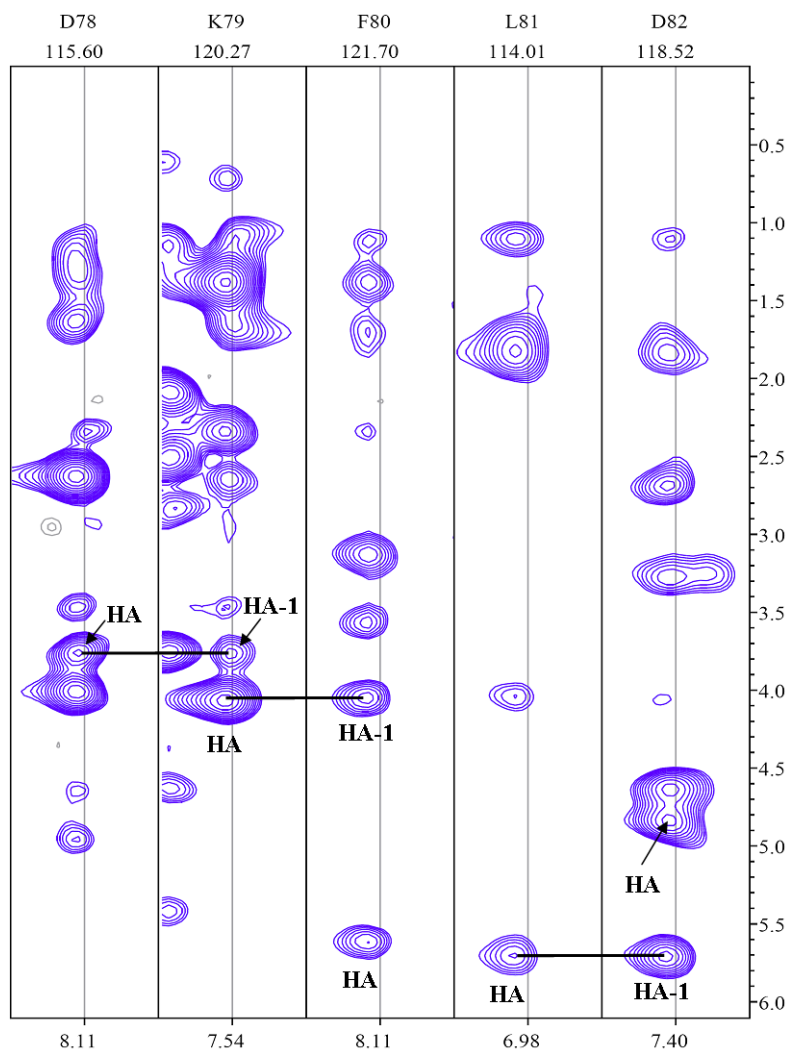


Figure 3.23.: Strips from a ^{15}N NOESY HSQC spectrum of BLA in the presence of 5 eq. Ca^{2+} (4M urea, 50mM Tris/HCl pH 7 at 35°C) illustrating the assignment of HA resonances.

3. Kinetic refolding of bovine α -lactalbumin (BLA)

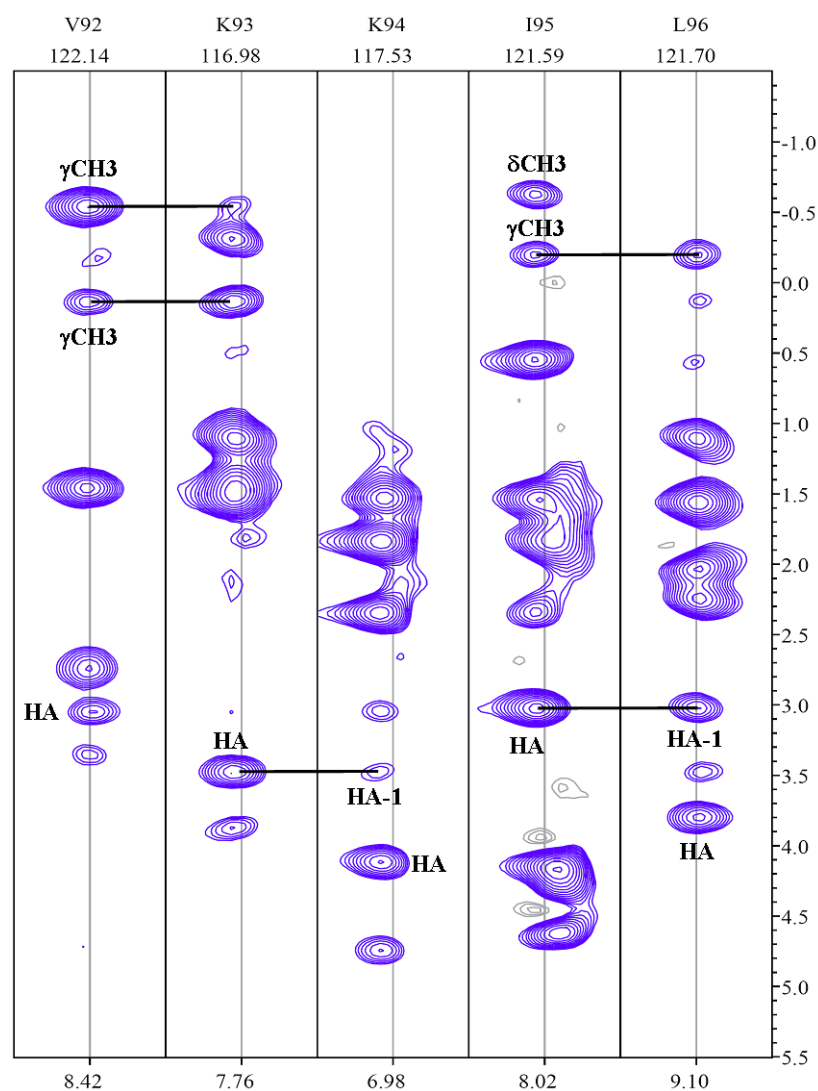


Figure 3.24.: Strips from a ^{15}N NOESY HSQC spectrum of BLA in the presence of 5 eq. Ca^{2+} (4M urea, 50mM Tris/HCl pH 7 at 35°C) illustrating the assignment of side-chain aliphatic and backbone HA resonances.

3.3.5. Folding kinetics by time-resolved NMR

Kinetics of singly resolved peaks

Refolding kinetics were also characterised by time-resolved 1D NMR spectroscopy. Refolding initiation within the NMR spectrometer was accomplished by releasing Ca^{2+} from the photo-caged compound DMN ($K_D \sim 5 \cdot 10^{-9}$ M, $K_{bind} (= 1/K_D) \sim 2 \cdot 10^8$ M $^{-1}$) [233, 234] by short laser pulses as described previously (cf. Fig. 3.7) [44, 51]. The affinity of Ca^{2+} for the photo products is about 5 orders of magnitude lower than for DMN [233]. Ca^{2+} release from DMN and binding to BLA are therefore essentially irreversible. Folding of BLA was monitored by recording consecutive single-scan 1D ^1H NMR spectra. In order to improve the time resolution between each NMR acquisition point, a variable time delay τ_{var} was inserted between the laser pulse and the acquisition of the first spectrum after the laser pulse (Fig. 3.25).

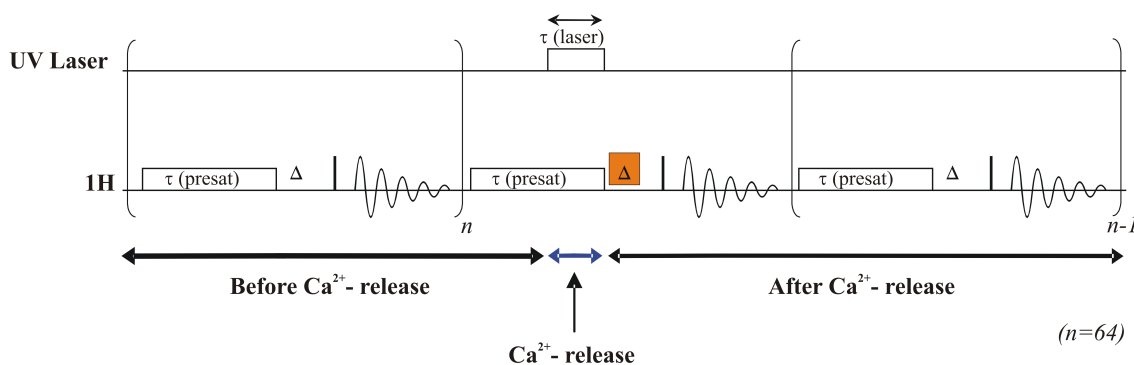


Figure 3.25.: Schematic representation of the pulse sequence used for the acquisition of time-resolved NMR data. 64 single scans are recorded before protein refolding initiation for equilibration purposes. Two equivalents of Ca^{2+} ions are subsequently released from the caged compound DM-nitrophen by a short laser pulse ($\tau_{laser} = 200\text{ms}$). Refolding detection is performed by the acquisition of a second set of 64 single scans. In order to accurately define the initial refolding reaction in subsequent data analysis a variable delay τ_{var} (Δ in orange-marked area) between the Ca^{2+} release and the first scan after the release is employed. Delays which were used in separate real-time experiments are $\tau_{var} = 0\text{ms}$, 100ms , 200ms , and 300ms . The delay between two subsequent single scans is 638ms . The signal from the denaturant urea was suppressed by a selective low power rectangular pulse of $\tau_{presat} = 500\text{ms}$ duration.

The start of data acquisition was consequently varied in separate experiments with fresh

3. Kinetic refolding of bovine α -lactalbumin (BLA)

samples. Refolding kinetics were thus effectively recorded in an interleaved manner which is feasible since laser-induced folding initiation enables acquisition of highly reproducible NMR data. The ^1H spectrum of the initial state is characterized by a large bulk peak centered at around 0.9 ppm whereas the spectrum of the final state shows several high-field shifted resonances indicative of methyl and methylene groups which make part of the highly structured hydrophobic core of a folded protein (Fig. 3.26a).

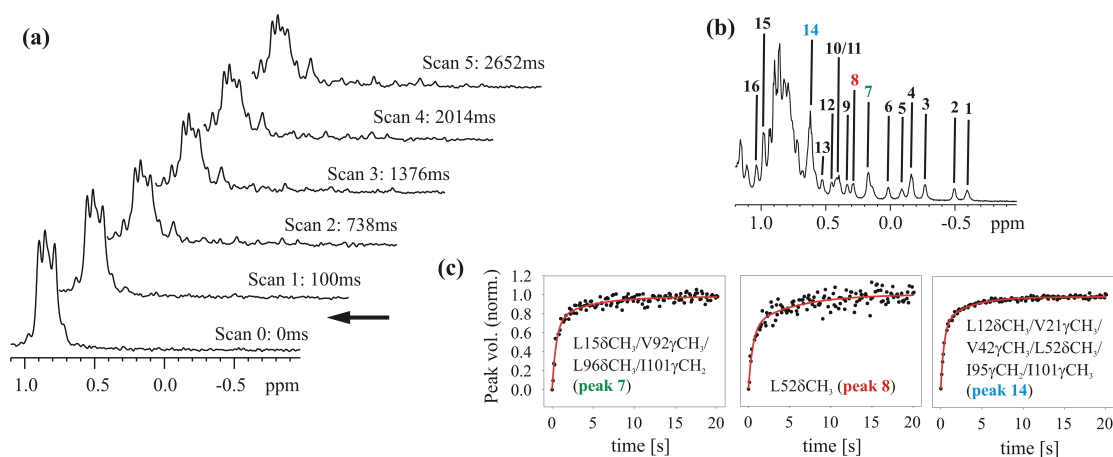


Figure 3.26.: (a) A single real-time NMR refolding experiment performed in 4M urea pH 7 at 35°C ($\tau_{var} = 0\text{ms}$, cf. Fig. 3.25). Formation of folded BLA upon release of two eq. Ca^{2+} from the photo-caged compound DMN (indicated by the arrow) is detected in form of high-field shifted resonances in the methyl/methylene group region of the 1D ^1H NMR spectrum. (b) Methyl/methylene group region of the 1D ^1H NMR spectrum of BLA after refolding. Analyzed peaks are indicated (1-16). (c) Built-up curves of representative NMR resonances (peaks 7, 8, and 14). The color coding is the same as in (b). Red lines represent double exponential fits to the data (cf. Table 3.4 for corresponding refolding rate constants and refolding phase amplitudes).

Thus, this approach makes it possible to monitor the N state population during refolding or, more specifically, the formation of a hydrophobic core.

Fig. 3.26a shows a stacked plot of single-scan 1D spectra of a single refolding experiment. Resonance assignment of all probes analysed (Fig. 3.26b) was performed as described in the previous section. Site-specific kinetic information was extracted by correlating peak volumes vs. time. NMR kinetic traces were fit using single and double exponential functions. For all of the analysed probes (Fig. 3.26b) it was found that double exponentials yield higher correlation

coefficients R^2 (Table 3.4). Three representative kinetic traces are given in Fig. 3.26c. For peaks 4-16, an F statistical analysis shows that a double exponential function fits the data significantly better than a single exponential function (Table 3.4) [218,219]. The fact that this is not the case for peaks 1-3 is most likely due to insufficient signal-to-noise. Refolding rate constants for the double exponential fits range from $1.4 \pm 0.4 \text{ s}^{-1}$ to $5.9 \pm 1.1 \text{ s}^{-1}$ (fast phase) and from $0.11 \pm 0.08 \text{ s}^{-1}$ to $0.4 \pm 0.1 \text{ s}^{-1}$ (slow phase) (Table 3.4).

3. Kinetic refolding of bovine α -lactalbumin (BLA)

Table 3.4.: Summary of fitting parameters of probes 1-16 (cf. Fig. 3.26b) in the upfield methyl/methylene group region of the ^1H 1D spectrum of BLA. Kinetic traces of probes 1-3 and 4-16 were fit using single and double exponential functions, respectively. Reported errors are fitting errors. F values were derived from an F test comparing single exponential and double exponential fits. Critical F values to be considered are $F_{0.01,2,125} = 4.78$ and $F_{0.05,2,125} = 3.07$.

Peak no.	k_1 [s^{-1}]	k_2 [s^{-1}]	A_1 [%]	A_2 [%]	R^2 (single)	R^2 (double)	F
1	0.65 ± 0.12	-	-	-	0.4300	0.4353	0.5764
2	0.68 ± 0.08	-	-	-	0.6290	0.6344	0.9072
3	0.86 ± 0.07	-	-	-	0.7617	0.7667	1.3295
4	1.61 ± 0.24	0.25 ± 0.08	72.9 ± 7.5	24.9 ± 7.0	0.9001	0.9243	19.8215
5	2.18 ± 1.14	0.22 ± 0.15	68.6 ± 17.7	37.7 ± 15.6	0.4624	0.4986	4.4847
6	1.38 ± 0.26	0.14 ± 0.09	76.3 ± 8.2	24.4 ± 5.7	0.7982	0.8293	11.2942
7	1.91 ± 0.22	0.23 ± 0.06	73.1 ± 4.7	24.9 ± 4.2	0.8992	0.9365	36.4375
8	1.83 ± 0.35	0.18 ± 0.06	66.2 ± 6.5	33.6 ± 5.3	0.7998	0.8574	25.0232
9	1.64 ± 0.45	0.22 ± 0.09	62.6 ± 10.9	31.9 ± 9.7	0.7762	0.8108	11.3423
10	1.82 ± 0.31	0.23 ± 0.07	69.7 ± 6.8	28.5 ± 6.1	0.8526	0.8928	23.2266
11	2.13 ± 0.44	0.30 ± 0.09	67.8 ± 8.7	30.7 ± 8.2	0.8390	0.8749	17.7807
12	1.73 ± 0.34	0.11 ± 0.08	78.4 ± 6.5	27.4 ± 4.6	0.7215	0.7753	14.8434
13	2.23 ± 0.68	0.28 ± 0.08	56.2 ± 9.8	41.0 ± 9.2	0.7870	0.8306	15.9661
14	2.41 ± 0.11	0.30 ± 0.02	69.6 ± 1.8	28.7 ± 1.7	0.9462	0.9909	304.5036
15	5.30 ± 0.48	0.23 ± 0.03	72.8 ± 2.0	25.4 ± 1.8	0.7991	0.9346	128.523
16	5.89 ± 1.07	0.41 ± 0.07	63.2 ± 4.5	40.7 ± 6.5	0.7762	0.8822	55.7531

The double exponential refolding behaviour underlines that BLA refolding is complex and involves the population of additional states apart from the apo-*MG* and *N* states. Significant differences in the refolding rate constants can, in general, not be deduced as they are associated with relatively large errors. A close inspection of the data, however, reveals that the fast phase in the built-up of peaks 15 and 16 displays a rate constant which is larger by a factor of 2-3 as compared to the fast phase of the remainder of the probes whereas the slow phase seems not to be affected. This difference in rate constant is reflected, e.g., in the large differences in the normalised peak volumes between probes 14 and 15 within the first 500ms as can be seen from the superposition of the corresponding kinetic traces in Figs. 3.27a,b.

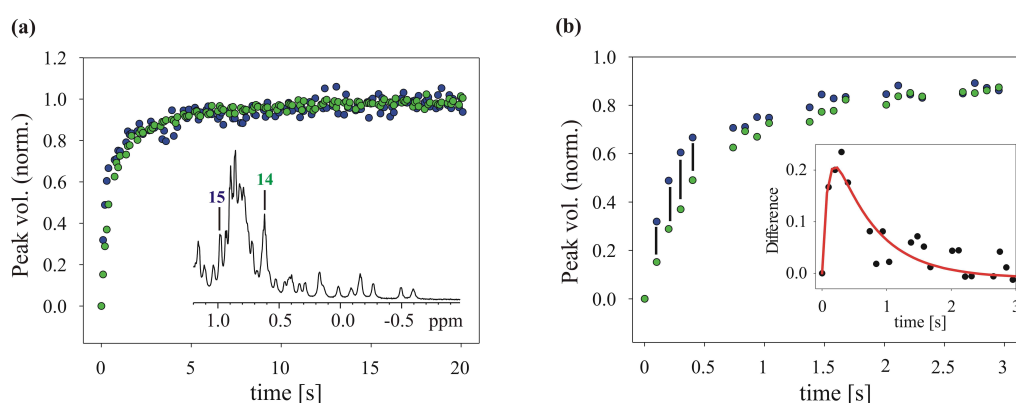


Figure 3.27.: (a) Superposition of the peak volumes of probes 14 (green) and 15 (blue) (b) Zoom into the initial stage of refolding for probes 14 and 15. The differences in peak volume between the two probes (indicated by vertical bars) most likely report on the formation of a folding intermediate as seen in the inset (the solid red line shows a double exponential fit to the data). Because of signal-to-noise considerations, probe 14 and probe 15 were favoured over probes 4-13 and probe 16, respectively.

The plot of the difference in peak volume for the two probes (inset in Fig. 3.27b) suggests the superposition of an aliphatic signal at about the chemical shift of probe 15 in the folded state which is rapidly observed after the release of Ca^{2+} , reaches its maximum intensity after about 300ms and disappears again within 2s after refolding initiation. This proposed signal may stem from a transiently populated folding intermediate which is formed on the millisecond timescale after refolding initiation and vanishes again as the *N* state accumulates. The time course can be analysed using a double exponential function giving rise to rate constants

3. Kinetic refolding of bovine α -lactalbumin (BLA)

of $k_1 = 11.5 \pm 6.0 \text{ s}^{-1}$ and $k_2 = 1.5 \pm 0.4 \text{ s}^{-1}$ for signal built-up and decay, respectively. These values are in accord with results from optical spectroscopy (cf. Table 3.2). The reasoning for probe 15 may also apply to probe 16 whereas, at the same time, such a behaviour may not be inferred from the kinetic rate constants of probes 4-14. This indicates that the proposed folding intermediate, on the basis of side-chain aliphatic resonances, is largely unfolded as it cannot be detected in the region of high-field shifted resonances indicative of tertiary intramolecular contacts. Global fitting of probes 4-14 was also performed with rate constants and amplitudes as global and local fitting parameters, respectively. Probes 15 and 16 were excluded from the analysis due to the apparently different refolding behaviour (see preceding paragraph). Global refolding rate constants were $k_1 = 1.9 \pm 0.1 \text{ s}^{-1}$ and $k_2 = 0.2 \pm 0.03 \text{ s}^{-1}$ being in good agreement with the results from the separate fits to each kinetic trace. Global fitting revealed significant differences in refolding amplitudes among the probes (Table 3.5). Signal overlap in the upfield spectral region, however, precludes a reliable interpretation at present.

Table 3.5.: Global fitting of the kinetic traces of methyl/methylene probes 4-14 to a double exponential function with rate constants (k_1 , k_2) and amplitudes (A_1 , A_2) as global and local fitting parameters, respectively. Globally fitted rate constants are $k_1 = 1.9 \pm 0.1 \text{ s}^{-1}$ and $k_2 = 0.2 \pm 0.03 \text{ s}^{-1}$.

Peak no.	A_1 [%]	A_2 [%]
4	72.2 ± 3.4	26.4 ± 4.2
5	70.9 ± 4.2	35.1 ± 4.3
6	66.1 ± 4.0	32.7 ± 4.2
7	74.0 ± 3.8	24.1 ± 4.2
8	62.9 ± 4.1	35.6 ± 4.3
9	61.0 ± 4.0	33.8 ± 4.2
10	69.8 ± 3.9	28.7 ± 4.2
11	74.7 ± 3.8	24.6 ± 4.2
12	71.0 ± 4.0	29.7 ± 4.2
13	63.6 ± 4.1	34.6 ± 4.3
14	78.0 ± 3.7	20.8 ± 4.1

Kinetics of bulk spectral regions

Refolding was further investigated by analysing the time evolution of the peak volume for three different bulk regions of the 1D ^1H spectrum. Fig. 3.28 shows the respective correlations of peak volumes vs. time for the methyl/methylene group region encompassing probes 4-14 (-0.19 to 0.67ppm, cf. Fig. 3.26b), for part of the backbone amide region (8.64 to 9.59ppm), and for the region comprising the methyl/methylene groups in the initial state (0.75 to 0.92ppm).

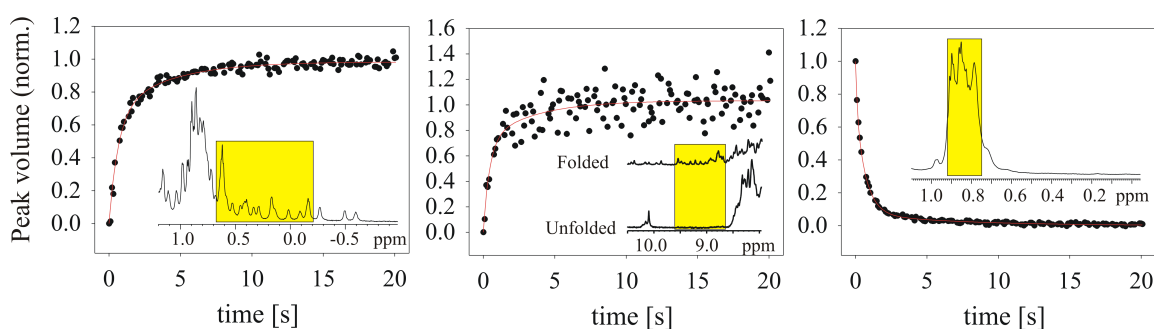


Figure 3.28.: Built-up of folded-state resonances in the upfield aliphatic (a) and downfield backbone amide (b) regions of the 1D ^1H spectrum of BLA. (c) Time course of the volume of the bulk peak summarising aliphatic groups in the initial state. Red lines represent double (a,b) and triple (c) exponential fits to the data, respectively. Appropriate fitting functions were chosen on the basis of an F test statistical analysis. The areas marked in yellow indicate the spectral parts for which the time evolution of peak volumes was determined.

