Investigation of the kinetics of protein folding and the ensemble of conformations in non-native states of proteins by liquid NMR spectroscopy

by

Julia Wirmer

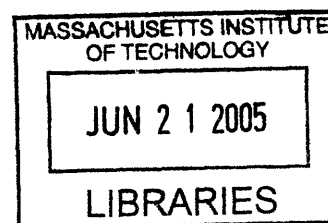Dipl. in Chemistry, J. W. Goethe University, Frankfurt, Germany, 2000

SUBMITTED TO THE DEPARTMENT OF CHEMISTRY IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN CHEMISTRY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June, 2005

Author_____ December 17th, 2004
Department of Chemistry
DATE

Certified by_____
Harald Schwalbe
Professor of Chemistry
Thesis Supervisor

Accepted by_____
Robert W. Field
Professor of Chemistry
Chairman, Department Committee on Graduate Students

This doctoral thesis has been examined by a committee
of the Departement of Chemistry as follows:

Professor Robert G. Griffin_____

                                                                                         Chair

Professor Barbara Imperiali_____

Professor Daniel S. Kemp_____

Professor Harald Schwalbe_____

                                                                                   Thesis Supervisor

Investigation of the kinetics of protein folding and the ensemble of conformations in non-native states of proteins by liquid NMR spectroscopy

by Julia Wirmer

Abstract

For a complete description of protein folding dynamics and the structure of the folded state, of unfolded and of non-native states of proteins and the kinetics of protein folding from the unfolded state to the folded state have to be determined. The focus of this PhD thesis was the development of novel NMR **methodologies** to study protein folding using NMR spectroscopy. This has been achieved by studying **three model proteins** ubiquitin, α-lactalbumin and lysozyme in their folded and especially in their unfolded states. The proteins were chosen, because α-lactalbumin and lysozyme are two proteins with nearly identical fold but different function and ubiquitin is a very stable protein without disulfide bonds.

Methodologies

- A new NMR pulse sequence for the determination of $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants in proteins was developped. The method is based on J-modulated HSQCs and can be applied to folded as well as unfolded proteins. The new coupling constants report on backbone ϕ and ψ angles.

- Residual structure and long-range interactions in unfolded proteins can be detected by a new combination of site directed non-conservative mutagenesis and NMR analysis. Identification of long-range interactions is a) based on the analysis and interpretation of R_2 relaxation rates, for which models have been derived and b) based on NMR diffusion data which directly correlate to the compactness of a given protein.

- A method to study laser triggered kinetics of protein folding by time-resolved photo-CIDNP NMR was developed. Two lasers were coupled into an NMR spectrometer: one for initiation of folding by releasing ions from photo-labile chelators with dead times as low as 200ms, and one for induction of photo-CINDP NMR. The method can not only be applied to ion induced kinetics of protein folding, but is generally applicable to kinetics of biomacromolecules such as proteins and RNA that involve photo-protected cofactors as could be shown in our laboratory.

Investigations on ubiquitin

- The newly developed NMR method for the determination of $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants was applied to folded ubiquitin. Analysis revealed the dependence of the coupling constants on the backbone conformation, predominantly on ψ, making the pulse sequence and the parameterization developed here a valuable new tool for the determination of the backbone conformation in folded as well as in unfolded proteins.

- Unfolded ubiquitin was investigated using scalar coupling constants and ^{1}H,^{15}N relaxation data. The experimental data agree well with models proposed to describe unfolded states of proteins as a statistical coil where dynamics are governed solely by segmental motions. Unfolded ubiquitin is thus a good model for a protein without detectable residual structure in its unfolded state.

Investigations on lysozyme
- Residual structure in the unfolded states of non conservative single point mutants (A9G, W62G, W62Y, W111G and W123G) of hen lysoszyme was monitored by chemical shift measurements, ^{15}N transverse relaxation rates and particularly diffusion constants. Long-range interactions between hydrophobic clusters of unfolded lysozyme were observed. Single point mutations dramatically alter the overall compactness of the unfolded state.

Investigations on α-lactalbumin
- Isotope labeled bovine α-lactalbumin (BLA) was expressed heterologously using a new construct with a His-tag and a trypsin cleavage site. The sequence and stability of the obtained BLA is identical to the wild type protein. This makes it a perfect construct to study kinetics of folding and unfolded states of BLA.
- ^{13}C,^{15}N isotope labeled unfolded BLA was assigned applying standard and non-standard NMR assignment experiments. Residual secondary structure was identified near the N- and the C-terminus of unfolded BLA, in regions belonging to the α-domain in the folded state, suggesting a possible folding nucleus.
- This unfolded state of BLA was furthermore compared to the unfolded state of human lactalbumin (HLA) and lysozyme based on residual structure and ^{15}N relaxation data. The unfolded states vary considerably for the three proteins, which possess very similar structures in their native state. The structural ensemble in the unfolded states of proteins are determined by the primary sequence of the protein and even smallest single point mutations as found between HLA and BLA can change the conformation of the unfolded state considerably.
- The Ca^{2+}-triggered folding kinetics of BLA under constant denaturant (4M urea) has been investigated by laser induced release of Ca^{2+}-ions from a photolabile chelator within the NMR spectrometer and subsequent photo-CIDNP signal detection. A folding intermediate possessing a tyrosine residue in a non-native conformation was detected 200ms after initiation of folding. Therefore, parts of the polypeptide chain in the β-domain of BLA sample non-native conformations, while a hydrophobic core is formed.

The findings in the kinetic investigations are in line with the detected residual structure. Refolding of amino acids involved in non-native clusters in the intermediates has to proceed the correct folding and therefore constitutes a rate limiting step on the Ca^{2+}-induced refolding of α-lactalbumin.

Thesis Supervisor: Harald Schwalbe
Title: Professor of Chemistry

Für meine Eltern

Acknowledgements

I am deeply indebted to Harald Schwalbe who has been a great supervisor. His contributions to my education and scientific thinking cannot be counted. I thank him for all the great advices he gave me on my projects and for encouraging my increasing independence towards the end of this thesis. He has a great ability to create an extraordinary scientific atmosphere and a unique social environment in his group. He believes in us and supports us (not only in terms of research) where he can.

I am grateful to my thesis committee, Robert G. Griffin, Barbara Imperiali and Daniel S. Kemp, for interest in my work, unproblematic organization of meetings and for taking the time to read this thesis.

I would like to express my gratitude to Christian Schlörb, Vijayalaxmi Manoharan, Judith Klein-Seetharaman, Ryoma Hirano, Tadashi Ueda, Taiji Imoto for the most successful collaboration on the unfolded lysozyme project. Michele Vendruscolo and Robert B. Best taught me how to do MD simulations on unfolded proteins; I am very grateful for the insight into the amazing field of MD simulations on protein folding I obtained during the time in Oxford in Michele's lab. I would like to thank Till Kühn who shared his enthusiasm about the α-lactalbumin refolding project and filled me in, when I started. Kai Schlepkkow and Holger Berg are also gratefully acknowledged for the collaboration on the α-lactalbumin refolding project and for continuing with it. The contribution of R_2 relaxation rates by Christina Redfield allowing for a comparison of different α-lactalbumins and lysozymes is greatly appreciated. I would like to thank Wolfgang Peti for the fruitful collaboration on the unfolded ubiquitin project. I thank Stefan Dröse, Ulrich Brandt, Bernd Ludwig and Thomas Kleinschroth for the collaboration on the bc_1 project and the insights into the world of membrane proteins.

I really enjoyed being in Boston and I would like to thank all people that made this time special for me. The MIT and the magnet lab is a scientifically most stimulating environment. Thanks to Peter Allen and Ajay Thacker for keeping the spectrometers running and special thanks to Chris Turner and Dave Ruben for help with implementing laser pulse sequences on the home-build spectrometers.
In addition, I would like to thank Bob for offering hazelnut flavored coffee and for the Monday morning coffee and donuts tradition. I thank Dierk Blechschmitt for teaching me how to express

protein, and for giving advice in "allen Lebenslagen". I am grateful to Aaron for his friendship and for the numerous weekend afternoons, we spent exploring Boston.

I would like to thank all present and past members of the Schwalbe group for an amazing and educational time. Thanks for all the scientific and non-scientific discussions in the seminar room, for amazing group trips and nights out. Work in the Schwalbe group in Frankfurt would not have been possible without the help of some special persons. First, I would like to thank Anna Paulus; she is always helpful, despite the fact that she is completely overworked with all the paper work that we and other people produce. Gottfried Zimmermann was irreplaceable for the installation of the laser and is always very helpful with any technical problem. Finally, I would like to thank Christian Schlörb, Dierk Blechschmidt, Elke Duchardt, Jan Ferner and Steffen Grimm for the maintenance of the computer systems.

I am grateful to Jens Wöhnert for huge advice in the biolab, for the afternoon coffee break, for his cynical humor and for endless discussions ranging from my projects, to prospective post doc groups and live in general. Assistance at the spectrometer by Christian Richter is gratefully acknowledged. I am still amazed by Christian's enthusiasm if it comes to potential projects, his family and his house. Elke Stirnal really helped with HPLC purifications. I thank her for personal conversations. Special thanks to the probably most helpful person of the group, Christian Schlörb. Good luck with your project.

One of the greatest things in this group is the girls' lab. I thank you all for a marvelous atmosphere and the proof that something like a girl's lab can work out! Karla, I like your quiet temper that changes into jolly cheerfulness if you know people better and if there is something to celebrate. I would like to thank Emily for being a person "mit der man Pferde stehlen kann" and I am grateful to Aph for being so girlish. I shared an office with Elke during my whole PhD thesis. It will be strange, when she has left for her post doc. I will miss the discussions about really everything, the snappy comments, and her feeding concept.

I would like to thank my parents and my brothers for their love and support and for their interest in my work. I always like to answer yes to the occasional question: "Are the proteins still folding?" and I am particularly thankful that you accept and support all my decisions and that I can always come to the Atzelbergstrasse to find a relaxing atmosphere.

I would like to thank my husband Stefan to start and to believe in a relationship over the ocean. I cannot thank you enough for all your love and support.

Publications

(1) <u>Julia Wirmer</u>, Till Kühn & Harald Schwalbe 'Millisecond Time Resolved Photo-CIDNP NMR Reveals a Non-native Folding Intermediate on the Ion Induced Refolding Pathway of Bovine α-Lactalbumin' *Angew. Chem. Int. Ed. Enlg.* **40**, 4248-4251 (2001).

(2) <u>Julia Wirmer</u> & Harald Schwalbe 'Angular Dependence of $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants measured in J-modulated HSQCs.' *J. Biomol. NMR* **23**, 47-55 (2002).

(3) Judith Klein-Seetharaman, Maki Oikawa, Shaun B. Grimshaw, <u>Julia Wirmer</u>, Elke Duchardt, Tadashi Ueda, Taiji Imoto, Lorna J. Smith, Christopher M. Dobson & Harald Schwalbe 'Long-Range Interactions Within a Nonnative Protein' *Science* **295**, 1719-1722 (2002).

(4) <u>Julia Wirmer</u>, Christian Schlörb, Judith Klein-Seethamaran, Ryoma Hirano, Tadashi Ueda, Taiji Imoto & Harald Schwalbe 'Modulation of Compactness and Long-range Interactions of Unfolded Lysozyme by Single Point Mutations' *Angew. Chem. Int. Ed. Engl.* **43**, 5780-5785 (2004).

(5) <u>Julia Wirmer</u>, Christian Schlörb & Harald Schwalbe 'Conformation and Dynamics of Nonnative States of Proteins studied by NMR Spectroscopy' *Protein Folding Handbook*, ed. by Johannes Buchner & Thomas Kiefhaber, WILEY-VCH, Weinheim (2005).

(6) Harald Schwalbe & <u>Julia Wirmer</u> 'Protein Misfolding Disease: Overview of liquid and solid state high resolution NMR studies' *Handbook of Modern Magnetic Resonance,* ed. by Graham Webb, Kluwer Academic Publishers, London (2005).

Table of Contents

List of Abreviations

1D, 2D, 3D	One dimensional, two dimensional, three dimensional
^{1}H, ^{13}C, ^{15}N	NMR active nuclei in proteins
A9G	A9G mutant of hen lysozyme
AcN	Acetonitrile
BLA	Bovine α-lactalbumin
BLA+3	BLA with additional three amino acids at the N-terminus
CD	Circular dichroism
CIDNP	Chemically induced nuclear polarization
CSI	Chemical shift index
D	Diffusion constant
DMN	Dimethoxynitrophen
DNA	Desoxy-ribonucleic acid
E. coli	*Escherichia coli*
EDTA	Ethylene-diamine-tetra-acetic acid
F	Folded state
FMN	Flavin-mononucleotid
FPLC	Fast protein liquid chromatography
GdnHCl	Guanidinium hydrochloride
hetNOE	Heteronuclear nuclear Overhauser effect
HEWL	Hen lysozyme
HISBLA	BLA with an additional N-terminal poly histidine sequence
HLA	Human α-lactalbumin
HPLC	High performance liquid chromatography
HSQC	Heteronuclear single quantum correlation
Hz	Hertz
I	Intermediate state
IPTG	Isopropylthiogalactoside
k	Kinetic rate constant
LB	Luria broth
M	molar
M9	Minimal medium
MALDI-TOF	Matrix-assisted laser desorption/ionization time of flight
MG	Molten globule state
MHz	Spectrometer frequency in mega hertz
MW	Molecular weight
Ni-NTA	Nickel-nitrilotriacetic acid
^{n}J	Scalar coupling constant with n=1, 2 or 3
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser effect spectroscopy
NS	Number of scans
OD	Optical density
pdb	Protein data bank
pHISBLA	Plasmid coding for HISBLA
pI	Isoelectric point

ppm	Parts per million
R_1	Longitudinal relaxation rate
R_2	Transverse relaxation rate
rc	Random coil
RD	Recyle delay
RDC	Residual dipolar coupling
R_g	Radius of gyration
R_h	Radius of hydration
RNA	Ribonucleic acid
S^2	Order parameter
SDS-PAGE	Sodium dodecylsulfate polyacrylamide gel electrophoresis
$-S^{ME}$	Methylated cysteine residues in a protein
T_2	Transverse relaxation time
TFA	Trifluoroacetic acid
TOCSY	Total correlation spectroscopy
Tris	Tris (hydroxymethyl) aminomethane
U	Unfolded state
UV	Ultra violet
VIS	Visible
W111G	W111G mutant of hen lysozyme
W123G	W123G mutant of hen lysozyme
W62G	W62G mutant of hen lysozyme
W62Y	W62Y mutant of hen lysozyme
WT	Wild type
WT-S^{ME}	Cysteine methylated hen lysozyme
α, β, γ	Position of atoms within a given amino acids
α-LA	α-lactalbumin
δ	Chemical shift
ϕ, ω, ψ	Protein backbone torsion angles
θ	Elipticity
τ_c	Overall correlation time
τ_e	Internal correlation time

Chapter 1

Introduction

1 Introduction

1.1 Protein folding

1.1.1 Protein folding and misfolding *in vivo*

Protein folding is one of the key steps in protein biosynthesis, where genetic information is transformed into biological activity. The term protein folding in general describes the process by which an unfolded protein (U) gains its native conformation (N). During this process, intermediates (I) might be populated (Figure 1).
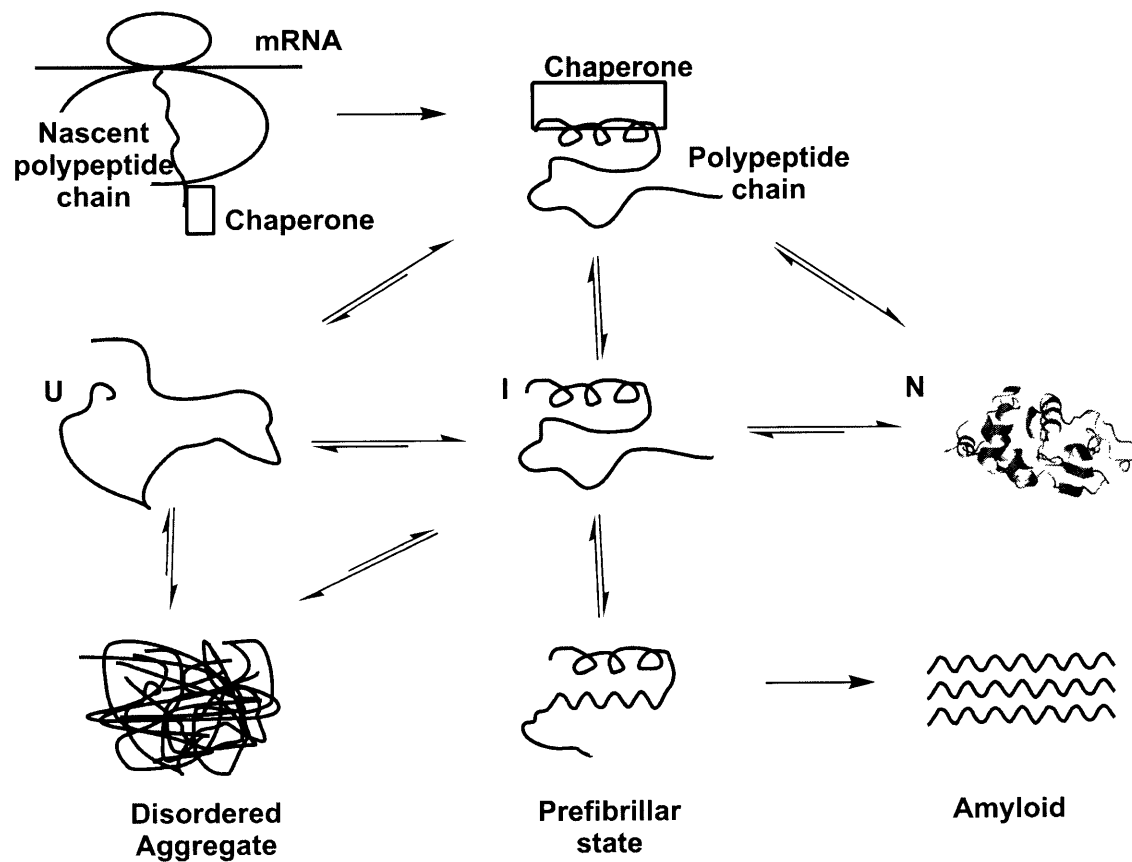
Figure 1: Schematic view of protein folding and misfolding in the cell. Upper part: Protein synthesis on the ribosome. The mRNA and the nascent polypeptide chain are indicated as well as a chaperone that binds to the nascent polypeptide chain. Upon completion of the synthesis the polypeptide chain is still bound to a chaperone. Middle part: unfolded state (U), intermediate or partially folded state (I) and native state (N) of a protein, here the native hen lysozyme structure (Vaney et al., 1996) is shown. Lower part: disordered aggregate, prefibrillar state and ordered aggregate, called amyloid.

However, since the living cell is a very crowded environment with an effective protein concentration as high as 300mg/ml, an "unprotected" unfolded protein is more prone to aggregate than to fold in the cellular environment (Zimmerman and Trach, 1991). Therefore, the nascent polypeptide chain is recognized by a chaperone already during protein synthesis (Hartl and Hayer-Hartl, 2002). Chaperones prevent aggregation and in some cases also actively aid protein folding. Proteins adopt their native conformation within milliseconds to seconds after the finished synthesis of the polypeptide chain.

As depicted in Figure 1, protein folding is not always productive (Dobson, 2003). Folding intermediates or even the native state can unfold due to e.g. cellular stress such as chemicals, or non-physiological temperatures. The cell usually responds to this stress by producing increasing amounts of specific protective proteins. However, if the protective mechanisms fail, unfolded and intermediate states of proteins may aggregate, or prefilbrillar states can be formed that ultimately result in the formation of amyloids. These misfolded species of proteins lead to diseases such as Alzheimer's, Parkinson's and BSE (Dobson, 2004).

Protein folding and misfolding can be studied *in vitro* by a variety of biophysical methods. *In vitro* studies of protein folding intend to investigate aspects of protein folding that are intrinsic to properties of the polypeptide chain. This PhD thesis investigates aspects of protein folding using NMR spectroscopy.

1.1.2 Protein folding *in vitro*

In 1960, the first *in vitro* experiments of protein folding were done on Ribonuclease A (RNase A) (Anfinsen et al., 1961). Incubation of RNase A in 8M urea and a reducing agent resulted in enzymatically inactive protein without disulfide bridges; a state that is called "unfolded" (U) today. The protein refolded upon removal of the reducing agent and urea, and regained its enzymatic activity. The observation of Anfinsen indicated that the three dimensional structure of a protein is a property of its amino acid sequence and that folding is an autonomous and spontaneous process.

In 1969, Levinthal speculated about the discrepancy between the time scales for a random conformational search and the rapidness of protein folding. Protein folding would take - even for a small protein of 150 amino acids - longer than the age of the universe if folding involved a

19

random search of all possible conformations. Therefore, protein folding must follow a predetermined folding pathway (Levinthal, 1969).

It is the aim of this thesis to contribute to the understanding of this predetermined folding pathway. For a description of productive protein folding, the native state as well as the non-native states and the kinetics of protein folding have to be described. The folded (native) state is well characterized for a large number of proteins. In contrast, kinetics of protein folding and non-native states of proteins are not well understood so far.

1.1.2.1 Native state of proteins

Folded state: For most proteins the active native state of the respective protein is the folded state. This folded state is characterized by one conformation around which minor fluctuations may occur. The folded state of proteins is the best-characterized state of proteins. Currently (as of November 2004) 27969 three dimensional structures of folded proteins are deposited in the protein data bank (Berman et al., 2000). These structures can be classified into ≈ 500 known folds (Holm and Sander, 1998). However, estimates for the total number of naturally occurring folds range between 1000 and 10000 (Chothia, 1992; Wolf et al., 2000). Based on the amount of native protein structures available, the prediction of the three dimensional fold from the sequence is still an emerging field (Lee et al., 2004a; Lee et al., 2004b; Rost et al., 2004; Skolnick et al., 1993).

Natively unfolded proteins: Recently, a number (126 as of 2003) (Tompa, 2003) of natively unfolded proteins have been identified. These natively unfolded proteins are extremely flexible, non-compact, and reveal little if any secondary structure under physiological conditions. Most natively unfolded proteins are regulators in key cellular processes. These can be divided into five different classes (Tompa, 2002). The first class is the class of entropic chains, while all other classes bind to either proteins, RNA, DNA, or a range of small ligands. They thereby either store ligands, inhibit functionality, assemble multiprotein complexes, or regulate functionality by phosphorylation. Natively unfolded proteins are implied in the development of a number of neurodegenerative diseases including Alzheimer's disease, Down's syndrome and Parkinson's disease to name a few (Uversky et al., 2000). They are predicted to be ubiquituous in the proteom (Tompa, 2002; Uversky, 2002), and algorithms available as a web-program (http://dis.embl.de/) have been developed to predict protein disorder (Linding et al., 2003). According to predictions, 35-51% of eucaryotic proteins have at least one long (> 50 residues) disordered region, and 11%

of proteins in the Swiss-Prot database and between 6 and 17% of proteins encoded by various genomes are probably fully disordered (Tompa, 2002).

1.1.2.2 Non-native states of proteins

Non-native states of proteins are not only the starting point of protein folding, but are also implied in misfolding, transport through membranes, protein turnover and degradation. The term non-native states of proteins is usually used to describe unfolded or partially folded states. *In vitro* proteins are usually unfolded by denaturant and/or pH. In addition, a number (126 as of 2003) (Tompa, 2003) of natively unfolded proteins that mostly fold upon ligand binding have been identified (section 1.1.2.1). In contrast to the folded state of a protein, non-native states cannot be described by a single conformation but as an ensemble of rapidly inter-converting conformers. The individual members of this ensemble may differ substantially in their structural and dynamical properties, and different parts of the polypeptide chain may change conformation at different rates (Smith et al., 1996b). Previous studies showed that the ensemble of conformers in the non-native state of proteins sample the preferred regions of the Ramachandran ϕ,ψ space (Smith et al., 1996b). This sampling, however, can be restricted by varying degrees of residual structure. Residual secondary structure elements were identified in unfolded conformations of a significant number of investigated proteins (Blanco et al., 1998; Fiebig et al., 1996; Klein-Seetharaman et al., 2002; Lietzow et al., 2002; Schwalbe et al., 1997; Shortle and Abeygunawardana, 1993; Shortle and Ackerman, 2001; Wirmer et al., 2004; Wong et al., 1996). In agreement with the protein folding funnel theory (section 1.1.2.4), residual structure in unfolded proteins could channel protein folding into different distinct pathways along the protein folding funnel. In this thesis the unfolded states of α-lactalbumin, lysozyme and ubiquitin are investigated using NMR spectroscopy (chapters 3, 4 and 6).

1.1.2.3 Aggregated states

Aggregated states of proteins are the result of protein misfolding. Two sorts of aggregated states occur *in vivo*: 1) disordered aggregates that are usually rapidly removed by cellular proteases and 2) ordered amyloid aggregates that are not digested by proteolysis. These insoluble amyloids are deposited as fibrils or plaques in different tissues and organs of the body. There they lead to diseases such as Alzeimer's, Parkinson's and BSE. Recently, atomic structures of amyloid fibrils have been solved using solid state NMR spectroscopy (Jaroniec et al., 2002; Jaroniec et al., 2004;

Petkova et al., 2004; Petkova et al., 2002). The fibrillar form of an amyloidogenic peptide (transthyrethin105-115) for example displays an extended β-sheet conformation with its backbone and side chain torsion angles close to their optimal values for this secondary structure element. In addition, long-range order could be detected that is generally associated with crystalline materials (Jaroniec et al., 2002; Jaroniec et al., 2004). Earlier investigations revealed very similar fibrillar structural characteristics for all amyloids. The fibrils are mostly long, unbranched and often twisted, with a core of β-sheets running perpendicular to the fibril axis (Jimenez et al., 1999; Sunde and Blake, 1997).

1.1.2.4 Kinetics of protein folding (and misfolding)

It is of key interest whether the kinetics of protein folding proceed through few or even one well-defined folding pathway or whether a large number of independent pathways lead to the native structure (Bachmann and Kiefhaber, 2001). The most general model of protein folding describes the energy landscape of a foldable protein as a many-dimensional rough funnel with a free energy gradient towards the native structure (Bryngelson et al., 1995; Onuchic et al., 1995). Different folding pathways are possible along this funnel; local minima within the funnel give rise to folding intermediates. These intermediates could be on pathway or off pathway acting as folding traps. Probing of this folding funnel *in vitro* is often carried out by rapid changes from a denaturing environment into a native environment (e.g. by dilution of denaturant). Two state folders (U→F) have been identified using these methods. On the other hand proteins have been identified that populate intermediates (U→I→F) during folding (Baldwin and Rose, 1999). A folding intermediate can be either a distinct conformation or a highly dynamic ensemble of different conformations. However, there are few methods that provide sufficient structural insight into folding intermediates with high temporal and site resolution to allow their detailed characterization.

Here, a folding intermediate on the Ca^{2+}-induced folding pathway of bovine α-lactalbumin has been characterized using time-resolved photo-CIDNP NMR (see chapter 7).

1.2 Background: Biophysical characterization of non-native proteins using NMR spectroscopy

1.2.1 Characteristic NMR spectra of different folding states of proteins

Characteristic one dimesional (1D) ^{1}H NMR spectra showing the methyl group regions of the three thermodynamic stable states of the small (14.2kDa, 123 amino acids) protein α-lactalbumin are shown in Figure 2. In the native state (Figure 2A), narrow line widths and a large chemical shift dispersion are observed. Particularly in the region between -0.5 and 0.5ppm, well-separated signals corresponding to single methyl groups are observed. In contrast, small chemical shift dispersion is observed in the molten globule state (MG) and in the unfolded state of proteins, due to averaging between different conformers. Whereas the conformational averaging is fast in the unfolded state, leading to sharp signals (Figure 2C), slow conformational averaging is observed in the molten globule state, as reflected by broad signals (Figure 2B).
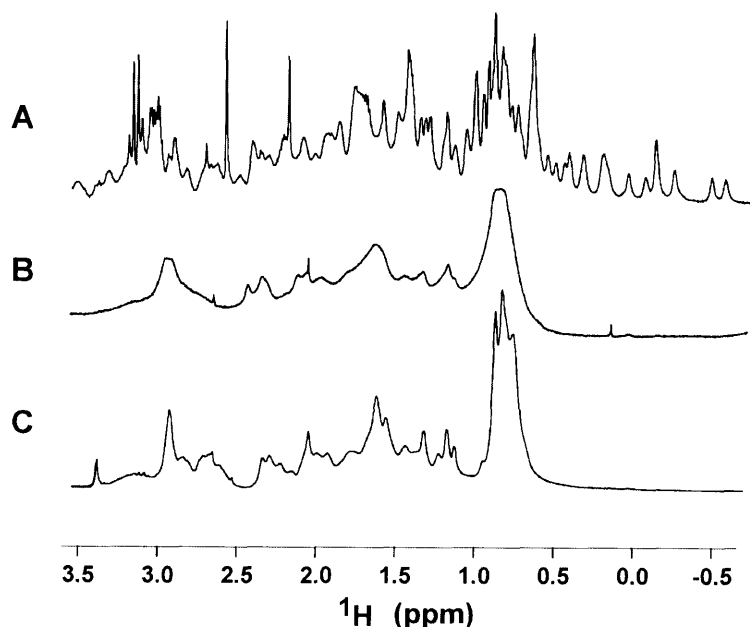
Figure 2: 1D ^{1}H-NMR spectra of the methyl group region showing the differences in chemical shift dispersion and line width for the three states of α-lactalbumin: A) folded state, B) MG and C) unfolded state. All spectra were recorded at 600MHz at 35°C.

The same characteristics in ^{1}H chemical shift dispersion and line width are observed in ^{1}H,^{15}N correlation spectra of the three states (Figure 3). However, in contrast to the ^{1}H chemical shift dispersion, the chemical shift dispersion in the ^{15}N dimension is intrinsically larger, even for the unfolded state of proteins. This larger chemical shift dispersion leads to resolved signals in correlation spectra of unfolded proteins as shown in Figure 3C. Each backbone amide gives rise to a single well separated resonance in the spectrum. It is this property of the ^{15}N nucleus that is the key to the possibility of investigating unfolded proteins at atomic resolution using NMR spectroscopy.

In contrast, the investigation of highly dynamic states such as the MG is not possible. Most amide signals are broadened beyond detection (Figure 3B) due to intermediate exchange.
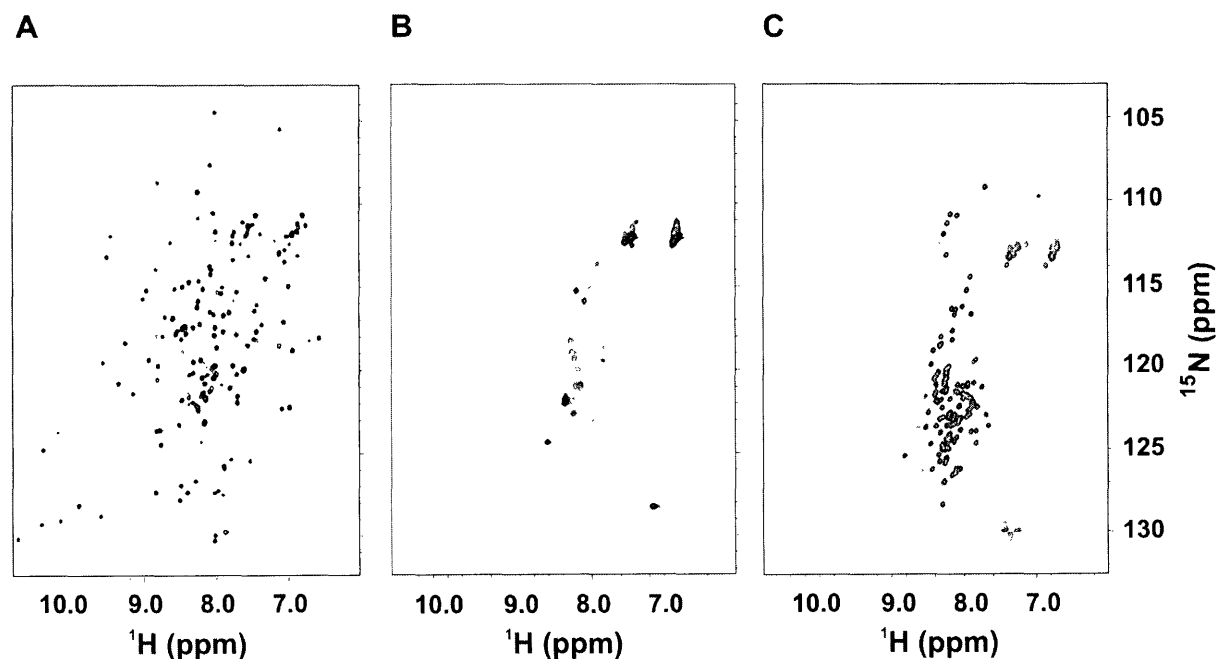
Figure 3: Two dimensional (2D) ^{1}H,^{15}N correlation experiments of the three different states of α-lacatlbumin. A) native state, B) MG, C) unfolded state. All spectra were recorded at 600MHz at 20°C.

1.2.2 Backbone assignment

There are two ways to assign the backbone of isotope labeled proteins. ^{15}N labeled proteins can be assigned based on distances through space using NOESY- and TOCSY-HSQC spectra (Marion et al., 1989a; Marion et al., 1989b). ^{13}C,^{15}N labeled proteins are typically assigned based on through bond connections using scalar coupling constants.

1.2.2.1 Backbone assignment of ^{13}C,^{15}N labeled unfolded proteins

Backbone assignment of ^{13}C,^{15}N labeled proteins is based on correlation experiments that exploit through bond coupling constants. These connect each amide resonance to the C_α, C_β and CO of its own ($C_{\alpha i}$, $C_{\beta i}$ and CO_i) and of its preceding residue ($C_{\alpha(i-1)}$, $C_{\beta(i-1)}$ and $CO_{(i-1)}$). In the HNCACB experiment (Wittekind and Mueller, 1993) (Figure 4B), each amide is connected with its own C_α and C_β as well as with the C_α and the C_β of the preceding residue. In contrast, only the $C_{\alpha(i-1)}$, $C_{\beta(i-1)}$ resonances appear in the CBCA(CO)NH experiment (Muhandiram and Kay, 1994) (Figure 4A). Thus, the CBCA(CO)NH experiment can be used to differentiate between signals arising from the own or the preceding residue, while the HNCACB is used for sequential assignment. C_α and C_β chemical shifts depend on the type of amino acid on one hand and on the conformation and the environment of the respective residue on the other hand.
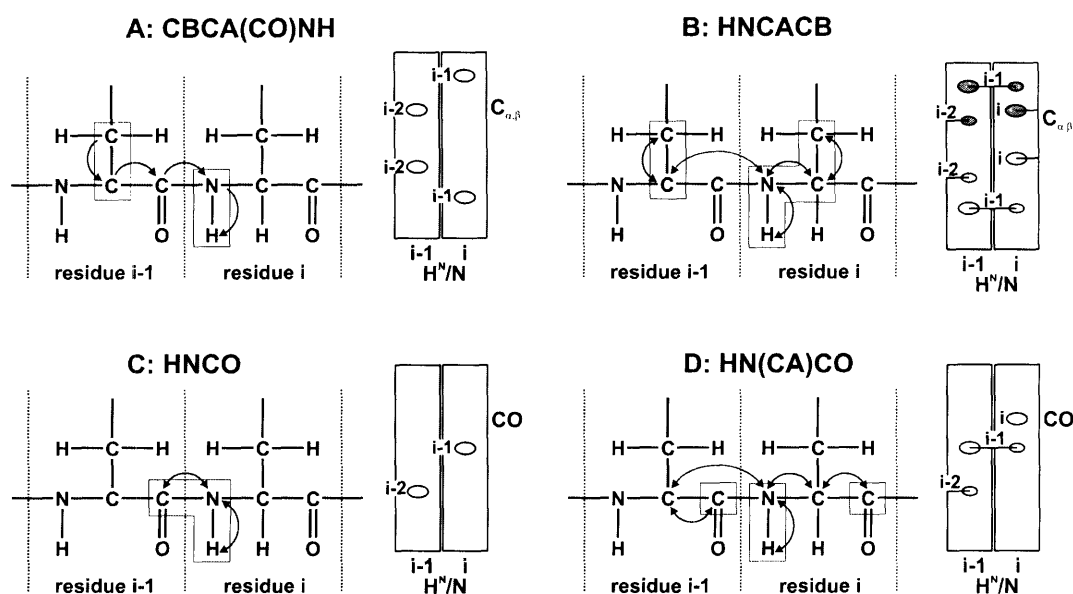
Figure 4: Transfer pathways of standard 3D NMR assignment experiments (Clubb et al., 1992; Grzesiek and Bax, 1992; Muhandiram and Kay, 1994; Wittekind and Mueller, 1993). Arrows indicate magnetization transfer pathways via ^{1}J and ^{2}J coupling constants. Pointers indicate the direction of the transfer. Boxed nuclei are nuclei that evolve chemical shift during the pulse sequence. At the right part of each panel schematic drawings of so-called strip plots are shown. H^N/N indicates either H^N or N chemical shifts (both possible), $C_{\alpha,\beta}$ indicates C_α and C_β chemical shifts and CO indicates carbonyl chemical shifts. The respective residue that gives rise to a peak is indicated with i, i-1 and i-2. Positive peaks are indicated by open ovals, while negative peaks are indicated by ovals filled in gray.

Due to fast conformational averaging the amino acid side chain type solely influences the C_α and C_β chemical shifts in unfolded proteins. This allows the straightforward identification of side chains with a characteristic chemical shift pattern for amino acids such as threonine, serine, alanine and glycine. However, there are often multiple possibilities for the sequential assignment due to repetitions of one or more residues in the sequence, or because a number of amino acids, such as e.g. glutamine, arginine, methionine and lysine have very similar C_α and C_β chemical shifts in unfolded proteins. Unambiguous assignment of unfolded states of proteins is only possible by combining the HNCACB, CBCA(CO)NH based approach with either the NOESY-HSQC and TOCSY-HSQC based assignment approaches (see next section) or with carbonyl-based assignment procedures as used in this work (see chapter 5). The chemical shift dispersion of carbonyl carbons does not depend as strongly on the amino acid type as the C_α and the C_β chemical shifts. A combination of HNCO and HN(CA)CO can therefore be used to simplify assignment in unfolded proteins. The same principles as seen for the combination of HNCACB and CBCA(CO)NH are true for the combination of HN(CA)CO (Clubb et al., 1992) (Figure 4D) and the HNCO (Grzesiek and Bax, 1992) (Figure 4C) experiments. Each amide in the HN(CA)CO is connected to its own CO and to the CO of the preceding residue. The HN(CA)CO experiment can be used to sequentially connect the amides while the HNCO where a given amide is connected only to its $CO_{(i-1)}$ is used to differentiate between CO and $CO_{(i-1)}$.

1.2.2.2 Backbone assignment of ^{15}N labeled unfolded proteins

^{15}N labeled proteins can be assigned using a combination of 3D-NOESY-HSQC and 3D-TOCSY-HSQC (Marion et al., 1989a; Marion et al., 1989b) spectra based on distances of protons through space. Signals of all protons in a spin system, here one amino acid residue (Figure 5A), give rise to a signal in the TOCSY spectrum as drawn in Figure 5C. In a $^{1}H,^{1}H$-NOESY spectrum all protons within a 5Å sphere of a selected proton (in NOESY-HSQC spectra H^N protons as depicted in Figure 5B are chosen) give rise to a signal. Figure 5D shows a schematic drawing of strips of hypothetical NOESY-HSQC spectra of the depicted peptide in Figure 5A and B. Intra-residual and i, i+1 inter-residual cross-signals can be detected in the NOESY spectrum. Due to the flexibility of unfolded proteins, conformations are sampled where all side chain protons of the preceding residue (i-1) are nearer to the H^N proton of residue i, than 5Å (different to the illustration in Figure 5B). These conformations contribute super-proportionally to the signal

observed in the spectrum due to the $1/r^6$ dependence of the NOE (section 1.2.3.3). Signals arising from its own amino acid can be identified by comparison with a TOCSY spectrum (Figure 5C). In addition, TOCSY spectra of the different amino acid types are very characteristic; the identification of a specific amino acid is often straightforward. The disadvantage of the approach has its origin in the small chemical shift dispersion observed in unfolded proteins. Unambiguous assignment is often not possible due to severe peak overlap.
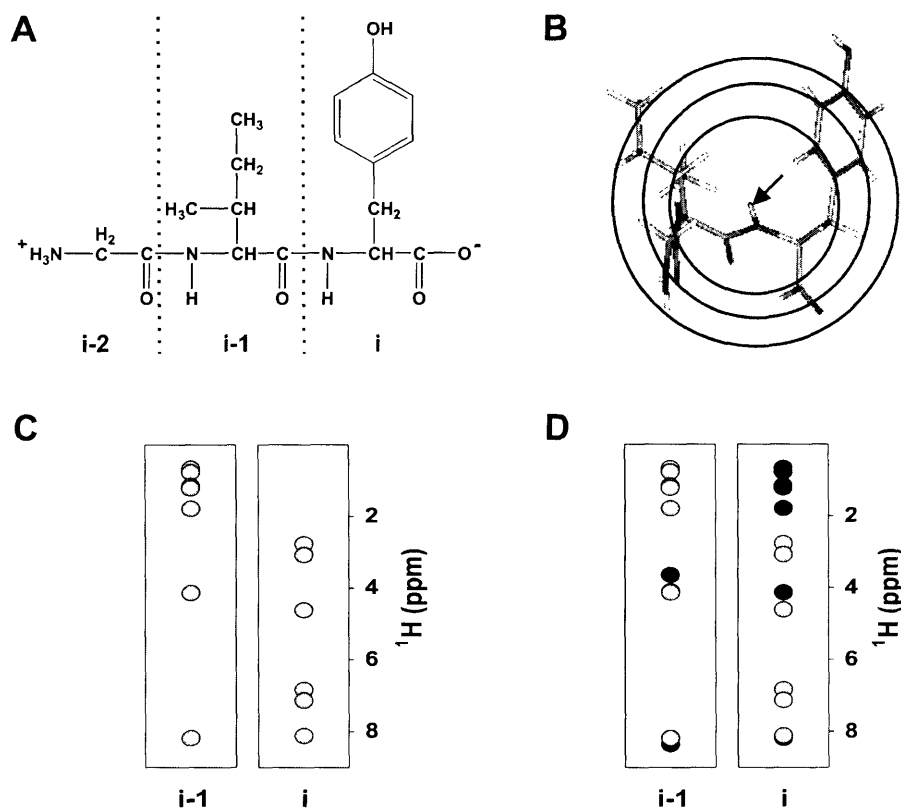
Figure 5: A) Chemical structure of the tripeptide: Gly-Ile-Tyr; Tyr is labeled as residue i, Ile is labeled as residue i-1 and Gly is labeled as i-2. Dashed lines separate the residues. B) Structure of the same tripeptide in the α-helical conformation, the H^N of the Tyr (residue i) is indicated by an arrow, circles indicate 3Å, 4Å and 5Å spheres around this H^N. C) Schematic drawing of so-called strip plots of a TOCSY spectrum of residues i (Tyr) and i-1 (Ile): signals arise from all protons of the respective residue. D) Schematic drawing of strip plots of a NOESY spectrum of residues i (Tyr) and i-1 (Ile). All protons within a 5Å sphere around the H^N of the respective residue give rise to a signal: signals filled in gray in the drawing arise from the own residue, while black signals in the drawing indicate signal arising from the preceding (i-1) residue. Color coding does not correspond to signs in the spectrum.

1.2.3 Structural information derived from NMR Parameters

The most important NMR parameters are summarized in Table 1, and are discussed in the subsections of this section with the main focus on the parameters used in this thesis. All parameters are discussed with an emphasis on the interpretation in non-native states and particularly the unfolded state of proteins.

NMR Parameter	Conformational dependence
chemical shift δ (ppm)	multiple torsion angles: ϕ, φ, ω, χ_1
^{n}J couplings (Hz)	single torsion angles via Karplus equations
homonuclear NOEs (a.u.)	distances, dependence on correlation time and motional properties
heteronuclear relaxation (Hz)	motional properties, dependence on τ_c, S^2, τ_e
residual dipolar couplings RDC (Hz)	overall shape, dynamics, S
H/D exchange (s^{-1})	exchangeable H^N
diffusion (m^2/s)	radius of hydration (R$_h$)
photo-CIDNP	accessible Trp, Tyr, His

Table 1: Overview of NMR parameters and their conformational dependence.

The interpretation of NMR parameters in unfolded states of proteins differs from the interpretation in the native state and requires specific models. Three different models are currently used for the prediction of properties in the unfolded state of proteins and are presented here. The models provide a framework for the interpretation of experimental data for non-native states of proteins.

Random coil model: The random coil model is based on the assumption that the random coil state of a protein is best described as a polymer consisting of twenty different monomers, the amino acids. The polymer possesses no structure except that inherent to a small unit. The unit can be defined by a single amino acid, thus local conformational preferences would be sampled according the low energy ϕ, ψ conformations in Ramachandran space (Fiebig et al., 1996; Schwalbe et al., 1997; Smith et al., 1996a; Smith et al., 1996b). On the other hand the unit can be defined by two amino acids (i and i-1) as suggested by recent experimental data (Peti et al., 2001). The random coil model consists of ϕ, ψ populations derived from coil regions of 402 proteins in the protein data bank (pdb). These ϕ, ψ populations fall within the preferred regions of the Ramanchandran space. (Serrano, 1995; Smith et al., 1996a; Smith et al., 1996b; Swindells et al., 1995). Torsion angle dependent properties in unfolded states of proteins such as scalar (J)

coupling constants (section 1.2.3.2) can be predicted using the random coil model based on this distribution. The random coil model has been applied for the investigation of unfolded ubiquitin (chapter 3).

Random coil peptides: Some properties in unfolded states of proteins cannot be predicted using the above introduced random coil model, as torsion angle dependencies are not parameterized (e.g. chemical shifts, section 1.2.3.1). However, amino acid specific and structure depend variations of the properties are often observed. These properties are often predicted based on experimental data derived from measurements on so-called random coil peptides which are small and unstructured (Wishart et al., 1995a).

Polymer theory: Some properties of unfolded proteins are independent of the nature and conformation of the amino acid side chains, such as the radius of hydration (section 1.2.3.5), residual dipolar couplings (RDCs, section 1.2.3.8) or heteronuclear relaxation data (section 1.2.3.4). These properties are predicted treating the protein as a homopolymer based on polymer theory (Flory, 1953; Flory, 1969; Louhivuori et al., 2003; Schwalbe et al., 1997; Wilkins et al., 1999).

1.2.3.1 Chemical shifts

The value of the chemical shift (δ) is sensitive to the environment of the observed nuclei. This environment depends on the amino acid side chain and on the secondary structure. Unfolded proteins are ensembles of rapidly inter-converting conformers, whose average local structure is in between α-helical and β-sheet conformation. This is reflected by their resonances in the spectra: 1) Chemical shifts measured in unfolded proteins are in between chemical shifts expected in α-helical conformations (δ_α) and β-sheet conformations (δ_β). This leads to limited dispersion in their respective NMR spectra. 2) Only one sharp set of signals is observed for the ensemble of conformations, indicating that the exchange between the conformations in unfolded proteins ($k_{\alpha\beta}$) is fast compared to the chemical shift difference ($\Delta\delta_{\alpha,\beta}$): $\Delta\delta_{\alpha,\beta} < k_{\alpha,\beta}$.

Comparison of the experimental chemical shifts δ^{exp} with chemical shifts predicted for a completely unstructured polypeptide δ^{rc} ($\Delta\delta = \delta^{exp} - \delta^{rc}$) allows delineation of parts of the polypeptide chain with (some) residual structure. For the identification of residual secondary structure elements, four consecutive residues have to show significant deviations towards α-helical or β-sheet conformations. (Wishart and Sykes, 1994; Wishart et al., 1992).

Reference values for random coil chemical shifts (δ^{rc}) have been determined by measurements on so-called random coil peptides (Bundi and Wüthrich, 1979; Merutka et al., 1995; Schwarzinger et al., 2001; Wishart et al., 1995a). The most reliable methods for the detection of residual secondary structure use H^α and $^{13}C^\alpha$ chemical shifts since neighboring effects on ^{15}N, H^N and ^{13}CO chemical shifts are not completely understood (Yao et al., 1997). The identification of residual secondary structure in proteins by comparison to random coil chemical shifts is one of the most used and most reliable tools for the identification of residual structure (Blanco et al., 1998; Fiebig et al., 1996; Lietzow et al., 2002; Schwalbe et al., 1997; Shortle and Abeygunawardana, 1993; Wong et al., 1996), and is used in this thesis (chapter 3, 4 and 5).

1.2.3.2 Scalar coupling constants

Scalar coupling constants (^nJ), particularly vicinal (^3J) coupling constants, depend on the intervening torsion angle of the connecting spins and can therefore be used to determine protein conformations. Empirical Karplus relations (Karplus, 1963) parameterize the angular dependence of a given coupling constant. These Karplus equations are not single-valued but give up to four degenerate angles for a measured coupling constant. Table 2 lists coupling constants and their angular dependence.

Most coupling constants report on the angle ϕ, while only very few coupling constants are available for the determination of the torsion angle ψ. ψ however is directly determined by the secondary structure. Therefore, a new method for the determination of ψ based on $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants has been developed and is described in chapter 2 of this thesis (see also Table 2). Analysis of coupling constants in unfolded proteins is performed by comparison with predictions from the random coil model (section 1.2.3) (Fiebig et al., 1996; Schwalbe et al., 1997; Smith et al., 1996a; Smith et al., 1996b) or with experimental values of random coil peptides (Plaxco et al., 1997). $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ and $^3J(H^N,H^\alpha)$ coupling constants in unfolded ubiquitin have been determined in chapter 3 of this thesis.

Coupling constant	Karplus parameterization	reference
$^3J(H^N,H_\alpha)$	$^3J = 6.51 \cos^2(\phi\text{-}60°) - 1.76 \cos(\phi\text{-}60°) + 1.60$	(Vuister et al., 1993)
	$^3J = 6.40 \cos^2(\phi\text{-}60°) - 1.40 \cos(\phi\text{-}60°) + 1.90$	(Pardi et al., 1984)
	$^3J = 6.60 \cos^2(\phi\text{-}60°) - 1.30 \cos(\phi\text{-}60°) + 1.50$	(Ludvigsen et al., 1991)
	$^3J = 7.90 \cos^2(\phi\text{-}60°) - 1.05 \cos(\phi\text{-}60°) + 0.65$	(Schmidt et al., 1999)
	$^3J = 6.64 \cos^2(\phi\text{-}60°) - 1.43 \cos(\phi\text{-}60°) + 1.86$	(Wang and Bax, 1996)
$^3J(H^N,C')$	$^3J = 4.01 \cos^2(\phi) + 1.09 \cos(\phi) + 0.07$	(Wang et al., 1995)
	$^3J = 4.02 \cos^2(\phi) + 1.12 \cos(\phi) + 0.07$	(Wang and Bax, 1996)
$^3J(H^N,C_\beta)$	$^3J = 4.70 \cos^2(\phi\text{+}60°) - 1.50 \cos(\phi\text{+}60°) - 0.20$	(Bystrov, 1976)
	$^3J = 2.78 \cos^2(\phi\text{+}60°) - 0.37 \cos(\phi\text{+}60°) + 0.03$	(Wang and Bax, 1996)
$^3J(H_\alpha,C')$	$^3J = 4.50 \cos^2(\phi\text{+}120°) - 1.30 \cos(\phi\text{+}120°) - 1.20$	(Bystrov, 1976)
	$^3J = 3.72 \cos^2(\phi\text{+}120°) - 1.71 \cos(\phi\text{+}120°) + 1.07$	(Löhr et al., 1997)
	$^3J = 3.62 \cos^2(\phi\text{-}60°) - 2.11 \cos(\phi\text{-}60°) + 1.29$	(Wang and Bax, 1996)
$^3J(C',C_\beta)$	$^3J = 1.61 \cos^2(\phi\text{-}120°) - 0.66 \cos(\phi\text{-}120°) + 0.26$	(Hu and Bax, 1997)
	$^3J = 1.28 \cos^2(\phi\text{-}120°) - 1.02 \cos(\phi\text{-}120°) + 0.30$	(Hu and Bax, 1998)
	$^3J = 2.54 \cos^2(\phi\text{-}120°) - 0.55 \cos(\phi\text{-}120°) + 0.37$	(Löhr et al., 1997)
$^3J(C',C')$	$^3J = 1.33 \cos^2(\phi) - 0.88 \cos(\phi) + 0.62$	(Hu and Bax, 1996)
	$^3J = 1.57 \cos^2(\phi) - 1.07 \cos(\phi) + 0.49$	(Hu and Bax, 1996)
$^1J(N_i,C_{\alpha i})$	$^1J = 9.51 - 0.98 \cos(\psi_i) + 1.70 \cos^2(\psi_i)$	(Wirmer and Schwalbe, 2002)
	$^1J = 8.65 - 1.21 \cos(\psi_i) + 2.85 \cos^2(\psi_i)$	(Ding and Gronenborn, 2004)
$^2J(N_i,C_{\alpha(i-1)})$	$^2J = 7.82 - 0.17 \cos(\phi_{(i-1)}) - 0.64 \cos^2(\phi_{(i-1)}) -$ $- 1.39 \cos(\psi_{(i-1)}) - 0.37 \cos^2(\psi_{(i-1)})$	(Wirmer and Schwalbe, 2002)
	$^2J = 7.85 - 1.52 \cos(\psi_{(i-1)}) - 0.66 \cos^2(\psi_{(i-1)})$	(Ding and Gronenborn, 2004)

Table 2: Overview of Karplus parameterizations for different ^1J, ^2J and ^3J coupling constants.