Refolding on the level of backbone amide and high-field shifted methyl/methylene resonances is characterized by double exponential kinetics with refolding rate constants on the order of $k_1 = 2 \text{ s}^{-1}$ and $k_2 = 0.3 \text{ s}^{-1}$ (cf. Table 3.6) mirroring the results obtained for methyl/methylene probes 4-14. The best fit of the decrease in peak volume for the aliphatic resonances in the initial state, however, is yielded by using a triple exponential function giving rise to refolding rate constants $k_1 = 7.8 \pm 1.1 \text{ s}^{-1}$, $k_2 = 1.5 \pm 0.1 \text{ s}^{-1}$, and $k_3 = 0.1 \pm 0.01 \text{ s}^{-1}$.

3. Kinetic refolding of bovine α -lactalbumin (BLA)

Table 3.6.: Results from double and triple exponential fits, respectively, to the data shown in Fig.

3.28

Probe	k_1 [s^{-1}]	A_1 [%]	k_2 [s^{-1}]	A_2 [%]	k_3 [s^{-1}]	A_3 [%]
Aliphatics (<i>N</i>)	1.5 ± 0.1	66.5 ± 4.5	0.26 ± 0.04	31.9 ± 4.2	-	-
Backbone amides	2.3 ± 0.8	71.3 ± 14.6	0.29 ± 0.15	32.1 ± 13.7	-	-
Aliphatics (<i>U</i>)	7.8 ± 1.1	25.7 ± 2.8	1.5 ± 0.1	66.8 ± 2.6	0.14 ± 0.01	17.4 ± 0.5

3.4. Discussion

3.4.1. BLA refolding is (more) complex

Comparison of the data accumulated within in the course of this thesis with data found in the literature [221, 222] indicates that the conformational transition observed in equilibrium and kinetic studies actually corresponds to a population shift from the apo-*MG* state to the *N* state. The transition from the kinetic *MG* to the native state has been shown to proceed highly cooperatively without the population of additional intermediate states [13, 78, 80, 209].

The kinetics of the Ca^{2+} -induced folding transition of BLA were studied herein by a combination of stopped-flow fluorescence and CD as well as time-resolved NMR experiments. The analysis of the kinetic data shows that BLA refolding exhibits triple exponential kinetics as seen in stopped-flow fluorescence and far-UV CD experiments (Fig. 3.11). Time-resolved NMR also reveals triple exponential refolding kinetics when the bulk peak around 0.9 ppm indicative of aliphatic proton resonances in the initial state is monitored (Fig. 3.28c). The rate constants characterising the three phases coincide well with those obtained from the triple exponential fits to the stopped-flow data showing that the same processes are being detected in the stopped-flow and in the time-resolved NMR experiments. Signal intensity buildup for several probes in the upfield aliphatic and downfield backbone amide spectral regions indicating formation of the *N* state are, however, characterised by apparent double exponential kinetics. Compared to the optical and NMR unfolded bulk data the fastest phase is not detected. The lack of this fastest reaction is most likely due to its low amplitude and to limitations in signal-to-noise. As a result, the slow phase exhibits an amplitude of around 30% which is about 20% more than expected on the basis of the triple exponential fits to the other data. Consistently, 20% is about the amplitude of the fastest reaction which is missed in most NMR kinetics of single peaks (cf. Table 3.7). These results clearly show that BLA refolding is more complex

than anticipated and seems to involve additional conformational states which have escaped detection so far. Triple exponential kinetics for BLA refolding have only been reported in a single independent study. Here, the transient population of intermediate states apart from the *MG* intermediate was suggested [235].

3. Kinetic refolding of bovine α -lactalbumin (BLA)

Table 3.7.: Summary of kinetic parameters for the Ca^{2+} -induced refolding of BLA as determined from a combination of time-resolved fluorescence, CD, NMR, and interrupted refolding experiments.

Probe	k_1 [s^{-1}]	A_1 [%]	k_2 [s^{-1}]	A_2 [%]	k_3 [s^{-1}]	A_3 [%]
Trp fluorescence	4.3 ± 0.04	30.8 ± 0.8	2.0 ± 0.01	60.0 ± 0.8	0.29 ± 0.001	9.8 ± 0.1
Far-UV CD (235nm)	9.5 ± 4.3	8.9 ± 3.4	2.4 ± 0.1	74.9 ± 3.4	0.38 ± 0.01	15.7 ± 0.7
NMR (unfolded bulk peak, 0.75 - 0.92ppm)	7.8 ± 1.1	25.7 ± 2.8	1.5 ± 0.1	66.8 ± 2.6	0.14 ± 0.01	7.4 ± 0.5
NMR (aliphatic groups, -0.19 - 0.67ppm)	- ^a	- ^a	1.5 ± 0.1	66.5 ± 4.5	0.26 ± 0.04	31.9 ± 4.2
NMR (aliphatic groups, probes 4-14, global fit)	- ^a	- ^a	1.9 ± 0.1	- ^b	0.2 ± 0.03	- ^b
NMR (backbone amide groups, 8.64 - 9.59 ppm)	- ^a	- ^a	2.3 ± 0.8	71.3 ± 14.6	0.29 ± 0.15	32.1 ± 13.7
NMR (proposed folding intermediate, 0.98ppm)	11.5 ± 6.0	-	1.5 ± 0.4	-	-	-
Interrupted refolding	4.7 ± 3.4	27 ± 26	1.5 ± 0.3	70 ± 27	- ^a	- ^a

^aThis phase has not been detected, most likely due to its low amplitude.

^bRefer to Table 3.5

3.4.2. BLA folds *via* parallel pathways

Interrupted refolding experiments [183] reveal that native protein is formed in the two fastest kinetic phases and clearly rule out a sequential folding pathway (Fig. 3.15). Together with the kinetic NMR data for probe 15 of the folded state (Fig. 3.27) indicating the transient formation of a folding intermediate, the data support a triangular folding mechanism with the formation of the native state and of an intermediate (*I*) occurring in the fastest reaction and interconversion of the intermediate to the native state taking place in the second fastest reaction [236] (Fig. 3.29).

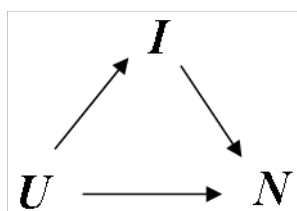


Figure 3.29.: Triangular folding mechanism involving parallel pathways for formation of the *N* state.

Since the triangular folding mechanism gives rise to only two apparent rate constants for fundamental reasons, fast formation of the native state and formation of the intermediate occur with the same observable rate constant [236, 237]. Close inspection of the kinetic data summarised in Table 3.7 reveals, however, that there is a certain spread in the rate for the fastest reaction. This finding might indicate that the rates characterising the transitions from the apo-*MG* to the *I* and *N* states, respectively, differ slightly. The slowest reaction observed with optical probes and NMR is most likely due to slow prolyl isomerisation reactions during folding of apo-*MG* molecules with non-native Xaa-Pro peptide bonds [78]. This interpretation is supported by the fact that the rate of this phase is independent of the urea concentration (Fig. 3.13a), which is a specific feature of proline isomerisation in unfolded states of proteins [231, 232]. The slow folding phase is not detected in the interrupted refolding assay, likely because of an insufficient number of data points for long delays t_i .

The investigation of the effect of the Ca^{2+} concentration on the refolding kinetics (Fig. 3.14) shows a strong dependence for k_1 whereas k_2 and k_3 are only little affected and not all

3. Kinetic refolding of bovine α -lactalbumin (BLA)

affected, respectively. Acceleration of BLA refolding as a response to an increase in the Ca^{2+} concentration has already been demonstrated [78,80,210] and attributed to the reduction of the activation free energy in the conversion of the *MG* to the native state. It is shown here for the first time that BLA refolding is actually more complex and involves parallel pathways. Folding phase k_1 showing a strong dependence on the Ca^{2+} concentration does not only correspond to the conversion of the apo-*MG* to the *N* state but to the formation of a folding intermediate from the apo-*MG* as well. The relative influence of Ca^{2+} on the two separate steps can, however, not be delineated at present. Increasing the Ca^{2+} concentration has only a slight effect on the formation of the *N* state from the intermediate (k_2) and no effect at all on the presumed isomerisation reaction in the apo-*MG* (k_3).

3.4.3. BLA and HEWL share a common folding mechanism

Comparison of BLA folding to the folding of the highly homologous protein lysozyme shows that the two proteins share a common folding mechanism. According to the triangular folding scheme (Fig. 3.29, *U* denotes the respective unfolded state), the majority of refolding molecules with native Xaa-Pro peptide bonds fold through a folding intermediate accounting for 80 % of the lysozyme [183] and for 70 ± 27 % (Fig. 3.15 and Table 3.7) for the BLA molecules. The remaining molecules fold directly to the native state. Extensive NMR analyses of unfolded lysozyme and single point mutants thereof have established the existence of multiple hydrophobic clusters [64–67]. An essential role plays the W62W63 motif located at the interface of the α - and β -subdomains in the native structure of lysozyme exerting a stabilising effect on hydrophobic clusters in the α -domain via a long-range interaction in the unfolded state [65,66]. The possible role of this interaction in guiding the protein toward the folded state by mediating intersubdomain interactions has recently been discussed [68]. In contrast to lysozyme, the corresponding I59W60 motif in the β -domain of BLA seems not to be involved in any long-range interaction with the α -domain in the unfolded state. Wirmer *et al.* have addressed the possibility that the task of the W62W63 motif of mediating intersubdomain contacts might be fulfilled by the Ca^{2+} binding site in BLA [68]. Inspection of the 2.2 Å crystal structures of apo- and holo-BLA is insightful in this respect. In comparison to the holo protein, the apo protein exhibits a more expanded binding loop due to charge repulsion between five aspartyl residues in this region of the structure. By virtue of the C73-C91 disul-

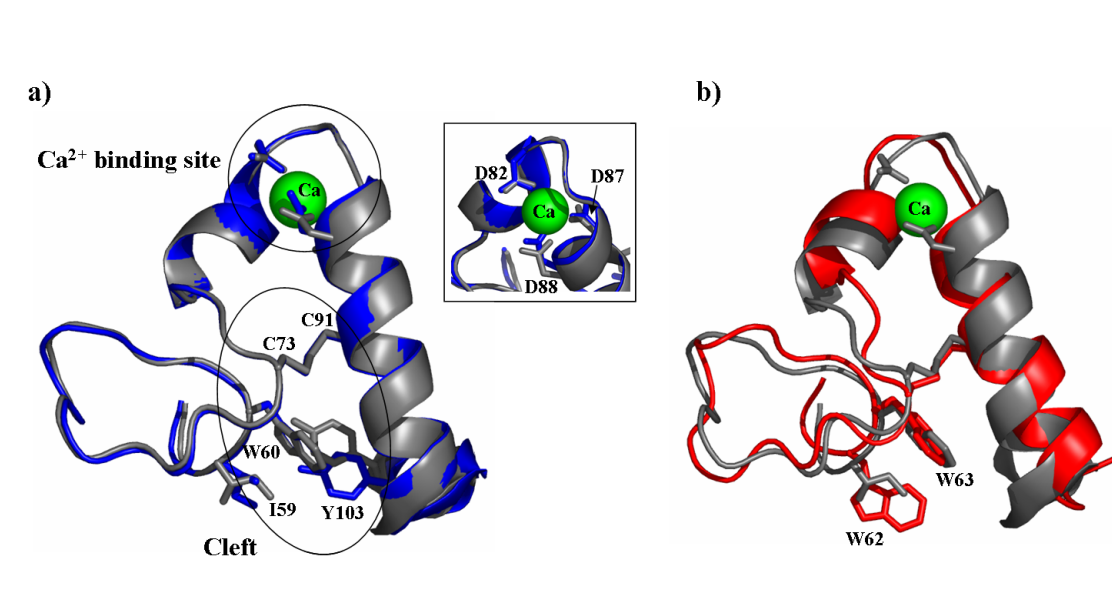


Figure 3.30.: Structural alignment based on C^α atomic coordinates sketching the subdomain interface of BLA and HEWL. (a) apo- (blue) and holo-BLA (grey). The Ca^{2+} ion is indicated as a green sphere. Disruption of intersubdomain hydrophobic interactions involving Y103 due to charge repulsion in the Ca^{2+} binding loop is transmitted to the cleft region by virtue of the C73-C91 disulfide bond connecting the α - and β -subdomains. The inset shows the Ca^{2+} binding site (top view) with carboxylate groups of D82, D87, and D88 being part of the Ca^{2+} -liganding moieties. (b) holo-BLA (grey) and lysozyme (red, first structure of NMR ensemble PDB 1E8L [238]). W62 and W63 (corresponding to I59 and W60 in BLA) are involved in long-range interactions with residues in the α -subdomain in the unfolded state. Graphics were created using PyMOL [239].

side bond connecting the two subdomains, this expansion is transmitted to the cleft region between the subdomains so that the cleft slightly widens. As a result, it is seen that native intersubdomain hydrophobic interactions involving Y103 are disrupted (Y103 pointing more downwards as compared to holo state, see Fig. 3.30a) [200].

The structure of the apo protein has been determined at high ionic strength (2M ammonium sulfate). It is expected that, at low ionic strength conditions as chosen in our study, subdomain separation and subsequent disruption of hydrophobic interactions will be even more pronounced. Because of charge neutralisation, addition of Ca^{2+} will then lead to less expansion of the binding loop and cleft regions promoting intersubdomain interactions, such as those involving Y103, to occur. The fact that the Ca^{2+} -induced refolding of BLA is herein

3. Kinetic refolding of bovine α -lactalbumin (BLA)

shown to involve parallel pathways with the population of an intermediate state on the predominating pathway, as it is also seen for lysozyme, [183] now strongly supports the notion that the guiding role of the long-range interaction between the α - and β -subdomains in the unfolded state of lysozyme is replaced by the Ca^{2+} binding site in the case of BLA [68].

3.4.4. Advances in NMR methodology

The combination of a relatively short dead-time and laser-induced folding initiation enables the acquisition of highly reproducible NMR data in a short time window with good time digitisation. A detailed characterisation of the Ca^{2+} -induced refolding process involving the population of a folding intermediate on the millisecond timescale is therefore possible. Longer dead-times of 1-5s in mixing-based NMR folding experiments of BLA may be the reason why this folding intermediate has been missed in other studies [40,53,79,80]. The data also underline the general power of time-resolved NMR spectroscopy for probing conformational transitions in complex biomolecules with atomic resolution. High quality data at protein concentrations as low as $100\mu\text{M}$ can be obtained. The usage of the jump-return-echo excitation scheme [217] in combination with cryoprobe technology is crucial in this respect. The possibility to acquire NMR data in a highly reproducible fashion enables to measure folding kinetics in an interleaved manner. Folding rate constants of up to 8 s^{-1} can thus be obtained. These are the highest rate constants for protein folding transitions which have been measured up to now using time-resolved 1D NMR techniques. Even higher rate constants would become measurable if the dead-time of the experiment could be reduced. Initial attempts to more efficiently illuminate the NMR sample using specifically designed glass fibre ends [240] or 1mm probes, however, did not lead to any improvement in terms of experimental dead-times (data not shown).

The time-resolved NMR data complement the findings based on the stopped-flow data emphasising the importance of time-resolved NMR in deciphering folding mechanisms. From a more general point of view, the experimental setup opens up the possibility of studying structural transitions in many different types of ion-dependent biomolecular reactions with site-specific resolution. This is important as many processes on the biomolecular and cellular levels are crucially dependent on the presence of cofactors such as Ca^{2+} or other inorganic ions [241,242].

3.4.5. Future prospects

It is obvious that future studies will mainly focus on the *I* state which can clearly be inferred from the present data but for which a more detailed characterisation is highly desirable. It might, e.g., turn out to be helpful to use amino-acid-specific ^{13}C aliphatic labels in combination with X filter NMR techniques [243]. 1D spectra would become significantly simpler and resonances as indicators of the *I* state might become transiently observable. Extension of the kinetic studies to 2D NMR approaches would benefit from the much higher resolution as compared to 1D spectra helping to identify resonances characterising the *I* state. This, however, would be challenging as the intermediate is populated on a millisecond timescale. Realisation would require the use of the SOFAST technique [244] in conjunction with acquisition of a single (kinetic) 2D spectrum [40] in the course of refolding. Another possibility to gain further insight into the intermediate is to tune experimental conditions in interrupted refolding experiments such that unfolding of the *I* state becomes measurable. This approach has already been successfully applied to RNaseA [182] and HEWL [236] for non-NMR detection where the rates for formation and subsequent disappearance of folding intermediates could be determined. However, interrupted refolding experiments could also be undertaken in an NMR context. In a first step, refolding could be initiated in the conventional way by releasing Ca^{2+} from a caged compound, followed by a second step mixing the refolding protein with unfolding buffer [50]. Since NMR provides site-specific resolution, data of such type would usefully complement global data based on tryptophan fluorescence.

Moreover, it will be insightful to measure the refolding kinetics on the well-resolved tryptophan indole protons in the very downfield part of the 1D spectrum and to compare the results with those obtained using stopped-flow fluorescence where the signal is an average over all tryptophans. The nature of the slow folding phase could be addressed by application of the classical stopped-flow double-jump test probing for slow isomerisation around Xaa-Pro peptide bonds in the unfolded state [15].

Although first attempts to more efficiently illuminate the NMR sample have failed (see previous section), it may still be possible to reduce the experimental dead-time. One possibility constitutes the use of novel Ca^{2+} -caging compounds based on the nitrodibenzofuran (NDBF) chromophore [245] which might allow to more rapidly release the amount of Ca^{2+} needed. Alternatively, the time-resolution could be improved by using a mixing device [50] for folding

3. *Kinetic refolding of bovine α -lactalbumin (BLA)*

initiation which might, however, compromise data reproducibility because of the possibility of mixing artefacts.

4. Photo-CIDNP studies on bovine α -lactalbumin in the presence of paramagnetic lanthanides

4.1. Introduction

4.1.1. Photochemically induced dynamic nuclear polarisation (Photo-CIDNP)

The term 'chemically induced dynamic nuclear polarisation' (CIDNP) refers to a non-Boltzmann distribution of nuclear spin state populations in the products of thermal and (photo)chemical reactions involving the transient formation of radicals. CIDNP manifested as emissions and absorption enhancements in NMR spectra [246, 247]. Only two years after its discovery, the radical pair mechanism was proposed which could account for the majority of the experimental observations [248–250]. According to this mechanism, the origin of the CIDNP phenomenon relies on nuclear-spin-state-dependent differences in the rates of the triplet-to-singlet transition of a radical pair which forms in the course of a particular reaction. Large polarisations are generated which can be detected in the reaction products in the form of emissive and enhanced absorptive resonances in the NMR spectrum.

Whereas initial investigations provided mechanistic insight into chemical reactions, applications to proteins showed that this technique was useful for structural studies providing information on surface properties of proteins [251]. In a typical experiment, an argon ion laser is used to illuminate the sample solution within the NMR active volume containing the protein and a small-molecule dye, typically a flavin derivative. The dye molecule (F, cf. Fig. 4.1)

4. Photo-CIDNP studies on bovine α -lactalbumin in the presence of paramagnetic lanthanides

is excited to a triplet state which may then react with certain solvent-accessible amino acid side-chains to form a triplet state radical pair (for reviews see references [252–254]).

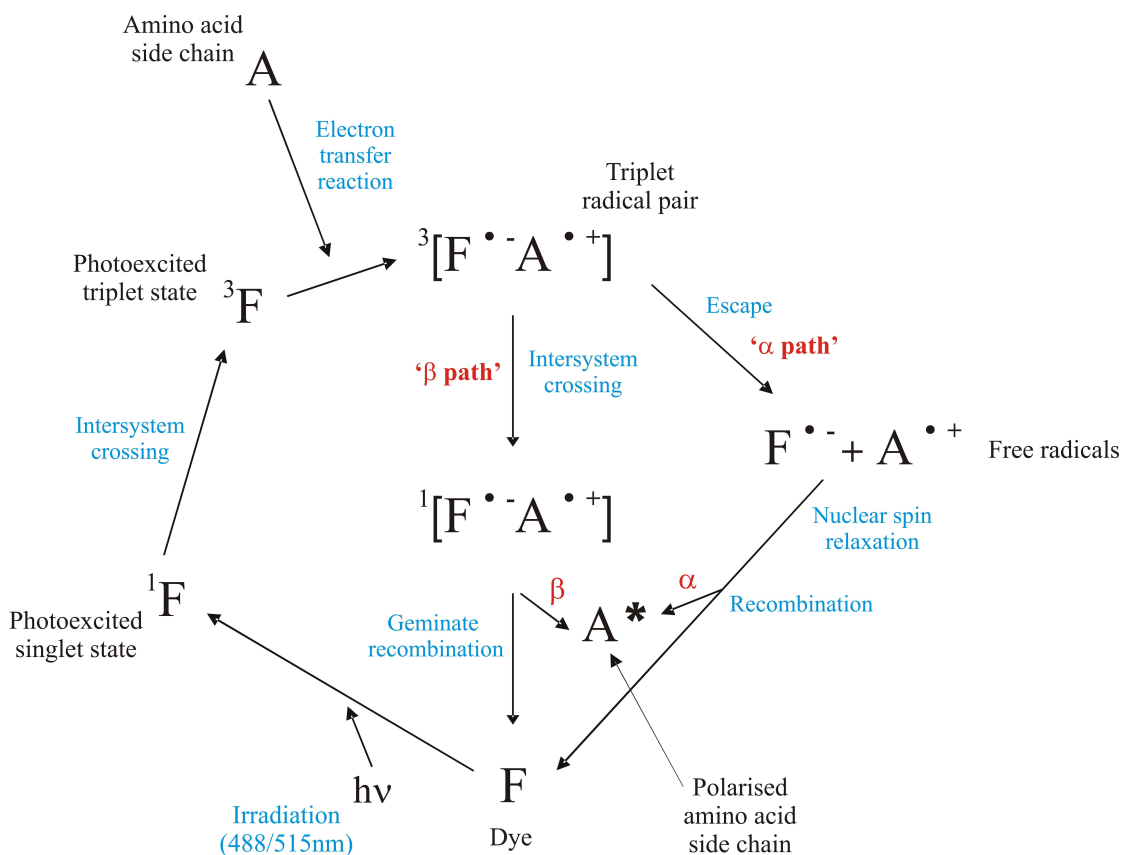


Figure 4.1.: Cyclic reactions during a photo-CIDNP experiment of a protein. The ' α and β paths' refer to the nuclear spin sorting process in the triplet radical pair (see text for details). The figure was adapted from Mok and Hore [254].