1.2.3.3 Homonuclear NOEs

NOE interactions between protons form the basis of the NMR structure determination in folded proteins. Distance information can be obtained from through space mediated transfers in NOESY experiments that reveal cross peaks between two protons (H_1 and H_2) closer than 5Å. The cross-relaxation rate between two protons that gives rise to observable cross peaks in NOESY spectra is defined by the following equation:

$$1) \qquad \sigma^{NOE}_{H_1,H_2} = \frac{d^2\tau_c}{5}\left[-1 + \frac{6}{1+4\omega_0^2\tau_c^2}\right]$$

Where ω_H is the Lamor frequency of ^1H, d describes the constant for dipolar interaction, with $d = \mu_0 h \gamma_H^2 /(\sqrt{8}r_{H1,H2}^3 \pi^2)$, μ_0 is the permeability of the vacuum, h is Planck's constant, γ_H is the

gyromagnetic ratio of ^{1}H and $r_{H1,H2}$ is the distance between H_1 and H_2. The NOE intensity depends to the 6th inverse power of the distance between two protons ($\sigma \sim 1/r_{1,2}^6$). Intensities of NOE cross peaks in non-native proteins in addition reflect an average of conformations with various distributions of ϕ, ψ and χ_1 angles and various proton-proton distances. Conformations with shorter proton-proton distances contribute more than proportional to the cross-peak intensity (section 1.2.2.2).

1.2.3.4 Heteronuclear relaxation rates

Backbone dynamics in proteins can be monitored by ^{15}N amide heteronuclear relaxation measurements. These heteronuclear relaxation rates depend on the overall rotational tumbling of the molecule (τ_c) and are sensitive to motions on a subnanosecond time scale and to slow conformational exchange in the millisecond time scale (Wagner, 1993). A typical set of ^{15}N relaxation data includes longitudinal (R_1) relaxation rates, transverse (R_2) or rotating frame ($R_{1\rho}$) relaxation rates and the heteronuclear NOE (hetNOE):

2) $$R_1 = \frac{d^2}{4} \left[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N) \right] + c^2 J(\omega_N)$$

3) $$R_2 = \frac{d^2}{8} \left[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N) \right]$$
$$+ \frac{c^2}{6} \left[4J(0) + 3J(\omega_N) \right] + R_{ex}$$

4) $$hetNOE = 1 + \left(d^2 / 4R_1 \right) \left(\gamma_N / \gamma_H \right) \left[6J(\omega_H + \omega_N) - J(\omega_H + \omega_N) \right]$$

In these equations, ω_H and ω_N are the Lamor frequencies of ^{1}H and ^{15}N, respectively. γ_i is the gyromagnetic ratio of nucleus i, $J(\omega)$ is the spectral density function at frequency ω. R_{ex} is the parameter accounting for conformational exchange contributions, d and c describe the dipolar and the chemical shift anisotropy (CSA) interaction, respectively ($d = \mu_0 h \gamma_N \gamma_H / (8 r_{NH}^3 \pi^2)$ and $c = \omega_N \Delta\sigma_N / \sqrt{3}$), μ_0 is the permeability of the vacuum, h is Planck's constant, r_{NH} is the internuclear distance between N and H, $\Delta\sigma_N$ is the nitrogen CSA.

The dependence of the heteronuclear relaxation parameters on the overall rotational correlation time τ_c in the absence of internal motions and assuming isotropic tumbling is shown in Figure 6. While the longitudinal relaxation rate R_1 exhibits a maximum, R_2 increases monotonously with

τ_c. This difference arises from the dependence of R_2 but not of R_1 on a J(0) term (equations 2, 3 and 4). J(0) is negligible at lower τ_c values but becomes more and more relaxation relevant for higher τ_c values. For macromolecules, τ_c is usually larger than 2ns. Consequently, small R_1 but large R_2 values are expected.
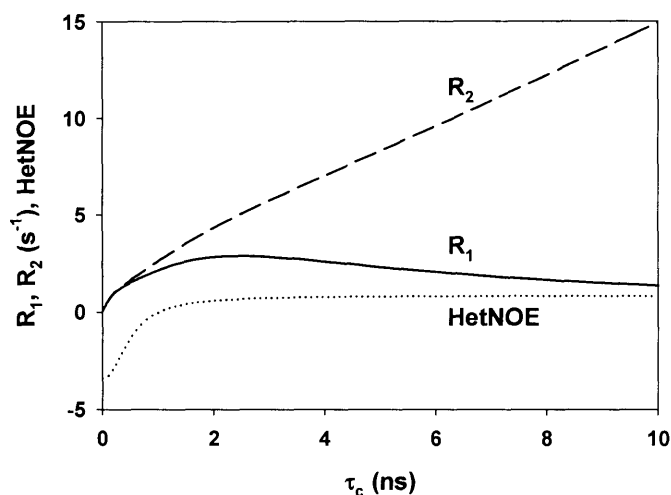
Figure 6: Dependence of ^{15}N R_1 (solid line), R_2 (dashed line) and hetNOE (dotted line) on the rotational correlation time (τ_c). Calculations were done based on the assumption of a rigid molecule tumbling isotropically.

A more detailed analysis of relaxation rates in folded states of proteins is based on the model-free formalism (Lipari and Szabo, 1982a; Lipari and Szabo, 1982b). This formalism assumes that the overall motion of a protein (τ_c) and internal motions of amides are on different time scales. This allows the independent analysis of the internal motions at a known τ_c. Internal motions of a given amide are hereby described by an order parameter (S^2) depicting the amplitude of the motion and - depending on the type of motion - a time scale of the motion (τ_e) or an exchange term (R_{ex}) or both. Figure 7 shows the dependence of R_2 relaxation rates on the internal motions as calculated using the model-free approach. The measured R_2 are smaller for less rigid states. This generalized assumption is also true for relaxation rates in the unfolded state. However, unfolded states are not characterized by one overall motion and the separation of internal and overall motions does not apply. Consequently the model-free approach cannot be used for the analysis of these states.
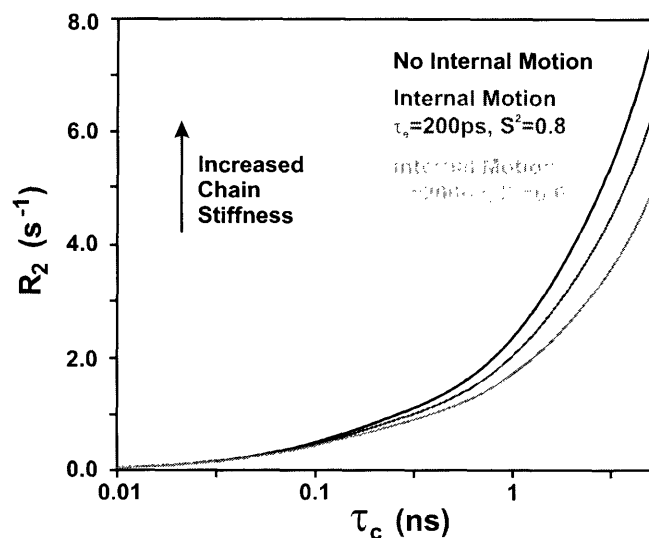
Figure 7: ^{15}N heteronuclear R_2 relaxation rates as function of the correlation time τ_c depending on internal motions as calculated using the Lipari-Szabo approach (Lipari and Szabo, 1982a; Lipari and Szabo, 1982b). Internal motions are characterized by an internal correlation time τ_e and the order parameter S^2.

R$_2$ relaxation rates in unbranched unfolded proteins (without disulfide bridges): A description of unfolded proteins without disulfide bridges (subsequently called unbranched) can be performed using the segmental motion model (Schwalbe et al., 1997). This model is based on polymer theory and assumes that the relaxation properties of a given amide are only governed by segmental motions of parts of the polypeptide chain, independent from the overall tumbling of the polymer. Parts of the chain that are not covalently linked do not interact with each other and the nature of a neighboring residue has no influence on the relaxation properties. The mathematical expression of the segmental motion model is shown in Equation 5. The influence of any neighboring residue (j) decays exponentially as the distance (in number of peptide bonds) from a given residue (i) increases.

5)
$$R_2^{rc}(i) = R_{int} \sum_{j=1}^{N} e^{-\frac{|i-j|}{\lambda_0}}$$

Where R_{int} is the intrinsic relaxation rate (depends also on the temperature and viscosity of the solution), λ_0 is the persistence length of the polypeptide chain (in numbers of residues) and N is the total chain length of the polypeptide.

34

In agreement with the segmental motion model, relaxation rates in unfolded proteins (Frank et al., 1995; Logan et al., 1994; Schwarzinger et al., 2002)) and in other polymers (Allerhand and Hailstone, 1972), have lower values at the termini, while they approach a plateau at the middle of the sequence (Figure 8, Equation 5). In this thesis (chapter 3) it is shown that ubiquitin is very well described by the segmental motion model and is thus a very good model for an unfolded protein without residual structure.

However, careful analysis of relaxation data in other proteins (see chapter 4, (Klein-Seetharaman et al., 2002; Schwalbe et al., 1997; Wirmer et al., 2004)) reveals regions, where significant positive deviations from the segmental motion model appear. These deviations indicate the presence of residual structural elements in the unfolded state disrupting the pure segmental motions. The deviations are best described by Gaussian clusters as developed in this thesis (chapter 4) (Klein-Seetharaman et al., 2002; Wirmer et al., 2004). Thus, a model is proposed that includes two components, the segmental motion part (R_2^{rc}, Equation 5) plus a Gaussian term (R_2^{clus}), as shown in Equation 6.

$$6) \qquad R_2^{un}(i) = \underbrace{R_{int}\sum_{j=1}^{N}e^{-\frac{|i-j|}{\lambda_0}}}_{R_2^{rc}} + \underbrace{\sum_{cluster}R_{cluster}\,e^{-0.5\left(\frac{(i-x_{cluster})}{\lambda_{cluster}}\right)^2}}_{R_2^{clus}}$$

These Gaussian clusters are characterized by the position of the cluster in the protein as residue number ($x_{cluster}$), the width of the cluster ($\lambda_{cluster}$) and a distinct relaxation rate for each cluster ($R_{cluster}$). Identification of clusters of residual structure in the unfolded state can be done using the formula above. However, it is not possible to draw conclusions about the nature of the residual structure and its stabilization. Residual structure could be either stabilized by short-range interactions only or additionally by long-range interactions. A combination of R_2 with mutational studies as performed in this thesis (chapter 4), however, can be used to identify the nature of interactions in the unfolded state.
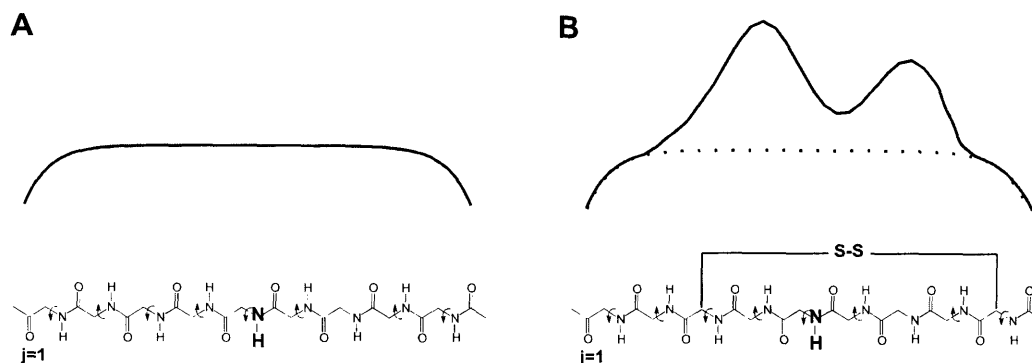
Figure 8: Relaxation profile expected for A) an unbranched chain (Equation 5) and B) for a branched polymer chain (Equation 7). The number of neighboring residues along the chain influences the relaxation property of a given residue.

R₂ relaxation rates in branched unfolded proteins (with disulfide bridges): The interpretation of R_2 rates in unfolded proteins with disulfide bridges (subsequently called branched) is more complicated. The relaxation properties of a given amide are not only influenced by the sequentially neighboring residues, but also by residues that are in proximity via disulfide bridges. Therefore a topological distance matrix, dm_{ij} counting the shortest path from residue i to residue j is used. In dm_{ij} a disulfide bond is treated in the same way as a peptide bond in connecting two residues. Furthermore, exchange processes are taking place around disulfide bridges. For a complete description of R_2 relaxation rates it is therefore necessary to include a term describing these exchange processes as shown in Equation 7 (Collins et al., submitted; Schwalbe et al., 1997).

$$7) \qquad R_2^{exp}(i) = R_{int} \sum_{j=1}^{N} e^{-\frac{dm_{ij}}{\lambda_0}} + R_{exch} \sum_{k=1}^{N_{cys}} e^{-\frac{|i-Cys_k|}{\lambda_2}}$$

1.2.3.5 Diffusion constants

Dimensions of spherical proteins can be determined using NMR diffusion measurements since the hydrodynamic radius (R_h) of a protein is inversely proportional to its diffusion constant (D). Knowing the diffusion constant, the R_h can be extracted using the Stokes-Einstein equation, requiring the knowledge of the exact diffusion constant, temperature and viscosity of the solution. An easier approach for the determination of diffusion constants is by comparison to an internal

probe with known hydrodynamic radius such as dioxan (R_h= 2.12Å) (Wilkins et al., 1999) according to Equation 8:

$$8) \quad R_h^{prot} = \frac{D_{diox}}{D_{prot}} R_h^{diox}$$

Measurement of diffusion constants is usually done by PFG (pulse field gradient) NMR experiments performed with a PG-SLED (pulse gradient simulated echo longitudinal encode-decode) sequence varying the strength of bipolar gradient pulses for diffusion (Wu et al., 1995a). The result of such an experiment is shown in Figure 9. The signals decay with increasing strength of gradients; the decay rate is higher for the small molecule dioxane (B) than for the protein (A). As the ensemble of conformers in the denatured state is a distribution of polypeptides that have different hydrodynamic radii, an average of the hydrodynamic radii present in a sample is obtained from the diffusion measurements. Thus, changes in this distribution are reflected by changes in the measured R_h as shown in chapter 4. The R_h of unfolded non-disulfide bridged proteins under highly denaturing conditions (high concentration of urea or guanidinium hydrochlorid; extreme pH) can also be correlated to the length of the polypeptide chain (Wilkins et al., 1999) by treating the protein as a homopolymer (Flory, 1953; Flory, 1969).
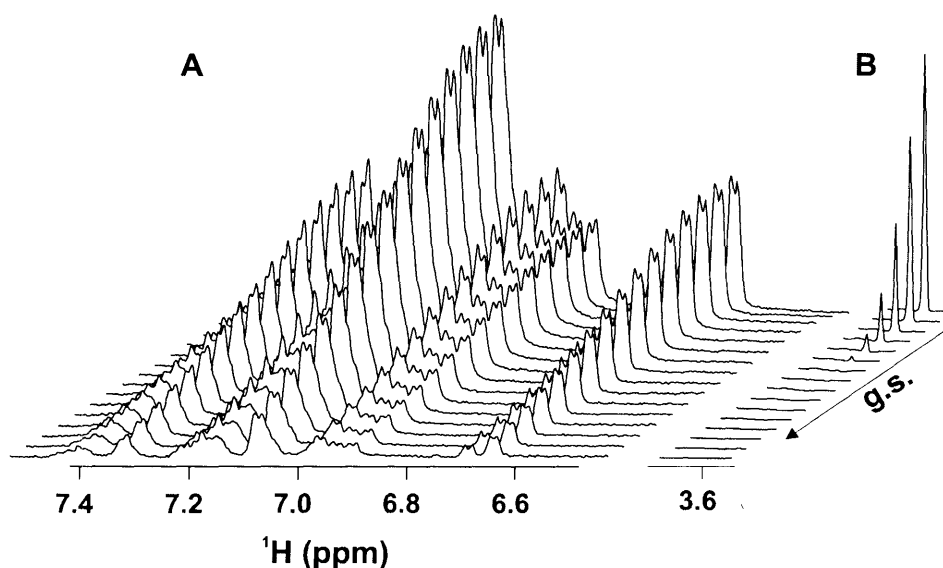
Figure 9: Example for a DOSY spectrum: Signal intensities in a ^{1}H 1D NMR spectrum vary as function of gradient strength (g.s.). A) lysozyme mutant W62G B) dioxan.

37

1.2.3.6 Photo-CIDNP

Accessible aromatic side chains in proteins can be monitored using the photo-CIDNP (Chemically Induced Dynamic Nuclear Polarization) NMR technique (Hore and Broadhurst, 1993; Kaptein, 1982). The photo-CIDNP NMR technique relies on the temporary interaction of a photo-excited dye – mostly a flavin such as FMN (Figure 10) - with an aromatic amino acid side chain (Trp, Tyr and His). Applying the photo-CIDNP technique to a protein results in signal enhancement as well as a better resolution in the aromatic region of the NMR spectrum (Figure 10).
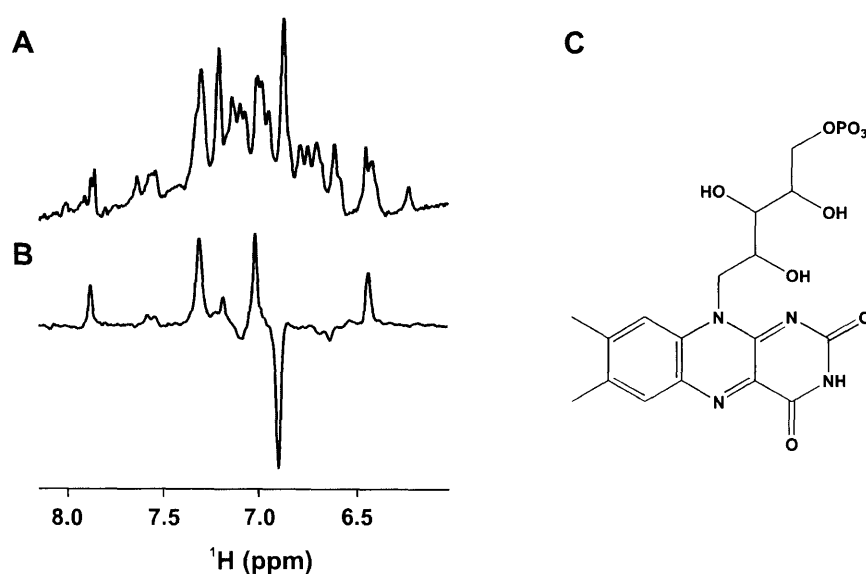
Figure 10: Photo-CIDNP of proteins. A) Aromatic region of a conventional NMR spectrum of folded bovine α-lactalbumin, number of scans (NS)=128 (see also Chapter 7.3.4). B) Photo-CIDNP NMR spectrum of folded bovine α-lactalbumin, NS=1. C) Chemical structure of flavine-mononucleotide (FMN).

Intensities of photo-CIDNP signals depend on the polarization efficiency of the amino acid (Trp > Tyr >> His >>> Met) and on the accessibility of the amino acid. Emissive photo-CIDNP signals are observed for Tyr residues, whereas the signals of Trp and His are absorptive. Therefore, the sign of a signal reports on the type of amino acid (for more information see (Kaptein, 1971)).

One particular advantage of photo-CIDNP NMR of unfolded states is the better resolution as only a few residues are polarized. Using an unlabeled protein, it is therefore often possible to assign

peaks to the respective amino acid side chain and furthermore to follow changes in accessibility in different unfolded states (Broadhurst et al., 1991). Furthermore, photo-CIDNP NMR is being used in kinetic investigations of protein folding to monitor folding intermediates (Hore et al., 1997; Maeda et al., 2000; Wirmer et al., 2001) as done in this work (chapter 7). Besides the enhanced resolution, better time resolution compared to normal NMR methods can be achieved due to the fact that the interscan delay depends only on the relaxation time of an electron (~10ms).

1.2.3.7 Paramagnetic spin labels

Paramagnetic spin labels in proteins can be used to test long-range order or contacts. Spin labels such as nitroxide cause a significant enhancement of relaxation rates of resonances that are within a distance of 15Å from the label (Kosen et al., 1986; Schmidt and Kuntz, 1984). Spin labels can be attached to a number of residues containing reactive side chains such as Cys (Hubbell et al., 1996), His (Schmidt and Kuntz, 1984) or Lys (Kosen et al., 1986) or to the N-terminus of the protein (Kosen et al., 1986).

In folded proteins, distances can be extracted from the differential line broadening of the protein with a diamagnetic spin label compared to the protein with a paramagnetic spin label (Kosen et al., 1986; Matthews, 1995; Schmidt and Kuntz, 1984). A qualitative picture is obtained in non-native states of protein, as they are composed of fluctuating conformations (Gillespie and Shortle, 1997a; Gillespie and Shortle, 1997b; Lietzow et al., 2002; Yi et al., 2000). Long-range interactions were identified e.g. in highly denatured myglobin (Lietzow et al., 2002), in staphylococcal nuclease (Gillespie and Shortle, 1997a; Gillespie and Shortle, 1997b) and in protein L (Yi et al., 2000). Introduction of new binding sites for paramagnetic ions (Franz et al., 2003; Nitz et al., 2003; Nitz et al., 2004; Wohnert et al., 2003) and tagging those to non-native states of proteins will allow detailed investigations using the paramagnetic effects.

1.2.3.8 Residual dipolar couplings

Residual dipolar couplings (RDCs) arise from dipolar interactions between two nuclei close in space. The size of these dipolar interactions depends on the angle (θ) between the vector connecting the two nuclei and the magnetic field by $\cos^2\theta$. Interactions are averaged out in solution due to rapid tumbling. However, the interactions can be reintroduced by partial alignment of a biomacromolecule. Either externally by liquid crystalline media (Gaemers and

Bax, 2001; Tjandra and Bax, 1997a), phages (Tolman et al., 1995) and strained polyacrylamide gels (Sass et al., 2000; Tycko et al., 2000), or internally by binding of paramagnetic ions to the protein (Banci et al., 1997; Barbieri et al., 2002) or to specifically designed binding sites (Wohnert et al., 2003). Theoretical calculations (Louhivuori et al., 2003) treating unfolded proteins as a random-flight polypeptide revealed that for long polypeptides, short segments of the chain are uncorrelated and all orientations in space are equally populated even under aligning conditions. Thus no dipolar couplings would be observed for such a long unfolded protein in a random coil state. Short unfolded proteins with chains up to 100-200 amino acids in contrast, will give rise to observable residual couplings, as shorter polypeptide chains are restrained by the covalent structure of the chain and thus each segment is coupled and not free.

1.2.3.9 H/D exchange

Hydrogen exchange methods can give insight into protein structure and dynamics by looking at the exchange of amide protons with solvent protons. In general, protons that are involved in stable hydrogen bonds or are buried within the protein core are protected from exchange with the solvent. If either the buffer contains D_2O or the protein has been deuterated, the exchange of protons for deuterons or vice versa can be monitored by NMR. Hydrogen exchange techniques are valuable for the investigation of protein folding kinetics (Baldwin, 1993; Miranker et al., 1991; Radford et al., 1992a; Roder and Wuthrich, 1986), folding intermediates and unfolded states (Buck et al., 1994; Englander and Mayne, 1992; Robertson and Baldwin, 1991; Roder et al., 1985). They provide information on the degree and dynamics of secondary structure formation depending on the formation of stable hydrogen bonds.

1.3 Background: Characterization of folding kinetics using NMR spectroscopy

Kinetics of protein folding can be monitored using a large range of biophysical techniques such as time-resolved stopped-flow CD and fluorescence spectroscopy as well as H/D exchange methods in combination with NMR or ESI-MS (Dobson et al., 1994; Kim and Baldwin, 1982; Kim and Baldwin, 1990). While most of these techniques provide a global picture of the folding process, NMR in contrast is capable of investigating proteins at atomic resolution. Slow folding kinetics (>hours, minutes), e.g. kinetics that involve prolyl-isomerization (Blum et al., 1978; Kautz and Fox, 1993; Koide et al., 1993; Wishart et al., 1993) can be initiated outside the NMR spectrometer, and kinetics can be monitored by subsequent collection of 1D NMR or (fast) 2D NMR spectra. However, most kinetics are fast, and therefore new techniques for folding initiation have been introduced recently (Balbach et al., 1996; Balbach et al., 1995; Frieden et al., 1993; Hoeltzli and Frieden, 1995; Kuhn and Schwalbe, 2000). (i) Folding can be initiated by rapid injection of unfolded protein into folding buffer within the NMR tube in the spectrometer (Balbach et al., 1995; Frieden et al., 1993; Hoeltzli and Frieden, 1995; Maeda et al., 2000) or within a special mixing probe(Spraul et al., 1996). (ii) Alternatively, folding can be initiated photochemically by laser irradiation within the NMR spectrometer: This method can either be applied to proteins that carry a photosensitive group (Derix et al., 2003), or folding can be initiated by release of cofactors from photo-labile chelators (Kuhn and Schwalbe, 2000) or from photo-labile precursors (Wenter et al., submitted). An example for photochemical folding initiation is described in chapter 7 of this work. There, α-lactalbumin refolding is initiated by Ca^{2+} release from the photo-labile EDTA derivative DM-Nitrophen.

41

1.4 Model systems

Unfolded states and kinetics of protein folding have been investigated in this thesis using ubiquitin, lysozyme and α-lactalbumin. The properties of the three proteins are shortly summarized in the following sections. For an insight into the emerging field of unfolded proteins, Table 3 summarizes a fraction of unfolded proteins that have been investigated using heteronuclear NMR spectroscopy.

Protein unfolded by denaturant	Citation	Natively unfolded protein	Citation
α-lactalbumin	(Ramboarina and Redfield, 2003)	Anti-sigma factor FlgM	(Daughdrill et al., 1998)
Apomyoglobin	(Schwarzinger et al., 2002)	Antitermination protein N	(Mogridge et al., 1998)
Barnase	(Wong et al., 2000)	Colicin translocation domain	(Collins et al., 2002)
beta(2)-microglobulin	(Katou et al., 2002)	Dessication-related protein	(Lisse et al., 1996)
acyl-coenzyme a binding protein	(Lindorff-Larsen et al., 2004)	Engrailed homeodomain	(Mayor et al., 2003)
bovine β-lactoglobulin	(Kuwata et al., 2001)	extracellular domain of beta-dystroglycan	(Bozzi et al., 2003)
chemotactic protein CheY	(Garcia et al., 2001)	IA3 (aspartic proteinase inhibitor)	(Green et al., 2004)
drkN SH3 domain	(Crowhurst and Forman-Kay, 2003)	Negative factor, NEF protein	(Geyer et al., 1999)
FK506 binding protein	(Logan et al., 1994)	N-terminal domain of p53	(Zeev-Ben-Mordehai et al., 2003)
glutaredoxin 3	(Nordstrand et al., 1996)	Prion protein N-terminal part	(Liu et al., 2000)
Hen lysozyme	(Wirmer et al., 2004)	Pro peptide of subtilisin	(Buevich et al., 2001)
HIV-1 protease tethered dimmer	(Bhavesh et al., 2003)	Prothymosin α	(Gast et al., 1995)
Protein L	(Yi et al., 2000)	Snc1, cytoplasmic domain	(Fiebig et al., 1999)
Reduced high-potential iron-sulfur protein	(Bentrop et al., 1999)	Staphylococcal nuclease, Δ131Δ fragment	(Choy et al., 2003)
Ubiquitin	(Mizushima et al., 2004)	β-Tubulin,394-445 fragment	(Jimenez et al., 1999)

Table 3: Denatured proteins and natively unfolded proteins investigated by heteronuclear NMR spectroscopy. Only one recent publication is shown per protein. A more complete list of refenrences, also containing more proteins can be found in (Wirmer et al., 2005).