Upon transfer of an electron (or, alternatively, a hydrogen atom) a triplet state radical pair is formed. The electron spins of the two radicals are initially correlated but the difference in their precession frequencies will cause a triplet-to-singlet transition to occur (intersystem crossing). After the transition has occurred, the radicals may recombine giving the dye molecule and the particular amino acid side-chain in their ground states (recombination path). The crucial point in the triplet-to-singlet interconversion is the presence of significant hyperfine interactions of an electron spin with nuclear spins in a radical since the difference in the precession frequencies of the electron spins depends on the spin state level of the nuclear spin.

Let us assume that there is one nucleus in the $A^{\bullet+}$ radical coupling with the electron spin

through a hyperfine interaction and that the difference in the precession rates of the two electron spins is larger if the nucleus is in the β state. The probability for intersystem crossing will consequently be higher if the nucleus in the radical pair is in the β state. On the other hand, the probability that the radicals separate before intersystem crossing occurs is higher if the nucleus is in the α state. The free radicals may then recombine to give the amino acid side-chain and the dye molecule in their ground states (escape path). The recombination and escape paths yield β and α spin polarisations, respectively. Since the reaction is apparently cyclic as can be seen in Fig. 4.1 one may intuitively assume that α and β spin polarisations exactly cancel in the reaction product and that the effect of the nuclear spin sorting process in the triplet state radical pair cannot be observed. This is, however, not the case. If the triplet state radical separates to give free radicals in solution, the polarised nucleus in the side-chain radical has considerable time to be relaxed by the electron spin before radical recombination takes place. This relaxation process yields a net excess of β spin polarisation in the reaction product. The sign of the respective signal depends, amongst other things, on the sign of the hyperfine coupling in the radical and on the sign of the difference in g factors of the two radicals [255].

The cyclic character of the photochemical reaction is an important prerequisite for probing surface properties of proteins, since it guarantees the ground state and not some (photo)chemically modified form of the protein to be reached. The reversibility of the reaction is advantageous in that several experiments may be performed on a single sample. This allows both to improve the signal-to-noise ratio and to reduce the amount of protein needed. It should be noted, however, that the reaction scheme shown in Fig. 4.1 is not perfectly cyclic as some photobleaching of the dye cannot be avoided.

Due to the intrinsically low sensitivity of NMR, it is especially attractive to exploit the CIDNP effect as it provides signal enhancements of one to two orders of magnitude as compared to conventional NMR experiments. Moreover, photo-CIDNP NMR spectra are much simpler since, in the case of proteins, only nuclei in the side-chains of tryptophan, tyrosine, and histidine are significantly polarisable and therefore detectable. In general, tryptophan and histidine give rise to absorptive enhancements whereas tyrosine shows emissive signals. An example is shown in Fig. 4.2.

Spectra are normally acquired with and without illumination of the sample, respectively,

4. Photo-CIDNP studies on bovine α -lactalbumin in the presence of paramagnetic lanthanides

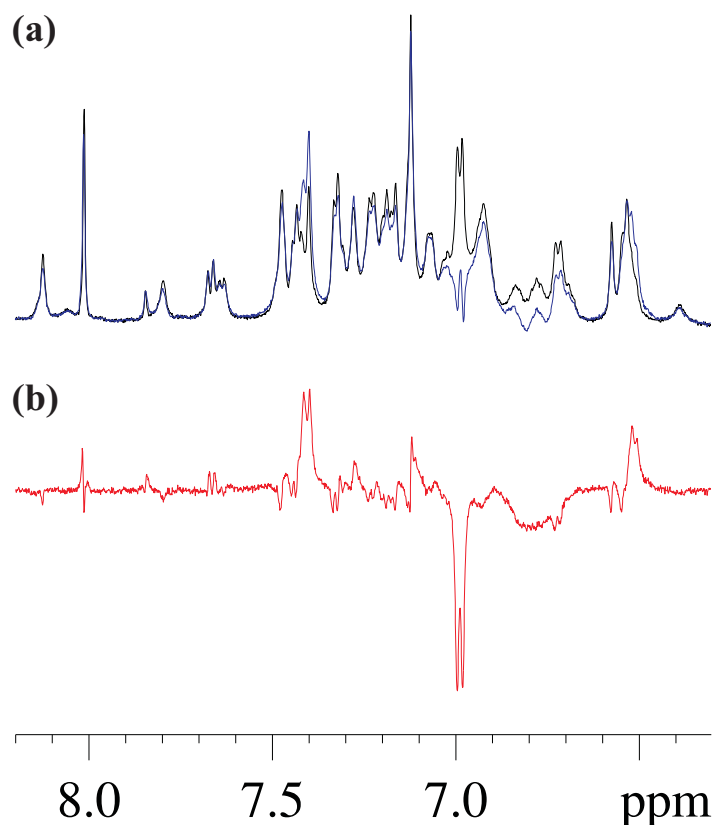


Figure 4.2.: Generation of a photo-CIDNP NMR spectrum (red, (b)) as the difference spectrum of the 'light' (blue) and 'dark' (black) spectra (a). Shown is the aromatic part of the spectrum for a 1mM BLA sample in the presence of 1.5 eq. Ca^{2+} . Further experimental conditions: 0.2mM flavin I, 25mM citric acid pH 7, 35°C.

before the ^1H excitation pulse under otherwise identical conditions. The difference of the 'light' and 'dark' spectra gives the photo-CIDNP spectrum.

Polarisable side-chains need be solvent-accessible for encounters with excited dye molecules to occur. Side-chains buried within the interior of the protein are consequently not polarisable and cannot be observed in photo-CIDNP NMR spectra. This makes photo-CIDNP a useful technique to study surface properties of proteins. The hydrophobic nature of polarisable amino acid side-chains has been exploited in studies of protein folding where important insights into changes in hydrophobic clustering during folding of proteins such as lysozyme and bovine α -lactalbumin could be obtained [42, 44].

4.1.2. Paramagnetic lanthanides as shift and relaxation reagents and their effect on CIDNP

If a paramagnetic lanthanide ion Ln(III) binds to a ligand, the induced change in the chemical shift of a nucleus in the ligand, δ , can be dissected as follows [256]:

$$\delta = \delta^d + \delta^c + \delta^{\text{pcs}} \quad (4.1)$$

where δ^d , δ^c , and δ^{pcs} denote the diamagnetic shift, contact shift, and pseudocontact shift contributions, respectively. Diamagnetic shifts are usually small, except for the donor atoms of the ligand. They may arise, e.g., from conformational changes upon binding of the Ln(III) ion and would also be observed for binding of a non-paramagnetic ion, although the value may be different for different ions. In general, the diamagnetic contribution can be calculated by determining the difference $\Delta\delta$ between the shifts of a given nucleus in the Ln(III)^{para}- and Ln(III)^{dia}-bound states of the ligand, where Ln(III)^{para} and Ln(III)^{dia} denote particular paramagnetic and diamagnetic lanthanide ions, respectively.

The contact shift results from a through-bond transmission of unpaired electron spin density of the Ln(III) ion to the nucleus in question. Donor nuclei are usually largely affected. The effect decreases rapidly, however, upon increasing the number of bonds between Ln(III) and a particular nucleus. The pseudocontact shift in turn is due to a change in the local magnetic field of a given nucleus caused by the large magnetic moment of the Ln(III) ion. If the resonating nucleus is separated from the metal ion by at least four bonds and there are no π bonds in between and, additionally, one has corrected for the diamagnetic contribution, the contact shift can be considered negligible and any observable shift will be pseudocontact in nature [257]. δ^{pcs} values can then be determined and have been shown to be useful restraints in NMR protein structure determination [257].

The binding of a Ln(III) ion to a target molecule also results in enhancements of both the longitudinal ($1/T_1$) and transverse ($1/T_2$) relaxation rates of nuclei in the ligand. Generally, this effect consists of inner-sphere and outer-sphere contributions the latter of which are relatively small and, therefore, often neglected. Neglecting the outer-sphere contributions and having corrected for diamagnetic contributions (see above), the relaxation rate of a given nucleus in the ligand, $1/T_i$, will be defined as follows [256]:

4. Photo-CIDNP studies on bovine α -lactalbumin in the presence of paramagnetic lanthanides

$$\frac{1}{T_i} = \frac{1}{T_{i,c}} + \frac{1}{T_{i,p}} + \frac{1}{T_{i,\chi}} \quad (4.2)$$

where $1/T_{i,c}$, $1/T_{i,p}$, and $1/T_{i,\chi}$ refer to the contact, dipolar, and Curie contributions to the relaxation rate, respectively. Due to the $1/r_6$ dependence (r is the distance between the nucleus and the Ln(III) ion), relaxation data yield information on geometrical parameters of nuclei in the proximity of Ln(III).

At the time when the CIDNP phenomenon was discovered, Bargon already investigated the influence of paramagnetic additives on the thermal decomposition of dibenzoylperoxide in cyclohexanone [258]. It was shown that the addition of paramagnetic Fe(III) resulted in a decrease in the intensity of the emissive resonance of the reaction product benzene. At a sufficiently high concentration of Fe(III), the emissive character of this resonance completely vanished. This observation was interpreted as due to a shorter relaxation time in the presence of Fe(III). The same author later on demonstrated that, in order to simplify the analysis of CIDNP spectra of organic compounds, the shifting nature of paramagnetic lanthanides such as Pr(III) or Eu(III) could be exploited to selectively shift the resonance lines of CIDNP-showing reaction products [259]. The choice of sufficiently low concentrations of the shift reagents gave only small reductions of the CIDNP intensities. In a later study by Hore and Kaptein, the effect of paramagnetic lanthanides on CIDNP was also investigated in the case of proteins [260]. The photo-CIDNP spectrum of hen egg white lysozyme in the presence of one equivalent of Gd(III) showed selective broadening of resonances in comparison to the spectrum in the absence of Gd(III). Changes in chemical shifts of CIDNP-active nuclei could not be observed.

It may be summarised at this point that (i) CIDNP is generally observable in the presence of paramagnetic additives, (ii) CIDNP resonances may decrease in intensity, and (iii) CIDNP resonances may shift or broaden depending on the nature of a particular Ln(III) ion.

4.1.3. Lanthanide-induced refolding of bovine α -lactalbumin

Paramagnetic lanthanide ions have long been successfully used in NMR structural studies of Ca^{2+} binding proteins [261–263]. Since they exhibit similar ionic radii and a similar coordination chemistry in comparison to Ca^{2+} , they can substitute for Ca^{2+} and yield valuable

structural restraints such as δ^{pcs} due to their favourable paramagnetic properties. It is generally accepted that no major structural changes occur upon substitution of Ca^{2+} with Ln(III).

Binding of Ln(III) ions to bovine α -lactalbumin has also been demonstrated. Several studies indicated that Ln(III) ions such as Tb(III), Eu(III), Gd(III), Yb(III), Pr(III), and Dy(III) indeed bind at the Ca(II) site in BLA [264–266]. More specifically, it was shown that the fluorescence changes taking place upon Ln(III) binding to apo-BLA are identical to those which have been observed in the case of Ca(II) [264]. This finding indicates that Ln(III) binding leads to stabilisation of the *N* state of the protein.

It has been shown that the mere addition of Ca^{2+} at a constant urea concentration initiates folding of BLA [44, 51, 52]. By comparison with literature data, it was concluded that the observed folding transition corresponds to folding of the apo-*MG* population [52]. As can be seen in Fig. 3.26, the chemical shift dispersion in the upfield aliphatic region of the 1D ^1H NMR spectrum is quite limited. This makes it difficult to extract kinetic rate constants for a representative number of residues throughout the protein. Substitution of Ca(II) with a suitable paramagnetic Ln(III) ion should principally result in significant pseudocontact shifts which would improve dispersion of chemical shifts. Since the report by Murakami *et al.* indicates that, in this study, (i) apo-BLA mainly adopts the apo-*MG* conformation and (ii) subsequent binding of Ln(III) ions has very similar effects as compared to Ca(II) binding [264], recent studies on the Ca^{2+} -induced folding of BLA (see above, [52]) may well be extended by substitution of Ca(II) with Ln(III) aiming at an improved chemical shift dispersion in time-resolved NMR studies.

Initial studies indeed showed that BLA folding may be induced at 4M urea concentration by the addition of Yb(III) which could be followed by stopped-flow fluorescence spectroscopy. In addition, the 1D ^1H NMR spectrum of BLA in the presence of Yb(III) displayed the chemical shift dispersion which is characteristic for a well-folded protein (Fig. 4.3, [81]).

This chapter deals with steady-state photo-CIDNP investigations of BLA. These shall provide a basis for real-time photo-CIDNP studies on the Yb(III)-induced folding in the future. Corresponding studies on the Ca(II)-induced folding identified a folding intermediate in the form of a transiently observable tyrosine side-chain [44, 81].

4. Photo-CIDNP studies on bovine α -lactalbumin in the presence of paramagnetic lanthanides

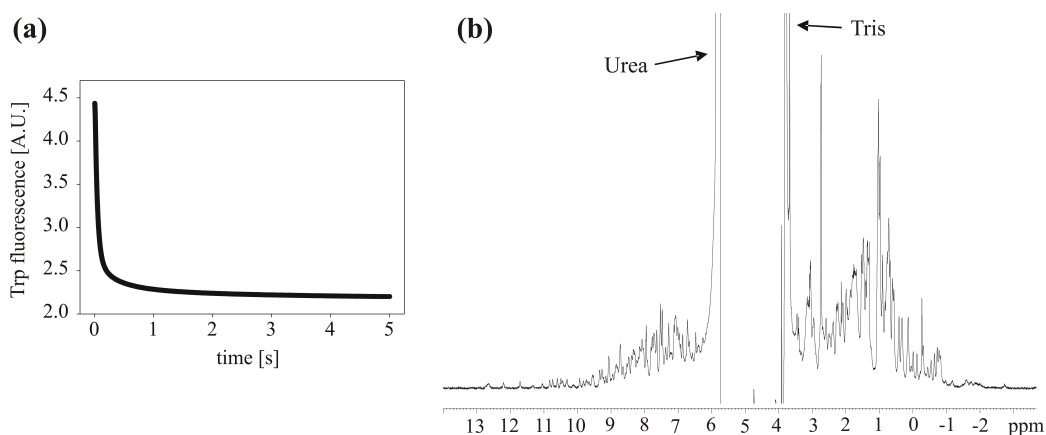


Figure 4.3.: Yb(III)-induced folding of BLA (100 μ M apo-BLA, 200 μ M Yb(III), 4M urea, 50mM Tris/HCl pH7, 35 $^{\circ}$ C). (a) Stopped-flow fluorescence spectroscopy. (b) 1D 1 H NMR spectroscopy (spectrum unreferenced).

4.2. Materials and Methods

4.2.1. Photo-CIDNP NMR spectroscopy

Photo-CIDNP NMR spectra were acquired in the group of Prof. Peter J. Hore in the Physical and Theoretical Chemistry Laboratory at Oxford University, UK, together with Iain J. Day and Lars T. Kuhn. Spectra were taken using a 600MHz Varian Inova NMR spectrometer whereas sample illumination was accomplished using an Argon ion laser (Spectra Physics, model 2085-25 S) in an experimental setup similar to that depicted in Fig. 3.7 [267]. The laser pulse power was usually adjusted to values between 4 and 5 W at the end of the optical fibre and the illumination time was typically 100ms. Photo-CIDNP NMR spectra were generated as the difference between the 'light' and the 'dark' spectra. For each spectrum, 16 transients were recorded. All experiments were performed at 35 $^{\circ}$ C.

Two sets of experiments were conducted. The first set constitutes the investigation of the influence of the presence of paramagnetic Yb(III) ions on photo-CIDNP of the three polarisable amino acid side-chains. For this, N-acetylated versions of tryptophan, tyrosine, and histidine were employed. Specifically, experimental conditions were 10% D₂O, 0.2mM flavin mononucleotide (FMN), 5mM YbCl₃, and 5mM N-Acetyltryptophan, N-Acetyltirosine, and N-Acetylhistidine, respectively. A three-second interval between successive laser pulses was

implemented. In the second set of experiments, photo-CIDNP spectra of bovine α -lactalbumin were taken in 25mM citric acid pH 7 (90% D₂O). Differences between apo-BLA and the Ca²⁺- and Yb³⁺-bound states of BLA, respectively, were investigated. Moreover, the influences of the denaturant urea and of the type of flavin dye used (FMN and flavin I, respectively (Fig. 4.4, [253])) were studied.

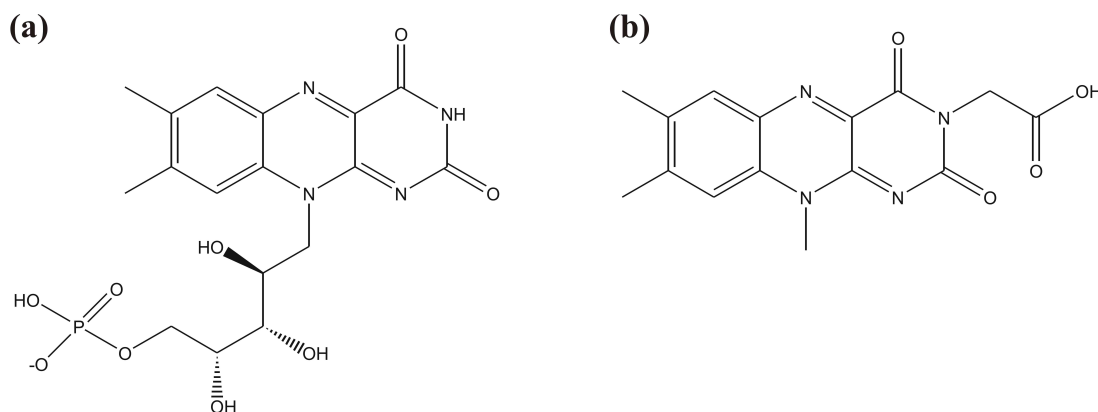


Figure 4.4.: Chemical structures of flavin dyes FMN (a) and flavin I (3-N-carboxymethylflavin) (b).

In respective experiments, concentrations employed were 1mM BLA, 1.5mM CaCl₂, 1.5mM YbCl₃, 0.2mM FMN, 0.2mM flavin I, 2M urea, and 0.1mM 1,4-dioxane (chemical shift reference substance). The interval between successive laser pulses was at least 30 seconds.

4.2.2. Organic synthesis of flavin I

The protocol for the two-step organic synthesis of the CIDNP dye flavin I follows earlier described procedures [268, 269].

Synthesis of 3-Ethoxycarbonylmethylflavin (2)

In the first reaction, lumiflavin (1) (200mg, 0.78mmol) was dissolved in 20ml dry dimethyl formamide (DMF) using a dried flask. This led to an orange-coloured suspension. After the addition of K₂CO₃ (100mg, 0.72mmol), ethylbromoacetate (100 μ l, 0.9mmol) was added stepwise. The reaction mixture was stirred at room temperature for two days (it is important that the reaction is performed under an Argon atmosphere to prevent hydrolysis of the reaction

4. Photo-CIDNP studies on bovine α -lactalbumin in the presence of paramagnetic lanthanides

product (2)). The solvent (DMF) was removed using both a rotary evaporator and an oil pump (40°C water bath temperature) yielding a brown solid. This was followed by dissolving the reaction product (2) in dichloromethane (DCM). Water was added and the organic phase was subsequently extracted three times with DCM in conjunction with brine (saturated NaCl solution) to improve phase separation. The organic phases were combined and dried using MgSO₄ which was removed afterwards by filtration. DCM was evaporated and the reaction product was finally separated from residual traces of lumiflavin by HPLC (RPC18AQ column) using a mixture of 62% H₂O and 38% CH₃CN (pH 2). 3-Ethoxycarbonylmethylumiflavin (2) was obtained as a solid after lyophilisation.

When purifying (2), it is essential to avoid usage of bases such as NaOH as they lead to decomposition of the tricycle [268].

Synthesis of 3-N-carboxymethylumiflavin (3)

In the second reaction, 3-Ethoxycarbonylmethylumiflavin (2) (65mg, 0.19mmol) was dissolved in 20ml concentrated HCl. The solution was refluxed for six hours. Dilution with H₂O (dest.) and cooling (4°C) over the weekend resulted in precipitation of flavin I (3) in the form of orange-red crystals.

4.3. Results

4.3.1. Photo-CIDNP of N-Acetylhistidine, N-Acetyltryptophan, and N-Acetyltyrosine in the presence of ytterbium(III)

A thorough literature survey reveals only a single paper reporting on photo-CIDNP experiments in the presence of paramagnetic Ln(III) employing proteins [260]. Therefore, in preparation for photo-CIDNP studies on BLA in the presence of paramagnetic Yb(III), experiments were conducted in which the influence of Yb(III) on photo-CIDNP of N-Acetyltryptophan, N-Acetyltyrosine, and N-Acetylhistidine, respectively, was investigated. These experiments are more easily interpretable and should provide a useful basis for the analysis of photo-CIDNP data in the context of proteins.

Fig. 4.5 shows the 1D ¹H and photo-CIDNP NMR spectra of N-Acetyltryptophan.