1.4.1 Ubiquitin

Ubiquitin is a very small protein (8.6kDa) consisting of 76 amino acid residues, without a prosthetic group or any disulfide bridges. It is present in all eukaryotic systems and a very important marker for protein degradation. Covalent binding of mostly more than one ubiquitin with its C-terminal glycine to lysine residues of other proteins labels proteins for degradation (Rechensteiner, 1988). The NMR structure of native ubiquitin is shown in Figure 11. The structure is composed of a five stranded β-sheet, an α-helix and two 3^{10} helices.
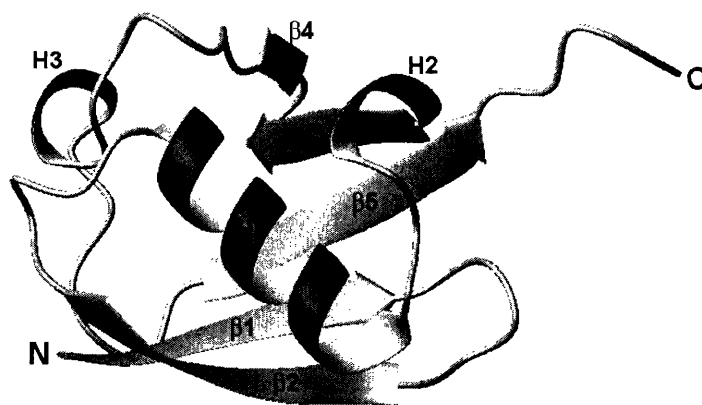
Figure 11: NMR structure of human ubiquitin (first conformer out of 10 deposited is shown). The picture was prepared using the program MOLMOL (Konradi et al., 1996) from the pdb file 1D3Z (Cornilescu et al., 1998). N and C termini are labeled as well as beta strands (β1-β5), the main α-helix (H1) and the two 3^{10} helices (H2 and H3).

Ubiquitin is very stable: The protein is folded at temperatures up to 80°C and between pH 1.2 to pH 13 (Jenson et al., 1980; Lenkinski et al., 1977; Nash and Jonas, 1997b). The so-called A-state (acid state) of ubiquitin is formed in the presence of 60% methanol at pH 2 (Brutscher et al., 1997; Wilkinson and Mayer, 1986). NMR spectroscopic investigations revealed that while the N-terminal part comprising β1, β2 and H1 (Figure 11) of the protein is conserved in the A-state, the structure of the C-terminal part becomes more flexible and changes dramatically. Chemical shift analysis revealed the conversion of the three β-strands β3-β5 into an α-helical conformation (Brutscher et al., 1997).

43

Unfolding of ubiquitin can be achieved in 8M urea at pH 2. All amide resonances have been assigned and comparison with the random coil model (Fiebig et al., 1996; Schwalbe et al., 1997; Smith et al., 1996a; Smith et al., 1996b) revealed a good correlation with the predicted parameters (Peti et al., 2000; Peti et al., 2001). In chapter 3 of this thesis additional investigations on the ensemble of conformers present in the unfolded state of ubiquitin have been performed using coupling constants (section 1.2.3.2) and relaxation data (see section 1.2.3.4). The results indicate that ubiquitin is a good model for a protein without any residual structure.

1.4.2 α-lactalbumin and lysozyme

1.4.2.1 The native states of α-lactalbumin and lysozyme

Hen lysozyme (HEWL) and α-lactalbumin (α-LA) are small model proteins whose native states superimpose very well (Figure 12), yet the proteins have entirely different functions. The Ca^{2+}-binding α-LA ($K_D=2*10^{-7}$ M at 37 °C) is a subunit of the lactose synthase (Hill and Brew, 1975; Permyakov and Berliner, 2000) which catalyzes the final step in lactose biosynthesis (formation of a β(1→4) glucopyranosyl linkage). Lysozyme is a functioning enzyme by itself. It destroys cell walls by cleavage of a β(1→4) glucopyranosyl linkage (McKenzie and White, 1991). The native structures of the proteins are extremely well described. The lysozyme crystal structure was the first structure of an enzyme solved (Blake et al., 1962; Blake et al., 1965) and the structures of α-LA (here bovine (BLA) and human α-LA (HLA) are included) have been solved alone as well as in complex with other proteins (Chandra et al., 1998; Harata et al., 1999; Pike et al., 1996; Ramakrishnan and Qasba, 2001; Ramakrishnan et al., 2001). Lysozyme is 129 residues long and has a pI of 9.3. α-LA is an acidic protein (pI=4.7) consisting of 124 residues. Both proteins consist of two domains, the α-domain spans from residue 1 to 35 and from residue 85 to 129 in lysozyme (residues 1 to 39 and residues 82-123 in α-LA), the β-domain involves residues 36 to 84 in lysozyme (residues 40 to 81in α-LA). Four disulfide bridges are present in the native structure; two within the α-domain, one in the β-domain and one connecting the two domains (Figure 12).

Another difference between α-lactalbumin and lysozyme is the propensity for amyloid formation. Variants of human lysozyme are known that cause familial amyloidosis due to a I56T or D67H

mutation (Pepys et al., 1993). In contrast, amyoloidogenic variants of α-lactalbumin that are implicated in any amyloidogenic disease are not yet known.
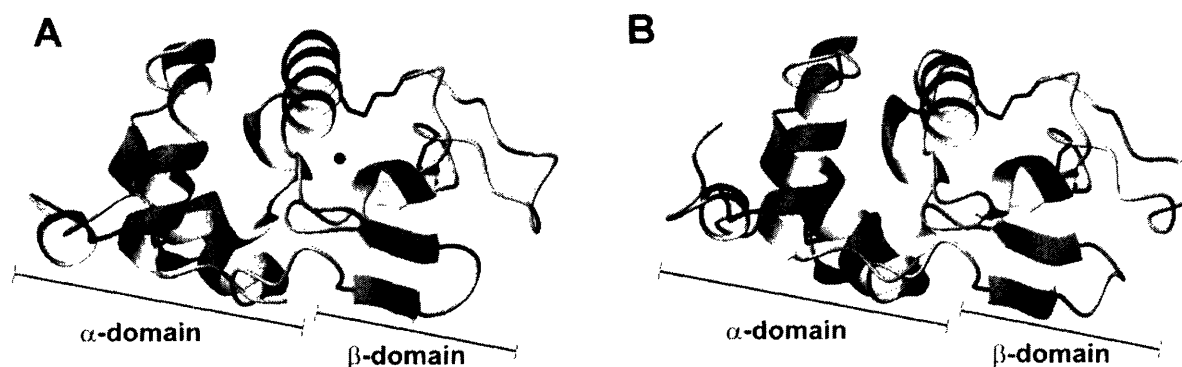
Figure 12: Structures of α–LA (A) and HEWL (B). Structures were prepared using MOLMOL (Konradi et al., 1996) using the pdb-files 1HFZ (Pike et al., 1996) and 193L (Vaney et al., 1996) for bovine α-LA and HEWL, respectively. Disulfide bridges and Ca^{2+}-ion (in α-LA) are indicated.

1.4.2.2 Non-native states of α-lactalbumin

α-lactalbumin has remarkable folding characteristics. In addition to the unfolded state, a second non-native state, the so-called molten globule (MG), can be stabilized and studied under acidic conditions (pH 2) (Dolgikh et al., 1981). This MG is partially folded and highly dynamic. It is stable at pH 2 in the presence as well as in the absence of disulfide bridges (Redfield et al., 1999), and has been proposed to resemble a folding intermediate. The MG state is a highly dynamic state where secondary structure (CD, H/D exchange (Kuwajima, 1996; Schulman et al., 1995)) is present (predominantly in the α-domain of the protein), while no tertiary interactions are formed. Compared to the ensemble of structures present in unfolded states of proteins, exchange processes between different conformations in the MG are slow. This is reflected by the absence of detectable NMR signals (Figure 3B). Residue specific investigation of the MG is thus not possible by direct detection of NMR signals. However, appearance of sharp NMR signals during thermal, pressure or urea induced unfolding is possible (Lassalle et al., 2003; Ramboarina and Redfield, 2003; Redfield et al., 1999; Schulman et al., 1997; Wijesinha-Bettoni et al., 2001). Residues located in the β-domain of the protein unfold first, before the stable core formed by segments from the N- and the C- terminal unfolds.

Complete unfolding of oxidized α-LA judged by the appearance of all backbone resonances is only possible in 8M GdnHCl at 50°C (Schulman et al., 1997; Wijesinha-Bettoni et al., 2001). The

45

reduced protein (absence of disulfide bridges) unfolds in the presence of 8M urea as shown in this work (chapter 6) for bovine α-lactalbumin (BLA) and in the literature for human α-lactalbumin (HLA) (Redfield et al., 1999). Structural preferences of these completely unfolded states of α-LA are investigated in chapter 6 of this thesis.

1.4.2.3 Refolding of α-lactalbumin with intact disulfide bridges

The refolding of α-LA is of particular interest, as it has been proposed that the MG state of α-LA resembles a folding intermediate. Indeed, previous kinetic experiments using time-resolved NMR and time-resolved photo-CIDNP in the groups of Dobson and Hore revealed a close similarity between the MG and a folding intermediate. This folding intermediate is populated after a rapid collapse in the dead time of kinetic experiments during the slow (> 120 s) refolding of bovine α-lactalbumin (BLA) by tenfold dilution from 6M guanidinium hydrochloride (GdnHCl) (Forge et al., 1999; Maeda et al., 2000). The folding intermediate possesses a high degree of secondary structure in the α–domain of the protein, while it lacks distinct tertiary interactions. The second folding phase, following the initial collapse is highly cooperative; there are no signs for the population of an additional distinct folding intermediate. This folding pathway is significantly accelerated in the presence of Ca^{2+}-ions (Forge et al., 1999), indicating that Ca^{2+}-ions stabilize native contacts in partially folded conformations and hereby reduce energy barriers during the second folding phase.

A different approach for the study of refolding of BLA relies on the stabilizing effect of Ca^{2+}-ions against denaturants such as GdnHCl and urea, high temperature or pressure (Dolgikh et al., 1981). Fast folding at constant urea concentration can be induced by the addition of Ca^{2+}-ions (Kuhn and Schwalbe, 2000). Folding proceeds via an intermediate with rate constants of $k_1=11.4s^{-1}$, $k_2=1.3s^{-1}$ and $k_1=8.9s^{-1}$, $k_2=1.0s^{-1}$, respectively. Refolding of methyl-groups monitored by 1D proton NMR in contrast follows slower single-exponential kinetics ($0.16s^{-1} \leq k \leq 0.33s^{-1}$) (Kuhn and Schwalbe, 2000). The formation of secondary structure precedes the formation of the exact position of methyl groups. Here, the intermediate on the Ca^{2+} induced folding pathway of BLA has been investigated by time-resolved photo-CIDNP NMR (see chapter 7). The intermediate populated 200-300ms after initiation of refolding has no characteristics of a MG.

1.4.2.4 Oxidative refolding of α-lactalbumin

Two oxidative folding pathways are known in α-LA, depending on the presence or absence of Ca^{2+}-ions. In the absence of Ca^{2+}, yields of oxidative folding are relatively low, and it is not possible to identify distinct folding intermediates populated along the pathway. Instead scrambled disulfide intermediates with two, three and four non-native and native disulfide bridges are populated during the course of folding (Chang, 2004; Chang and Li, 2002). In contrast, folding proceeds via two well-defined intermediates containing only native disulfide bridges in the presence of Ca^{2+}-ions. In the first intermediate (II-A), the two native disulfide bridges involving the β-domain of α-LA are formed, in the 2nd intermediate (III-A) the [28-111] disulfide bridge is formed as well (Chang, 2004; Chang and Li, 2002). Interestingly, if folding is started from the isolated II-A in the absence of Ca^{2+}-ions, the native disulfide bridges unravel and folding proceeds via the scrambled intermediates as in the case starting from a completely unfolded state.
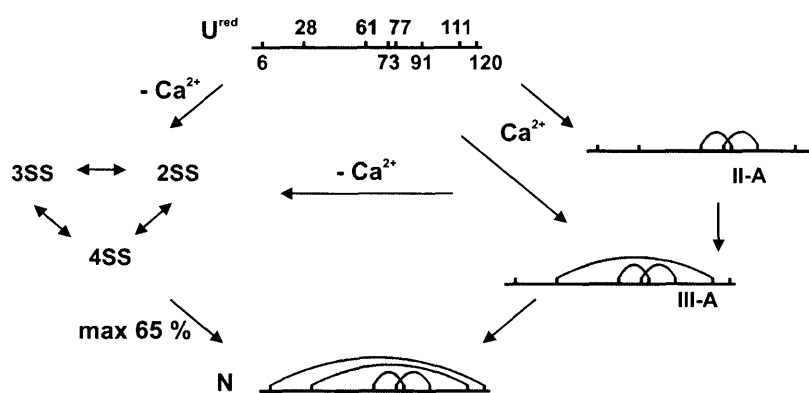
Figure 13: Oxidative folding pathway of α-LA. The protein chain is depicted as a straight horizontal line, cysteines as short vertical lines. The positions of the cysteines are labeled in the scheme depicting the reduced unfolded state (U^{red}). Scrambled disulfide intermediates are indicated as 2SS, 3SS and 4SS, according to the number of disulfides formed. The names of the distinct intermediates are according to the names given in the original publications. Gray shaded areas indicate the β-domain of the BLA, the rest of the chain forms the α-domain in folded BLA.

1.4.2.5 Non-native states of lysozyme

Unfolded HEWL is one of the best studied unfolded proteins. It has been studied under a range of denaturant conditions (trifluoroethanol (TFE), urea) in the presence and absence of disulfide bridges using NMR spectroscopy (Buck et al., 1993; Buck et al., 1994; Buck et al., 1995; Buck et al., 1996; Evans et al., 1991; Hennig et al., 1999; Klein-Seetharaman et al., 2002; Radford et al., 1992a; Schwalbe et al., 1997; Wirmer et al., 2004).

Secondary structure is present in oxidized (with disulfides) lysozyme in 70% TFE (pH 2) while extensive tertiary interactions are absent (Buck et al., 1993; Buck et al., 1995; Buck et al., 1996). Six helices are formed as demonstrated by H/D exchange, J coupling constants and chemical shift deviations. Only one of the six helices (helix 3), the least pronounced one, is found in the β-domain of the protein, all other helices are in the α-domain and coincide with helices in the native structure.

The urea denatured oxidized state (8M urea, pH 2) is significantly less structured. Small areas exhibiting residual helical secondary structure (Schwalbe et al., 1997) coincide with the areas identified in the TFE denatured state. However, the propensity for the helical secondary structure is drastically reduced and amides are not protected from exchange (Buck et al., 1994). Relaxation data of the urea denatured oxidized state show that motions are significantly restricted in this state (Schwalbe et al., 1997). Due to the contribution of disulfide exchange on one side and residual structure on the other side, delineation of relaxation data in the oxidized unfolded state is difficult.

This problem is circumvented in the disulfide reduced state of lysozyme. This state is unfolded in urea, as well as in water (see chapter 4) (Klein-Seetharaman et al., 2002; Lee and Atassi, 1973; Schwalbe et al., 1997; Wirmer et al., 2004). Residual secondary structure in this reduced state of lysozyme corresponds well with the residual structure found in the oxidized urea denatured state. Six hydrophobic clusters where motions are restricted due to residual interactions are present (Klein-Seetharaman et al., 2002; Schwalbe et al., 1997; Wirmer et al., 2004). These clusters correspond loosely to the areas identified in the TFE state. In this work the reduced state in water has been further investigated: A combination of mutations and relaxation data revealed the presence of long-range interactions between the six hydrophobic clusters (Klein-Seetharaman et al., 2002; Wirmer et al., 2004) (see chapter 4).

1.4.2.6 Refolding of hen lysozyme with intact disulfide bridges

Refolding kinetics by dilution of denaturant of oxidized lysozyme have been studied using a variety of biophysical techniques ranging from absorbance, CD, fluorescence, ANS binding and mass spectrometry to NMR (Kiefhaber, 1995; Wildegger and Kiefhaber, 1997) (results obtained before 1994 are reviewed in (Dobson et al., 1994)). Two main folding pathways exist during refolding of HEWL: a) Approximately 80% of the proteins fold via an intermediate populated after a hydrophobic collapse (<1ms). This intermediate slowly converts to the native state (>1.5s). The α-helical domain of HEWL is structured in this intermediate, while the β-domain is unstructured (Dobson et al., 1994). b) Approximately 20% of the proteins fold through a faster pathway with no observable intermediate (Kiefhaber, 1995; Wildegger and Kiefhaber, 1997).

1.4.2.7 Oxidative refolding of hen lysozyme

Oxidative refolding of HEWL proceeds via scrambled two disulfide species (2SS). The 2SS intermediates fold fast into three intermediates possessing each three native disulfide bridges (des-[6-127], des-[76-94] and des-[64-80], Figure14A) (van den Berg et al., 1999). One disulfide bridge ([30-115]) is present in all three intermediates. Formation of the last disulfide bridge is slow for des-[76-94] where the disulfide connecting the α- and the β-domain has to be formed, while it is fast for the other two intermediates.
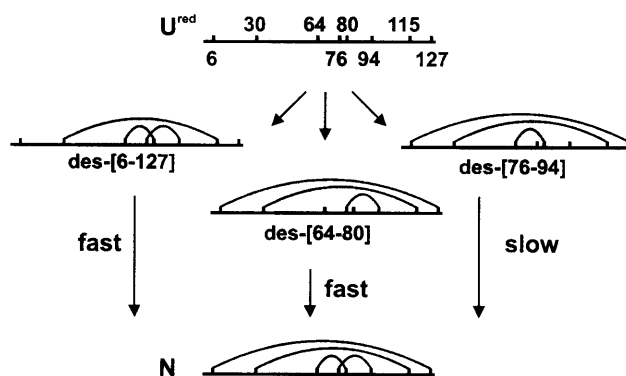
Figure14: Oxidative folding pathway of HEWL. The protein chain is depicted as a straight horizontal line, cysteins as short vertical lines. Positions of the cysteins are labeled in the scheme depicting the reduced unfolded state (U^{red}). The names of the distinct intermediates are according to the names given in the original publications. Gray shaded areas indicate the β-domain of HEWL, the rest of the chain forms the α-domain in folded HEWL

1.5 Outline and general discussion

1.5.1 Outline

Chapter 2 describes the development of a new pulse sequence for the measurement of $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants. The method is based on J-modulated HSQCs and the parameterization of the coupling constants using folded ubiquitin is performed. The value of the coupling constants depends on the protein backbone angle ψ (and ϕ).

$$R_2^{rc}(i) = R_{int} \sum_{j=1}^{N} e^{-\frac{|i-j|}{\lambda_0}}$$
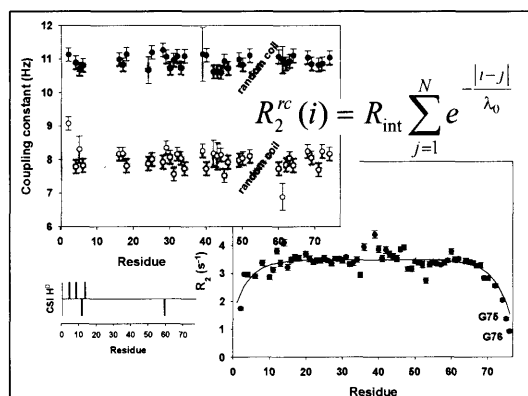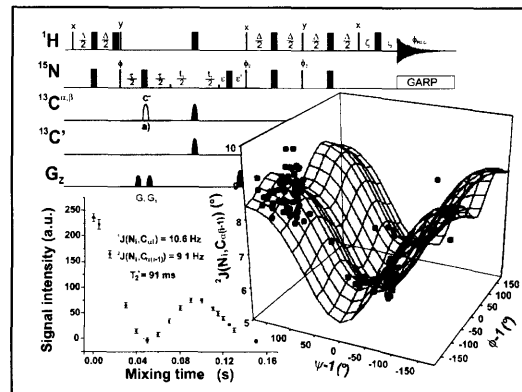
In **chapter 3** $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants are measured in unfolded ubiquitin using the pulse sequence developed in chapter 2. It is shown that ubiquitin is a good model protein for a protein without any residual structure based on chemical shift perturbations, scalar coupling constants and heteronuclear relaxation rates.

The disulfide reduced and methylated state of unfolded lysozyme has been investigated in **chapter 4**. A new method based on relaxation and diffusion rates and single point mutations is introduced for the identification of long-range interactions in non-native states of proteins. Non-native and native long-range interactions are present in the unfolded protein.

EQLTKCEVFRE
LKDLKGYGGVSLPE
WVCTTFHTSGYDTQ
AIVQNNDSTEYGLF
QINNKIWCKDDQNP
HSSNICNISCDKFL
DDDLTDDIMCVKKI
LDKVGINYWLAHKA
LCSEKLDQWLCEKL
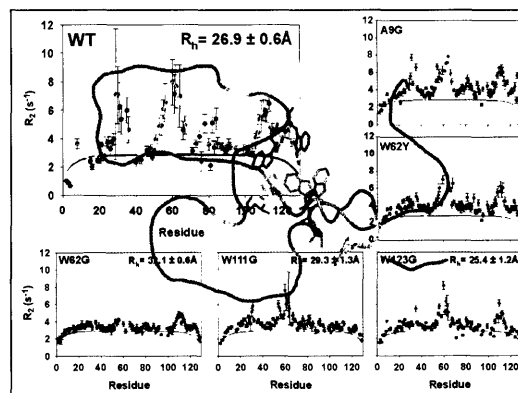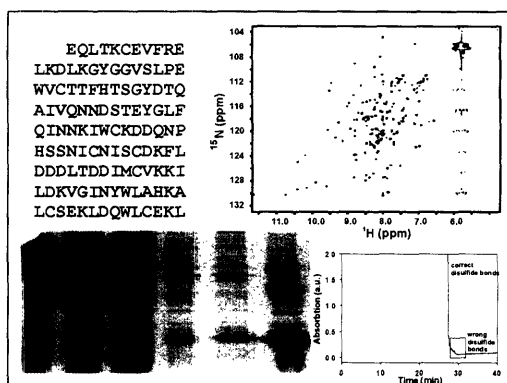
In **Chapter 5** the expression of bovine α-lactalbumin in *E.coli* using a His-tag and a trypsin cleavage site is shown. The heterologously expressed protein has the same sequence and stability as the wild-type protein isolated from cows' milk.

Chapter 6 describes the results of the assignment of unfolded $^{13}C,^{15}N$ labeled disulfide reduced and methylated bovine α-lactalbumin. A comparison of residual structure in the unfolded states of α-lactalbumin and lysozyme reveals that unfolded states of the structurally homologous proteins differ considerably.
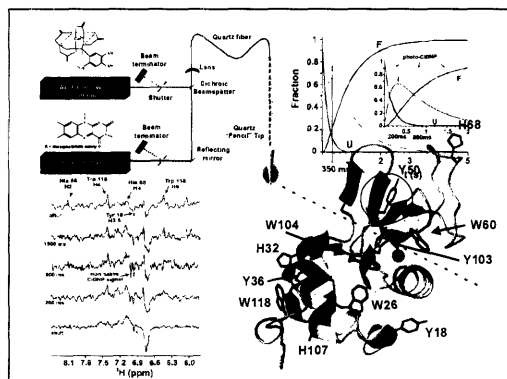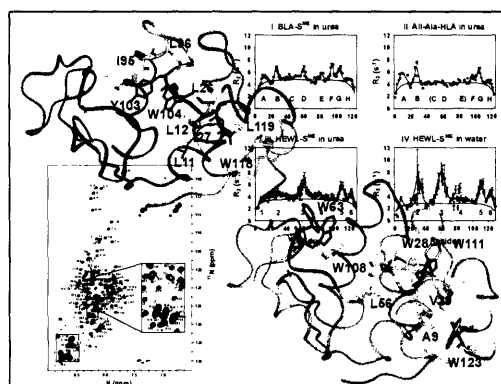
The last chapter, **chapter 7**, deals with the development and application of a new method for the investigation of laser induced folding kinetics using photo-CIDNP NMR on the Ca^{2+} induced folding pathway of bovine α-lactalbumin. A distinct folding intermediate possessing a tyrosine residue in non-native environment could be detected 200ms after initiation of folding.

1.5.2 General discussion of accomplishments

1.5.2.1 NMR studies to aspects of protein folding

Complementary to a variety of other biophysical techniques, NMR spectroscopy allows the residue specific and/or even atom specific investigation of rapidly inter-converting conformers. Local structural parameters such as residual secondary structure, averaged torsion angles, accessibility of aromatic side chains, motional restrictions and exchange protection of amides can be derived using chemical shifts, homonuclear NOEs, scalar coupling constants, photo-CIDNP, relaxation data and H/D exchange. In addition, global parameters such as the hydrodynamic radius (R_h) can be determined by NMR diffusion measurements. Both the local and the global parameters can be easily obtained from NMR measurements of non-modified isotope labeled unfolded proteins.

In contrast, the identification of non-local interactions is not as straight forward. Long-range interactions can be identified by differential line broadening in the presence of spin labels. This approach requires the attachment of a spin-label to one side chain within the protein, mostly requiring mutations. Mutations and the spin-label might in the worst case perturb residual long-range interactions in the unfolded state. Here a new approach based on R_2 relaxation data and non-conservative single point mutations has been developed (chapter 4). The approach relies on the identification of deviations from random coil relaxation rates. Areas that show large deviations are deliberately perturbed by non-conservative single point mutations. Long-range interactions are indicated by differences in R_2 relaxation rates far away in sequence from the mutation site.

Both approaches for the identification of long-range interactions in unfolded proteins require the mutation of the wild-type protein. Consequently the identification of long-range interactions is more complicated than the identification of local interactions. However, using the combination of NMR and mutations it is possible to identify these interactions and gain insight into unfolded states of proteins.

While NMR is the best method for the characterization of unfolded states of proteins, the applicability of NMR spectroscopic methods to kinetics of protein folding is limited. The acquisition of a single transient 1D NMR spectrum takes approximately 50-100ms, a necessary

recovery delay between experiments lasts at least 500ms. Therefore only kinetics of relatively slowly folding proteins (sec) can be detected. However, only NMR is capable to detect aspects of folding intermediates at atomic resolution. Here experiments were designed allowing for the characterization of a folding intermediate. The major challenge lies in the efficient fast initiation of the kinetic experiment by either fast mixing devices or photolabile compounds (chapter 7). A second challenge lies in the choice of NMR experiment for signal detection. The chosen experiment is ideally short with good signal to noise and reveals maximum structural information, such as accessibility, chemical shifts or torsion angles. In this thesis, a method has been developed for testing the accessibility of aromatic residues during protein folding and in intermediates as fast as 200ms after folding initiation. This could be achieved by a combination of laser induced folding initiation and photo-CIDNP signal detection (chapter 7).

1.5.2.2 Unfolded states of proteins

Unfolded states of proteins are ensembles of rapidly inter-converting conformers. Properties of a perfectly unfolded state can be predicted by the random coil model. The random coil model describes the protein as a polymer consisting of twenty different monomers, the amino acids. The polymer possesses no structure except that inherent to its different monomers (Fiebig et al., 1996; Schwalbe et al., 1997; Smith et al., 1996a; Smith et al., 1996b). Properties of unfolded ubiquitin (chapter 3) agree with predictions from the random coil model. This makes the protein a good model for a protein in the random coil state on the one hand and allows testing of properties expected for a random coil state on the other hand. Thus using ubiquitin it could be shown that relaxation properties in random coil states of proteins are independent from the amino-acid sequence, rather than depending on side chain volumes or hydrophobicities.

Ubiquitin is the exception rather than the rule. While the overall properties of unfolded proteins are very similar to the random coil state, residual secondary and even tertiary structure are observed in non-native proteins.