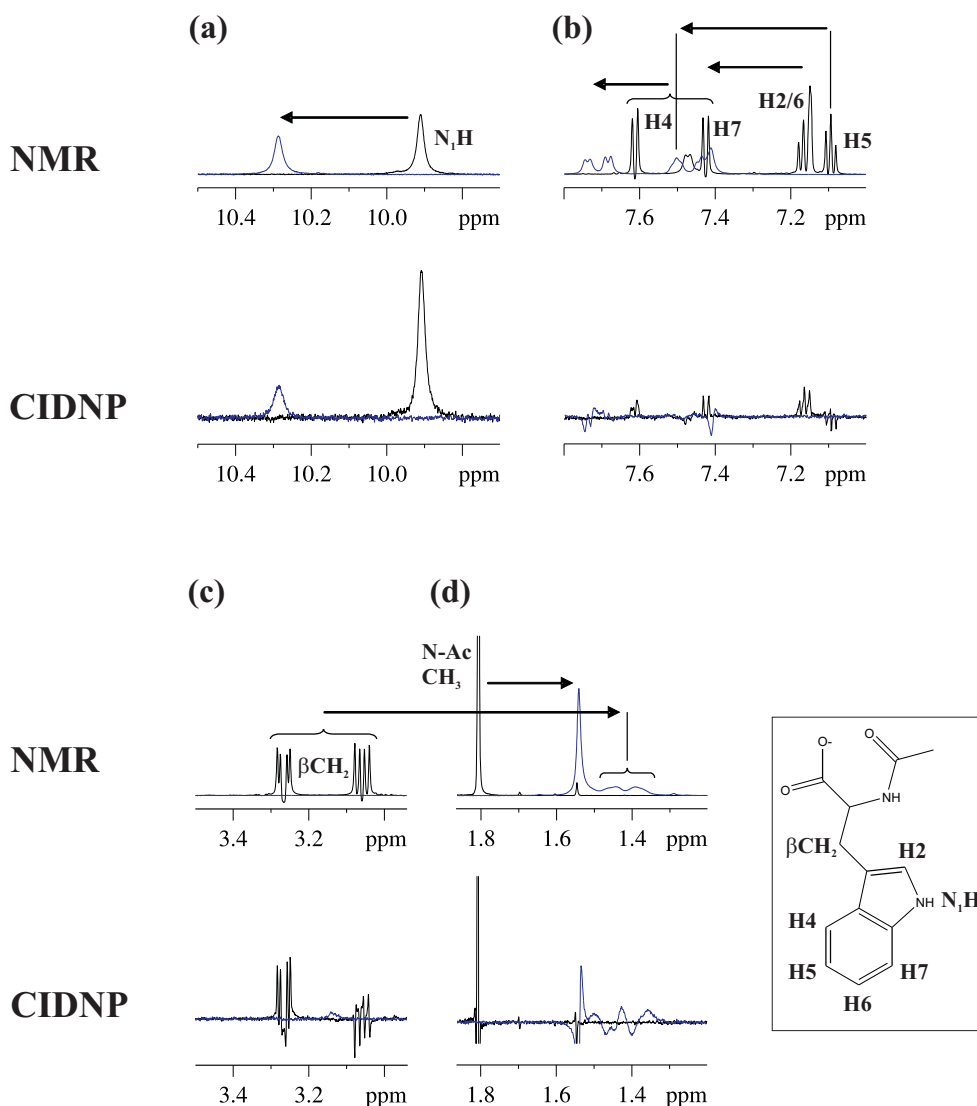


Figure 4.5.: 1D ^1H and photo-CIDNP NMR spectra of N-Acetyltryptophan in the absence (black) and in the presence (blue) of Yb(III) , respectively. Shown are the downfield (top) and upfield (bottom) spectral regions. Overlaid spectral parts are shown at the same noise level. Changes in chemical shifts upon addition of Yb(III) are indicated by arrows. The inset displays the chemical structure of N-Acetyltryptophan including labels for relevant H atoms. Signals from nuclei different from those indicated in the inset have been cut off for illustrative purposes if needed.

4. Photo-CIDNP studies on bovine α -lactalbumin in the presence of paramagnetic lanthanides

All polarisable H atoms experience significant changes in their chemical shifts due to the presence of Yb(III). The indole and aromatic protons exhibit downfield shifts between 0.19ppm and 0.41ppm whereas the methylene protons show a very large upfield shift (-1.74ppm) (Table 4.1).

Table 4.1.: Changes in chemical shifts and polarisation behaviour for the polarisable H atoms of N-Acetyltryptophan.

Atom(s)	$\Delta\delta (= \delta_{\text{para}} - \delta_{\text{dia}})$ [ppm]	CIDNP _{dia}	CIDNP _{para}
N ₁ H	0.38	as expected	attenuated
H4	0.19	as expected	absorptive and emissive components
H7	0.19	comparable to H4	absorptive and emissive components
H2/6	0.26	as expected	emissive for upfield part
H5	0.41	emissive	no enhancement
β CH ₂	-1.74	upfield part partly emissive, downfield part absorptive	absorptive and emissive components

Additionally, all signals display significant line broadening. The CIDNP data are only of limited use. In the absence of Yb(III), the N₁H, H2, H4, and H6 protons show absorptive enhancements as it is expected. However, the results for the H5, H7, and β methylene protons are inconsistent with literature data [253, 254]. In principle, the H5 and H7 protons do not show any enhancement and the β CH₂ protons exhibit an emissive enhancement. This not in accordance with the spectra shown in Fig. 4.5 (see also Table 4.1). On this basis, the CIDNP results for the paramagnetic situation can only be of preliminary nature. In general, when Yb(III) is present, CIDNP signals seem to be attenuated and/or to consist of absorptive as well as emissive components.

Fig. 4.6 depicts the results obtained for N-Acetyltyrosine.

The presence of Yb(III) leads to considerable changes in chemical shifts, namely, downfield shifts for the aromatic H2,6 and H3,5 protons (0.03ppm and 0.14ppm, respectively) and upfield shifts for the β methylene protons (-0.56ppm and -0.50ppm, respectively). Moreover, significant line broadening is observed for respective resonances. The photo-CIDNP spectrum in the diamagnetic case exhibits the features which are characteristic for tyrosine: absorptive enhancements for the H2,6 and β methylene protons and a large emissive enhancement for the H3,5 protons [253, 254]. Upon addition of Yb(III), enhancements become significantly attenuated. Interestingly, the CIDNP signal for the H2,6 protons indicates absorptive as well

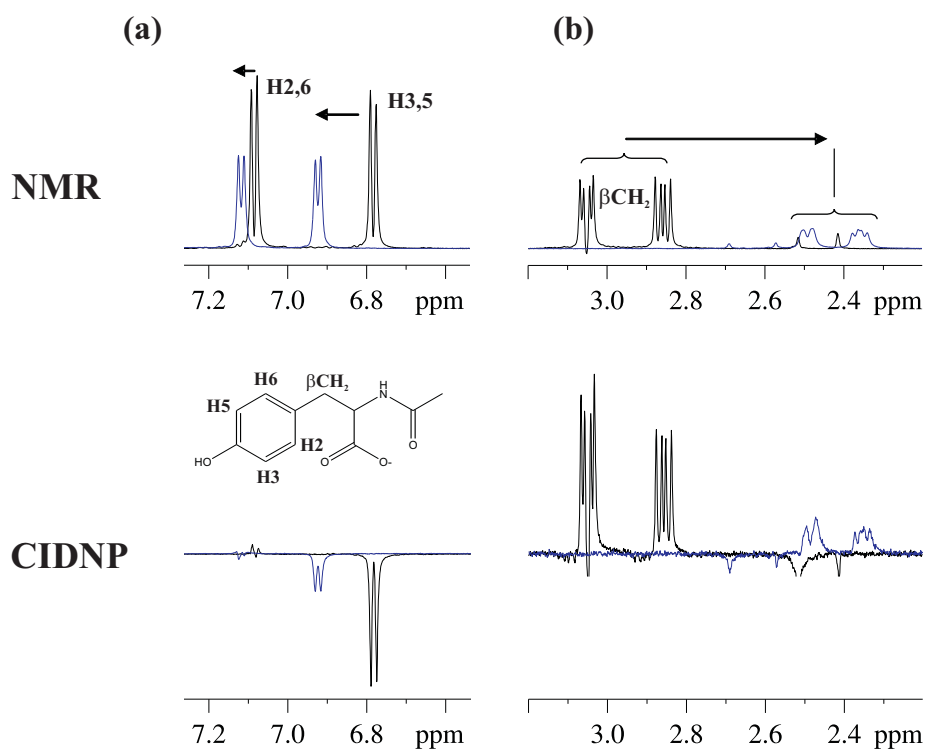


Figure 4.6.: 1D ^1H and photo-CIDNP NMR spectra of N-Acetyltyrosine in the absence (black) and in the presence (blue) of Yb(III), respectively. Shown are the downfield (left) and upfield (right) spectral regions. Overlaid spectral parts are shown at the same noise level. Changes in chemical shifts upon addition of Yb(III) are indicated by arrows. The inset displays the chemical structure of N-Acetyltyrosine including labels for relevant H atoms. Signals from nuclei different from those indicated in the inset, potentially arising from FMN or lowly concentrated impurities, have been cut off for illustrative purposes if needed.

4. Photo-CIDNP studies on bovine α -lactalbumin in the presence of paramagnetic lanthanides

as emissive components even though the enhancement is very weak (Table 4.2).

Table 4.2.: Changes in chemical shifts and polarisation behaviour for the polarisable H atoms of N-Acetyltyrosine. β CH(a) and β CH(b) indicate the downfield and upfield multiplets of the β methylene protons, respectively.

Atom(s)	$\Delta\delta$ ($= \delta_{\text{para}} - \delta_{\text{dia}}$) [ppm]	CIDNP _{dia}	CIDNP _{para}
H2,6	0.03	as expected	absorptive and emissive components
H3,5	0.14	as expected	attenuated
β CH(a)	-0.56	as expected	attenuated
β CH(b)	-0.50	as expected	attenuated

The results obtained for N-Acetylhistidine are shown in Fig. 4.7.

While the downfield shifts experienced by the H2 and H4 protons due to the presence of Yb(III) are very small (0.014ppm and 0.005ppm, respectively), the β methylene protons display a large corresponding upfield shift (-0.81ppm). In addition, all signals experience significant line broadening. The CIDNP spectrum in the absence of Yb(III) is characteristic for N-Acetylhistidine: strong absorptive enhancements for the H2 and H4 protons and a weaker emissive enhancement for the β CH₂ protons [253,254]. In the presence of Yb(III), the polarisations for the aromatic protons almost vanish. The signal pattern for the β methylene protons suggests absorptive as well as emissive contributions while the enhancement is much weaker as in the diamagnetic case (Table 4.3).

Table 4.3.: Changes in chemical shifts and polarisation behaviour for the polarisable H atoms of N-acetyl histidine.

Atom(s)	$\Delta\delta$ ($= \delta_{\text{para}} - \delta_{\text{dia}}$) [ppm]	CIDNP _{dia}	CIDNP _{para}
H2	0.014	as expected	almost vanished
H4	0.005	as expected	almost vanished
β CH ₂	-0.81	as expected	attenuated (absorptive and emissive components)

The results for N-Acetyltyptophan, N-Acetyltyrosine, and N-Acetylhistidine can be summarised as follows: upon addition of Yb(III), downfield and upfield shifts are observed for

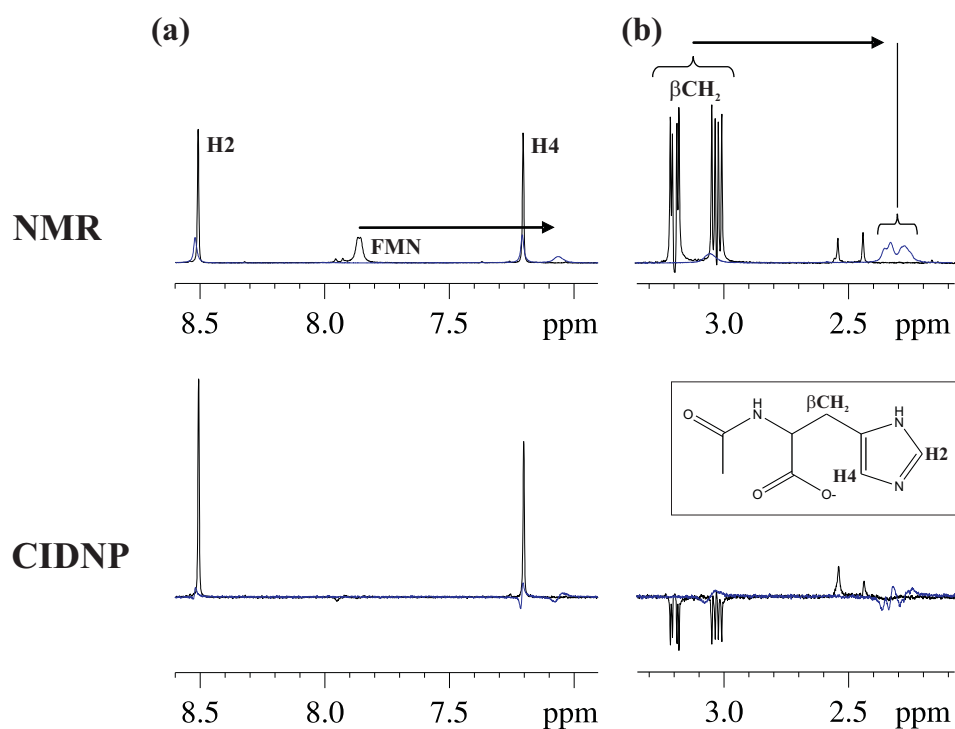


Figure 4.7.: 1D ^1H and photo-CIDNP NMR spectra of N-Acetylhistidine in the absence (black) and in the presence (blue) of Yb(III), respectively. Shown are the downfield (left) and upfield (right) spectral regions. Overlaid spectral parts are shown at the same noise level. Changes in chemical shifts upon addition of Yb(III) are indicated by arrows. The inset displays the chemical structure of N-Acetylhistidine including labels for relevant H atoms. Signals from nuclei different from those indicated in the inset may arise from FMN or lowly concentrated impurities.

4. Photo-CIDNP studies on bovine α -lactalbumin in the presence of paramagnetic lanthanides

the aromatic (and indolic) and β methylene protons, respectively. Moreover, resonances of CIDNP-active nuclei become significantly broadened. Although the CIDNP results for N-Acetyltyptophan are somewhat questionable, it can further be concluded that Yb(III) leads to markedly reduced CIDNP enhancements.

4.3.2. Organic synthesis of flavin I

Photo-CIDNP experiments in the presence of Yb(III) by Kuhn and Hore using FMN as the photosensitiser have shown that insoluble complexes of Yb(III) and FMN may form under certain experimental conditions hindering the acquisition of photo-CIDNP spectra. Since the negatively charged phosphate group moiety appears to play an important role in complex formation [270], FMN was replaced in further experiments by the flavin derivative 3-N-carboxymethylflavin (flavin I) which has a different side-chain chemistry preventing complex formation with Yb(III) (Fig. 4.4). Flavin I has long been known as a photosensitiser for photo-CIDNP experiments and successfully applied to the study of surface properties of proteins [251, 252, 271–273].

Fig. 4.8 shows the reaction scheme which was followed when synthesising flavin I. It is slightly different from the procedure given in section 4.2.2 (see below).

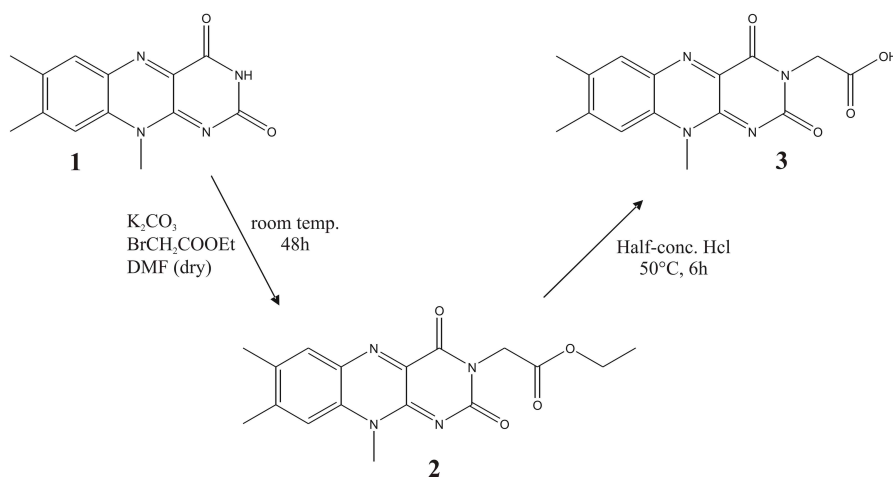


Figure 4.8.: Two-step organic synthesis of the CIDNP dye flavin I (3-N-carboxymethylflavin (3)).

NMR spectra of (2) and (3) were measured at 250MHz and not referenced.

Synthesis of 3-Ethoxycarbonylmethylumiflavin (2)

^1H NMR (CDCl_3): 8.07 (s, 1H), 7.44 (s, 1H), 4.87 (s, 2H), 4.23 (q, $^3J = 7.2$ Hz, 2H), 4.14 (s, 3H), 2.56 (s, 3H), 2.46 (s, 3H), 1.29 (t, $^3J = 7.2$ Hz, 3H).

MALDI-MS: 343.28 (found), 342.35 (expected).

R_f (DCM:MeOH=9:1) = 0.7.

Yield: 128mg (48%).

It turned out that 20ml DMF are not sufficient to dissolve 200mg lumiflavin. Several syntheses of 3-Ethoxycarbonylmethylumiflavin (2) indicated that at least the 5-fold amount of DMF is needed to fully dissolve lumiflavin. This might be important in order to increase the yield of the reaction for which a value of 85% has been reported [269]. Moreover, more than 50ml of DCM are needed for dissolving the reaction product after reaction completion (being the first step in the sequence of purification steps).

Synthesis of 3-N-carboxymethylumiflavin (3)

^1H NMR (MeOD): 8.01 (s, 1H), 7.84 (s, 1H), 4.79 (s, 2H), 4.17 (s, 3H), 2.61 (s, 3H), 2.50 (s, 3H).

ESI-MS(-): 313.1 (found), 314.3 (expected).

R_f (ddH₂O (pH2):CH₃CN=7:3) = 0.2 (here, a reverse phase TLC was performed (RP8, Merck)).

Yield: 8.7mg (15%).

When following the procedure given in section 4.2.2 (refluxing for 6h using concentrated HCl), precipitation of (3) was very limited giving yields below 7%. Upon choosing milder reaction conditions (stirring for 6h at 50°C using half-concentrated HCl, Fig. 4.8), precipitation was still limited. However, about 9mg of flavin I could be obtained after the reaction solution had been lyophilised. Optimisation of the experimental conditions should principally result in higher reaction yields. This was not pursued since about 30 μg are already sufficient for one NMR sample having 0.2mM flavin I concentration.

4.3.3. Photo-CIDNP investigations on bovine α -lactalbumin using flavin I

Comparison between FMN and flavin I

In order to prevent undesirable complex formation with Yb(III) [270], flavin I was used instead of FMN as the photosensitiser in photo-CIDNP experiments on bovine α -lactalbumin. First, differences between FMN and flavin I in probing the protein's surface properties were examined. Fig. 4.9 shows the aromatic part of the ^1H NMR spectra of apo-BLA in the presence of FMN and flavin I, respectively, along with corresponding CIDNP spectra.

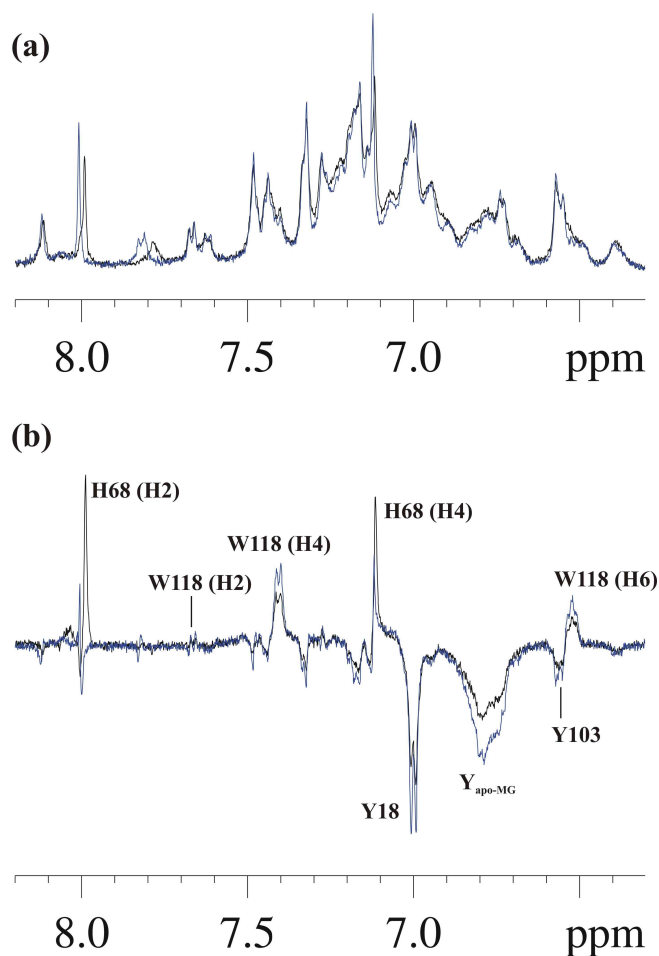


Figure 4.9.: 1D ^1H (a) and photo-CIDNP (b) NMR spectra of apo-BLA using FMN (black spectra) and flavin I (blue spectra), respectively, as the photosensitiser.

The ^1H NMR spectra indicate that the structure of the protein is not affected irrespective

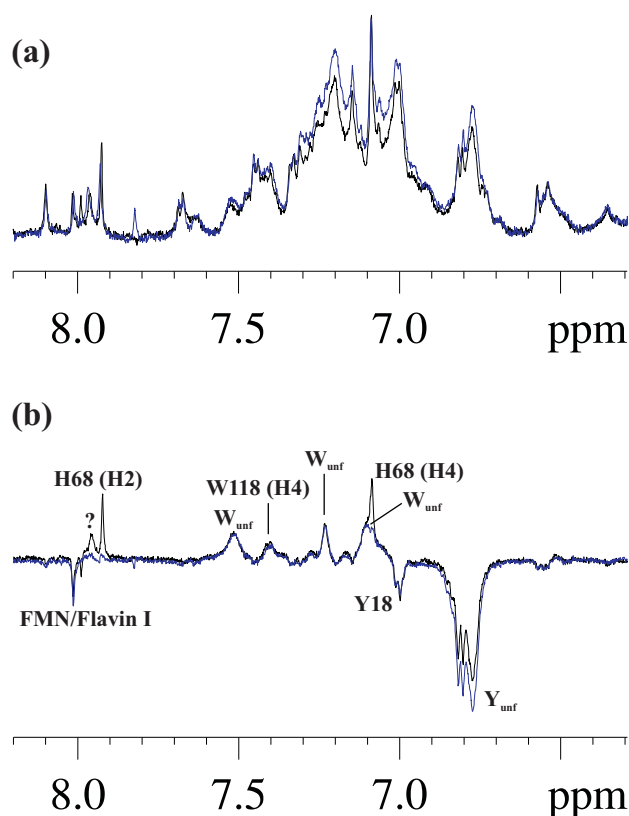


Figure 4.10.: 1D ^1H (a) and photo-CIDNP (b) NMR spectra of apo-BLA in the presence of 2M urea using FMN (black spectra) and flavin I (blue spectra), respectively, as the photosensitiser.

of the CIDNP dye present. Slight chemical shift differences at about 7.8ppm and 8.0ppm are most likely due to the different dyes employed. The CIDNP spectra show that the two dyes do not behave identically although the overall spectral signature is the same. Interestingly, when flavin I is used, tryptophan and tyrosine residues are more strongly polarised whereas polarisation of histidine side-chains is markedly reduced.

The situation is different if urea (2M) is present (Fig. 4.10).

Again, the overall CIDNP spectral pattern is similar. However, tryptophan side chain nuclei are equally strongly polarised irrespective of the dye used. When going from FMN to flavin I, enhanced polarisation for tyrosine residues is only detected for the population of unfolded molecules in the sample (Y_{unf}) but not for the native state population indicated by emissive polarisation for Y18 (for peak assignments and interpretations in terms of distributions of conformations see below). A very clear difference between the two spectra is the complete

4. Photo-CIDNP studies on bovine α -lactalbumin in the presence of paramagnetic lanthanides

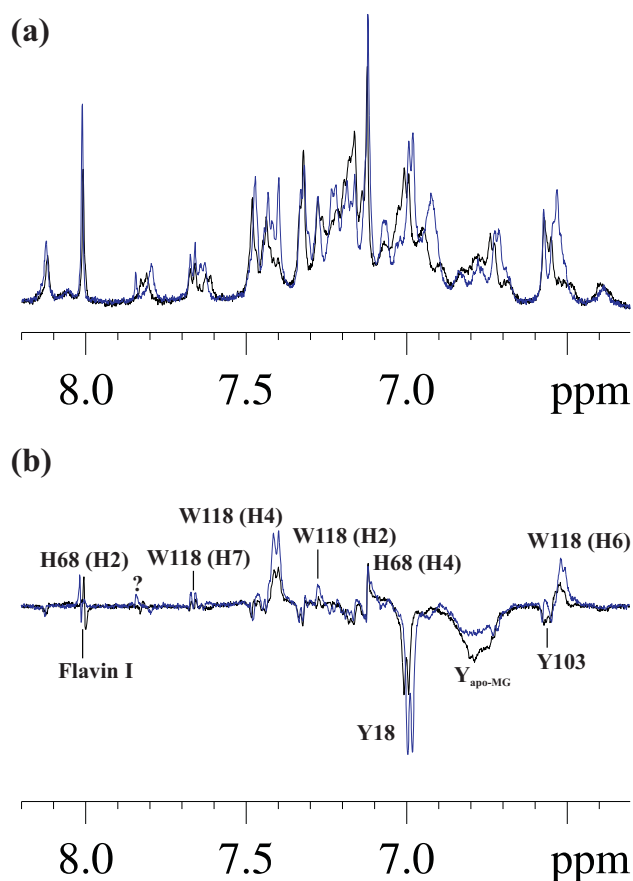


Figure 4.11.: Effect of Ca^{2+} binding on the 1D ^1H (a) and photo-CIDNP (b) NMR spectra of BLA using flavin I as the photosensitiser. Black and blue spectra indicate the apo state and holo state of the protein, respectively.

absence of histidine polarisation if flavin I is used. It is concluded from Figs. 4.9 and 4.10 that FMN and flavin I react in a similar fashion with exposed aromatic amino acid side-chains of BLA. The overall spectral pattern does not change although significant differences in polarisation of CIDNP-active nuclei may occur.

Influences of Ca^{2+} , Yb^{3+} , and urea on the exposure of aromatic amino acid side-chains in bovine α -lactalbumin using flavin I as the CIDNP photosensitiser

Fig. 4.11 displays the ^1H and photo-CIDNP NMR spectra of apo- and Ca^{2+} -bound α -lactalbumin.

It can be seen that significant changes occur throughout the aromatic region of the spectrum

if Ca^{2+} is bound to the protein (Fig. 4.11a). These changes are also evident from the CIDNP spectra of the two states (Fig. 4.11b).

It is known that the apo protein forms an *MG* at neutral pH under low salt conditions [223] which, at 35°C, will be in an equilibrium with the *N* state [224]. These conditions are fulfilled in the presented studies and the coexistence of the apo-*MG* state and the *N* state is therefore also evident from the CIDNP spectrum of apo-BLA shown in Fig. 4.11b (black spectrum). Most of the observable resonances correspond to the native state of BLA, namely to H68 (H2 and H4 at 8.00ppm and 7.12ppm, respectively), W118 (H4, H6, and H7 at 7.41ppm, 6.52ppm, and 7.66ppm, respectively), Y18 (H3,5 at 7.00ppm) [43–45, 273, 274], and Y103 (H3,5 at 6.56ppm) [275]. The broad emissive resonance around 6.74–6.82ppm instead is characteristic of tyrosine H3,5 protons in the apo-*MG*. These protons correspond to Y18, Y36, and Y50 which are known to be exposed to solvent in all types of BLA molten globules [46]. The fact that the chemical shift range for the H3,5 protons in Y18, Y36, and Y50 in the *A* state molten globule has been reported to be 6.73–6.76ppm (measured at 35°C) further seems to support this assignment [276].

As mentioned, the binding of Ca^{2+} induces significant changes in the photo-CIDNP spectrum. Since Ca^{2+} is known to stabilise the native state of the protein [199], binding of Ca^{2+} leads to a shift of the conformational equilibrium from the apo-*MG* state towards the *N* state. This is reflected by stronger polarisations for CIDNP-active nuclei in Y18 and W118 in the native state. Additionally, the H2 proton of W118 at 7.28ppm appears. At the same time, polarisation for tyrosine H3,5 protons in the apo-*MG* is markedly weaker. Comparison of the spectra of the apo and holo states of the protein also suggests that Ca^{2+} binding is accompanied by slight changes in chemical shifts for certain H nuclei (H2 of H68, H3,5 of Y18, and H6 of W118).

The influence of Yb^{3+} on the structure of BLA was also investigated (Fig. 4.12).

Inspection of the ^1H and photo-CIDNP spectra unexpectedly shows that the presence of Yb^{3+} does not induce any change in the structure of the protein apart from a systematic upfield shift of all the resonances. This finding could be attributed to the citrate buffer molecules which are largely deprotonated at pH 7 and therefore strongly bind Yb^{3+} ions leaving protein molecules unbound. The citrate signal in the upfield region of the ^1H NMR spectrum was no longer observable due to severe line broadening (data not shown). The same observations

4. Photo-CIDNP studies on bovine α -lactalbumin in the presence of paramagnetic lanthanides

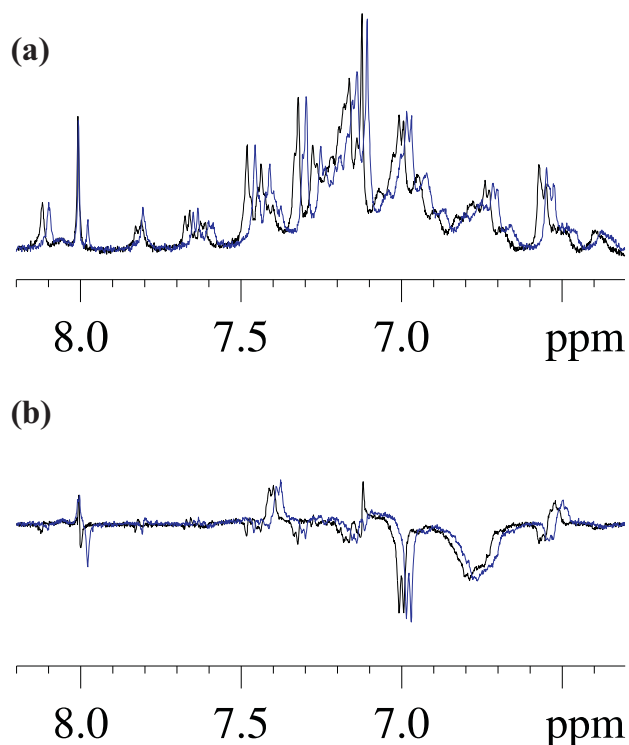


Figure 4.12.: 1D ^1H (a) and photo-CIDNP (b) NMR spectra of BLA in the absence (black spectra) and in the presence of Yb^{3+} (blue spectra).

were made when the influence of Yb^{3+} on BLA was investigated in the presence of urea (2M) (data not shown).

Fig. 4.13 compares the spectra of the apo protein in the absence of urea and in the presence of 2M urea.

The presence of 2M urea induces significant changes in the structure of the protein as is evident from the ^1H NMR spectra shown in Fig. 4.13a. The photo-CIDNP spectrum of apo-BLA when urea is absent is the one shown in Fig. 4.11b. Assignments of resonances characteristic for the N state population have therefore simply been transferred. The presence of 2M urea, conditions where the unfolded (U), MG , and N states are expected to be populated [221], has three effects. First, resonances of the N state either show much weaker polarisations as it is the case for Y18, Y103, and W118 or even completely vanish which holds true for H68. Second, several resonances showing considerable absorptive enhancements appear in the CIDNP spectrum which can be assigned to (a) tryptophan residue(s) in the U state [276]. Respective resonances appear at 7.52ppm, 7.23ppm, and 7.10ppm (indicated by W_{unf} in Fig. 4.13b).

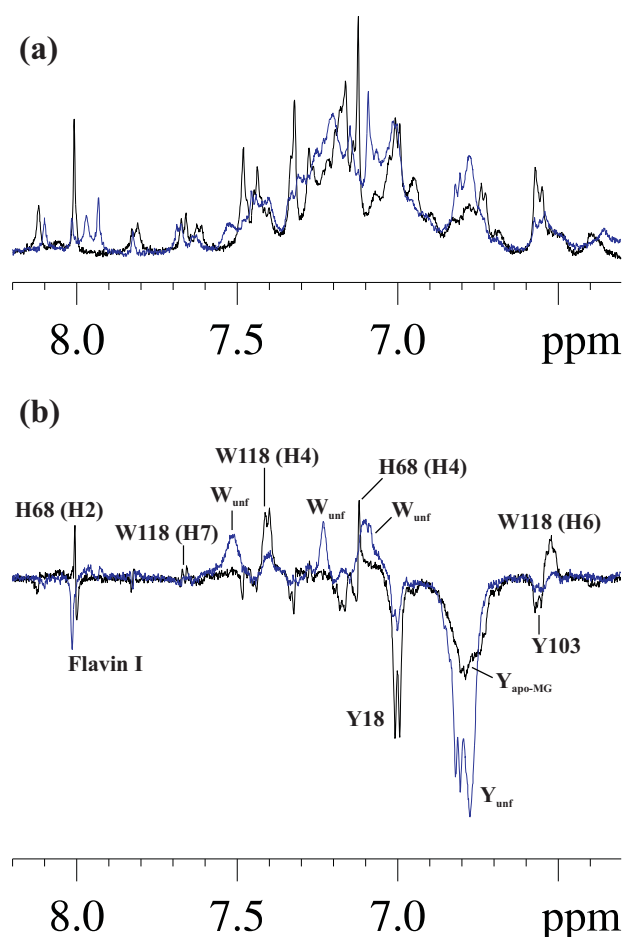


Figure 4.13.: Effect of urea on the 1D ^1H (a) and photo-CIDNP (b) NMR spectra of apo-BLA using flavin I as the photosensitizer. Black and blue spectra indicate the protein in the absence of urea and in the presence of urea, respectively.

Third, the character of the strong emissive resonance around 6.80ppm changes significantly. The sharpening of this resonance area and its slight downfield shift as compared to the apo protein in the absence of urea indicate both a significant fraction of unfolded molecules and a reduced amount of *MG* molecules in the sample. Comparison with the *A* state *MG* suggests that the emissive resonance with a chemical shift range of 6.76-6.82ppm corresponds to the H3,5 protons of Y18, Y36, and Y50 in the *U* state [276]. The photo-CIDNP spectra in Fig. 4.13 therefore show a significant shift in the conformational distribution from the *N* state to the *U* state and a likely concomitant reduction in the *MG* state population upon addition of 2M urea.

4. Photo-CIDNP studies on bovine α -lactalbumin in the presence of paramagnetic lanthanides

In the following, the effect of urea was further studied for the Ca^{2+} -bound protein (Fig. 4.14).

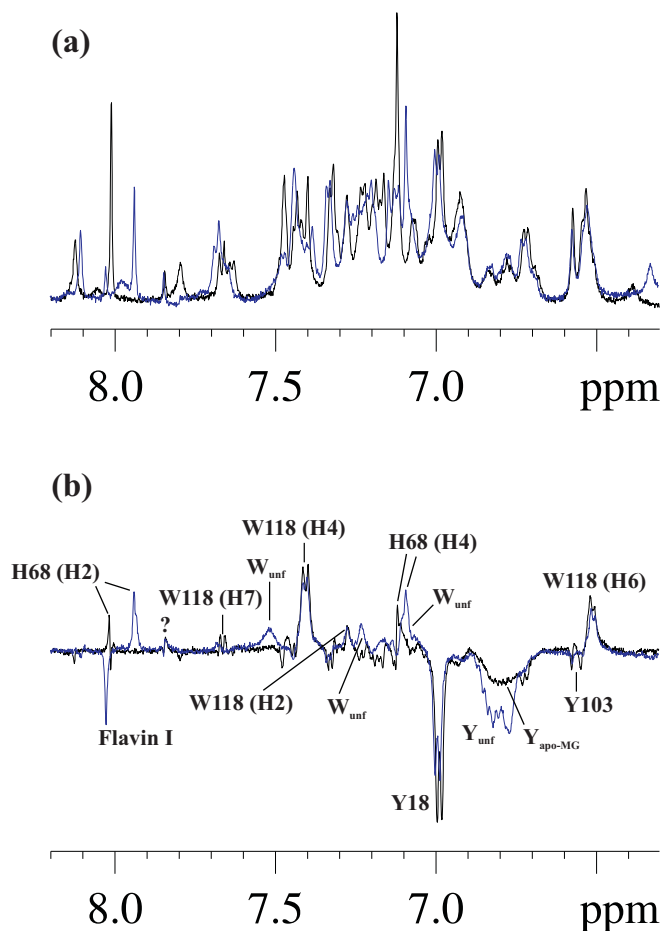


Figure 4.14.: Effect of urea on the 1D ^1H (a) and photo-CIDNP (b) NMR spectra of Ca^{2+} -bound BLA using flavin I as the photosensitiser. Black and blue spectra indicate the protein in the absence of urea and in the presence of urea, respectively.

Structural changes are already evident from the ^1H NMR spectra but more pronounced when comparing the photo-CIDNP NMR spectra. Generally, the same reasoning applies as for the apo protein (cf. Fig. 4.13). Urea leads to a shift from the N state to the U state as reflected in the respective resonances. The changes are, however, not as pronounced underlining the strongly stabilising effect which is being exerted by Ca^{2+} on native BLA. As opposed to the apo protein, this might be the reason why polarisation for the H2 and H4 protons of H68 is still detectable when urea is present. Simultaneously, the spectra suggest that urea leads to

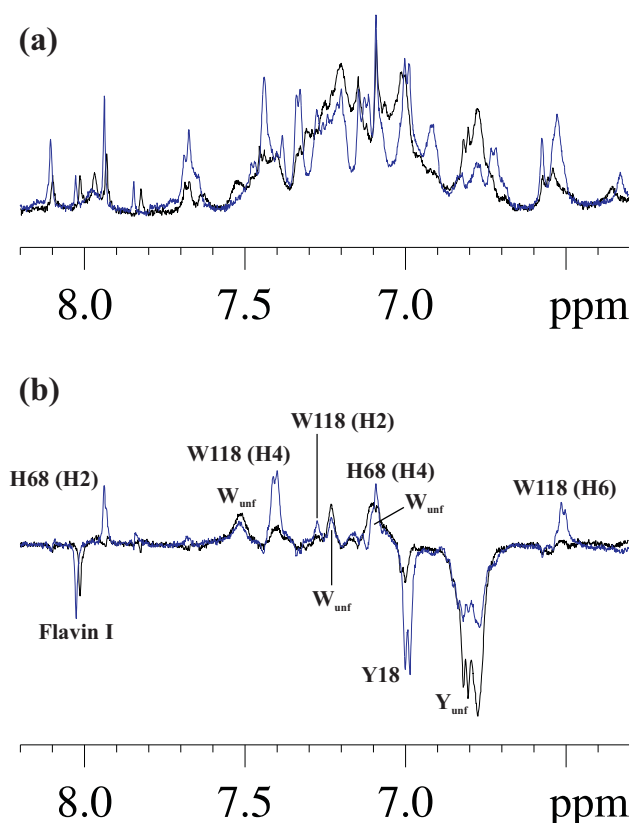


Figure 4.15.: Effect of Ca^{2+} binding on the 1D ^1H (a) and photo-CIDNP (b) NMR spectra of BLA in the presence of 2M urea using flavin I as the photosensitiser. Black and blue spectra indicate the apo state and holo state of the protein, respectively.

significant upfield shifts for these protons, namely from 8.00ppm to 7.92ppm (H2) and from 7.12ppm to 7.09ppm (H4).

Fig. 4.15 shows the effect of Ca^{2+} binding to BLA when urea (2M) is present.

Ca^{2+} binding has marked influences on the protein's structure as can be seen from both the ^1H and photo-CIDNP NMR spectra. Spectra shown are identical to those given in Figs. 4.13 and 4.14, respectively. Assignments were accordingly transferred. The overlay clearly illustrates the aforementioned stabilising effect of Ca^{2+} on the structure of the *N* state leading to a reduction in the *U* state population.

Table 4.4 summarises chemical shifts of CIDNP-active nuclei and their assignments to specific residues in the respective conformations.

Table 4.4.: Chemical shifts of CIDNP-active nuclei of BLA and assignments to specific residues in the unfolded (*U*), molten globule (*MG*), and native (*N*) states.

Conformation	Amino acid	H atom(s)	δ [ppm]
U	Y18, Y36, Y50	H3,5	6.76-6.82
U	W	-	7.10
U	W	-	7.23
U	W	-	7.52
apo-MG	Y18, Y36, Y50	H3,5	6.74-6.82
N	Y18	H3,5	7.00
N	H68	H2 ^a	8.00
N	H68	H2 ^b	7.92
N	H68	H4 ^a	7.12
N	H68	H4 ^b	7.09
N	Y103	H3,5	6.56
N	W118	H2	7.28
N	W118	H4	7.41
N	W118	H6	6.52
N	W118	H7	7.66

^aUrea absent^bUrea present

4.4. Discussion

N-Acetyltryptophan, N-Acetyltyrosine, and N-Acetylhistidine were used to study the basic effects of Yb(III) on CIDNP of the three amino acids which are known to be polarisable. ¹H NMR spectra show that Yb(III) leads to significant changes in chemical shifts which are downfield and upfield in nature for the aromatic (and indolic) and aliphatic protons, respectively. Significant line broadening has been observed for the resonances of all CIDNP-active nuclei. The presence of Yb(III) moreover brings about markedly reduced CIDNP enhancements. These observations are in good agreement with earlier work [258–260]. The results obtained for N-Acetyltryptophan are only of preliminary nature since its CIDNP features for

the diamagnetic case could not be fully reproduced [253, 254]. Further experiments need to be conducted in order to clarify whether certain CIDNP signals in the paramagnetic situation truly exhibit absorptive as well as emissive components as has been observed here for all amino acids. Generally, photo-CIDNP experiments on the single amino acids in the presence of Yb(III) are feasible albeit with less sensitivity.

The flavin dye 3-N-carboxymethylflavin (flavin I) was synthesised in this work to replace FMN which, under certain conditions, has been shown to form insoluble complexes with Yb(III) hindering the acquisition of photo-CIDNP data [270]. Available protocols for the two-step organic synthesis of flavin I [268, 269] turned out to be of only limited use. Since the synthesis of the dye was successfully accomplished and synthesised amounts were sufficient for about 300 NMR samples, no serious attempts aiming at improving reaction yields were undertaken.

Preliminary photo-CIDNP studies employing bovine α -lactalbumin and using flavin I as the photosensitiser show that flavin I can well be used as the triplet-state-radical-pair-generating dye which is in line with earlier investigations [251, 252, 271–273]. Comparison between FMN and flavin I reveals that the two dyes give the same overall spectral pattern although significant differences become obvious with respect to polarisation of certain CIDNP-active nuclei. Assignments for native state resonances show that Y18, H68, and W118 are accessible to solvent [43–45, 273, 274]. Interestingly, the spectra clearly indicate that Y103 is another residue which is solvent-accessible in the native state [275]. The photo-CIDNP spectra in the absence of urea can be interpreted in terms of a conformational equilibrium between the apo-*MG* state and the *N* state [46, 223, 224, 276]. If 2M urea are added, a significant fraction of *U* state molecules in the sample becomes additionally apparent [221, 276]. The analysis of the effect of Ca^{2+} binding shows its strongly stabilising effect on the *N* state of the protein [199]. The effect of Yb^{3+} binding could not be analysed since citrate was mistakenly used as the buffering agent. Yb^{3+} strongly binds to citrate at pH 7 impeding ion binding to the protein. Citrate might be replaced by cacodylate in future studies. Studies by Hore and Kaptein suggest that photo-CIDNP should be principally observable for the Yb^{3+} -bound protein [260].

A. Appendix

A.1. Heterologous expression and purification of ¹⁵N-labelled bovine α -lactalbumin

A. Appendix

Table A.1.: ¹⁵N-M9 minimal medium.

Component	Final concentration
MgSO ₄	2mM
CaCl ₂	0.1mM
Na ₂ HPO ₄	48mM
KH ₂ PO ₄	22mM
NaCl	8.6mM
¹⁵ NH ₄ Cl	18.4mM
Thiamine	0.01%
FeCl ₃	30μM
Trace elements	1μM
Ampicillin	200ug/ml

Table A.2.: Buffers for solubilisation of inclusion bodies.

Sonication buffer	Washing buffer	Solubilisation buffer
50mM Tris pH 8 (before addition of EDTA)	20mM Tris pH 8 (before addition of EDTA)	50mM Tris pH 8.5
25% sucrose	1% Triton X-100	300mM NaCl
1mM EDTA	1mM EDTA	8M urea
		20mM β-mercaptoethanol

Table A.3.: Buffers for purification of His₆-α-lactalbumin *via* a Ni-NTA column.