Residual secondary structure has not only been detected in lysozyme and α-lactalbumin (chapter 4 and chapter 6), but also in a number of other unfolded proteins such as apomyoglobin (Yao et al., 2001), barnase (Arcus et al., 1995), protein G (Frank et al., 1995; Sari et al., 2000), acyl-CoA-binding protein (ACBP) (Fieber et al., 2004), SH3 of drk (Blanco et al., 1998) and α-synuclein (Bussell and Eliezer, 2001). Interestingly, the observed residual secondary structure is

mostly α-helical and often coincides with hydrophobic residues. Furthermore, a correlation between residual α-helical structure in unfolded states of proteins (e.g. α-lactalbumin (chapter 6), hen lysozyme (chapter 4) and apomyoglobin (Yao et al., 2001)) and positions of secondary structure elements in the folded state of proteins has been observed, suggesting possible initial starting nuclei for protein folding.

In addition, residual long-range interactions are a feature observed in unfolded proteins. Long-range interactions could be identified in unfolded myoglobin (Lietzow et al., 2002), staphylococcal nuclease (Gillespie and Shortle, 1997a; Gillespie and Shortle, 1997b), protein L (Yi et al., 2000) and ACBP (Lindorff-Larsen et al., 2004) using spin labels. Here, long-range interactions have been detected in unfolded lysozyme using a combination of relaxation data, diffusion measurements and single point mutations (chapter 4).

1.5.2.3 Kinetics of protein folding

Kinetics of protein folding vary considerably for different proteins. Small proteins can be divided into two classes. Class I proteins populate folding intermediates during folding (U→I→F) and class II proteins are two-state folders (U→F) (Baldwin and Rose, 1999). α-lactalbumin belongs to class I. A folding intermediate is populated during the Ca^{2+} induced folding pathway at constant urea concentration (chapter 7) as well as during refolding by dilution from 6M GdnHCl (Forge et al., 1999; Maeda et al., 2000). The folding intermediates populated during the different refolding experiments differ from each other. While a distinct folding intermediate is observed on the Ca^{2+} induced pathway (chapter 7), the folding intermediate on the dilution induced pathway resembles a MG (Forge et al., 1999; Maeda et al., 2000). The comparison of the two folding kinetics reveals that different folding pathways lead to the native state, an observation also found for folding pathways of other proteins, such as lysozyme (Kiefhaber, 1995; Wildegger and Kiefhaber, 1997) or Im9 (Friel et al., 2004). Observation of the folding intermediate populated on the Ca^{2+} induced folding pathway reveals another interesting feature of protein folding. One part of the protein (the α-domain) is folded in the intermediate, while non-native structure is observed in the β-domain. This observation is in line with the hypothesis of hierarchical folding, assuming folding as a process where structures that are near in sequence and marginal in stability interact to produce intermediates ultimately resulting in the native structure (Baldwin and Rose, 1999).

1.5.2.4 Implications of residual structure in unfolded proteins for protein folding and misfolding

Unfolded states of proteins in general are the starting point of protein folding. Residual structure in unfolded states may be the nucleus for either productive folding or unproductive folding. The conformational space in unfolded proteins is a lot larger than the conformational space for folded proteins. Only $\approx$ 500 different folds have been classified in folded proteins. E.g. lysozyme and α-lactalbumin have very similar native structures and an identical fold in their native states, yet residual structure in their unfolded states differs considerably (chapter 6). The size of the conformational space in unfolded proteins becomes obvious in α-synuclein, a natively unfolded protein that forms amyloids in Parkinson's disease. Unfolded states of two early onset mutants (more amyloidogenic than the wild type) differ considerably from the wild type (Bussell and Eliezer, 2001), suggesting that the differences in the residual structure in the early onset mutants are nuclei for amyloid formation. Moreover, single point mutations considerably change the conformational ensemble of unfolded proteins, as observed for hen lysozyme (chapter 4) and ACBP (Fieber et al., 2004). In contrast, the native structure remains unchanged. This result suggests that unfolded states rather than folded states of proteins have to be investigated to gain insight into amyloid formation. Amyloidogenic mutants of human lysozyme possess identical native structures to the wild type, while mutations are found in a region that are presumably important for stabilization of long-range interactions in the unfolded state.

The importance of residual structure as nucleus for productive or unproductive folding is furthermore confirmed by comparison of the results on α-lactalbumin in this thesis (chapter 6 and 7). Residual structure in unfolded (reduced and methylated) α-lactalbumin was identified primarily in the α-helical domain of the protein (chapter 6). Kinetic refolding experiments of oxidized α-lactalbumin revealed the presence of a folding intermediate where native contacts are formed in the α-helical domain, while the β-domain of the protein is still in a non-native conformation (chapter 7). It is likely that residual structure in the α-domain guides folding, as the formation of native contacts in the α-domain of the protein precedes the formation of the native β-domain.

55

Chapter 2

Angular dependence of $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants

measured by J-modulated HSQCs

2 Angular dependence of $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants measured by J-modulated HSQCs

2.1 Introduction

The aim of this chapter was to develop a method for the measurement of $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants in proteins and to derive a parameterization for the torsion angle dependence of the respective coupling constants. Torsion angle restraints that report on local conformation and conformational equilibria are often determined by measurement of ^3J coupling constants (Bystrov, 1976). While there are plenty of ^3J coupling constants that report on the torsion angle ϕ, very few coupling constants report on the torsion angle ψ (see also Table 2 in chapter 1), for the following reasons. (i) Fewer NMR active nuclei are accessible around the torsion angle ψ (Figure 15). (ii) Values of the three vicinal coupling constants around ψ $^3J(N,N)$, $^3J(N,C^\beta)$ and $^3J(N,H^\alpha)$ are very small, due to the involvement of the ^{15}N nucleus (which has a low gyromagnetic ratio) so that differences in coupling constants due to angular dependencies are barely larger than the measurement error.
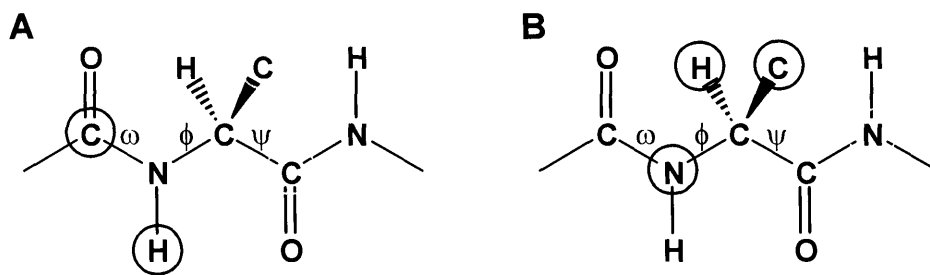
Figure 15: Definition of torsion angles in proteins. NMR active nuclei potentially involved in ^3J coupling constants around the torsion angle ϕ (A) and ψ (B) in a $^{13}C,^{15}N$ labeled protein.

The torsion angle ψ however, determines the protein backbone conformation, and is therefore the more interesting torsion angle to determine (Figure 16).

In contrast to ^3J coupling constants, the conformational dependence of ^1J (Juranic and Macura, 2001; Mierke et al., 1992) and ^2J coupling constants is less well characterized. Only few reports have investigated methods to measure $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants (Bystrov, 1976; Delaglio et al., 1991; Heikkinen et al., 2001; Permi and Annila, 2000; Permi et al., 2000).

This is surprising as a vast number of triple resonance experiments that are routinely used for the assignment of $^{13}C, ^{15}N$ labeled proteins, such as the HNCA experiment (Kay et al., 1990) rely on magnetization transfer via $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ couplings.

Bystrov reported values of $^1J(N_i,C_{\alpha i}) \sim 11Hz$ and $^2J(N_i,C_{\alpha(i-1)}) \sim 7Hz$ in peptides (Bystrov, 1976). Few (11) coupling constants in ubiquitin were published by the group of Kilpeläinen (Heikkinen et al., 2001). The most comprehensive study before the work presented here was done in the group of Bax (Delaglio et al., 1991): $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants were measured in staphylococcal nuclease, and their dependence on secondary structure was observed, however not parameterized. Subsequently, a torsion angle dependent parameterization for the coupling constants from *ab-initio* calculations was derived in the group of Markley (Edison et al., 1994a; Edison et al., 1994b). This parameterization is based on 25 parameter Fourier series, which may be over parameterized.
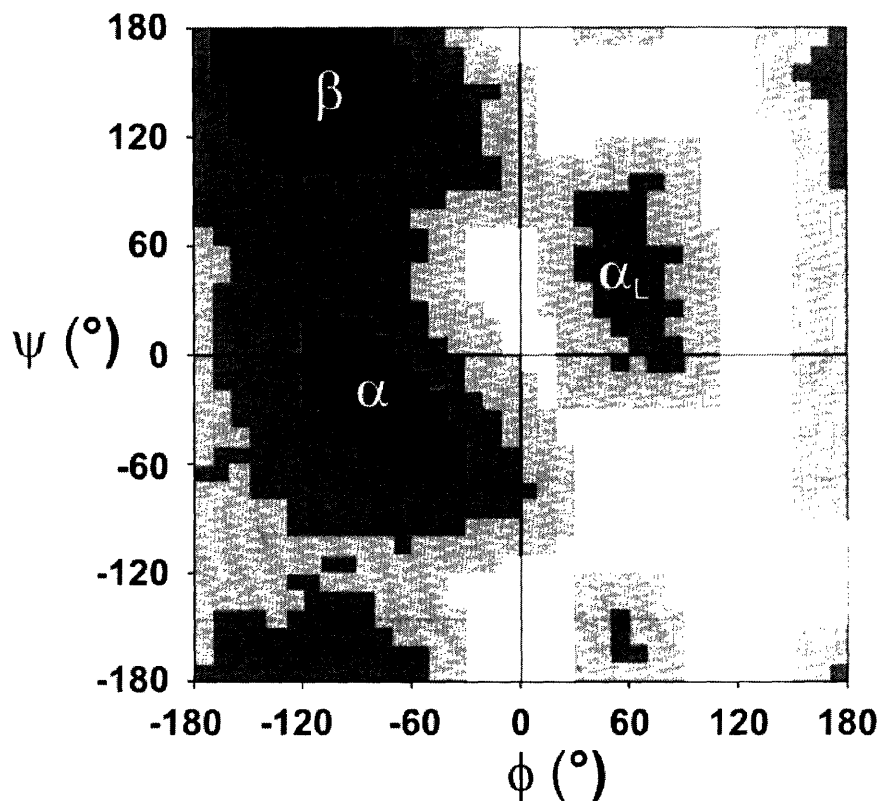
Figure 16: Ramachandran plot of all amino acids: The distribution of ϕ and ψ torsion angles in proteins. Secondary structure elements are labeled. The torsion angle ψ differs considerably for a side chain depending on the protein backbone conformation.

Methods to measure the coupling constants reported in the literature often use the E.COSY principle (Delaglio et al., 1991) and extract coupling constants from peak splittings in the ^{15}N dimension in a ^{1}H^{15}N correlation spectrum. As both couplings evolve simultaneously, the resolution of the coupling pattern is often poor and only few coupling constants can be extracted. Here, ^{1}J(N$_i$,C$_{\alpha i}$) and ^{2}J(N$_i$,C$_{\alpha(i-1)}$) coupling constants have been measured using J-modulated HSQCs (Billeter et al., 1992; Kuboniwa et al., 1994; Neri et al., 1990; Tjandra and Bax, 1997b; Tjandra et al., 1996; Vuister et al., 1993). These J-modulated experiments are less sensitive to limited resolution in the ^{15}N dimension. They are furthermore easy to implement and to automate and do not require prior knowledge of coupling constants for optimal suppression of unwanted peaks.

The conformational dependence of the coupling constants has been analyzed. In agreement with previous data in the literature, a dependence of the ^{1}J(N$_i$,C$_{\alpha i}$) and ^{2}J(N$_i$,C$_{\alpha(i-1)}$) coupling constants on the conformation of the protein backbone was observed. Based on the data by Delaglio (Delaglio et al., 1991) for staphylococcal nuclease and the data measured here for ubiquitin, improved Karplus-type relations are derived. It is found that ^{1}J(N$_i$,C$_{\alpha i}$) shows a weak dependence on ψ and that ^{2}J(N$_i$,C$_{\alpha(i-1)}$) depends on both angles $\phi_{(i-1)}$ and $\psi_{(i-1)}$.

2.2 Materials and Methods

2.2.1 NMR sample

^{13}C,^{15}N labeled human ubiquitin was purchased from VLI Research (Malvern, USA) and used without further purification. The NMR sample used contained 2mM ubiquitin in 30mM acetate buffer at pH 4.7 in 90% H$_2$O, 10% D$_2$O.

2.2.2 NMR measurements and data processing

All NMR measurements were recorded on a four-channel Bruker DRX600 spectrometer, equipped with an actively shielded z-gradient probe at a temperature of 25°C. Pulsed field gradients were used for coherent order selective coherence transfer, making use of sensitivity enhancement for all HSQC-type experiments (Schleucher et al., 1994). 64 complex points with an acquisition time of 65.534ms for ^{15}N (ω_1) and 2048 complex points with an acquisition time of 256ms were recorded for ^{1}H^{N} at a frequency of 600MHz.

All spectra were processed with Xwinnmr (Bruker, Karlsruhe), applying a 90° shifted sinebell function for apodization and zerofilling. 2D matrices consisted of 2048 x 1024 real points after fourier transformation, while only the $^1H^N$ region of the spectra was retained.

Data analysis (peak picking and integration) was done using the program Felix 98.0 (Biosym/MSI).

Fitting of NMR intensities was done using the program Origin. Fitting to determine the angular dependence of coupling constants was done using the program Sigma Plot.

2.3 Results and Discussion

2.3.1 Pulse sequence

Figure 17 shows the pulse sequence for the measurement of $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants.
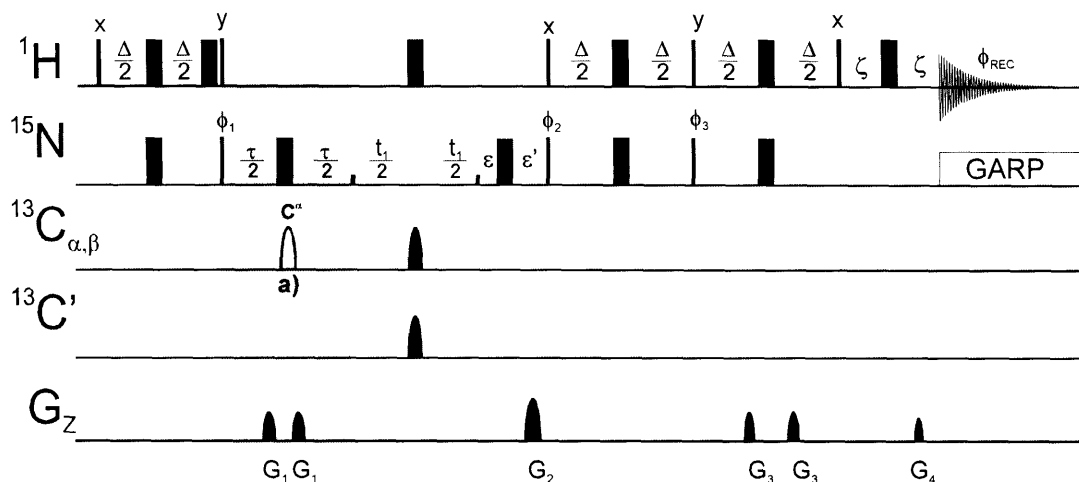
Figure 17: Pulse scheme for measurement of $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants: Carrier positions in the present work were 4.707ppm for ^1H (H_2O on resonance), 178.221ppm for $^{13}C'$, 58.678ppm for $^{13}C_{\alpha,\beta}$ and 116.77ppm for ^{15}N. A trim pulse of 1ms was applied at high proton pulse power, high power proton pulses were applied with a field strength of 34.7kHz. ^{15}N decoupling during acquisition employed a 2.3kHz GARP (Shaka et al., 1985) field, while high power ^{15}N pulses were applied with a field strength of 6.6kHz. Narrow and wide pulses denote 90° and 180° flip angles, respectively and unless indicated the phase is x. Carbon 180° pulses were implemented as phase modulated G3 inversion pulses (Emsley and Bodenhausen, 1990) of durations 1000μs for the selective C_α inversion pulse in the middle of τ and 512μs for decoupling inversions pulses. Delay durations: Δ=5.5ms, τ=20μs-150ms, ε and ε' were adjusted to avoid chemical shift evolution for t_1=0, ζ=1.507ms. Phase cycling: ϕ_1=x -x, ϕ_2=x x -x -x, ϕ_3= -y -y y y, ϕ_{rec}= x -x -x x. For each t_1 value, echo- and antiecho coherences where obtained by recording data sets where the phase ϕ_3 and gradient G_2 were inverted. Sine shaped gradient durations and amplitudes were: G_1=0.5ms (20G/cm); G_2=1ms (40G/cm); G_3=0.5ms (15G/cm); G_4=1ms (-4.05G/cm). Pulse scheme for the determination of T_2^*, omitting the selective C_α pulse (labeled with a)) in the middle of τ.

The experiment is based on a J-modulated sensitivity enhanced ^{1}H,^{15}N-HSQC (Billeter et al., 1992; Kuboniwa et al., 1994; Neri et al., 1990; Tjandra and Bax, 1997b; Tjandra et al., 1996; Vuister et al., 1993). ^{1}J(N$_i$,C$_{\alpha i}$) and ^{2}J(N$_i$,C$_{\alpha(i-1)}$) coupling constants evolve during a mixing time τ (gray shaded area in the pulse sequence of Figure 17) by selective recoupling, before the evolution of the nitrogen chemical shift during t$_1$. Selective recoupling during τ is achieved by applying a selective 180° pulse on C$_\alpha$ and a 180° pulse on N in the middle of the delay. Signal intensities in the HSQC spectra are therefore modulated by both coupling constants:

$$9) \qquad I(\tau) \propto \cos(^1J\pi\tau)\cos(^2J\pi\tau)$$

(with I (τ)=signal intensity depending on the delay τ, ^{1}J=^{1}J(N$_i$,C$_{\alpha i}$) and ^{2}J=^{2}J(N$_i$,C$_{\alpha(i-1)}$). However since τ is varied in between the experiments, transverse relaxation is also active during τ, which needs to be taken into account for the determination of the coupling constants. A full description of processes active during τ is therefore given as:

$$10) \qquad I(\tau) = A\cos(^1J\pi\tau)\cos(^2J\pi\tau)e^{-\tau/T_2^*}$$

The successful recoupling of the ^{1}J(N$_i$,C$_{\alpha i}$) and ^{2}J(N$_i$,C$_{\alpha(i-1)}$) coupling and therefore the modulation of the signal intensities according to Equation 10 relies on the selective inversion of the C$_\alpha$ carbons. Due to the limited chemical shift dispersion in the C$_\alpha$ region, the pulse sequence cannot be applied to measure couplings for couplings that involve either serines or threonines due to chemical shift overlap of C$_\alpha$ and C$_\beta$. Neither can the pulse sequence be applied to measure coupling constants to glycines as glycines are not inverted by the selective C$_\alpha$ pulse.

2.3.2 Determination of coupling constants from peak heights

A series of HSQCs with τ=20μs to 150ms was measured for native ubiquitin (17 points). Figure 18 shows experimental peak intensities and data fitting using a four parameter fit fitting A, ^{1}J, ^{2}J and T$_2^*$ (Equation 10) to the data points measured for residues Arg 42 and Glu 51. The intensity modulation could reliably be fitted.
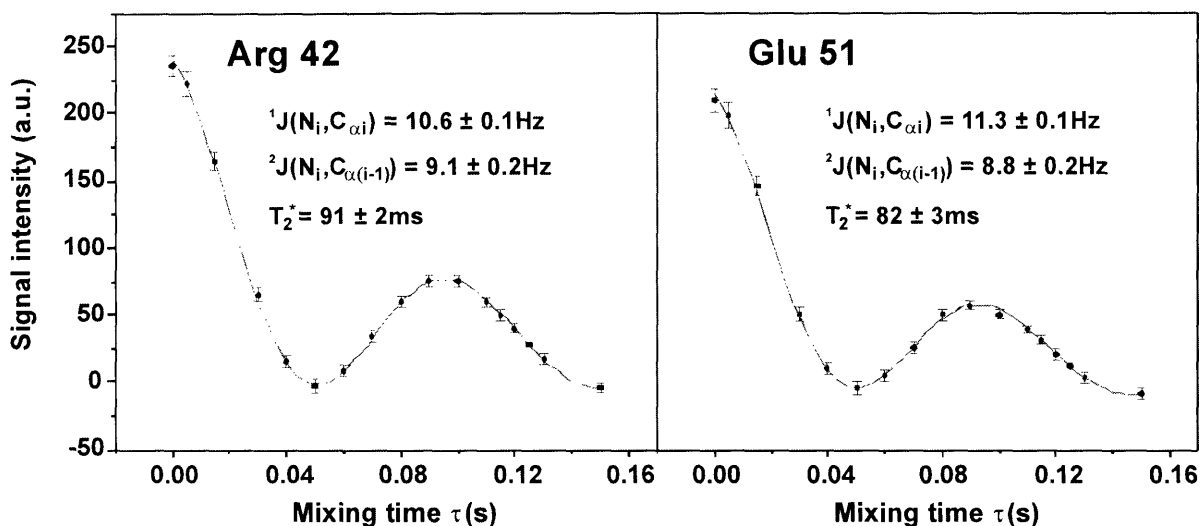
Figure 18: Experimental peak intensities (black diamonds) and fitting the equation $I=A\cos(\pi\,^1J\tau)\cos(\pi\,^2J\tau)e^{-\tau/T2^*}$ to them (gray line) for the cross peaks of Arg 42 and Glu 51 in folded ubiquitin. Error bars for the peak intensities were deducted from the standard deviation of 15 noise peaks in each spectrum. The quality of the fit is very good ($R^2=0.9993$ for Arg 42 and $R^2=0.9986$ for Glu 51). Fitting was performed using the program Origin. Examples shown here were chosen for comparison with the values of coupling constants measured by (Heikkinen et al., 2001).

To decrease the number of fitting parameters, T_2^* values were determined in an independent series of experiments for native ubiquitin by omitting the C_α selective pulse in the pulse sequence shown in Figure 17. T_2^* values were obtained by fitting peak intensities to a mono-exponential decay. Comparison of the values for T_2^* derived from fitting to Equation 10 with the independently determined T_2^* revealed that the fitted and the independent T_2^* values agree within the error margins.

2.3.3 $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ are sensitive to secondary structure elements

$^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants were determined for 38 residues (out of 45 theoretically possible values) in folded ubiquitin. The following averaged values were found: $^1J(N_i,C_{\alpha i})=10.8\pm0.8Hz$ and $^2J(N_i,C_{\alpha(i-1)})=7.7\pm1.3Hz$. The coupling constants measured here compare well (average observed deviation <0.3Hz) with the 11 coupling constants published by Heikkinen et al. (2001). The size of $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants presented, correlates well with the location of secondary structure elements in ubiquitin (Figure 19). Both $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants are substantially larger in β-sheet regions

$(^1J(N_i,C_{\alpha i})=11.2\pm0.5Hz$ and $^2J(N_i,C_{\alpha(i-1)})=8.4\pm0.7Hz)$ than in α-helices $(^1J(N_i,C_{\alpha i})=9.9\pm0.2Hz$ and $^2J(N_i,C_{\alpha(i-1)})=6.3\pm0.7Hz)$.
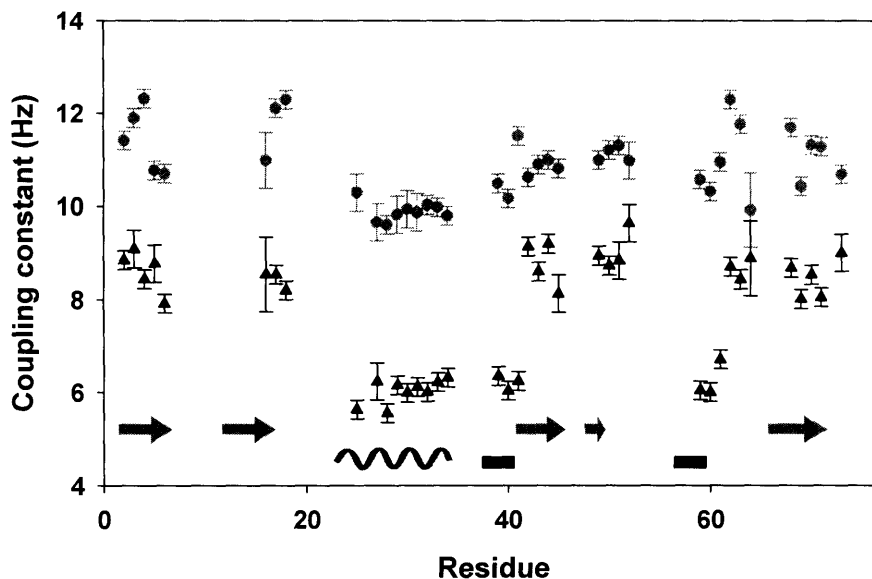
Figure 19: Correlation of $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants of folded ubiquitin with the residue. Gray circles represent $^1J(N_i,C_{\alpha i})$ and black triangles correspond to $^2J(N_i,C_{\alpha(i-1)})$. Secondary structures as determined by the algorithm implemented in MOLMOL (Konradi et al., 1996) are also shown (arrow: β-sheets 2-6, 12-16, 41-45, 48-49, 66-71, helix: α-helix 23-34, black bar: 3_{10}-helix 38-40, 57-59). Error bars are according to the deviations given in the fit *2 (Figure 18)

2.3.4 Angular dependence of $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants

$^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants for residues in staphylococcal nuclease (Delaglio et al., 1991) and in ubiquitin are analyzed in the following section. The structure dependence of the coupling constants in both proteins is in close agreement.

2.3.4.1 Analysis of $^1J(N_i,C_{\alpha i})$ coupling constants

The correlation of $^1J(N_i,C_{\alpha i})$ coupling constants on the torsion angles ϕ and ψ of ubiquitin (Cornilescu et al., 1998) and staphylococcal nuclease (Loll and Lattman, 1989) was investigated (116 coupling constants in total). Table 4 summarizes some important fitting results.

A Karplus correlation of $^1J(N_i,C_{\alpha i})$ depending on the torsion angle ϕ could not be observed as indicated by $R^2=0.32$ for a fit using Equation 11 (Table 4). A much better correlation is observed

for a Karplus correlation of $^1J(N_i,C_{\alpha i})$ with the torsion angle ψ (Equation 12) in Table 4): the statistical significance of the fit is described by a correlation coefficient R=0.85 (R^2=0.72). The probability that this nonlinear correlation results from random uncertainties is negligible ($P<10^{-4}$).

angle	$^1J(N_i,C_{\alpha i})=$	NP	R	R^2	Parameters	E. n°
ϕ	$A + B\cos(\phi) + C\cos^2(\phi)$	3	0.56	0.32	A=10.5792, B=-1.1616, C= 0.3752	11
ψ	$A + B\cos(\psi) + C\cos^2(\psi)$	3	0.85	0.72	A=9.5098, B=-0.9799, C=1.7040	12
ϕ,ψ	$A + B\cos(\psi) + C\cos^2(\psi) + D\cos(\phi) + E\cos^2(\phi)$	5	0.86	0.73	A=9.5456, B=-0.9093, C=1.6250 D=-0.1923, E=0.0376,	13
ϕ,ψ	$\sum_i^M \sum_j^N \left\{ \begin{array}{l} \cos(i\phi)\cos(j\psi) + \cos(i\phi)\sin(j\psi) \\ + \sin(i\phi)\cos(j\psi) + \sin(i\phi)\sin(j\psi) \end{array} \right\}$	25	0.81	0.66	(Edison et al., 1994a; Edison et al., 1994b)	14

Table 4: Different Karplus-type equations used for fitting of the experimental $^1J(N_i,C_{\alpha i})$ coupling constants depending on the torsion angle. Number of fitting parameters (NP) is given as well as the correlation coefficients R and R^2 the fitting parameters and an equation number (E.n°).