IMAC I	IMAC II	IMAC III
50mM Tris pH 8.5	50mM Tris pH 8.5	50mM Tris pH 8.5
300mM NaCl	500mM NaCl	300mM NaCl
4M urea	4M urea	4M urea
5mM β-mercaptoethanol	5mM β-mercaptoethanol	5mM β-mercaptoethanol
		500mM imidazole

A.1. Heterologous expression and purification of ¹⁵N-labelled bovine α-lactalbumin

Table A.4.: Buffer X

For refolding	Stock
10mM Tris pH 8.5	500mM Tris pH 8.5
50mM NaCl	2.5M NaCl
1mM CaCl ₂	50mM CaCl ₂

Table A.5.: Refolding buffers

Step 1	Step 2	Step 3
Buffer X	Buffer X	Buffer X
2.5mM red. glutathione	2.5mM red. glutathione	
0.5mM ox. glutathione	0.5mM ox. glutathione	
4M urea		

Table A.6.: Buffers for ion exchange chromatography

Buffer A	Buffer B
20mM NaH ₂ PO ₄ pH 6.5	20mM NaH ₂ PO ₄ pH 6.5
10mM NaCl	1M NaCl

A.2. NMR resonance assignments

Table A.7.: Backbone assignment for Ca²⁺-loaded BLA under refolding conditions (4M urea, 50mM Tris/HCl pH 7, 35°C)

Residue	HN [ppm]	HA[ppm]	N [ppm]
Q2	8.93		124.12
T4	7.89	4.64	108.08
K5	8.75	3.53	121.05
C6	8.59	4.61	112.80
E7	7.64	4.12	121.92
V8	8.14	3.14	120.93
F9	7.94		117.86
R10	7.30	4.11	114.67
E11	8.54	4.09	117.20
L12	7.83	4.14	117.53
K13	6.86	3.84	119.36
D14	9.17	4.56	118.63
L15	8.20	4.14	116.65
K16	7.66	3.70	122.47
G17	8.66		118.30
Y18	8.19	4.17	126.86
G19	8.84		118.52
G20	7.95		105.12
V21	7.01	3.76	122.58
S22	8.14	4.59	122.69
L23	9.54		120.13
P24		4.55	
E25	7.09	3.78	113.58
W26	8.12	4.53	123.79
V27	8.77		124.12
C28	7.66		120.27

A.2. NMR resonance assignments

Residue	HN [ppm]	HA[ppm]	N [ppm]
T29	8.46		117.97
T30	8.78	3.64	114.56
F31	7.80		125.66
H32	7.36	4.09	115.99
T33	7.86	4.17	115.38
S34	8.18	4.56	111.38
G35	7.86		115.55
Y36	7.03	3.88	106.00
D37	7.05	5.14	118.96
Q39	8.31	4.65	115.22
A40	6.90	3.98	122.58
I41	7.96	4.28	120.82
V42	8.31	4.25	127.63
Q43	8.78	4.81	127.74
N44	8.71		123.90
N45	8.91	4.63	118.74
D46	8.52	4.65	117.20
S47	7.56		113.47
T48	8.50		119.18
E49	8.71	4.62	124.67
Y50	8.42	4.19	118.08
G51	9.39		112.37
L52	10.15	4.00	124.11
F53	8.76		109.29
Q54	7.96	3.63	117.09
I55	8.07	3.65	123.35
N56	7.50	5.94	125.76
N57	8.42	5.31	118.52
K58	8.87	4.35	119.94
I59	7.55		111.60

A. Appendix

Residue	HN [ppm]	HA[ppm]	N [ppm]
W60	6.50	4.54	118.41
C61	7.93	5.79	112.09
K62	8.20	4.95	122.47
D63	8.43	4.92	128.07
D64	8.19		115.22
Q65	7.93	4.18	119.07
N66	8.01	5.32	114.45
S69	7.04		113.24
S70	8.10		122.25
N71	8.02	3.64	116.98
I72	8.74	3.60	120.05
C73	9.46	4.23	113.68
N74	7.93		120.60
I75	9.29	4.16	121.15
S76	7.72	4.97	120.16
C77	8.48	4.97	123.79
D78	8.11	3.76	115.60
K79	7.54	4.07	120.27
F80	8.11	5.62	121.70
L81	6.98	5.70	114.01
D82	7.40	4.83	118.52
D83	7.99	4.37	112.80
D84	7.38	4.91	118.19
L85	9.50	4.65	128.95
T86	8.51	3.68	115.66
D87	8.10	4.07	121.15
D88	10.35	3.93	125.10
I89	8.67	2.79	117.20
M90	7.48		118.95
C91	7.83	4.74	118.08

A.2. NMR resonance assignments

Residue	HN [ppm]	HA[ppm]	N [ppm]
V92	8.42	3.05	122.14
K93	7.76	3.47	116.98
K94	6.98	4.12	117.53
I95	8.02	3.02	121.59
L96	9.10	3.80	121.70
D97	8.35	4.44	118.30
K98	7.95	4.48	118.30
V99	8.89	4.46	115.44
G100	8.19		109.73
I101	8.27	4.31	117.53
N102	8.27	4.64	116.86
Y103	7.60	3.76	120.38
W104	8.05		117.42
L105	8.89		118.52
A106	9.00	4.04	116.43
H107	8.02		114.89
K108	8.48		115.43
A109	7.48		117.75
L110	7.80	4.59	114.45
C111	7.29	5.10	116.32
S112	8.05	4.53	111.05
E113	7.67	4.56	118.40
K114	8.79	4.09	118.41
L115	8.41	4.57	119.18
D116	8.34		117.53
Q117	7.95	4.05	115.77
W118	7.76	4.53	121.37
L119	6.93	4.23	115.44
C120	6.59	4.24	118.63
K122	8.00	4.18	121.48

A. Appendix

Residue	HN [ppm]	HA[ppm]	N [ppm]
L123	7.82	4.27	130.05

Table A.8.: Partial side-chain assignment for Ca²⁺-loaded BLA under refolding conditions (4M urea, 50mM Tris/HCl pH 7, 35°C)

Residue	γCH_2 [ppm]	γCH_3 [ppm]	δCH_3 [ppm]	H ϵ N [ppm]	N ϵ [ppm]
V8		0.41 ^a			
L12			0.62		
L15			-0.27, 0.17		
V21		0.02, 0.62			
W26				10.45	129.83
V42		0.62			
L52			0.29, 0.62		
I55	0.33		-0.09		
I59		-0.16			
W60				9.87	128.51
I72		0.41 ^a	-0.16		
I75			0.41 ^a		
I89			0.53		
V92		-0.49, 0.17			
I95		-0.16	-0.60		
L96			0.17		
V99		-0.16			
I101	0.17	0.62	-0.09		
W104				11.61	130.49
W118				10.69	130.49

^aAliphatic groups of V8 γCH_3 , I72 γCH_3 , and I75 δCH_3 could not be unambiguously assigned to peaks 10 and 11 at 0.40ppm and 0.42ppm, respectively. The given chemical shift of 0.41ppm is an 'average' chemical shift.

B. References

- [1] Young, J. C.; Agashe, V. R.; Siegers, K.; Hartl, F. U. *Nat Rev Mol Cell Biol* **2004**, *5*, 781–791.
- [2] Anfinsen, C. B.; Haber, E.; Sela, M.; White, F. H. *Proc Natl Acad Sci U S A* **1961**, *47*, 1309–1314.
- [3] Anfinsen, C. B. *Science* **1973**, *181*, 223–230.
- [4] Levinthal, C. *Mössbauer Spectroscopy in Biological Systems Proceedings, Univ. of Illinois Bulletin* **1969**, *67*, 22–24.
- [5] Schechter, A. N.; Chen, R. F.; Anfinsen, C. B. *Science* **1970**, *167*, 886–887.
- [6] Epstein, H. F.; Schechter, A. N.; Chen, R. F.; Anfinsen, C. B. *J Mol Biol* **1971**, *60*, 499–508.
- [7] Ikai, A.; Tanford, C. *Nature* **1971**, *230*, 100–102.
- [8] Tsong, T. Y.; Baldwin, R. L.; Elson, E. L. *Proc Natl Acad Sci U S A* **1971**, *68*, 2712–2715.
- [9] Creighton, T. E. *Prog Biophys Mol Biol* **1978**, *33*, 231–297.
- [10] Kim, P. S.; Baldwin, R. L. *Annu Rev Biochem* **1982**, *51*, 459–489.
- [11] Ptitsyn, O. B. *Adv Protein Chem* **1995**, *47*, 83–229.
- [12] Arai, M.; Kuwajima, K. *Adv Protein Chem* **2000**, *53*, 209–282.
- [13] Ikeguchi, M.; Kuwajima, K.; Mitani, M.; Sugai, S. *Biochemistry* **1986**, *25*, 6965–6972.

B. References

- [14] Jennings, P. A.; Wright, P. E. *Science* **1993**, *262*, 892–896.
- [15] Brandts, J. F.; Halvorson, H. R.; Brennan, M. *Biochemistry* **1975**, *14*, 4953–4963.
- [16] Schmid, F. X.; Baldwin, R. L. *Proc Natl Acad Sci U S A* **1978**, *75*, 4764–4768.
- [17] Schmid, F. X. In *Prolyl Isomerization in Protein Folding*; Buchner, J., Kiefhaber, T., Eds.; Wiley-VCH, Weinheim, 2005; Vol. 2, pp 916–945.
- [18] Udgaonkar, J. B.; Baldwin, R. L. *Nature* **1988**, *335*, 694–699.
- [19] Roder, H.; Elöve, G. A.; Englander, S. W. *Nature* **1988**, *335*, 700–704.
- [20] Radford, S. E.; Dobson, C. M.; Evans, P. A. *Nature* **1992**, *358*, 302–307.
- [21] Wolynes, P. G.; Onuchic, J. N.; Thirumalai, D. *Science* **1995**, *267*, 1619–1620.
- [22] Onuchic, J. N.; Wolynes, P. G.; Luthey-Schulten, Z.; Socci, N. D. *Proc Natl Acad Sci U S A* **1995**, *92*, 3626–3630.
- [23] Radford, S. E. *Trends Biochem Sci* **2000**, *25*, 611–618.
- [24] Baldwin, R. L. *J Biomol NMR* **1995**, *5*, 103–109.
- [25] Dill, K. A.; Chan, H. S. *Nat Struct Biol* **1997**, *4*, 10–19.
- [26] Dobson, C. M. *Nature* **2003**, *426*, 884–890.
- [27] Mandelkow, E. M.; Mandelkow, E. *Trends Cell Biol* **1998**, *8*, 425–427.
- [28] Uversky, V. N.; Li, J.; Fink, A. L. *J Biol Chem* **2001**, *276*, 10737–10744.
- [29] Raffin, R.; Dieckman, L. J.; Szpunar, M.; Wunschl, C.; Pokkuluri, P. R.; Dave, P.; Stevens, P. W.; Cai, X.; Schiffer, M.; Stevens, F. J. *Protein Sci* **1999**, *8*, 509–517.
- [30] Hammarström, P.; Jiang, X.; Hurshman, A. R.; Powers, E. T.; Kelly, J. W. *Proc Natl Acad Sci U S A* **2002**, *99 Suppl 4*, 16427–16432.
- [31] Fersht, A. R.; Daggett, V. *Cell* **2002**, *108*, 573–582.
- [32] Daggett, V.; Fersht, A. R. *Trends Biochem Sci* **2003**, *28*, 18–25.

- [33] Schmid, F. X.; Baldwin, R. L. *J Mol Biol* **1979**, *135*, 199–215.
- [34] Roder, H.; Wüthrich, K. *Proteins* **1986**, *1*, 34–42.
- [35] Bycroft, M.; Matouschek, A.; Kellis, J. T.; Serrano, L.; Fersht, A. R. *Nature* **1990**, *346*, 488–490.
- [36] Kautz, R. A.; Fox, R. O. *Protein Sci* **1993**, *2*, 851–858.
- [37] Kiefhaber, T.; Labhardt, A. M.; Baldwin, R. L. *Nature* **1995**, *375*, 513–515.
- [38] Hoeltzli, S. D.; Frieden, C. *Proc Natl Acad Sci U S A* **1995**, *92*, 9318–9322.
- [39] Bann, J. G.; Pinkner, J.; Hultgren, S. J.; Frieden, C. *Proc Natl Acad Sci U S A* **2002**, *99*, 709–714.
- [40] Balbach, J.; Forge, V.; Lau, W. S.; van Nuland, N. A.; Brew, K.; Dobson, C. M. *Science* **1996**, *274*, 1161–1163.
- [41] Killick, T. R.; Freund, S. M.; Fersht, A. R. *Protein Sci* **1999**, *8*, 1286–1291.
- [42] Hore, P. J.; Winder, S. L.; Roberts, C. H.; Dobson, C. M. *J Am Chem Soc* **1997**, *119*, 5049–5050.
- [43] Maeda, K.; Lyon, C. E.; Lopez, J. J.; Cemazar, M.; Dobson, C. M.; Hore, P. J. *J Biomol NMR* **2000**, *16*, 235–244.
- [44] Wirmer, J.; Kühn, T.; Schwalbe, H. *Angew Chem, Int Ed* **2001**, *40*, 4248–4251.
- [45] Lyon, C. E.; Suh, E.-S.; Dobson, C. M.; Hore, P. J. *J Am Chem Soc* **2002**, *124*, 13018–13024.
- [46] Mok, K. H.; Nagashima, T.; Day, I. J.; Hore, P. J.; Dobson, C. M. *Proc Natl Acad Sci U S A* **2005**, *102*, 8899–8904.
- [47] Korzhnev, D. M.; Salvatella, X.; Vendruscolo, M.; Nardo, A. A. D.; Davidson, A. R.; Dobson, C. M.; Kay, L. E. *Nature* **2004**, *430*, 586–590.
- [48] Mittermaier, A.; Kay, L. E. *Science* **2006**, *312*, 224–228.

B. References

- [49] Religa, T. L.; Markson, J. S.; Mayor, U.; Freund, S. M. V.; Fersht, A. R. *Nature* **2005**, *437*, 1053–1056.
- [50] Mok, K. H.; Nagashima, T.; Day, I. J.; Jones, J. A.; Jones, C. J. V.; Dobson, C. M.; Hore, P. J. *J Am Chem Soc* **2003**, *125*, 12484–12492.
- [51] Kühn, T.; Schwalbe, H. *J Am Chem Soc* **2000**, *122*, 6169–6174.
- [52] Schlepckow, K.; Wirmer, J.; Bachmann, A.; Kiefhaber, T.; Schwalbe, H. *J Mol Biol* **2008**, *378*, 686–698.
- [53] Schanda, P.; Forge, V.; Brutscher, B. *Proc Natl Acad Sci U S A* **2007**, *104*, 11257–11262.
- [54] Frydman, L.; Scherf, T.; Lupulescu, A. *Proc Natl Acad Sci U S A* **2002**, *99*, 15858–15862.
- [55] Gal, M.; Schanda, P.; Brutscher, B.; Frydman, L. *J Am Chem Soc* **2007**, *129*, 1372–1377.
- [56] Kupce, E.; Freeman, R. *J Magn Reson* **2003**, *162*, 300–310.
- [57] Ardenkjaer-Larsen, J. H.; Fridlund, B.; Gram, A.; Hansson, G.; Hansson, L.; Lerche, M. H.; Servin, R.; Thaning, M.; Golman, K. *Proc Natl Acad Sci U S A* **2003**, *100*, 10158–10163.
- [58] Bierzynski, A.; Baldwin, R. L. *J Mol Biol* **1982**, *162*, 173–186.
- [59] Dobson, C. M.; Evans, P. A.; Williamson, K. L. *FEBS Lett* **1984**, *168*, 331–334.
- [60] Neri, D.; Billeter, M.; Wider, G.; Wüthrich, K. *Science* **1992**, *257*, 1559–1563.
- [61] Alexandrescu, A. T.; Abeygunawardana, C.; Shortle, D. *Biochemistry* **1994**, *33*, 1063–1072.
- [62] Smith, L. J.; Bolin, K. A.; Schwalbe, H.; MacArthur, M. W.; Thornton, J. M.; Dobson, C. M. *J Mol Biol* **1996**, *255*, 494–506.

- [63] Fiebig, K. M.; Schwalbe, H.; Buck, M.; Smith, L. J.; Dobson, C. M. *J Phys Chem* **1996**, *100*, 2661–2666.
- [64] Schwalbe, H.; Fiebig, K. M.; Buck, M.; Jones, J. A.; Grimshaw, S. B.; Spencer, A.; Glaser, S. J.; Smith, L. J.; Dobson, C. M. *Biochemistry* **1997**, *36*, 8977–8991.
- [65] Klein-Seetharaman, J.; Oikawa, M.; Grimshaw, S. B.; Wirmer, J.; Duchardt, E.; Ueda, T.; Imoto, T.; Smith, L. J.; Dobson, C. M.; Schwalbe, H. *Science* **2002**, *295*, 1719–1722.
- [66] Wirmer, J.; Schlörb, C.; Klein-Seetharaman, J.; Hirano, R.; Ueda, T.; Imoto, T.; Schwalbe, H. *Angew Chem, Int Ed* **2004**, *43*, 5780–5785.
- [67] Collins, E. S.; Wirmer, J.; Hirai, K.; Tachibana, H.; Ichi Segawa, S.; Dobson, C. M.; Schwalbe, H. *Chembiochem* **2005**, *6*, 1619–1627.
- [68] Wirmer, J.; Berk, H.; Ugolini, R.; Redfield, C.; Schwalbe, H. *Protein Sci* **2006**, *15*, 1397–1407.
- [69] Gillespie, J. R.; Shortle, D. *J Mol Biol* **1997**, *268*, 158–169.
- [70] Bertoncini, C. W.; Jung, Y.-S.; Fernandez, C. O.; Hoyer, W.; Griesinger, C.; Jovin, T. M.; Zweckstetter, M. *Proc Natl Acad Sci U S A* **2005**, *102*, 1430–1435.
- [71] Kristjansdottir, S.; Lindorff-Larsen, K.; Fieber, W.; Dobson, C. M.; Vendruscolo, M.; Poulsen, F. M. *J Mol Biol* **2005**, *347*, 1053–1062.
- [72] Shortle, D.; Ackerman, M. S. *Science* **2001**, *293*, 487–489.
- [73] Meier, S.; Strohmeier, M.; Blackledge, M.; Grzesiek, S. *J Am Chem Soc* **2007**, *129*, 754–755.
- [74] Meier, S.; Blackledge, M.; Grzesiek, S. *J Chem Phys* **2008**, *128*, 052204.
- [75] Bernadó, P.; Blanchard, L.; Timmins, P.; Marion, D.; Ruigrok, R. W. H.; Blackledge, M. *Proc Natl Acad Sci U S A* **2005**, *102*, 17002–17007.
- [76] Jha, A. K.; Colubri, A.; Freed, K. F.; Sosnick, T. R. *Proc Natl Acad Sci U S A* **2005**, *102*, 13099–13104.

B. References

- [77] Louhivuori, M.; Pääkkönen, K.; Fredriksson, K.; Permi, P.; Lounila, J.; Annala, A. *J Am Chem Soc* **2003**, *125*, 15647–15650.
- [78] Kuwajima, K.; Mitani, M.; Sugai, S. *J Mol Biol* **1989**, *206*, 547–561.
- [79] Balbach, J.; Forge, V.; van Nuland, N. A.; Winder, S. L.; Hore, P. J.; Dobson, C. M. *Nat Struct Biol* **1995**, *2*, 865–870.
- [80] Forge, V.; Wijesinha, R. T.; Balbach, J.; Brew, K.; Robinson, C. V.; Redfield, C.; Dobson, C. M. *J Mol Biol* **1999**, *288*, 673–688.
- [81] Schlepckow, K. M.Sc. thesis, Institute for Organic Chemistry and Chemical Biology, Johann Wolfgang Goethe University Frankfurt/Main, 2003.
- [82] Fink, A. L. *Curr Opin Struct Biol* **2005**, *15*, 35–41.
- [83] Dyson, H. J.; Wright, P. E. *Nat Rev Mol Cell Biol* **2005**, *6*, 197–208.
- [84] Prakash, S.; Matouschek, A. *Trends Biochem Sci* **2004**, *29*, 593–600.
- [85] Winter, J.; Jakob, U. *Crit Rev Biochem Mol Biol* **2004**, *39*, 297–317.
- [86] Pace, C. N.; Grimsley, G. R.; Scholtz, J. M. In *Denaturation of proteins by urea and guanidine hydrochloride*; Kiefhaber, T., Buchner, J., Eds.; Wiley-VCH, Weinheim, 2005; Vol. 1, pp 45–69.
- [87] Tanford, C.; Kawahara, K.; Lapanje, S. *J Biol Chem* **1966**, *241*, 1921–1923.
- [88] Kohn, J. E.; Millett, I. S.; Jacob, J.; Zagrovic, B.; Dillon, T. M.; Cingel, N.; Dothager, R. S.; Seifert, S.; Thiyagarajan, P.; Sosnick, T. R.; Hasan, M. Z.; Pande, V. S.; Ruczinski, I.; Doniach, S.; Plaxco, K. W. *Proc Natl Acad Sci U S A* **2004**, *101*, 12491–12496.
- [89] Wirmer, J.; Schlörb, C.; Schwalbe, H. In *Protein Folding Handbook*, 1st ed.; Wiley-VCH, Weinheim, 2005; pp 737–794.
- [90] Ohnishi, S.; Lee, A. L.; Edgell, M. H.; Shortle, D. *Biochemistry* **2004**, *43*, 4064–4070.
- [91] Fitzkee, N. C.; Rose, G. D. *Proc Natl Acad Sci U S A* **2004**, *101*, 12497–12502.

- [92] Shi, Z.; Chen, K.; Liu, Z.; Kallenbach, N. R. *Chem Rev* **2006**, *106*, 1877–1897.
- [93] Graf, J.; Nguyen, P. H.; Stock, G.; Schwalbe, H. *J Am Chem Soc* **2007**, *129*, 1179–1189.
- [94] Saupe, A.; Englert, G. *Phys Rev Lett* **1963**, *11*, 462–464.
- [95] Bothner-By, A. A.; Domaille, P. J.; Gayathri, C. *J Am Chem Soc* **1981**, *103*, 5602–5603.
- [96] Gayathri, C.; Bothner-By, A.; van Zijl, P.; MacLean, C. *Chem Phys Lett* **1982**, *87*, 192–196.
- [97] Tolman, J. R.; Flanagan, J. M.; Kennedy, M. A.; Prestegard, J. H. *Proc Natl Acad Sci U S A* **1995**, *92*, 9279–9283.
- [98] Kung, H. C.; Wang, K. Y.; Goljer, I.; Bolton, P. H. *J Magn Reson B* **1995**, *109*, 323–325.
- [99] Tolman, J. R.; Flanagan, J. M.; Kennedy, M. A.; Prestegard, J. H. *Nat Struct Biol* **1997**, *4*, 292–297.
- [100] Pike, A. C.; Brew, K.; Acharya, K. R. *Structure* **1996**, *4*, 691–703.
- [101] Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1111–1114.
- [102] Bax, A.; Tjandra, N. *J Biomol NMR* **1997**, *10*, 289–292.
- [103] Bax, A.; Kontaxis, G.; Tjandra, N. *Methods Enzymol* **2001**, *339*, 127–174.
- [104] Losonczi, J. A.; Andrec, M.; Fischer, M. W.; Prestegard, J. H. *J Magn Reson* **1999**, *138*, 334–342.
- [105] Zweckstetter, M.; Bax, A. *J Am Chem Soc* **2000**, *122*, 3791–3792.
- [106] Cornilescu, G.; Marquardt, J. L.; Ottiger, M.; Bax, A. *J Am Chem Soc* **1998**, *120*, 6836–6837.
- [107] Ottiger, M.; Bax, A. *J Biomol NMR* **1999**, *13*, 187–191.
- [108] Clore, G. M.; Gronenborn, A. M. *Proc Natl Acad Sci USA* **1998**, *95*, 5891–5898.
- [109] Hus, J. C.; Marion, D.; Blackledge, M. *J Mol Biol* **2000**, *298*, 927–936.