Correlation of $^1J(N_i,C_{\alpha i})$ with both torsion angles ϕ and ψ using five and 25 parameters respectively (Equation 13 and Equation 14) (Edison et al., 1994a; Edison et al., 1994b) show similarly good correlation coefficients as the ones observed for the three parameter fit that depends only on ψ in Equation 12.

F-statistics can be used to test whether the multi-dimensional models (Equations 13 and 14) with more adjustable parameters represent the experimental data better than the one dimensional model of Equation 12 (Bevington, 1969). F-values are calculated from number of data points (N), the numbers of parameters in the fit (n, with $n_2>n_1$) and the sum of squared errors between the experimental and the predicted data points (χ_1,χ_2) as following:

$$15) \qquad F = \frac{v_2(\chi_1 - \chi_2)}{(v_1 - v_2)\chi_2}$$

v is the statistical degrees of freedom for a certain model (v=N-n). An improvement of a given fit (compared to random statistical reduction of the χ value due to the incorporation of additional parameters) is indicated by an F-value that is larger than a so-called critical value for F-distributions $F_{0.01,N1,N2}$ (N_1=n_2-n_1, N_2=N-n_2) (Bronstein and Semendjajew, 1989).

The calculated F-value F=1.40 for a comparison of parameterizations of the three parameter ($f(\psi)$, Equation 12) and the five parameter fit ($f(\psi,\phi)$ Equation 13) is smaller than the corresponding critical value $F_{0.01,2,111} \approx 4.80$ (tabulated F values for N=128 and N=103 are $F_{0.01,2,125}=4.78$ and $F_{0.01,2,100}=4.82$, respectively). Therefore, the addition of the ϕ dependent parameters D and E do not change the quality of the fit and can therefore be omitted.

A comparison of the ψ dependent karplus equation (Equation 12) with the 25 paramter fit (Equation 14) derived from *ab-initio* calculations yields F=-11.43. This is considerably smaller than the critical F-value $F_{0.01,23,92} \approx 2.10$ ($F_{0.01,24,100}=1.98$, $F_{0.01,20,80}=2.23$). Therefore, the simple ψ dependent Karplus equation (Equation 12) describes the experimental data best. Figure 20 shows the curve resulting from the Karplus fit of Equation 12.
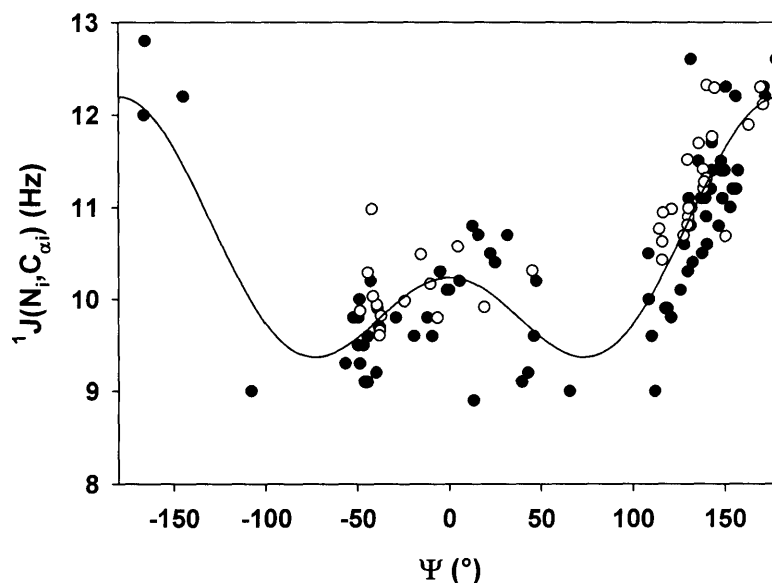
Figure 20: $^1J(N_i,C_{\alpha i})$ coupling constants for residues in folded ubiquitin (open circles) and folded staphylococcal nuclease, (Delaglio et al., 1991) (filled circles) as a function of ψ torsion angles. The correlation coefficient R for the least squares Karplus fit $f(\psi)=A+B\cos(\psi)+C\cos^2(\psi)$ is 0.85. Coefficients are A=9.5098, B=-0.9799 and C=1.7040 (Equation 12).

2.3.4.2 Analysis of $^2J(N_i,C_{\alpha(i-1)})$ coupling constants

Following Edison et al. (Edison et al., 1994a; Edison et al., 1994b), also the dependence of 113 $^2J(N_i,C_{\alpha(i-1)})$ coupling constants in ubiquitin and staphylococcal nuclease (Delaglio et al., 1991) were investigated. As observed for the $^1J(N_i,C_{\alpha i})$ coupling constants, a direct correlation of

$^2J(N_i,C_{\alpha(i-1)})$ with ϕ or $\phi_{(i-1)}$ (both angles were investigated, as two sequential amino acids are involved) cannot be observed. In contrast, the $^2J(N_i,C_{\alpha(i-1)})$ as a function of $\psi_{(i-1)}$ can be fitted to a Karplus equation as given in Equation 17 (R=0.91 (R^2=0.82), Table 5, Figure 21). The coefficients are A=7.6213, B=-1.3791 and C=-0.2067 and the probability that this correlation occurs by chance is negligible (P<10^{-4}) (Figure 21).
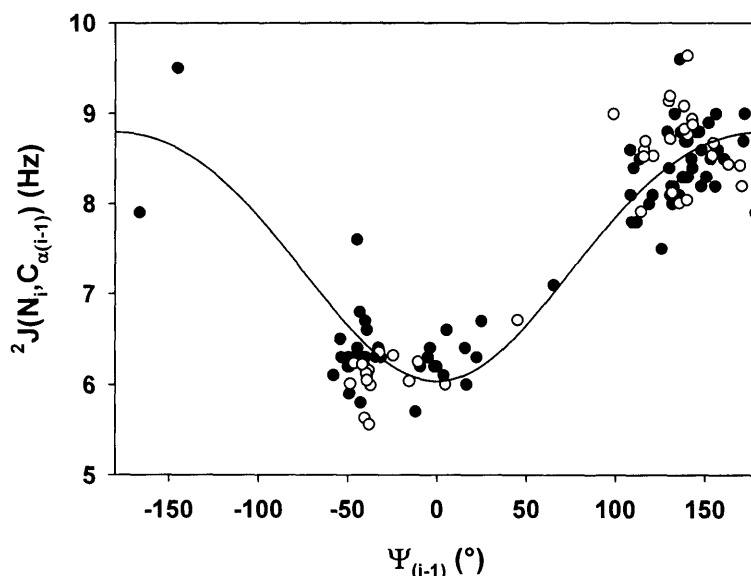
Figure 21: $^2J(N_i,C_{\alpha(i-1)})$ coupling constants of folded ubiquitin (open circles) and folded staphylococcal nuclease (Delaglio et al., 1991) (filled circles) as a function of $\psi_{(i-1)}$ torsion angles. The statistical significance of the Karplus relation between $\psi_{(i-1)}$ and $^2J(N_i,C_{\alpha(i-1)})$, f($\psi_{(i-1)}$)=A+Bcos($\psi_{(i-1)}$)+Ccos^2($\psi_{(i-1)}$) is described by R=0.91. The coefficients for the Karplus equation are A=7.6213, B=-1.3791 and C=-0.2067 (Equation 17).

The $\phi_{(i-1)}$ and $\psi_{(i-1)}$ dependent equation (Equation 18) was fitted using a least square optimization, yielding R=0.94 (R^2=0.88). The coefficients describing a surface are A=7.8163, B=-0.1717, C=-0.6408, D=-1.3892 and E=-0.3709. Applying F-statistics results in F=5.53 fulfilling the condition that F>F$_{0.01,2,108}$ as F$_{0.01,2,125}$=4.78, F$_{0.01,2,100}$=4.82. Therefore the five parameter equation (result of the fit is shown in Figure 22) describes the experimental data better than the Karplus fit, indicating that $^2J(N_i,C_{\alpha(i-1)})$ is not only dependent on $\psi_{(i-1)}$ but also on $\phi_{(i-1)}$.
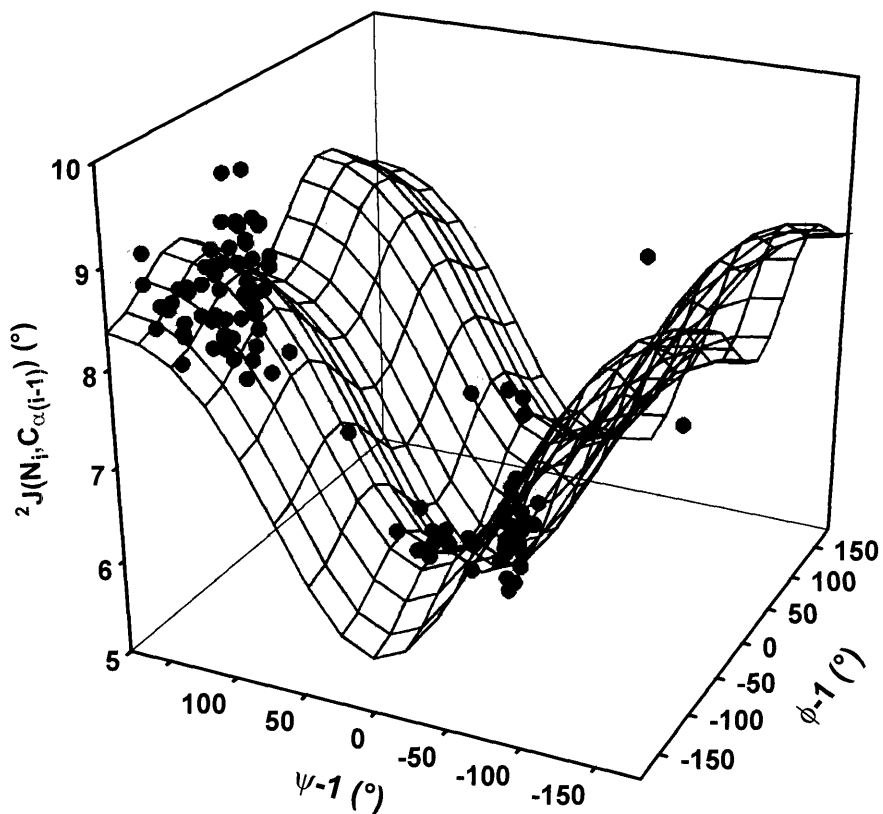
Figure 22: ^{2}J(N$_i$,C$_{\alpha(i-1)}$) coupling constants of residues of folded ubiquitin and folded staphylococcal nuclease (Delaglio et al., 1991) as a function of neighboring $\phi_{(i-1)}$ and $\psi_{(i-1)}$ torsion angles. The experimental coupling constant values were fitted to the five parameter combined Karplus Equation 18: $f(\phi_{(i-1)},\psi_{(i-1)})$=A+Bcos($\phi_{(i-1)}$)+Ccos^2($\phi_{(i-1)}$)+Dcos($\psi_{(i-1)}$)+Ecos^2($\psi_{(i-1)}$), giving rise to the depicted surface. The correlation coefficient R for the least square fit is 0.94, the coefficients are A=7.8163, B=-0.1717, C=-0.6408, D=-1.3892, E=-0.3709.

angle	$^2J(N_i,C_{\alpha(i-1)})=$	NP	R	R^2	Parameters	E. n°
$\phi_{(i-1)}$	$A + B\cos(180+\phi_{(i-1)})+C\cos^2(180+\phi_{(i-1)})$	3	0.61	0.37	A=7.6322, B=1.7123, C=-0.8173	16
$\psi_{(i-1)}$	$A + B\cos(\psi_{(i-1)})+C\cos^2(\psi_{(i-1)})$	3	0.91	0.82	A=7.6213, B=-1.3791, C=-0.2067	17
$\phi_{(i-1)},\psi_{(i-1)}$	$A + B\cos(\psi_{(i-1)})+C\cos^2(\psi_{(i-1)})$ $+ D\cos(\phi_{(i-1)})+E\cos^2(\phi_{(i-1)})$	5	0.94	0.88	A=7.8163, B=-0.1717, C=-0.6408, D=-1.3892, E=-0.3709	18
$\phi_{(i-1)},\psi_{(i-1)}$	$\sum_i^M \sum_j^N \left\{ \begin{array}{l} \cos(i\phi)\cos(j\psi)+\cos(i\phi)\sin(j\psi) \\ + \sin(i\phi)\cos(j\psi)+\sin(i\phi)\sin(j\psi) \end{array} \right\}$	25	0.79	0.63	(Edison et al., 1994a; Edison et al., 1994b)	19

Table 5: Different Karplus-type equations used for fitting of the experimental $^2J(N_i,C_{\alpha(i-1)})$ coupling constants depending on the torsion angle. Number of fitting parameters (NP) is given as well as the correlation coefficients R and R^2 the fitting parameters and an equation number (E.n°).

The results in this report were compared with the results of the analysis by Edison et al. (Edison et al., 1994a; Edison et al., 1994b). The correlation coefficient following Edison et al. (Edison et al., 1994a; Edison et al., 1994b) for staphylococcal nuclease alone and for the combined data set is R=0.79 (R^2=0.63). This is significantly worse than R=0.94 for Equation 18 (Table 5). The parameterization presented here (Equation 18) therefore describes the data a lot better and is a lot easier to handle, than the earlier parameterization.

2.3.5 Discussion

Parameterizations for the dependence of $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants on the protein backbone angles were derived. It was found that $^1J(N_i,C_{\alpha i})$ depends weakly on ψ, while the size of $^2J(N_i,C_{\alpha(i-1)})$ coupling constants depends on both ψ and ϕ (more precise on $\phi_{(i-1)}$ and $\psi_{(i-1)}$). Using this parametrization it is not possible to derive a single pair of protein backbone angles from measured $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants, due to the degeneracy of Karplus equations. However, it is possible to differentiate between different secondary structure elements from measured coupling constants. Furthermore, $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants were found to be a useful tool in cases of ambiguous predictions from TALOS (Cornilescu et al., 1999) as discussed below.

Talos compares experimental chemical shifts with a chemical shift database of known NMR structures. Based on this comparison (using the 10 best matching structures), angle predictions

are made by TALOS. While the majority of structural elements can be well predicted using TALOS, there is always a fraction of ambiguous predictions.

Using TALOS, 55 ϕ/ψ good angle predictions can be derived for ubiquitin. However, TALOS does not provide an unambiguous prediction for 19 residues. $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants were measured for 9 (8 respectively) of these 19 residues. Figure 23 shows the obtained angle pairs of ϕ and ψ from the TALOS prediction and the correct values of ϕ and ψ in the ubiquitin structure (residues Q40, K63 and N60), together with the prediction from the measured coupling constants.
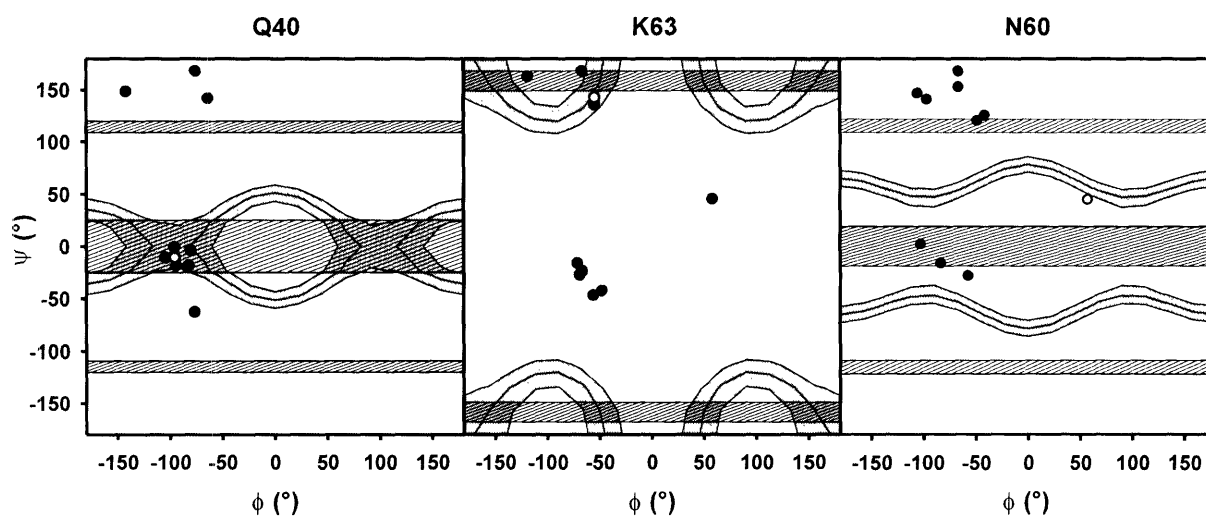
Figure 23: Ramachandran plot showing predicted backbone conformations for Q40, K63 and N60 in ubiquitin. Open circle presents the conformation in the ubiquitin structure, closed circles present conformations deducted by TALOS. Dashed and gray regions indicate possible conformations calculated from experimental $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants (including an error estimate of ±0.2).

Talos predictions cluster in two regions of the Ramachandran plot for Q40 (Figure 23). While a helical conformation is predicted five times, a sheet conformation is predicted three times. Experimental coupling constants of $^1J(N_i,C_{\alpha i})_{Q40}$=10.2±0.1Hz and $^2J(N_i,C_{\alpha(i-1)})_{Q40}$=6.0±0.1Hz clearly select the helical predictions in the helical region. Using coupling constant data, the derived ϕ and ψ angles for Q40 (ϕ=-92±10° and ψ=-10±8°) are in good agreement with the backbone angles in the ubiquitin structure (Cornilescu et al., 1998): ϕ=-96° and ψ=-10°. For K63, TALOS predicts pairs of ϕ and ψ in two clustered regions of the Ramachandran diagram. Comparison with prediction from coupling constants selects the region from the Ramachandran

diagram ($\phi=-80\pm34°$ and $\psi=155\pm17°$). This is in agreement with the structural data ($\phi=-50°$ and $\psi=143°$). For all other residues but N60, similar results are obtained. None of the TALOS predicted ϕ/ψ pairs is in agreement with the actual conformation of N60 in the ubiquitin structure. In contrast, the prediction made from $^2J(N_i,C_{\alpha(i-1)})_{I61}$ and $^1J(N_i,C_{\alpha i})_{N60}$ coupling constants includes the correct conformation in the α_L region of the Ramachandran diagram. Thus, the use of $^2J(N_i,C_{\alpha(i-1)})$ coupling constants in combination with TALOS can clarify the few ambiguous predictions from TALOS based on chemical shifts.

The importance of the above pulse sequence and particularly of the parameterization derived, is reflected by the renowned interest in the coupling constants after the publication of the method and parameterization presented here (Bhavesh et al., 2003; Ding and Gronenborn, 2004; Permi and Annila, 2004; Wienk et al., 2003). Thus, new pulse sequences were developed allowing a second independent determination of the coupling constants and the use of the coupling constants e.g. also in the manner of dipolar coupling constants for structure determination was furthermore consolidated (Ding and Gronenborn, 2004; Wienk et al., 2003).

2.4 Conclusion

A new method to measure $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants in proteins was introduced. The method is based on a J-modulated sensitivity enhanced HSQC and is therefore easy to implement. Furthermore, the method can be even used in systems with limited chemical shift resolution such as denatured proteins, as the required resolution is the required resolution of an HSQC only.

Measured $^2J(N_i,C_{\alpha(i-1)})$ coupling constants were found to be a valuable indicator to identify secondary structure elements in folded proteins. The dependence on protein backbone torsion angle conformations was quantified, and it was shown that simple models describe the observed coupling constants well. $^2J(N_i,C_{\alpha(i-1)})$ is correlated with the torsion angles $\phi_{(i-1)}$ and $\psi_{(i-1)}$. Also a weak structural dependence of $^1J(N_i,C_{\alpha i})$ coupling constants on the torsion angle ψ was found and parametrized. The ease of access make $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants an interesting parameter, together with chemical shift data (Cornilescu et al., 1999) for the rapid identification of secondary structure elements in proteins.

Chapter 3

Ubiquitin- Model for an unfolded protein without residual structure

3 Ubiquitin- model for an unfolded protein without residual structure

3.1 Introduction

The aim of this chapter is the investigation of unfolded ubiquitin and the derivation of a model for the interpretation of relaxation data in unfolded proteins without residual structure.

Often unfolded proteins investigated in the literature are not completely unfolded. However, residual structure has been observed by comparions of local NMR parameters with data expected for a random coil (Blanco et al., 1998; Fiebig et al., 1996; Klein-Seetharaman et al., 2002; Lietzow et al., 2002; Schwalbe et al., 1997; Shortle and Abeygunawardana, 1993; Shortle and Ackerman, 2001; Wirmer et al., 2004; Wong et al., 1996). Electrostatic interactions in unfolded states of proteins can be excluded as a reason for the residual structure in unfolded proteins, as investigations of unfolded proteins are usually carried out at pH 2, often in 8M urea. The presence of 8M urea depletes possible electrostatic interactions. Furthermore negatively charged amino acids present at neutral pH are neutralized by protonation at pH 2. Rather, these residual structure elements coincide with the presence of aromatic residues, particularly with tryptophans (Klein-Seetharaman et al., 2002; Neri et al., 1992; Ropson and Frieden, 1992; Saab-Rincon et al., 1996; Schwalbe et al., 1997; Wirmer et al., 2004). Ubiquitin has no tryptophan residues and in total only three aromatic residues (F4, Y59 and H68), suggesting that unfolded ubiquitin might possess less residual structure than other unfolded proteins. Ubiquitin therefore might serve as model for an unfolded protein without residual structure. This study shows that ubiquitin indeed is a good model for a protein in a random coil state by analysis of chemical shift deviations, $^3J(H^N,H^\alpha)$, $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants and ^{15}N relaxation data.

3.2 Materials and Methods

3.2.1 Sample preparation

5mg of uniformly $^{13}C,^{15}N$ labeled (^{15}N labeled) human ubiquitin (VLI Research, Malvern, PA) were dissolved in 300µl of 90% H_2O, 10% D_2O, 8M urea at pH 2.

3.2.2 Chemical shift deviations

Resonance assignment was taken from Peti et al. (Peti et al., 2001). Random coil chemical shifts (δ^{rc}) using the "Wishart peptide" scale (Wishart et al., 1995a) were extracted using the program NMRView (version 5.0.4) including sequence specific corrections (Schwarzinger et al., 2001). Chemical shift deviations ($\Delta\delta$) were calculated according to $\Delta\delta=\delta^{exp}-\delta^{rc}$ (δ^{exp}=experimental chemical shift taken from Peti (Peti et al., 2001)). Significant deviations are defined according to the chemical shift index (CSI) (Wishart and Sykes, 1994; Wishart et al., 1992).

3.2.3 NMR measurements

All NMR measurements were performed on a four-channel Bruker DRX 600 spectrometer equipped with an actively shielded TXI z-gradient probe at a temperature of 25°C.

$^{1}J(N_i,C_{\alpha i})$ and $^{2}J(N_i,C_{\alpha(i-1)})$ coupling constants were measured and determined as described in chapter 2, exept for the following differences: 14 delays between 20μs and 150ms were used for the evolution of the coupling, and 128 complex points were recorded in the indirect dimension at a spectral width of 1520Hz.

^{15}N longitudinal (R_1) and transverse (R_2) relaxation rates and the heteronuclear NOE (hetNOE) were measured as described in the literature (Akke and Palmer, 1996; Kay et al., 1989; Palmer and Case, 1992) using sensitivity enhancement in the back transfer (Schleucher et al., 1994). All spectra were acquired with 1024 points in the direct dimension at a sweep width (^{1}H) of 8389Hz, and 128 complex points in the indirect dimension (^{15}N) at a sweep width of 1520Hz. For the measurement of R_1 and R_2 relaxation rates nine experiments each were recorded using seven different mixing times between 30 and 1500ms (R_1) and 17 and 340ms (R_2 relaxation rates). Two experiments were recorded in an interleaved manner for the measurement of the hetNOE, one with proton presaturation (applied by a series of 120° high power proton pulses during the recycle delay) and one without proton presaturation. The recycle delay (RD) was 3s for the hetNOE experiment and 2s for R_1 and R_2 experiments.

Processing of the relaxation data and measurement of peak heights was carried out using the program Felix 98.0 (Biosym/MSI San Diego, CA). Data were fourier transformed applying zerofilling and apodization by a 90° shifted sinebell window function; only the $^{1}H^{N}$ region of the spectra was retained, the final matrix size was 1024*256 points. R_1 and R_2 relaxation rates were fitted as single exponential decays ($I=A*exp(-t/R)$) to the peak height data (I) with t=relaxation

delay, A=Amplitude and R=either R_1 or R_2 (Stone et al., 1993; Stone et al., 1992) (http://cpmcnet.columbia.edu/dept/gsas/biochem/labs/palmer/software.html).

The hetNOE was determined by dividing the peak heights in the spectra without presaturation by the ones with presaturation.

3.2.4 Hydrophobicity calculations

Hydrophobicity was calculated using the protscale tool from the ExPASy (Appel et al., 1994) (Expert Protein Analysis System) molecular biology server (http://us.expasy.org/cgi-bin/protscale.pl). The Abraham and Leo (Abraham and Leo, 1987) scale was applied. The window size (length of the interval used for profile computation in units of residues) was 7 (as this is the persistence length found in a number of unfolded proteins (Schwalbe et al., 1997; Schwarzinger et al., 2002)) and the weight at the edge of the window was set to 100 %; the hydrophobicity shown is normalized from 0-1. Table 6 shows the hydrophobicities of the respective amino acids according to the Abraham and Leo scale (Abraham and Leo, 1987). The scale was used rather than the often used Kyte and Doolitle scale (Kyte and Doolittle, 1982), as we found that this scale reflects the hydrophobicity of tryptophan side chains in unfolded proteins better.

residue	Hydrophobicity	residue	hydrophobicity	residue	hydrophobicity
W	2.56	M	1.1	T	-0.41
F	2.54	C	0.58	Q	-0.71
I	2.46	A	0.44	S	-0.84
L	2.46	E	-0.34	N	-1.32
V	1.73	D	-0.31	R	-2.42
Y	1.63	H	-0.01	K	-2.45
P	1.29	G	0		

Table 6: Hydrophobicities according to scale of Abraham and Leo (Abraham and Leo, 1987).

3.2.5 Radii of gyration of side chains

Radii of gyration (R_g) of side chains were used as given by Levitt (Levitt, 1976) (Table 7). The intrinsic correlation time τ of each amino acid was assumed to be proportional to R_g^3 based on the Stokes law (Schwarzinger et al., 2002).

residue	R_g (Å)	residue	R_g (Å)	residue	R_g (Å)
W	2.21	M	1.80	T	1.24
F	1.90	C	1.22	Q	1.75
I	1.56	A	0.77	S	1.08
L	1.54	E	1.77	N	1.45
V	1.73	D	1.43	R	2.38
Y	1.29	H	1.78	K	2.08
P	1.25 (2)	G	0.58		

Table 7: R_g of amino acid side chains as given by Levitt (Levitt, 1976). The decreased conformational freedom associated with proline is modeled by using a R_g of 2Å rather than the measured R_g of 1.25Å (Schwarzinger et al., 2002).

77

3.3 Results and Discussion

3.3.1 Secondary structure

3.3.1.1 Chemical shift deviations

H^α chemical shifts are very sensitive to secondary structure. Therefore, they can be used as a measure for the extent of residual secondary structure present in unfolded ubiquitin. Chemical shift data in ubiquitin (Peti et al., 2001) are very close to random coil chemical shifts (Wishart et al., 1995a) as expected for an unfolded protein. Only five residues deviate from areas of random coil chemical shifts: negative deviations are found for K11 and Y59, positive deviations are found for F4, L8 and I13. The averaged absolute deviation in ubiquitin is $<|\Delta\delta|>_{ubi}=0.04\pm0.04$, being significantly smaller than the averaged absolute deviation in unfolded lysozyme ($<|\Delta\delta|>_{hewl}=0.11\pm0.09$, chapter 4)). Averaged positive and negative deviations are in both cases in the same order of magnitude as the absolute deviation ($<\Delta\delta->_{ubi}=-0.04\pm0.05$, $<\Delta\delta+>_{ubi}=0.05\pm0.03$, $<\Delta\delta->_{hewl}=-0.11\pm0.10$ and $<\Delta\delta+>_{hewl}=0.08\pm0.09$). The differences in deviations are therefore not due to errors in referencing. While residual secondary structure has been identified in lysozyme (chapter 4) where deviations are found in more than three consecutive residues, the opposite is the case in ubiquitin. The averaged deviations in ubiquitin are very small and deviations are randomly distributed over the sequence. Unfolded ubiquitin does not possess any residual secondary structure as judged from chemical shift data.

3.3.1.2 $^3J(H^N,H^\alpha)$ coupling constants

Another tool for the determination of protein backbone angles in unfolded proteins are scalar coupling constants such as $^3J(H^N,H^\alpha)$. The coupling constants have not been measured here but data and discussion follow Peti et al., 2000 and are shown for completeness of data discussion. Figure 24 shows a graphic reproduced from the literature (Peti et al., 2000) where experimentally determined coupling constants in ubiquitin are compared in a residue specific manner with data predicted by the random coil model (Fiebig et al., 1996; Hennig et al., 1999; Schwalbe et al., 1997; Smith et al., 1996a; Smith et al., 1996b). A good correlation is observed but for histidine (H), phenylalanine (F), aspartate (D) and glutamate (N) residues. These differences are most likely due to the different pH and the different solvation of the measured data (pH2, 8M urea) in

comparison with the structural data used by the random coil model (pH7, no denaturant). Detailed analysis of the data takes the nature of the preceding residue (i-1) in account. On average higher coupling constants are observed when residue (i-1) is either β-branched or aromatic, reflecting the higher tendency to populate β-conformations.
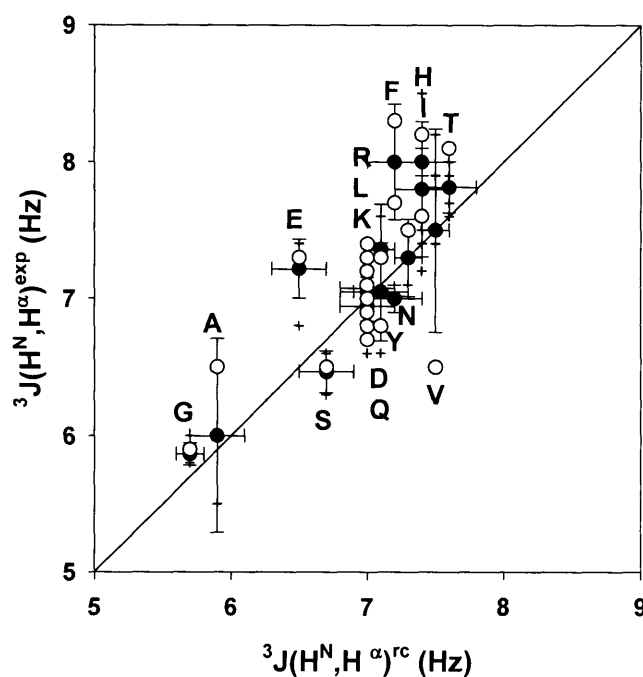
Figure 24: Residue specific correlation between experimental $^3J(H^N,H^\alpha)$ coupling constants (Peti et al., 2000) with predicted coupling constants from the random coil model (Fiebig et al., 1996; Smith et al., 1996b). The picture was reproduced from the literature (Peti et al., 2000). Residue types are indicated either on top or below the respective predicted value. Filled circles correlate the averaged experimental value, while crosshairs indicated experimental values for unbranched (i-1) residues and open circles indicate residues for either β-branched or aromatic (i-1) residues.

In conclusion, the experimental data agree very well with the predicted data for a random coil with a minor drawback. $^3J(H^N,H^\alpha)$ coupling constants only depend on the protein backbone angle φ. The angle, however, is very similar for either a residue in the α-helical conformation or in a β-sheet conformation or a statistically averaged conformation between the two as expected for a random coil. For a good differentiation between a protein in a random coil state and a protein where residual secondary structure elements are present, different coupling constants have to be measured.

79

3.3.1.3 $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants

In chapter 2 a method for the determination of $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants was developed. It was shown that the coupling constants, in particular $^2J(N_i,C_{\alpha(i-1)})$ are dependent on the backbone conformation; predominantly on the protein backbone angle ψ, critical for the determination of secondary structure elements. Figure 25 shows the $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants in the denatured state of ubiquitin.
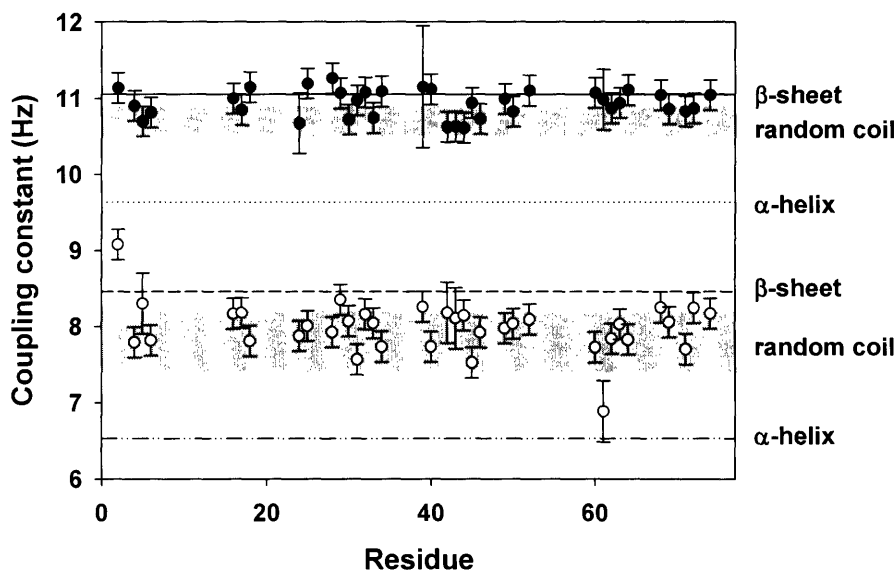
Figure 25: $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants in denatured ubiquitin as function of residue number. Black closed circles represent $^1J(N_i,C_{\alpha i})$ and black open circles represent $^2J(N_i,C_{\alpha(i-1)})$ coupling constants. Coupling constants expected in a random coil are indicated by gray areas. Coupling constants expected in β-sheets $^1J(N_i,C_{\alpha i})$ solid line, $^2J(N_i,C_{\alpha(i-1)})$ dashed line and α-helices $^1J(N_i,C_{\alpha i})$ dotted line, $^2J(N_i,C_{\alpha(i-1)})$ dashed and dotted line, are indicated. Errors are according to the deviations given in the fit *2 (chapter 2).

Uniform coupling constants are observed along the sequence ($^1J(N_i,C_{\alpha i})$=10.9±0.2Hz and $^2J(N_i,C_{\alpha(i-1)})$=8.0±0.3Hz). Mean values for the individual amino acid range from $^1J(N_i,C_{\alpha i})$=10.8±0.2Hz for Val, Ile and Leu to $^1J(N_i,C_{\alpha i})$=11.1±0.1Hz for Asp and Asn residues and from $^2J(N_i,C_{\alpha(i-1)})$=7.7±0.2Hz for Phe and Ile to $^2J(N_i,C_{\alpha(i-1)})$=8.2±0.1Hz for Asp, Arg, Val and His residues. The difference between the mean coupling constant values for the individual amino acid is therefore $\Delta^1J(N_i,C_{\alpha i})$=0.3Hz and $\Delta^2J(N_i,C_{\alpha(i-1)})$=0.5Hz. This is in the same order of

magnitude as the variations within the individual amino acids. A residue specific variation of the coupling constants is thus not observed.

Values for expected coupling constants in α-helices, β-sheet and random coil regions were calculated using the parameterization given in Equation 12 (f(ψ)) and Equation 18 (f(ϕ_{i-1},ψ_{i-1}) in chapter 2. For the calculation of values in α-helices and β-sheets averaged values of all proteins in the Ramachandran plot were used (Ramachandran and Sasisekharan, 1968). Mean values for the coupling constants in the random coil state were predicted using pairs of ϕ and ψ angles from 85 proteins of the pdb following Smith et al. (1996): $^1J(N_i,C_{\alpha i})^{rc}$=10.7±0.2Hz and $^2J(N_i,C_{\alpha(i-1)})^{rc}$=7.8±0.4Hz.

As shown in Figure 25, experimental coupling constants in denatured ubiquitin fall in between the range expected for either β-sheet or α-helix. Particularly, the experimental $^2J(N_i,C_{\alpha(i-1)})$ coupling constants agree remarkably well with the predicted coupling constants for a protein in the random coil state indicating that ubiquitin is unfolded at denaturing conditions. Experimental $^1J(N_i,C_{\alpha i})$ coupling constants are slightly higher than expected for the random coil state. This discrepancy can be attributed to the weaker dependence of the $^1J(N_i,C_{\alpha i})$ coupling constants on the backbone conformation and the lower quality of the parameterization, rather than a deviation from random coil behaviour. Thus, the observed coupling constants specifically the $^2J(N_i,C_{\alpha(i-1)})$ coupling constants are in good agreement with the model for the random coil state of a protein (Schwalbe et al., 1997; Smith et al., 1996; Hennig et al., 1999; Peti et al., 2000).

3.3.2 Relaxation

^{15}N heteronuclear relaxation rates of backbone amides of unfolded ubiquitin are shown in Figure 26. The relaxation data show only small variations over the sequence: Values approach a plateau at the middle of the protein while lower values are found at the termini of the sequence. This implies that the relaxation properties of a given amide are not influenced specifically by its neighbor, but just by the fact that it is part of the polypeptide chain. The only exception from this behavior is the R_2 relaxation rate of T9. A large deviation is observed, which is yet unexplained. Generally, R_2 rates in general show larger variations than the two other relaxation data (Schwalbe et al., 1997). Hence, they are a more sensible tool for the investigation of motions.

Relaxation data at the N-terminus of the protein are not as low as at the C-terminus. Particularly, the value of the hetNOE at the C-terminus is very low. This can be explained by the presence of

81

two consecutive glycine residues at the very C-terminus of ubiquitin. Glycine residues have no side chain and are more flexible than any other residue.
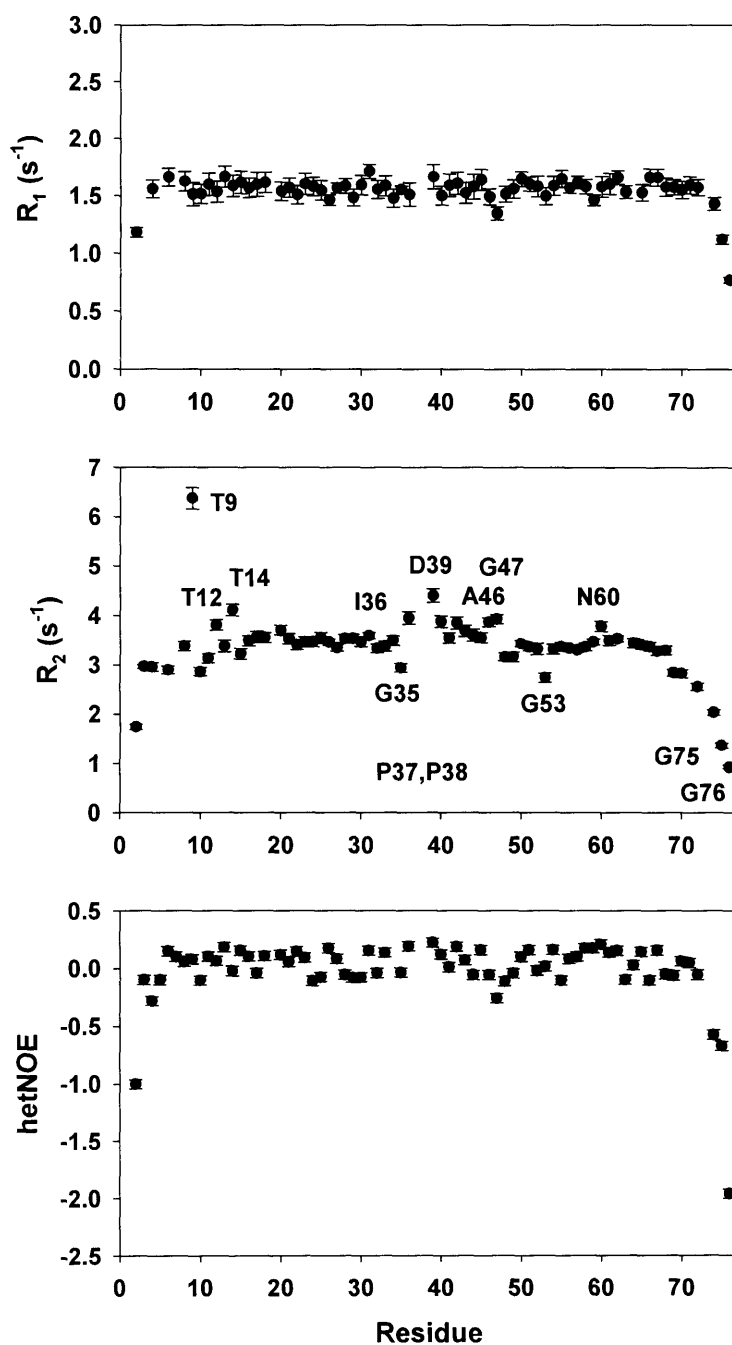
Figure 26: Backbone ^{15}N heteronuclear relaxation data of unfolded ubiquitin at 600 MHz. Top: R_1 relaxation rates in the range of 0-$3s^{-1}$, middle: R_2 relaxation rates in the range of 0-7; bottom: heteronuclear NOE in a range of -2.5-0.5a.u.

Plateau values in the central part of the sequence taking into account residue 6 to residue 71 are as follows: hetNOE=0.05 ± 0.11, R_1=$1.57\pm0.063s^{-1}$ and R_2=3.48 ± 0.50. All values are considerably lower than expected for a structured protein and reflect the unfolded character of the denatured ubiquitin: e.g. for a rigid protein a hetNOE near 1 is expected. The plateau value of the R_2s ($<R_2s>$=$3.48\pm0.5s^{-1}$) is very similar to plateau values found in other unfolded proteins in 8M urea. $<R_2s>$ equals $3.5s^{-1}$ in reduced and methylated lysozyme in urea (Schwalbe et al., 1997). From the relaxation data presented here it can be concluded that unfolded ubiquitin is a good model protein for unfolded states without residual structure. J coupling constants and chemical shifts deviations from unfolded states furthermore confirm this conclusion.

3.3.2.1 Interpretation of relaxation data

The interpretation of relaxation rates in unfolded proteins is different from the interpretation of relaxation data in folded proteins. Internal motions in folded proteins are on a different time scale than global motions of the protein. The motions therefore can be separated (Lipari and Szabo, 1982a; Lipari and Szabo, 1982b). This separation is not possible in unfolded state of proteins. Segments of the protein reorient independently from the whole protein. Different models based on segmental motions are currently discussed in the literature for the description of unfolded proteins (Schwalbe et al., 1997; Schwarzinger et al., 2002). Ubiquitin is a protein without residual structure in its unfolded state as shown in the previous section. This makes it a good system to test different models. In the following section, possible models are introduced and applied to the R_2 relaxation rates of unfolded ubiquitin. R_2 relaxation rates were chosen for the investigations since variations of relaxation data in unfolded states of proteins are most pronounced in R_2 relaxation rates as shown here and in other unfolded proteins (Schwalbe et al., 1997).

Model A: correlation with hydrophobicity. Residual structure and areas that exhibit increased relaxation rates often coincide with hydrophobic regions (Neri et al., 1992; Ropson and Frieden, 1992; Saab-Rincon et al., 1996). Thus, a correlation of R_2 relaxation rates in the unfolded state with the hydrophobicity of a given protein sequence could be imagined. However, in the case of ubiquitin variations of hydrophobicity as shown in Figure 27 have nothing in common with variations in R_2 relaxation rates. While the R_2 relaxation rates are rather uniform along the sequence, large variations are found in the hydrophobicities. Normalized hydrophobicities range from 0.33 to 0.75a.u. with a mean value of 0.53 ± 0.09a.u. Mean value and spread in hydrophobicities compare well with other proteins, e.g. lysozyme (mean: 0.51 ± 0.10a.u., spread:

0.25 to 0.74a.u.). A hydrophobicity dependence of relaxation data in unfolded proteins without residual structure can thus be excluded.
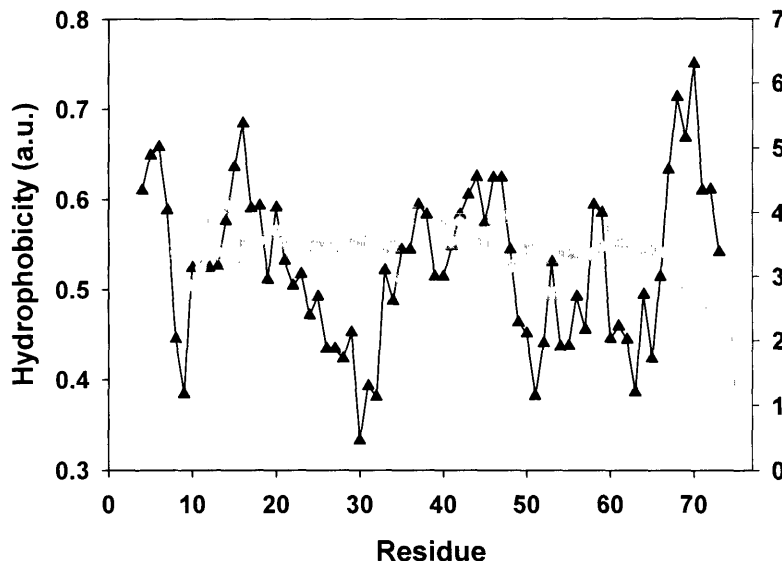
Figure 27: Hydrophobicity profile of ubiquitin in black linked triangles. Hydrophobicity was calculated from the sequence using the Abraham and Leo scale (Abraham and Leo, 1987) as described in the methods section. R_2 relaxation rates are shown in gray for comparison.

Model B: side chain volume depending segmental motions. Schwarzinger et al. (Schwarzinger et al., 2002) proposed a model which assumes segmental motions.These segmental motions are modulated by the side chain volumes of the amino acids constituting a protein:

$$(20) \qquad R_2^{Rg}(i) = k\sum_{j=1}^{N}\tau_j e^{-\frac{|i-j|}{\lambda_0}}$$

where τ_j=is proportional to the intrinsic correlation time of a residue j given as $\tau_j=(R_g)^3$ (R_g=radius of gyration of the respective amino acid side chain), N is the total chain length of the polypeptide, k is a scaling constant and λ_0 is the persistent length of the polypeptide chain (in numbers of residues). Schwarzinger et al., 2002 fixed λ_0 to 7 for all residues but for alanines and glycines (λ_0=2 was used instead). Here, uniform λ_0 was applied for all residues. λ_0 was varied in the fitting procedure.

Fitting of the experimental R_2 relaxation rates using the above equation is shown in Figure 28. The fitting results do not represent the experimental data well (R^2=0.39). The fitted line is

considerably lower than the experimental data in the N-terminal part of the protein (residues 3 to 26). Also the rest of the protein is not well described by the model. The fit is higher than the experimental data found for residues 29 to 35, for residues 48 to 58 and for the very C-terminus of the protein (residues 69 to 76). Predicted relaxation rates are lower than the experimental data for residues 60 to 68. In conclusion, the side chain volume dependent model does not describe the relaxation data in unfolded ubiquitin. The model cannot be used to predict relaxation data expected in a protein without residual structure in the unfolded state.
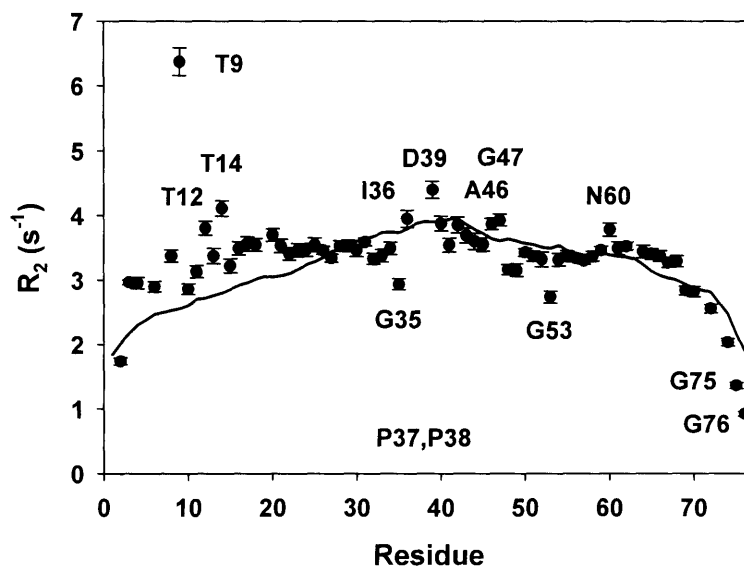
Figure 28: Experimental R_2 relaxation rates fitted with model B where R_2 data depend on the side chain volumes λ_0=7.69, k=0.05, R^2=0.39.

Model C: segmental motion model. The simplest model is the segmental motion model introduced by Schwalbe et al. (Schwalbe et al., 1997). The segmental motion model assumes that the influence of the neighboring residues decays exponentially as the distance (in number of peptide bonds) from a given residue increases. This is described by the following expression:

5)
$$R_2^{rc}(i) = R_{int} \sum_{j=1}^{N} e^{-\frac{|i-j|}{\lambda_0}}$$

where R_{int} is the intrinsic relaxation rate (depends also on the temperature and viscosity of the solution), λ_0 is the persistence length of the polypeptide chain (in numbers of residues) and N is the total chain length of the polypeptide.

In contrast to the volume dependent model, only few deviations are observed using the segmental motion model (model C). A best fit is found using an intrinsic relaxation rate of $R_{int}=0.4s^{-1}$ and a persistence length of $\lambda_0=4.34$ ($R^2=0.73$) (Figure 29). This persistent length is smaller than in other unfolded proteins that have $\lambda_0=7$ (Klein-Seetharaman et al., 2002; Schwalbe et al., 1997; Wirmer et al., 2004). This could be because ubiquitin is significantly smaller than the proteins studied in the literature (8.6kDa, versus 14.4kDa). However, the second best fit is found using $R_{int}=0.27s^{-1}$ and a persistence length of $\lambda_0=6.67$, which agrees with the literature value.
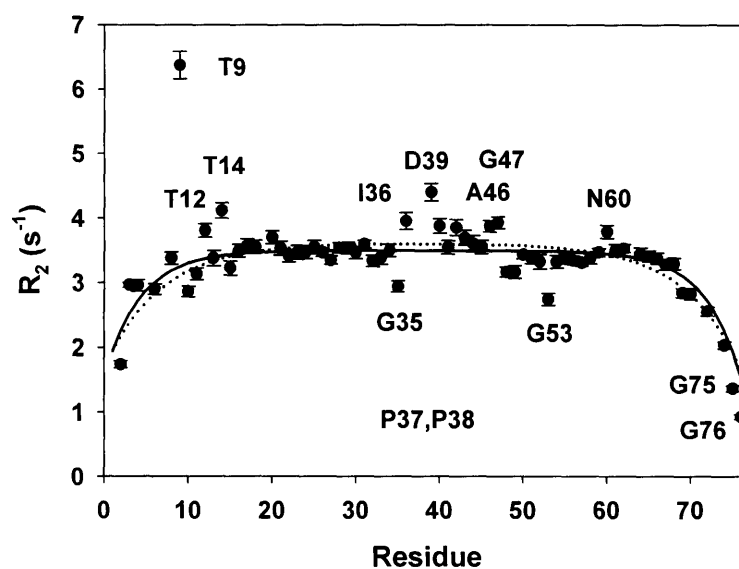
Figure 29: Experimental R_2 relaxation rates fitted with the segmental motion model (Equation 5) for the dynamics of an unbranched polymer chaindotted line: $R_{int}=0.27$, $\lambda=6.67$ ($R^2=0.73$); straight line $R_{int}=0.4$, $\lambda=4.34$ ($R^2=0.71$).

Both fits using the segmental motion model describe the R_2 relaxation rates reasonably well. Large deviations are found for T9, while small deviations are found for residues T12 and T14 and for residues I36 and D39. Two proline residues (P37, P38) flank I36 and D39. Earlier investigations on unfolded ubiquitin (Peti et al., 2000) revealed restricted ψ sampling of P38. Uniform $^3J(C_\alpha,C_\alpha)$ coupling constants at the random coil values were found for all residues in unfolded ubiquitin except for P19 and P38. Thus, the P37 and P38 most probably induce motional restrictions onto the area. T9, T12 and T14 are both threonine residues, suggesting that motional restrictions are found for these residues. The presence of additional threonine residues

(T22, T55 and T66) in ubiquitin that do not deviate from expected random coil relaxation rates however disproves this assumption. Altogether, few deviations from the segmental motion model are found. This suggests that the segmental motion model describes the protein very well.

Model D: combination of model B and C. A combination of the segmental motion model (model C) and the volume dependent model (model B) as shown in Equation 21 was also tested.

$$21) \qquad R_2(i) = R_{int} \underbrace{\sum_{j=1}^{N} e^{-\frac{|i-j|}{\lambda_0}}}_{\substack{segmental \\ motion}} + k \underbrace{\sum_{j=1}^{N} \tau_j e^{-\frac{|i-j|}{\lambda_0}}}_{\substack{volume \\ dependence}}$$

The best fit, however, was found for the R_g independent case (k=0), yielding the same equation as for the segmental motion model alone.

Comparison of the four models shown above reveals that the segmental motion model describes the relaxation data in unfolded states of proteins best. Relaxation rates in the unfolded state are therefore independent from the volume and hydrophobicity of the side chains involved, and are only governed by the segmental motions.