B. References

- [110] Ramirez, B. E.; Bax, A. *J Am Chem Soc* **1998**, *120*, 9106–9107.
- [111] Hus, J. C.; Marion, D.; Blackledge, M. *J Am Chem Soc* **2001**, *123*, 1541–1542.
- [112] Champier, L.; Sibille, N.; Bersch, B.; Brutscher, B.; Blackledge, M.; Covès, J. *Biochemistry* **2002**, *41*, 3770–3780.
- [113] Delaglio, F.; Kontaxis, G.; Bax, A. *J Am Chem Soc* **2000**, *122*, 2142–2143.
- [114] Garrett, D. S.; Seok, Y. J.; Peterkofsky, A.; Gronenborn, A. M.; Clore, G. M. *Nat Struct Biol* **1999**, *6*, 166–173.
- [115] Olejniczak, E. T.; Meadows, R. P.; Wang, H.; Cai, M.; Netteshein, D. G.; Fesik, S. W. *J Am Chem Soc* **1999**, *121*, 9249–9250.
- [116] Bolon, P. J.; Al-Hashimi, H. M.; Prestegard, J. H. *J Mol Biol* **1999**, *293*, 107–115.
- [117] Kuriyan, J.; Wilz, S.; Karplus, M.; Petsko, G. A. *J Mol Biol* **1986**, *192*, 133–154.
- [118] Osapay, K.; Theriault, Y.; Wright, P. E.; Case, D. A. *J Mol Biol* **1994**, *244*, 183–197.
- [119] Fischer, M. W.; Losonczi, J. A.; Weaver, J. L.; Prestegard, J. H. *Biochemistry* **1999**, *38*, 9013–9022.
- [120] Tian, F.; Al-Hashimi, H. M.; Craighead, J. L.; Prestegard, J. H. *J Am Chem Soc* **2001**, *123*, 485–492.
- [121] Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. *J Mol Biol* **1987**, *194*, 531–544.
- [122] Tolman, J. R.; Al-Hashimi, H. M.; Kay, L. E.; Prestegard, J. H. *J Am Chem Soc* **2001**, *123*, 1416–1424.
- [123] Lipari, G.; Szabo, A. *J Am Chem Soc* **1982**, *104*, 4546–4559.
- [124] Lipari, G.; Szabo, A. *J Am Chem Soc* **1982**, *104*, 4559–4570.
- [125] Meiler, J.; Prompers, J. J.; Peti, W.; Griesinger, C.; Brüschweiler, R. *J Am Chem Soc* **2001**, *123*, 6098–6107.

- [126] Peti, W.; Meiler, J.; Brüschweiler, R.; Griesinger, C. *J Am Chem Soc* **2002**, *124*, 5822–5833.
- [127] Tolman, J. R. *J Am Chem Soc* **2002**, *124*, 12020–12030.
- [128] Clore, G. M.; Schwieters, C. D. *J Am Chem Soc* **2004**, *126*, 2923–2938.
- [129] Bernadó, P.; Blackledge, M. *J Am Chem Soc* **2004**, *126*, 4907–4920.
- [130] Bernadó, P.; Blackledge, M. *J Am Chem Soc* **2004**, *126*, 7760–7761.
- [131] Brüschweiler, R.; Wright, P. E. *J Am Chem Soc* **1994**, *116*, 8426–8427.
- [132] Fernandes, M. X.; Bernadó, P.; Pons, M.; de la Torre, J. G. *J Am Chem Soc* **2001**, *123*, 12037–12047.
- [133] Azurmendi, H. F.; Bush, C. A. *J Am Chem Soc* **2002**, *124*, 2426–2427.
- [134] Almond, A.; Axelsen, J. B. *J Am Chem Soc* **2002**, *124*, 9986–9987.
- [135] Hansen, M. R.; Mueller, L.; Pardi, A. *Nat Struct Biol* **1998**, *5*, 1065–1074.
- [136] Clore, G. M.; Starich, M. R.; Gronenborn, A. M. *J Am Chem Soc* **1998**, *120*, 10571–10572.
- [137] Ferrarini, A. *J Phys Chem B* **2003**, *107*, 7923–7931.
- [138] Zweckstetter, M.; Hummer, G.; Bax, A. *Biophys J* **2004**, *86*, 3444–3460.
- [139] Zweckstetter, M. *Eur Biophys J* **2006**, *35*, 170–180.
- [140] Prosser, R. S.; Losonczi, J. A.; Shiyankovskaya, I. V. *J Am Chem Soc* **1998**, *120*, 11010–11011.
- [141] Barrientos, L. G.; Dolan, C.; Gronenborn, A. M. *J Biomol NMR* **2000**, *16*, 329–337.
- [142] Wu, B.; Petersen, M.; Girard, F.; Tessari, M.; Wijmenga, S. S. *J Biomol NMR* **2006**, *35*, 103–115.
- [143] Ackerman, M. S.; Shortle, D. *Biochemistry* **2002**, *41*, 3089–3095.

B. References

- [144] Ackerman, M. S.; Shortle, D. *Biochemistry* **2002**, *41*, 13791–13797.
- [145] Ohnishi, S.; Shortle, D. *Proteins* **2003**, *50*, 546–551.
- [146] Alexandrescu, A. T.; Kammerer, R. A. *Protein Sci* **2003**, *12*, 2132–2140.
- [147] Ding, K.; Louis, J. M.; Gronenborn, A. M. *J Mol Biol* **2004**, *335*, 1299–1307.
- [148] Fieber, W.; Kristjansdottir, S.; Poulsen, F. M. *J Mol Biol* **2004**, *339*, 1191–1199.
- [149] Mohana-Borges, R.; Goto, N. K.; Kroon, G. J. A.; Dyson, H. J.; Wright, P. E. *J Mol Biol* **2004**, *340*, 1131–1142.
- [150] Fredriksson, K.; Louhivuori, M.; Permi, P.; Annala, A. *J Am Chem Soc* **2004**, *126*, 12646–12650.
- [151] Meier, S.; Güthe, S.; Kiefhaber, T.; Grzesiek, S. *J Mol Biol* **2004**, *344*, 1051–1069.
- [152] Sallum, C. O.; Martel, D. M.; Fournier, R. S.; Matousek, W. M.; Alexandrescu, A. T. *Biochemistry* **2005**, *44*, 6392–6403.
- [153] Bertoncini, C. W.; Fernandez, C. O.; Griesinger, C.; Jovin, T. M.; Zweckstetter, M. *J Biol Chem* **2005**, *280*, 30649–30652.
- [154] Gebel, E. B.; Ruan, K.; Tolman, J. R.; Shortle, D. *J Am Chem Soc* **2006**, *128*, 9310–9311.
- [155] Binolfi, A.; Rasia, R. M.; Bertoncini, C. W.; Ceolin, M.; Zweckstetter, M.; Griesinger, C.; Jovin, T. M.; Fernández, C. O. *J Am Chem Soc* **2006**, *128*, 9893–9901.
- [156] Sibille, N.; Sillen, A.; Leroy, A.; Wieruszkeski, J.-M.; Mulloy, B.; Landrieu, I.; Lip-pens, G. *Biochemistry* **2006**, *45*, 12560–12572.
- [157] Dames, S. A.; Aregger, R.; Vajpai, N.; Bernado, P.; Blackledge, M.; Grzesiek, S. *J Am Chem Soc* **2006**, *128*, 13508–13514.
- [158] Ohnishi, S.; Kamikubo, H.; Onitsuka, M.; Kataoka, M.; Shortle, D. *J Am Chem Soc* **2006**, *128*, 16338–16344.

- [159] Meier, S.; Grzesiek, S.; Blackledge, M. *J Am Chem Soc* **2007**, *129*, 9799–9807.
- [160] Schlörb, C. Ph.D. thesis, Institute for Organic Chemistry and Chemical Biology, Johann Wolfgang Goethe University Frankfurt/Main, 2007.
- [161] Penkett, C. J.; Redfield, C.; Dodd, I.; Hubbard, J.; McBay, D. L.; Mossakowska, D. E.; Smith, R. A.; Dobson, C. M.; Smith, L. J. *J Mol Biol* **1997**, *274*, 152–159.
- [162] Avbelj, F.; Baldwin, R. L. *Proc Natl Acad Sci U S A* **2004**, *101*, 10967–10972.
- [163] Metropolis, N.; Ulam, S. *J Am Stat Assoc* **1949**, *44*, 335–341.
- [164] Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H. *J Chem Phys* **1953**, *21*, 1087–1092.
- [165] Bernadó, P.; Bertoncini, C. W.; Griesinger, C.; Zweckstetter, M.; Blackledge, M. *J Am Chem Soc* **2005**, *127*, 17968–17969.
- [166] Skora, L.; Cho, M.-K.; Kim, H.-Y.; Becker, S.; Fernandez, C. O.; Blackledge, M.; Zweckstetter, M. *Angew Chem Int Ed Engl* **2006**, *45*, 7012–7015.
- [167] Chandrasekhar, S. *Rev Mod Phys* **1943**, *15*, 1–89.
- [168] Louhivuori, M.; Fredriksson, K.; Pääkkönen, K.; Permi, P.; Annala, A. *J Biomol NMR* **2004**, *29*, 517–524.
- [169] Ernst, R.; Bodenhausen, G.; Wokaun, A. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*; Oxford University Press, 1987.
- [170] Landau, L.; Lifschitz, E. *Course of Theoretical Physics*; Pergamon Press, 1958; Vol. V.
- [171] Rayleigh, L. *Collected Papers*; Dover Publications, New York, 1964; Vol. 6.
- [172] Sanders, C. R.; Schwonek, J. P. *Biochemistry* **1992**, *31*, 8898–8905.
- [173] Rückert, M.; Otting, G. *J Am Chem Soc* **2000**, *122*, 7793–7797.
- [174] Tycko, R.; Blanco, F. J.; Ishii, Y. *J Am Chem Soc* **2000**, *122*, 9340–9341.

B. References

- [175] Sass, H. J.; Musco, G.; Stahl, S. J.; Wingfield, P. T.; Grzesiek, S. *J Biomol NMR* **2000**, *18*, 303–309.
- [176] Fersht, A. R. In *Kinetics of Protein Folding*; Freeman, New York, 1999; pp 541–572.
- [177] Hartridge, H.; Roughton, F. J. W. *Proc Roy Soc (London)* **1923**, *104A*, 376–394.
- [178] Chance, B. *J Franklin Inst* **1940**, *229*, 455–476,.
- [179] Berger, R. L.; Balko, B. *Rev Sci Instrum* **1968**, *39*, 486–493.
- [180] Berger, R. L.; Balko, B.; Chapman, H. F. *Rev Sci Instrum* **1968**, *39*, 493–498.
- [181] Welker, E.; Maki, K.; Shastry, M. C. R.; Juminaga, D.; Bhat, R.; Scheraga, H. A.; Roder, H. *Proc Natl Acad Sci U S A* **2004**, *101*, 17681–17686.
- [182] Schmid, F. X. *Biochemistry* **1983**, *22*, 4690–4696.
- [183] Kiefhaber, T. *Proc Natl Acad Sci U S A* **1995**, *92*, 9029–9033.
- [184] Hoang, L.; Bedard, S.; Krishna, M. M. G.; Lin, Y.; Englander, S. W. *Proc Natl Acad Sci U S A* **2002**, *99*, 12173–12178.
- [185] Huang, G. S.; Oas, T. G. *Proc Natl Acad Sci U S A* **1995**, *92*, 6878–6882.
- [186] Wang, M.; Tang, Y.; Sato, S.; Vugmeyster, L.; McKnight, C. J.; Raleigh, D. P. *J Am Chem Soc* **2003**, *125*, 6032–6033.
- [187] Sudmeier, J. L.; Pesek, J. J. *Inorg Chem* **1971**, *4*, 860–863.
- [188] Grimaldi, J.; Baldo, J.; McMurray, C.; Sykes, B. D. *J Am Chem Soc* **1972**, *94*, 7641–7645.
- [189] Couch, D. A.; Howarth, O. W.; Moore, P. *J Phys E* **1975**, *8*, 831–833.
- [190] Kühne, R. O.; Schaffhauser, T.; Wokaun, A.; Ernst, R. R. *J Magn Reson* **1979**, *35*, 39–67.
- [191] Green, D. B.; Lane, J.; Wing, R. M. *Appl Spectrosc* **1987**, *41*, 847–851.

- [192] McGee, W. A.; Parkhurst, L. J. *Anal Biochem* **1990**, *189*, 267–273.
- [193] McGarrity, J. F.; Prodolliet, J.; Smyth, T. *Org Magn Reson* **1981**, *17*, 59–65.
- [194] Balbach, J. *J Am Chem Soc* **2000**, *122*, 5887–5888.
- [195] Zeeb, M.; Rösner, H.; Zeslawski, W.; Canet, D.; Holak, T. A.; Balbach, J. *J Mol Biol* **2002**, *315*, 447–457.
- [196] Mizuguchi, M.; Kroon, G. J.; Wright, P. E.; Dyson, H. J. *J Mol Biol* **2003**, *328*, 1161–1171.
- [197] Wenter, P.; Fürtig, B.; Hainard, A.; Schwalbe, H.; Pitsch, S. *Angew Chem Int Ed Engl* **2005**, *44*, 2600–2603.
- [198] Fürtig, B.; Buck, J.; Manoharan, V.; Bermel, W.; Jäschke, A.; Wenter, P.; Pitsch, S.; Schwalbe, H. *Biopolymers* **2007**, *86*, 360–383.
- [199] Permyakov, E. A.; Berliner, L. J. *FEBS Lett* **2000**, *473*, 269–274.
- [200] Chrysinina, E. D.; Brew, K.; Acharya, K. R. *J Biol Chem* **2000**, *275*, 37021–37029.
- [201] Griko, Y. V.; Freire, E.; Privalov, P. L. *Biochemistry* **1994**, *33*, 1889–1899.
- [202] Veprintsev, D. B.; Permyakov, S. E.; Permyakov, E. A.; Rogov, V. V.; Cawthorn, K. M.; Berliner, L. J. *FEBS Lett* **1997**, *412*, 625–628.
- [203] Permyakov, E. A.; Morozova, L. A.; Burstein, E. A. *Biophys Chem* **1985**, *21*, 21–31.
- [204] Segawa, T.; Sugai, S. *J Biochem* **1983**, *93*, 1321–1328.
- [205] Kuwajima, K. *FASEB J* **1996**, *10*, 102–109.
- [206] Arai, M.; Kuwajima, K. *Fold Des* **1996**, *1*, 275–287.
- [207] Kuwajima, K.; Nitta, K.; Yoneyama, M.; Sugai, S. *J Mol Biol* **1976**, *106*, 359–373.
- [208] Kuwajima, K. *J Mol Biol* **1977**, *114*, 241–258.
- [209] Kuwajima, K.; Hiraoka, Y.; Ikeguchi, M.; Sugai, S. *Biochemistry* **1985**, *24*, 874–881.

B. References

- [210] Bushmarina, N. A.; Blanchet, C. E.; Vernier, G.; Forge, V. *Protein Sci* **2006**, *15*, 659–671.
- [211] Balbach, J.; Forge, V.; Lau, W. S.; Jones, J. A.; van Nuland, N. A.; Dobson, C. M. *Proc Natl Acad Sci U S A* **1997**, *94*, 7182–7185.
- [212] Laemmli, U. K. *Nature* **1970**, *227*, 680–685.
- [213] Wirmer, J. Ph.D. thesis, Massachusetts Institute of Technology, 2005.
- [214] Bodenhausen, G.; Ruben, D. J. *Chem Phys Lett* **1980**, *69*, 185–189.
- [215] Marion, D.; Driscoll, P. C.; Kay, L. E.; Wingfield, P. T.; Bax, A.; Gronenborn, A. M.; Clore, G. M. *Biochemistry* **1989**, *28*, 6150–6156.
- [216] Wishart, D. S.; Bigam, C. G.; Yao, J.; Abildgaard, F.; Dyson, H. J.; Oldfield, E.; Markley, J. L.; Sykes, B. D. *J Biomol NMR* **1995**, *6*, 135–140.
- [217] Sklenar, V.; Bax, A. *J Magn Reson* **1987**, *74*, 469–479.
- [218] Motulsky, H.; Christopoulos, A. In *Comparing models using the extra sum-of-squares F test*; Oxford University Press, 2004; pp 138–142.
- [219] Wirmer, J.; Schwalbe, H. *J Biomol NMR* **2002**, *23*, 47–55.
- [220] Bronstein, I. N.; Semendjajew, K. A. In *Tabellen spezieller Funktionen*; Grosche, G., Ziegler, V., Ziegler, D., Eds.; Verlag Harri Deutsch und Thun Frankfurt/M., 1989; pp 24–27.
- [221] Ikeguchi, M.; Kuwajima, K.; Sugai, S. *J Biochem* **1986**, *99*, 1191–1201.
- [222] Hamano, M.; Nitta, K.; Kuwajima, K.; Sugai, S. *J Biochem* **1986**, *100*, 1617–1622.
- [223] Dolgikh, D. A.; Gilmanshin, R. I.; Brazhnikov, E. V.; Bychkova, V. E.; Semisotnov, G. V.; SYu, V.; Ptitsyn, O. B. *FEBS Lett* **1981**, *136*, 311–315.
- [224] Vanderheeren, G.; Hanssens, I. *J Biol Chem* **1994**, *269*, 7090–7094.
- [225] Noyelle, K.; Dael, H. V. *J Inorg Biochem* **2002**, *88*, 69–76.

- [226] Santoro, M. M.; Bolen, D. W. *Biochemistry* **1988**, *27*, 8063–8068.
- [227] Kim, P. S.; Baldwin, R. L. *Annu Rev Biochem* **1990**, *59*, 631–660.
- [228] Garel, J. R.; Baldwin, R. L. *Proc Natl Acad Sci U S A* **1973**, *70*, 3347–3351.
- [229] Fox, R. O.; Evans, P. A.; Dobson, C. M. *Nature* **1986**, *320*, 192–194.
- [230] Pappenberger, G.; Bachmann, A.; Müller, R.; Aygün, H.; Engels, J. W.; Kiefhaber, T. *J Mol Biol* **2003**, *326*, 235–246.
- [231] Nall, B. T.; Garel, J.-R.; Baldwin, R. L. *J Mol Biol* **1978**, *118*, 317–330.
- [232] Schmid, F. X.; Baldwin, R. L. *J Mol Biol* **1979**, *133*, 285–287.
- [233] Kaplan, J. H.; Ellis-Davies, G. C. *Proc Natl Acad Sci U S A* **1988**, *85*, 6571–6575.
- [234] Ellis-Davies, G. C. R.; Kaplan, J. H. *J Org Chem* **1988**, *53*, 1966–1969.
- [235] Troullier, A.; Reinstädler, D.; Dupont, Y.; Naumann, D.; Forge, V. *Nat Struct Biol* **2000**, *7*, 78–86.
- [236] Wildegger, G.; Kiefhaber, T. *J Mol Biol* **1997**, *270*, 294–304.
- [237] Bachmann, A.; Kiefhaber, T. In *Kinetic mechanisms in protein folding*; Buchner, J., Kiefhaber, T., Eds.; Wiley-VCH, Weinheim, 2005; Vol. 1, pp 379–410.
- [238] Schwalbe, H.; Grimshaw, S. B.; Spencer, A.; Buck, M.; Boyd, J.; Dobson, C. M.; Redfield, C.; Smith, L. J. *Protein Sci* **2001**, *10*, 677–688.
- [239] DeLano, W. L. *The PyMOL User's Manual*; DeLano Scientific, Palo Alto, CA, USA, 2002.
- [240] Kuprov, I.; Hore, P. J. *J Magn Reson* **2004**, *171*, 171–175.
- [241] Burgoyne, R. D. *Nat Rev Neurosci* **2007**, *8*, 182–193.
- [242] Finney, L. A.; O'Halloran, T. V. *Science* **2003**, *300*, 931–936.
- [243] Otting, G.; Wüthrich, K. *J Magn Reson* **1989**, *85*, 586–594.

B. References

- [244] Schanda, P.; Brutscher, B. *J Am Chem Soc* **2005**, *127*, 8014–8015.
- [245] Momotake, A.; Lindegger, N.; Niggli, E.; Barsotti, R. J.; Ellis-Davies, G. C. R. *Nat Methods* **2006**, *3*, 35–40.
- [246] Bargon, J.; Fischer, H.; Johnsen, U. *Z Naturforsch* **1967**, *22a*, 1551–1555.
- [247] Ward, H. R.; Lawler, R. G. *J Am Chem Soc* **1967**, *89*, 5518–5519.
- [248] Closs, G. L. *J Am Chem Soc* **1969**, *91*, 4552–4554.
- [249] Kaptein, R.; Oosterhoff, J. L. *Chem Phys Lett* **1969**, *4*, 195–197.
- [250] Kaptein, R.; Oosterhoff, L. J. *Chem Phys Lett* **1969**, *4*, 214–216.
- [251] Kaptein, R.; Dijkstra, K.; Nicolay, K. *Nature* **1978**, *274*, 293–294.
- [252] Kaptein, R. In *Biological Magnetic Resonance*; Berliner, L. J., Reuben, J., Eds.; Plenum, New York, 1982; Vol. 4, Chapter Photo-CIDNP studies of proteins, pp 145–191.
- [253] Hore, P. J.; Broadhurst, R. W. *Prog Nucl Magn Reson Spectrosc* **1993**, *25*, 345–402.
- [254] Mok, K. H.; Hore, P. J. *Methods* **2004**, *34*, 75–87.
- [255] Kaptein, R. *Chem Commun* **1971**, 732–733.
- [256] Peters, J. A.; Huskens, J.; Raber, D. J. *Prog Nucl Magn Reson Spectrosc* **1996**, *28*, 283–350.
- [257] Bertini, I.; Luchinat, C.; Parigi, G.; Pierattelli, R. *Chembiochem* **2005**, *6*, 1536–1549.
- [258] Bargon, J.; Fischer, H. *Z Naturforsch* **1967**, *22a*, 1556–1562.
- [259] Bargon, J. *J Am Chem Soc* **1973**, *95*, 941–942.
- [260] Hore, P. J.; Kaptein, R. *Biochemistry* **1983**, *22*, 1906–1911.
- [261] Lee, L.; Sykes, B. D. *Biophys J* **1980**, *32*, 193–210.
- [262] Capozzi, F.; Cremonini, M. A.; Luchinat, C.; Sola, M. *Magn Reson Chem* **1993**, *31*, S118–S127.

- [263] Bertini, I.; Lee, Y.-M.; Luchinat, C.; Piccioli, M.; Poggi, L. *ChemBioChem* **2001**, *2*, 550–558.
- [264] Murakami, K.; Andree, P. J.; Berliner, L. J. *Biochemistry* **1982**, *21*, 5488–5494.
- [265] Bratcher, S. C.; Kronman, M. J. *J Biol Chem* **1984**, *259*, 10875–10886.
- [266] Kronman, M. J.; Bratcher, S. C. *J Biol Chem* **1984**, *259*, 10887–10895.
- [267] Scheffler, J. E.; Cottrell, C. E.; Berliner, L. J. *J Magn Reson* **1985**, *63*, 199–201.
- [268] Hemmerich, P. *Helv Chim Acta* **1964**, *47*, 465–475.
- [269] Ikeda, H.; Yoshida, K.; Ozeki, M.; Saito, I. *Tetrahedron Lett* **2001**, *42*, 2529–2531.
- [270] Kuhn, L. T. Ph.D. thesis, Oxford University, 2008.
- [271] Berliner, L. J.; Kaptein, R. *J Biol Chem* **1980**, *255*, 3261–3262.
- [272] Berliner, L. J.; Kaptein, R. *Biochemistry* **1981**, *20*, 799–807.
- [273] Improta, S.; Molinari, H.; Pastore, A.; Consonni, R.; Zetta, L. *Eur J Biochem* **1995**, *227*, 78–86.
- [274] Alexandrescu, A. T.; Broadhurst, R. W.; Wormald, C.; Chyan, C. L.; Baum, J.; Dobson, C. M. *Eur J Biochem* **1992**, *210*, 699–709.
- [275] Improta, S.; Molinari, H.; Pastore, A.; Consonni, R.; Zetta, L. *Eur J Biochem* **1995**, *227*, 87–96.
- [276] Shimizu, A.; Ikeguchi, M.; Sugai, S. *Biochemistry* **1993**, *32*, 13198–13203.

Zusammenfassung

Die vorliegende Dissertation widmet sich der Untersuchung der Proteinfaltung. Die Proteinfaltung stellt den letzten Schritt bei der Übersetzung genetischer Information in funktionale Proteine dar und beschreibt den Prozess, in dem die zunächst unstrukturierte Polypeptidkette nach ihrer Biosynthese am Ribosom ihre korrekte dreidimensionale Struktur bildet. Experimente aus den frühen 1960er Jahren haben gezeigt, dass die Information für die korrekte Struktur in der Aminosäuresequenz enthalten ist. Das Einnehmen der nativen Struktur eines Proteins ist von Wichtigkeit, da Struktur und Funktion eines Proteins eng miteinander verknüpft sind. Wird im Zuge der Faltung eines Proteins nicht die native Struktur gebildet (man spricht dann von Fehlfaltung), so kann dies zu einer Beeinträchtigung der Funktion des Proteins führen. Es kann seine Aufgabe im zellulären Kontext nicht mehr oder nur noch eingeschränkt wahrnehmen, was beispielsweise zu neurodegenerativen Erkrankungen führen kann. Aus medizinischer Sicht besteht demzufolge ein grosses Interesse an einem genauen Verständnis des Proteinfaltungsprozesses, um gezielter therapeutisch eingreifen zu können.

Die NMR-Spektroskopie hat sich als eine ausgezeichnete Methode erwiesen, um den Prozess der Proteinfaltung zu untersuchen. Da sie sowohl auf struktureller als auch dynamischer Ebene atomare Auflösung bietet, können einerseits Strukturen von verschiedenen Konformationen, die während der Faltung populiert werden, bestimmt und andererseits die Dynamiken, mit denen sie sich in einander umwandeln, detailliert untersucht werden.

Der Schwerpunkt der vorliegenden Arbeit besteht darin, dass bestehende NMR-Techniken zur Charakterisierung der Proteinfaltung weiter entwickelt wurden. Das erste Ergebniskapitel ist rein theoretischer Natur und entwickelt ein mathematisches Gerüst, das die Vorhersage dipolarer Restkopplungen für entfaltete Proteine auf der Grundlage polymerphysikalischer Prinzipien ermöglicht. Dipolare Restkopplungen sind seit 2001 für eine Reihe von denaturierten als auch intrinsisch unstrukturierten Proteinen gemessen worden. Auf Grund der

Zusammenfassung

Tatsache, dass die gemessenen Kopplungen das Ergebnis einer populationsgewichteten Mittelung darstellen, ist ihre Interpretation im strukturellen und dynamischen Sinne schwieriger im Vergleich zu gefalteten Proteinen. In einigen Fällen wurden aus den gemessenen Daten signifikante Anteile an sekundärer und tertiärer Reststruktur abgelesen. Für eine verlässliche Interpretation der Daten ist die Ausarbeitung einer Theorie für die partielle Ausrichtung entfalteter Proteine jedoch unerlässlich. Hier setzt die im ersten Kapitel dargestellte Arbeit an. Als Startpunkt wurde der von Annala und Mitarbeitern vorgeschlagene Ansatz gewählt. Eine ausführliche Auseinandersetzung mit der Originalpublikation ergab, dass sie Defizite aufweist. Diese Defizite wurden aufgegriffen und korrigiert. Erstens wurde für die Berechnung der Wahrscheinlichkeitsverteilung eines Polypeptidkettensegments eine Multiplikation anstelle einer Addition der entsprechenden Halbkettenverteilungen vorgenommen. Die Notwendigkeit der Multiplikation ergibt sich aus dem *random walk*-Charakter der Halbketten, der die Wahrscheinlichkeitsverteilungen voneinander unabhängig sein lässt. Zweitens wurden die Wahrscheinlichkeitsverteilungen dahingehend korrigiert, dass sie die Ausrichtung der Polypeptidkette in Anwesenheit von zwei Barrieren (anstelle von einer Barriere) beschreibt, was der Situation im Experiment deutlich näher kommt. Drittens wurde anstelle des 1D-*random walk*-Formalismus' der 3D-*random walk*-Formalismus implementiert. Dies bedingt, dass die Kettensegmente unabhängig von ihrer Orientierung in den drei Raumrichtungen eine einheitliche Länge aufweisen. Konnte in der Originalpublikation das zweite Legendre-Polynom für ein Kettensegment, das die Ausrichtung relativ zum Magnetfeld wiedergibt, nur numerisch ermittelt werden, so war es hier möglich, eine analytische Lösung zu erarbeiten. Die dipolare Restkopplung zwischen den Spins zweier Atomkerne P und Q berechnet sich demnach wie folgt:

$$D_{PQ} \approx -\frac{\mu_0 \hbar \gamma_P \gamma_Q}{4\pi^2 R_{PQ}^3} \frac{3 \cos^2 \alpha_{PQ} - 1}{2} \times \frac{1}{L} \left[\frac{4}{15\sqrt{\pi}} \sqrt{\frac{3}{2(N_1 + N_2)}} - \frac{1}{8\pi} \sqrt{\frac{1}{N_1 N_2}} \right], \quad (\text{B.1})$$
$$N = N_1 + N_2 + 1,$$

wobei N_1 , N_2 und L die Längen der Halbketten bzw. die Barrierendistanz anzeigen und die übrigen physikalischen Parameter ihre gewöhnliche Bedeutung haben. Nach dieser Gleichung lassen sich Abhängigkeiten von der Kettenlänge, von der Position innerhalb der Kette und von

der Barrierendistanz (entspricht der Konzentration des Ausrichtungsmediums) leicht ermitteln. Die sich daraus ableitenden Vorhersagen sind inklusive eines Vergleichs mit den Ergebnissen der Originalpublikation in Abb. B.1 dargestellt.

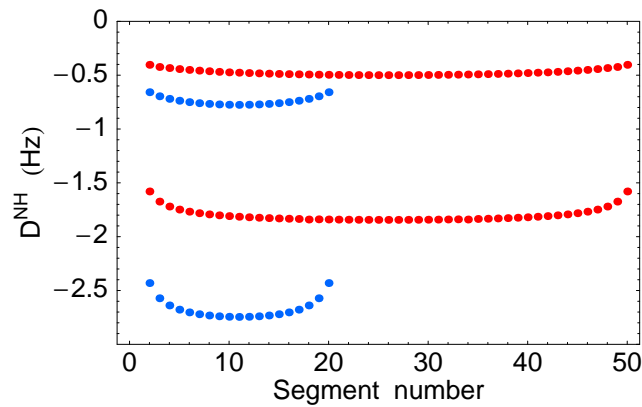


Figure B.1.: Vergleich von dipolaren NH-Restkopplungen als Funktion der Kettenposition für Kettenlängen von 21 (blau) und 51 (rot) Segmenten. Die oberen Profile wurden mit der Gleichung aus der Originalpublikation berechnet.

Die Vorhersagen auf der Basis des Modells sind, dass (i) dipolare Restkopplungen in der Mitte der Kette grösser sind als an den Enden der Kette, (ii) bei einer definierten Konzentration des Ausrichtungsmediums (Barrierendistanz) für kürzere Ketten grösser ausfallen als für längere Ketten und (iii) für eine abnehmende Konzentration des Ausrichtungsmediums im Betrag kleiner werden. Ein ausführlicher Vergleich mit Literaturdaten zeigt, dass dipolare Restkopplungen tatsächlich in der Mitte der Kette grösser sind als an den Kettenenden. Die Vorhersage bezüglich der Kettenlängenabhängigkeit wird durch experimentelle Daten ebenso grundsätzlich gestützt, allerdings lässt die geringe Anzahl verfügbarer experimenteller Daten nur eine vorläufige Aussage zu. Die erzielten Resultate stellen nicht nur eine deutliche Verbesserung des ursprünglichen Ansatzes dar, sondern erlauben auch, das Modell in der Zukunft weiter zu verfeinern, bspw. durch Einbezug langreichweitiger Wechselwirkungen.

Die Ergebniskapitel 2 und 3 fassen experimentell erzielte Resultate zusammen. Kapitel 2 widmet sich im Gegensatz zu Kapitel 1 nicht einem definierten konformationellen Zustand, der während der Proteinfaltung populiert wird, sondern untersucht die Kinetik, mit der sich der native Zustand eines Proteins aus dem entfalteten Zustand bildet. Hierfür wurde das bovine α -Lactalbumin als Modellprotein gewählt. Die Faltung von BLA ist mit verschiede-

Zusammenfassung

nen biophysikalischen Methoden untersucht worden und das gängige Modell für die Faltung des Proteins, das auf der Basis der Daten erstellt wurde, beschreibt die Bildung eines *MG*-Intermediates aus dem entfalteten Zustand und die Umwandlung des Intermediates in den nativen Zustand als einen schrittweisen Prozess.



Im Rahmen der hier vorgenommenen kinetischen Messungen konnten zeitaufgelöste Methoden derart weiter entwickelt werden, dass (i) die Proteinkonzentration auf $100\mu\text{M}$ abgesenkt werden konnte, (ii) Faltungsereignisse auf der Millisekundenzeitskala mit Ratenkonstanten von bis zu 8 s^{-1} detektiert werden konnten, und (iii) die kinetischen Daten eindeutig zeigten, dass die Faltung von BLA komplexer abläuft als bisher angenommen wurde. Gestützt durch *stopped flow*-Doppelsprung-Fluoreszenzmessungen konnte der folgende Faltungsmechanismus für BLA abgeleitet werden.

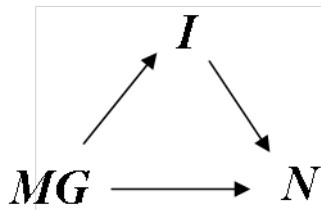


Figure B.2.: Triangular-Faltungsmechanismus für die Faltung von BLA. Es existieren parallele Faltungswege für die Bildung des nativen Proteins.

Besonders interessant hierbei ist, dass für das Hühnereiweisslysozym, das eine hohe Strukturhomologie zu BLA zeigt, ebenfalls dieser Triangular-Faltungsmechanismus gefunden wurde. Diese Faltungshomologie legt den Schluss nahe, dass der Calciumbindestelle bei der Faltung von BLA eine tragende Rolle bei der Ausbildung von Kontakten zwischen den Subdomänen zukommt, eine Funktion, die im Falle des Lysozyms durch langreichweitige Wechselwirkungen zwischen den Subdomänen im entfalteten Zustand erfüllt wird.

Die Existenz des in Abb. B.2 gezeigten Intermediates *I*, das nicht die Charakteristika eines *MG* zeigt, wurde erstmalig durch zeitaufgelöste Photo-CIDNP-NMR-Messungen nachgewiesen. Hierbei wurde Calcium als faltungsauslösendes Agens eingesetzt. Frühere Studien mit paramagnetischem Ytterbium anstelle von (diamagnetischem) Calcium hatten gezeigt, dass die

Faltung von BLA einem ähnlichen Verlauf folgt und die paramagnetischen Eigenschaften von Ytterbium zu einer erhöhten Dispersion der chemischen Verschiebung in NMR-Experimenten führen, was sich auf Grund verbesserter spektraler Auflösung positiv in orts aufgelösten kinetischen Studien auswirkt. Der Inhalt des 3. Ergebniskapitels gibt die Resultate wieder, die in Vorbereitung zeitaufgelöster Photo-CIDNP-NMR-Messungen in Anwesenheit von Ytterbium durchgeführt wurden. In einer ersten Reihe von Messungen wurde der Effekt von Ytterbium auf CIDNP der drei polarisierbaren Aminosäuren Tryptophan, Tyrosin und Histidin untersucht. In allen Fällen zeigte sich, dass (i) Ytterbium zu veränderten chemischen Verschiebungen der Aminosäureresonanzen führt, (ii) die betreffenden Resonanzen deutlich breitere Linien zeigen und (iii) CIDNP-basierte Signalverstärkungen merklich abgeschwächt wurden. Diese Beobachtungen sind in Abb. B.3 exemplarisch für N-Acetyltyrosin gezeigt.

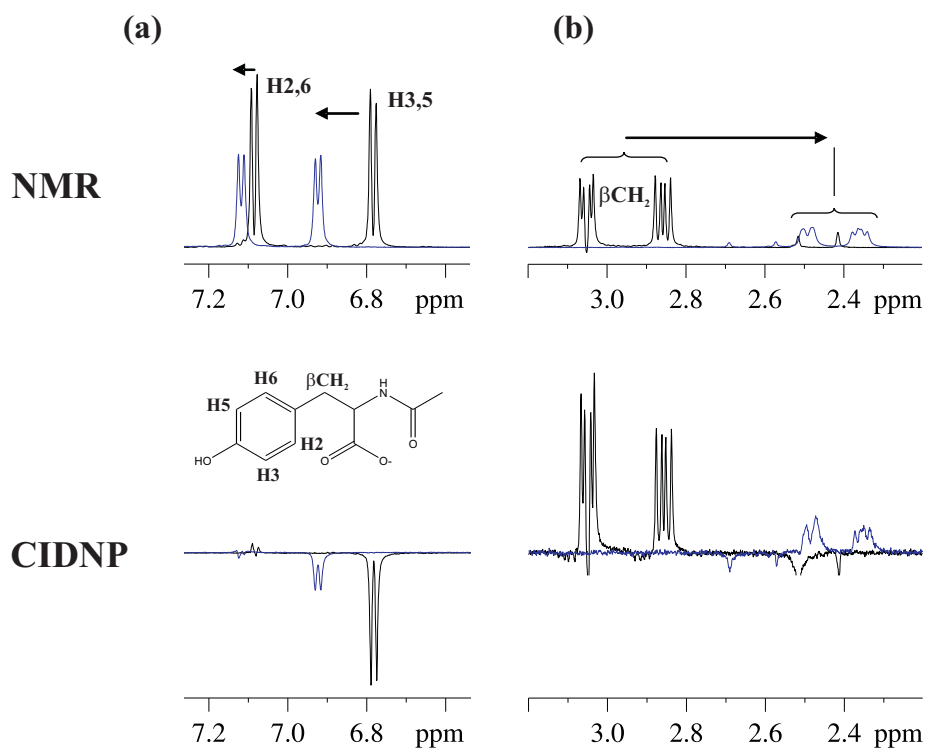


Figure B.3.: 1D, ^1H and Photo-CIDNP NMR-Spektren von N-Acetyltyrosin in Abwesenheit (schwarz) bzw. Anwesenheit (blau) von Ytterbium. Gezeigt sind die Tieffeld- (links) und Hochfeldregionen (rechts) des jeweiligen Spektrums.

Im zweiten Teil des Kapitels ist die erfolgreiche organische Synthese des CIDNP-Farbstoffs

Zusammenfassung

Flavin I dargestellt, der für Messungen an BLA anstelle von FMN verwendet wurde, weil, wie in unabhängigen Untersuchungen festgestellt (Doktorarbeit Lars T. Kuhn, Universität Oxford), Ytterbium mit FMN unlösliche Komplexe bilden kann. Der letzte Abschnitt fasst erste Photo-CIDNP-Messungen an BLA unter Gleichgewichtsbedingungen zusammen. Die in diesem Kapitel dargelegten Resultate erlauben nun, zeitaufgelöste Photo-CIDNP-NMR-Messungen in Gegenwart von Ytterbium durchzuführen.

Danksagung

Zunächst gilt mein Dank meinem Doktorvater Harald Schwalbe. Meinen besonderen Dank möchte ich dafür aussprechen, dass er mir die Möglichkeit gab, mich mit den beiden Themen auseinanderzusetzen, die mich seit der Grundvorlesung in Organischer Chemie fasziniert haben: NMR-Spektroskopie und Proteinfaltung. Weiterhin danke ich Harald Schwalbe für die Vielfalt in den bereitgestellten Projekten, die Möglichkeit kurzzeitiger Forschungsaufenthalte in England und der Schweiz, die ausgezeichnete Ausstattung mit Arbeitsmitteln sowie seine mir entgegengebrachte Geduld.

Als nächstes seien die (nahezu komplett) ehemaligen Mitglieder des Boys' room genannt: Tomislav Argirevic, Jan Ferner, Boris Fürtig, Jürgen Graf, Steffen Grimm, Martin Hähnke, Sergey Ilin und Christian Schlörb. Allen danke ich für vielfache Hilfe und die immer sehr spassige Zeit. Zahllose Momente werden mir, glaube ich, in bleibender Erinnerung bleiben.

Für eine ebenso spassige wie kurzweilige Zeit zum Ende der Doktorarbeit im Erdgeschoss des Instituts danke ich Hamid Nasiri, Jonas Noeske und Gottfried Zimmermann.

Mein besonderer Dank gilt auch Julia Wirmer, die mich in das α -Lactalbumin-Projekt einführte und mir über die Jahre eine hilfsbereite und kompetente Ansprechpartnerin war.

Darüberhinaus sei ehemaligen sowie momentanen Arbeitskreismitgliedern für eine gute Arbeitsatmosphäre und vielfältige Hilfe in allerlei Dingen gedankt: Karin Abarca-Heidemann, Hashim Al-Hashimi, Aphrodite Anastasiadis-Pool, Marie Anders-Maurer, Neda Bakhtiari, Katja Barthelmes, Holger Berk, Janina Buck, Florian Buhr, Emily Collins, Elke Duchardt, Christian Gerum, Friederike Heinicke, Fabian Hiller, Hendrik Jonker, Melanie Koschinat, Jitendra Kumar, Anna Lieblein, Tanja Machnik, Vijayalaxmi Manoharan, Daniel Mathieu, Hillary Moberly, Raja Muruga Poopathi, Gerd Nielsen, Senada Nozinovic, Stephan Rehm,

Danksagung

Anke Reining, Jörg Rinnenthal, Nicole Schmut, Robert Silvers, Sridhar Sreeramulu, Max Stadler, Jochen Stehle, Anna Wacker, Karla Werner und Jens Wöhnert.

Meinen Dank möchte ich auch Christian Richter und Sarah Mensch bzw. Elke Stirnal für Hilfe an den NMR-Spektrometern bzw. bei der Aufreinigung von Proteinproben aussprechen.

Anna Paulus und Elena Hartmann danke ich für verlässliche und kompetente Hilfe bei administrativen Angelegenheiten.

Weiterhin gilt mein Dank verschiedenen Kollaboratoren. So danke ich Andrey V. Solov'yov (FIAS) und Oleg I. Obolensky (NCBI, NIH, vormals FIAS) für die konstruktive und fruchtbare Zusammenarbeit im Zusammenhang mit der Vorhersage von dipolaren Restkopplungen in entfaltenen Proteinen, wodurch ich wertvolle Einblicke in die theoretische Behandlung biologischer Fragestellungen gewinnen konnte. An dieser Stelle sei auch Martin Blackledge (IBS, Grenoble) für hilfreiche Diskussionen gedankt. Auch Horst Stöcker als FIGSS-Direktor möchte ich für seine Unterstützung danken.

Darüberhinaus richtet sich mein Dank an Peter Hore (Universität Oxford, UK), der mir zweimal die Möglichkeit zu einem Forschungsaufenthalt in seiner Arbeitsgruppe gab. Diese wenn auch kurzen Aufenthalte wurden gerade durch persönliche Begegnungen bereichert. So möchte ich in diesem Zusammenhang Iain J. Day und Lars T. Kuhn für die gemeinsame Durchführung von Experimenten und Ken Hun Mok sowie Kiminori Maeda für Diskussionen danken.

Auch bedanke ich mich bei Thomas Kiefhaber (TU München, vormals Biozentrum Universität Basel), in dessen Labor ich unter Hilfe von Annett Bachmann Stopped-flow-Doppelsprung-Messungen vornehmen konnte, die für den Fortgang des α -Lactalbumin-Projektes sehr wichtig waren.

Zuguterletzt spreche ich meinen besonderen Dank meinen Eltern, meiner Familie und einigen guten Freunden aus, die mir wichtige Unterstützer waren und sind, insbesondere durch das positive soziale Umfeld, das sie mir geben.

Curriculum Vitae

KAI SCHLEPCKOW

Personal Data:

Date of birth: 5th June 1978
Place of birth: Karlsruhe (Germany)
Citizenship: German

PhD:

Since 09/03 PhD thesis under supervision of
Prof. Dr. Harald Schwalbe, Dissertation: "Development
of NMR Methods for the Study of Protein Folding"

Studies:

07/2003 Final examination with the degree "Diploma in Biochemistry"
01/03 – 07/03 Diploma thesis under supervision of
Prof. Dr. Harald Schwalbe, "Investigation of Folding Kinetics
of Model Proteins Using NMR Spectroscopy"
10/2000 Intermediate examination
10/98 – 07/03 Studies in Biochemistry

Civilian Service:

09/97 – 09/98 Surgery section of the Rotkreuzklinik Würzburg, Germany

Education:

1995 – 1997 Röntgen-Gymnasium Würzburg, Germany (secondary school)
1988 – 1995 Faust-Gymnasium Staufen, Germany (secondary school)
1984 – 1988 Primary school Heitersheim, Germany

Akademische Lehrer an der Johann Wolfgang Goethe-Universität Frankfurt am Main waren
(in alphabetischer Reihenfolge):

Prof. Dr. Hashim Al-Hashimi, Prof. Dr. Ernst Bamberg, Prof. Dr. Jürgen Bereiter-Hahn,
Prof. Dr. Joachim Engels, Prof. Dr. Karl-Dieter Entian, Prof. Dr. Dr. Hugo Fasold, Prof. Dr.
Jürgen Feierabend, Prof. Dr. Werner Gnatzy, Prof. Dr. Michael Göbel, Prof. Dr. Christian
Griesinger, Prof. Dr. Stephen K. Hashmi, Prof. Dr. Bernd O. Kolbesen, Prof. Dr. Bernd
Ludwig, Prof. Dr. Walter E. Müller, Prof. Dr. Lutz Nover, Prof. Dr. Heinz D. Osiewacz, Prof.
Dr. Roland Prinzinger, Prof. Dr. Reiner Protsch von Zieten, Prof. Dr. Hartmut Roskos, Prof.
Dr. Heinz Rüterjans, Prof. Dr. Michael Sattler, Prof. Dr. Harald Schwalbe, Prof. Dr. Jörg
Soppa, Prof. Dr. Robert Tampé, Prof. Dr. Matthias Wagner, Prof. Dr. Dr. h. c. Christian
Winter, Prof. Dr. Jens Wöhnert, Prof. Dr. Herbert Zimmermann

List of Publications

O.I. Obolensky, K. Schlepckow, H. Schwalbe, A.V. Solov'yov, Theoretical Framework for NMR Residual Dipolar Couplings in Unfolded Proteins, *J. Biomol. NMR* (2007) **39**, 1-16.

K. Schlepckow, J. Wirmer, A. Bachmann, T. Kiefhaber, H. Schwalbe, Conserved Folding Pathways of α -Lactalbumin and Lysozyme Revealed by Kinetic CD, Fluorescence, NMR, and Interrupted Refolding Experiments, *J. Mol. Biol.* (2008) **378**, 686-698.