3.4 Conclusion

The data presented here show that unfolded ubiquitin is a good model protein for unfolded states of proteins without any residual structure, thus for a protein in a random coil. $^3J(H^N,H^\alpha)$, $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants and chemical shift deviations reveal the absence of any secondary structure elements. Uniform relaxation data along the sequence indicate the absence of motional restriction caused by residual secondary structure and long-range interactions.

Comparison of different models used for the description of relaxation data in unfolded state revealed the predictive power of the very simple segmental motion model. Relaxation data in unfolded states without residual structure are best predicted independently from residue types involved, solely depending on segmental motions.

The segmental motion model can be used to predict residual structure in unfolded proteins simply by comparing experimental relaxation data with predictions from the segmental motion model.

Deviations from the segmental motions are due to residual structure rather than due to residue specific relaxation properties.

Chapter 4

Residual structure and long-range interactions in unfolded lysozyme

4 Residual structure and long-range interactions in unfolded lysozyme

4.1 Introduction

In this chapter, the ensemble of structures present in unfolded lysozyme was investigated using NMR spectroscopic methods in combination with single point mutagenesis.

Earlier investigations (Schwalbe et al., 1997) have revealed the presence of deviations from random coil behavior in the urea denatured state of lysozyme (pH 2, 8M urea). In contrast to the unfolded state of ubiquitin where no residual structure is found (chapter 3), chemical shift deviations and significant deviations from the R_2 relaxation rates expected for a random coil (segmental motion model, chapter 3) reveal the presence of residual secondary structure in lysozyme. Here, the ensemble of conformations present in the unfolded state where the disulfides are reduced and methylated has been investigated. While local residual secondary structure in unfolded states can be easily identified by chemical shift measurements or coupling constants, the identification of longer-range interactions is, in contrast, not as straight forward: due to the conformational averaging in unfolded proteins, it is mostly not possible to detect NOEs other than short or medium range NOEs. Here, a new method is developed allowing the identification of long-range interactions by a combination of mutations with NMR techniques, such as relaxation and diffusion rate measurements.

The expression and purification of single point mutants by Ryoma Hirano, Tadashi Ueda and Taiji Imoto was crucial for the work presented in this chapter.

4.2 Materials and Methods

4.2.1 Sample preparation/methylation

Construction of single point mutants (A9G, W62G, 62Y, W111G and W123G) of hen egg white lysozyme and expression and purification of ^{15}N labeled protein in *Saccharomyces cerevisae* was carried out by Ryoma Hirano, Tadashi Ueda and Taiji Imoto as described in previous publications (Ueda et al., 1994; Ueda et al., 1996; Ueda et al., 1990).

Reduced and methylated lysozyme (WT-S^{ME}) and mutants (A9G-S^{ME}, W62G-S^{ME}, W62Y-S^{ME}, W111G-S^{ME} and W123G-S^{ME}) were prepared following a procedure of Heinrikson (Heinrikson, 1971) from the folded oxidized ^{15}N labeled lysozyme.

2mg of ^{15}N labeled lysozyme (mutant) were dissolved in 0.5ml of denaturing solution (6M GdnCl, 0.25M Tris (pH 8.6), 3.3mM EDTA, 25% (v/v) acetonitrile). The solution was flushed with N_2 for 60 seconds, before a 12 fold excess over the total amount of -S groups (four disulfide bridges=eigth –S groups) of mercaptoethanol was added. After incubation at 50°C for 1h, the solution was cooled to 37°C and a 1.5 molar excess (over the total number of sulfhydryl groups in the protein) of methyl-*p*-nitrobenzenesulfonate in acetonitrile was added from a 0.6M stock solution. After 2h, the mixture was dialyzed against water (pH2).

4.2.2 NMR measurements

All NMR experiments were recorded at a temperature of 20°C on a four-channel Bruker DRX600 spectrometer equipped with a CryoProbe with z-gradients.

4.2.3 NMR assignment measurements

The NMR samples each contained between 60 to 100μM protein in water (10% D2O) pH 2. 3D NOESY-HSQC experiments (Marion et al., 1989a) using sensitivity enhancement for the back transfer (Schleucher et al., 1994) were recorded with a mixing time of 250ms. The recycle delay was set to 1.3s. Data sets were collected with 100, 48 and 1024 complex points in t_1, t_2 and t_3 respectively, with resolutions of 82Hz, 31.5Hz and 8.2Hz. Data processing was done using the program Xwinnmr version 3.5 (Bruker, Karlsruhe) with zero-filling of data sets to yield a final matrix size of 256*256*2048 points. Data analysis was performed using the program XEASY1.5 (Bartels et al., 1995).

4.2.4 Chemical shift analysis

The chemical shift deviations of the H^N and H^α chemical shifts of the methylated lysozyme mutants from random coil chemical shifts were determined as described in chapter 3.

C_α chemical shifts in WT-S^{ME} were obtained from the assignment of $^{13}C, ^{15}N$ labeled WT-S^{ME} by Christian Schlörb (Schlörb et al., in preparation).

4.2.5 R_2 measurements

Measurement, processing and extraction of R_2 relaxation rates was performed as described in chapter 3.

4.2.6 Diffusion measurements

Diffusion measurements were carried out in D_2O at pH 2 (pH corrected for D_2O). Protein concentrations ranged between 60 and 40μM and 300-200 μM dioxane was used as an internal standard. PFG (pulse field gradient) NMR diffusion measurements were performed with a PG-SLED (pulse gradient simulated echo longitudinal encode-decode) sequence using bipolar gradient pulses for diffusion (Jones et al., 1997; Wilkins et al., 1999; Wu et al., 1995a). The length of all the pulses and gradients in the sequence were held constant, whilst the strength of the diffusion gradients was varied between 2 and 95% in different experiments; each experiment was recorded with NS=64 and/or NS=128 and 32 experiments with different gradient strengths were acquired. The lengths of the diffusion gradients (6ms=2*3ms since a bipolar gradient was used, 0.3ms gradient recovery) and the stimulated echo (150ms) were optimized to give a total decay of between 80 and 90%. NMR spectra were acquired with 8k points and a spectral with of 8390Hz. Data processing and integration of signals and determination of diffusion rates (by fitting to a decay curve) were performed using the program Xwinnmr version 3.5 (Bruker, Karlsruhe).

The hydrodynamic radii of WT-S^{ME} and the single point mutants W62G-S^{ME}, W111G-S^{ME} and W123G-S^{ME} were determined by diffusion measurements. 1,4-dioxane was added to the protein samples as an internal reference with known hydrodynamic radius (2.12Å), since it has been shown that dioxane does not interact with proteins (Wilkins et al., 1999). Hydrodynamic radii (R_h) were thus calculated from NMR decay rates using the following formula:

$$8) \qquad R_h^{prot} = \frac{D_{diox}}{D_{prot}} R_h^{diox}$$

where D_{diox} and D_{prot} are the decay rates and R_h^{diox} and R_h^{prot} are the R_hs of dioxane and the protein respectively. Data in the literature (Wilkins et al., 1999) report aggregated species of methylated lysozyme (WT-S^{ME}) in exchange with the unfolded form. These are visible as an increase in the apparent R_h in the concentration range of 0.2-0.25mM. Here, aggregation is excluded by varying the protein concentration between 40 and 60μM for each mutant. The measured R_h (for the WT as well as the mutants) remained constant, indicating that no there was no aggregated protein present in the very dilute NMR-samples.

4.2.7 Circular dichroism measurements

CD spectra were recorded by Christian Schlörb (Schlörb et al., in preparation). The mean helix content was calculated using the approach of Rohl and Baldwin (Rohl and Baldwin, 1997).

4.2.8 Hydrophobicity predictions

Hydrophobicities were predicted as described in chapter 3.

4.2.9 MD simulations

MD simulations were carried out using the program CHARMM version c30a1 (Brooks et al., 1983) and the EEF1 potential (Lazaridis and Karplus, 1999). Simulations were run for 10ns at high temperature (700K), initiated from a linear, energy minimized structure (200 steps, also using CHARMM/EEF1).

100 of the generated structures were chosen, based on their radius of gyration, to sample the range of possible hydrodynamic radii (R_h). The hydrodynamic radii of the 100 chosen structures were calculated using HYDROPRO (de la Torre et al., 2000). The default settings and automatic extrapolation to zero bead size was used. Using the X-ray structure of hen lysozyme (Vaney et al., 1996) and the known experimental R_h (20.5Å) from the literature (Wilkins et al., 1999), the default settings of hydropro were tested prior to use. The structures shown were chosen based on their R_h.

4.3 Results and Discussion

4.3.1 Residual secondary structure in WT-S^{ME}

To obtain a model system for the study of non-native states of lysozyme, the protein was denatured and the four native disulfide bridges were reduced and methylated (indicated by –S^{ME}) and studied in water at pH 2. This reduced state of lysozyme remains unstructured even at near physiological pH in water (pH 6).

The H^N and H$^\alpha$ NMR chemical shifts of WT-S^{ME} are close to random coil chemical shifts, indicating that the reduced protein is unstructured in water at pH 2. Remaining chemical shift perturbations of WT-S^{ME} from chemical shifts measured in small unstructured peptides (Wishart et al., 1995a) are shown in Figure 30A and B. Significant deviations are found for Gly22 (H^N), for Val29 and Cys-S^{ME}30 (both H$^\alpha$), around Trp62/Trp63 (S60, R61 (H$^\alpha$); W62, W63 and C64 (H^N)) and around Trp108/Trp111 (V109, A110, W111,R112 (H$^\alpha$)). Large deviations from random coil chemical shifts are predominantly found for residues that follow tryptophan residues, e.g. W28-V29, W62-W63-C64, W108-V109 and W111-R112. These deviations remain if sequence dependent corrections (Schwarzinger et al., 2001) of random coil chemical shifts are applied (data not shown). Furthermore, regions of downfield shifted H$^\alpha$ resonances are present for residues 19 to 32, 58 to 64, 106 to 113 and 119 to 124, indicating a slight helical character for these regions. In order to compare the extent of residual helical character of the polypeptide chain as reported from NMR chemical shift measurements with predictions based on CD measurements, the helical character as defined by percentage of induced helicity (%hel) has been calculated. %hel was derived from a comparison of measured C$_\alpha$ chemical shifts (δC$_\alpha^{exp}$) in WT-S^{ME} to random coil chemical shifts (δC$_\alpha^{rc}$) (Wishart and Sykes, 1994; Wishart et al., 1992) and statistically derived chemical shifts in α-helices (δC$_\alpha^{hel}$) (Zhang et al., 2003) according to the formula:

$$22) \qquad \%hel = 100\, \frac{\delta C_\alpha^{exp} - \delta C_\alpha^{rc}}{\delta C_\alpha^{hel} - \delta C_\alpha^{rc}}$$

The percentage of induced helicity as function of the residue is shown in Figure 30C. High induced helicity is found for residues 8-12 and around Trp108/Trp111 and a mean helicity of 12.3% was calculated. This is in good agreement with the prediction of 14.4% helicity derived from CD measurements using the scale of Rohl and Baldwin (Rohl and Baldwin, 1997) (Figure 30D).
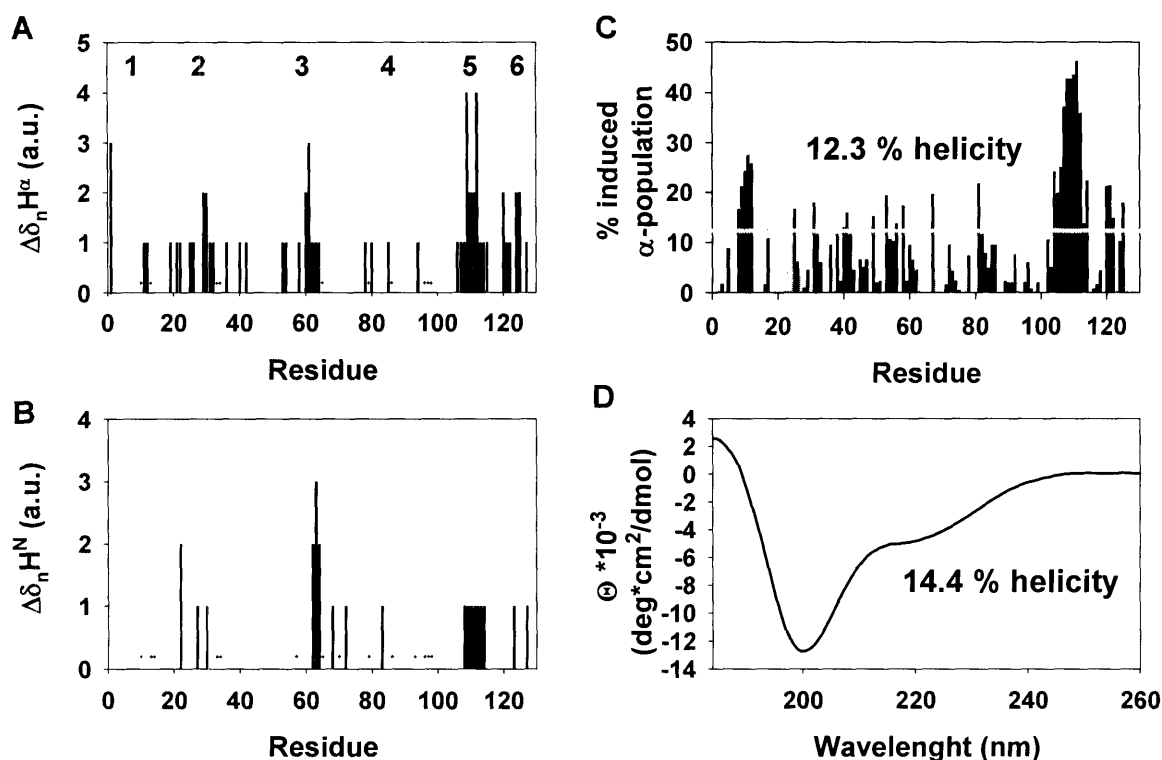
Figure 30: Residual secondary structure of WT-S^{ME} in water at pH 2. A) and B): Normalized H$^{\alpha}$ (A) and H^{N} (B) chemical shift perturbations in WT-S^{ME} from the chemical shifts measured of short unstructured peptides (Wishart et al., 1995a) ($\Delta\delta=\delta^{exp}-\delta^{rc}$), plotted as a function of protein sequence. The bars indicating the normalized chemical shift perturbations are normalized ($\delta\Delta_n$) according to the following scheme: H^{N} ($\delta\Delta_n$H^{N}=0 for $\delta\Delta$>-0.2, $\delta\Delta_n$H^{N}=1 for $\delta\Delta$<-0.2, $\delta\Delta_n$H^{N}=2 for $\delta\Delta$<-0.4, $\delta\Delta_n$H^{N}=3 for $\delta\Delta$<-0.6) and H$^{\alpha}$ ($\delta\Delta_n$H$^{\alpha}$=0 for $\delta\Delta$>-0.1, $\delta\Delta_n$H$^{\alpha}$=1 for $\delta\Delta$<-0.1, $\delta\Delta_n$H$^{\alpha}$=2 for $\delta\Delta$<-0.2, $\delta\Delta_n$H$^{\alpha}$=3 for $\delta\Delta$<-0.3, $\delta\Delta_n$H$^{\alpha}$=4 for $\delta\Delta$<-0.4). Comparisons with chemical shift indices from other sources (other peptides and databases) (Merutka et al., 1995; Schwarzinger et al., 2001) qualitatively yield the same picture. C) Induced α-population (=induced helicity) in WT-S^{ME} is shown as black (positive induced helicity) and gray (negative induced helicity) vertical bars as a function of residue number. A horizontal line indicates the mean induced helicity of 12.3%. D) CD spectrum of WT-S^{ME} with calculated helicity of 14.4%.

4.3.2 Deviations from random coil relaxation rates coincide with hydrophobic residues in lysozyme

The R_2 heteronuclear relaxation rates of WT-S^{ME} in water (pH 2) are shown in Figure 31. In addition, the predicted relaxation rates based on the segmental motion model (chapter 3, introduction) are shown (Figure 31). Regions of elevated relaxation rates that cannot be described by the segmental motion model (R_2^{rc}) are observed.
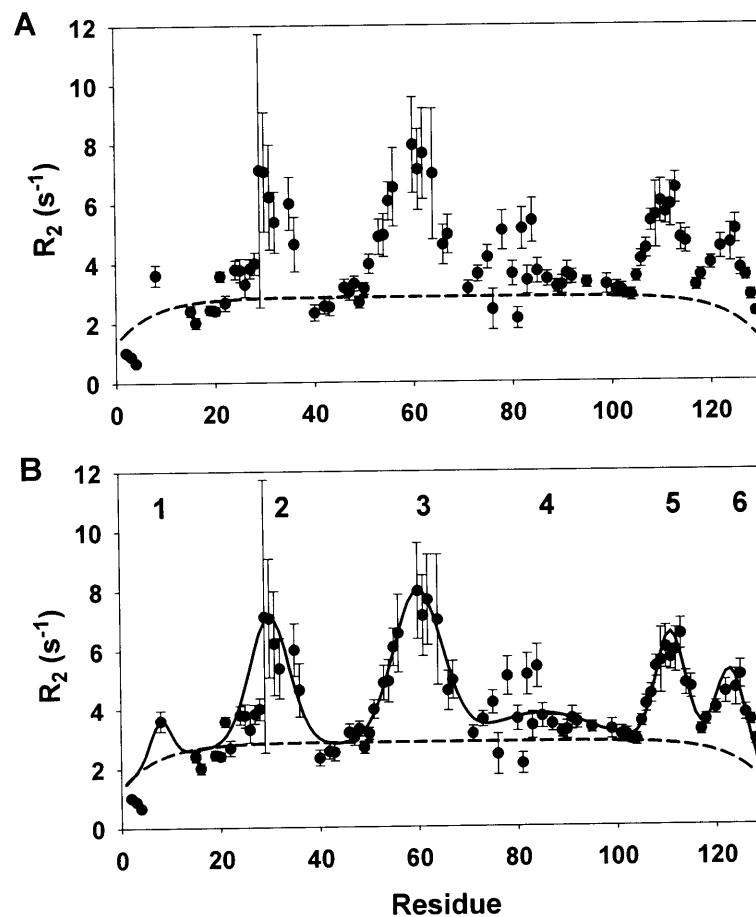
Figure 31: Transverse relaxation rates (R_2) of WT-S^{ME}. Experimental ^{15}N R_2 are shown as a scatter plot and the rates fitted by a model of segmental motion expected for a random coil with R_2^{rc} (R_{int}=0.2s^{-1} and λ_0=7) are shown as dashed lines in panel A and B. Black lines show the Gaussian fit. The clusters identified from Gaussian fitting are indicated by numbers 1 through 6 (panel B).

R_2 relaxation rates are sensitive to backbone motions on the subnanosecond time scale and to slow conformational exchange in the microsecond to millisecond time scale (Wagner, 1993). For reduced and methylated lysozyme however, comparison of R_2 and $R_{1\rho}$ experiments in urea exclude the contribution of pronounced slow conformational exchange to the measured R_2 rates (Schwalbe et al., 1997). The deviations from the segmental motion model are thus due to restrictions of backbone motions, either due to residual secondary structure and/or to long-range interactions in the unfolded state. A Gaussian distribution best describes the shape of the regions exhibiting deviations. Thus, a model is proposed that includes two components, the segmental motion part (R_2^{rc}) and a Gaussian term (R_2^{clus}), as shown in Equation 6:

$$6) \qquad R_2^{un}(i) = \underbrace{R_{int} \sum_{j=1}^{N} e^{-\frac{|i-j|}{\lambda_0}}}_{R_2^{rc}} + \underbrace{\sum_{cluster} R_{cluster} e^{-0.5\left(\frac{(i-x_{cluster})}{\lambda_{cluster}}\right)^2}}_{R_2^{clus}}$$

These Gaussian clusters are characterized by the position of the cluster in the protein (residue number) $x_{cluster}$, the width of the cluster $\lambda_{cluster}$ and a distinct relaxation rate for each cluster, $R_{cluster}$.

Six clusters can be identified using Equation 6 as shown in Figure 31B. These clusters have their centers at the following residues: 1) L8, 2) C30, 3) S60, 4) S85, 5) W111, and 6) W123. The clusters vary dramatically in their widths ($\lambda_{cluster}$) and their amplitude ($R_{cluster}$), as can be seen in Table 8.

cluster	$x_{cluster}$	$R_{cluster}$	$\lambda_{cluster}$
1	8	1.30	2
2	30	4.35	4
3	60	5.16	5
4	85	1.00	10
5	111	3.80	3
6	123	3.00	3

Table 8: Fitting results for the fitting of Gaussian clusters in WT-S^{ME}, using Equation 6. The final parameters for the segmental motion part (R_2^{rc}) are R_{int}=0.2s^{-1} and λ_0=7.

The highest cluster is cluster 3, with an amplitude of 5.16s^{-1}. Clusters 2, 5 and 6 are also quite pronounced with amplitudes between 3 and 4.35s^{-1}. The least pronounced cluster is cluster 4. The cluster is very flat ($R_{cluster}$=1) and very broad ($\lambda_{cluster}$=10), resulting in a barely visible cluster. Strikingly hydrophobic and specifically aromatic residues are present in most of these clusters: Two consecutive tryptophan residues (W62 and W63) are in the highest cluster, cluster 3.

Tryptophans are also found in cluster 2 (W28), cluster 5 (W108 and W111) and in cluster 6 (W123).

For a more detailed analysis of this finding, the distribution of hydrophobic amino acids in the lysozyme sequence is shown in Figure 32A. These data can also be fitted by Gaussian modeling to 6 distinct clusters, numbered 1 through 6 with maximal hydrophobicity values around residues (1) C6, (2) W28, (3) L56, (4) L83, (5) W108, and (6) W123. It is widely recognized that hydrophobic amino acids affect the folded state of a protein; mapping of the hydrophobic clusters onto the structure of lysozyme (Vaney et al., 1996), as shown in Figure 32B, therefore shows that the clusters ultimately result in the hydrophobic cores of the α- and β-domain of lysozyme.

However, comparison of the location of the hydrophobic clusters with data on the unfolded state of lysozyme reveals that the location of hydrophobic clusters not only has an impact on the folded state, but also on the unfolded states. (1) Residual non-random secondary structure elements tend to cluster at similar positions, in particular around Gly22, Val29, Trp62/Trp63, and Trp108/Trp111 (see text above and Figure 30). (2) The positions of the hydrophobic clusters overlaps to a remarkable extent with the average restrictions in conformational space in unfolded ensembles of lysozyme as identified by relaxation rate measurements (see text above and Figure 31). In contrast to the excellent correlation between the positions of the experimentally determined clusters of residual structure and those of all of the hydrophobic clusters, the intensities of the clusters do not match well for clusters 1 and 4, where barely elevated relaxation rates are found. Interestingly, these two clusters are the only two regions of predicted increased hydrophobicity (Figure 32) without tryptophan residues. (3) Experimentally determined folding core residues map to the central parts of clusters 1 through 4 (Buck et al., 1994; Miranker et al., 1991; Nash and Jonas, 1997a; Radford et al., 1992b).

98
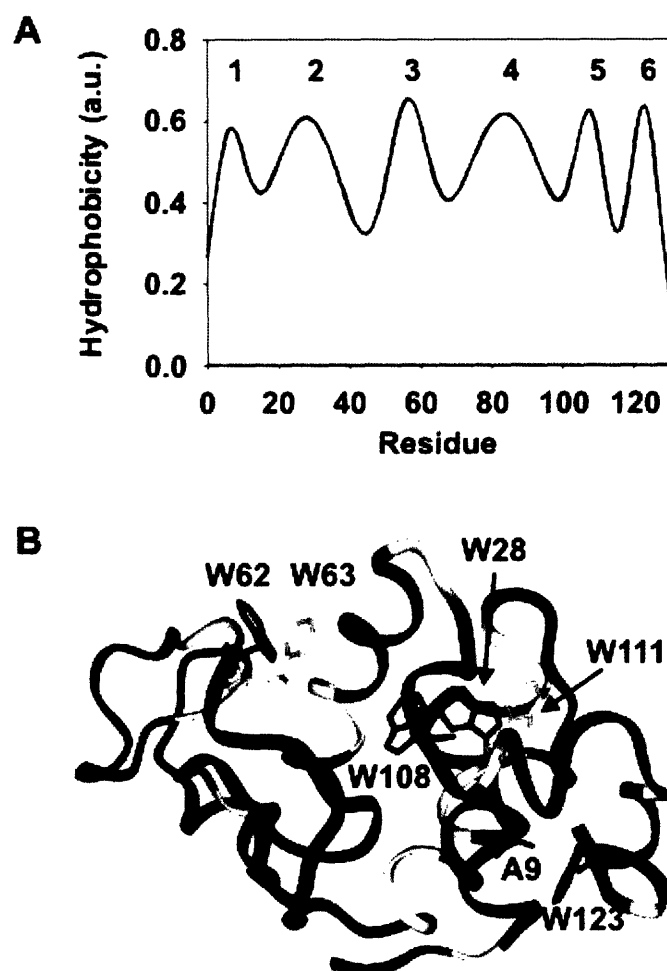
Figure 32: Distribution of hydrophobicities in lysozyme. A) Normalized (0-1) hydrophobicities in arbitrary units (a.u.) according to the hydrophobicity scheme of Abraham and Leo (Abraham and Leo, 1987). Gaussian least square fitting of hydrophobic clusters is shown as a black line, and clusters are numbered from 1 through 6. B) Position of Ala9 and all tryptophans in the native state. The lysozyme crystal structure (pdb: 193L (Vaney et al., 1996)) shown is colored according to the hydrophobicity values of the Gaussian fit (red, non-hydrophobic (0-0.46); blue, hydrophobic (0.52-1); for comparison: Glycine=0.49).

It is interesting to find out whether experimentally determined core folding residues are found in clusters 1 through 4 by pure coincidence or whether the clusters in unfolded lysozyme interact with each other and thereby produce a nucleus for protein folding. As unfolded proteins are averages of inter-converting fluctuating conformations, a single set of tertiary contacts in the unfolded state cannot be determined. Therefore, a new method is introduced to monitor the long-range interactions in unfolded proteins through a combination of non-conservative single point mutations and NMR measurements.

To investigate possible long-range interactions, the influence of two types of single point mutations on the compactness and interactions of clusters was studied. (1) replacement of tryptophan with either tyrosine or glycine (W62G/Y) and (2) replacement of a central amino acid with emphasis on tryptophans in clusters 1, 3, 5, and 6 with glycine (A9G, W62G, W111G, and W123G). Cluster 2 was not investigated as the Trp28 mutant protein was too unstable to obtain sufficient amounts of protein. Cluster 4 was not studied because the WT-S^{ME} already displays very little deviation from the random coil in this region.

4.3.3 Mutations do not affect the conformation of the native state

The positions of Ala9 and all the tryptophans in the native state of lysozyme are shown in Figure 32B. Figure 33 shows ^{1}H,^{15}N-HSQC spectra under native conditions (Schwalbe et al., 2001) of WT-S^{ME}, W62G-S^{ME}, A9G-S^{ME}, W111G-S^{ME} and W123G-S^{ME}. The number of amide signals and the chemical shift dispersion observed in all of the spectra indicates that all the mutants are folded under native conditions. Furthermore, only minimal changes in the chemical shifts of the mutants compared to those of WT-S^{ME} are observed; thus, none of the mutantions causes a significant change in the conformation of the native state.

100
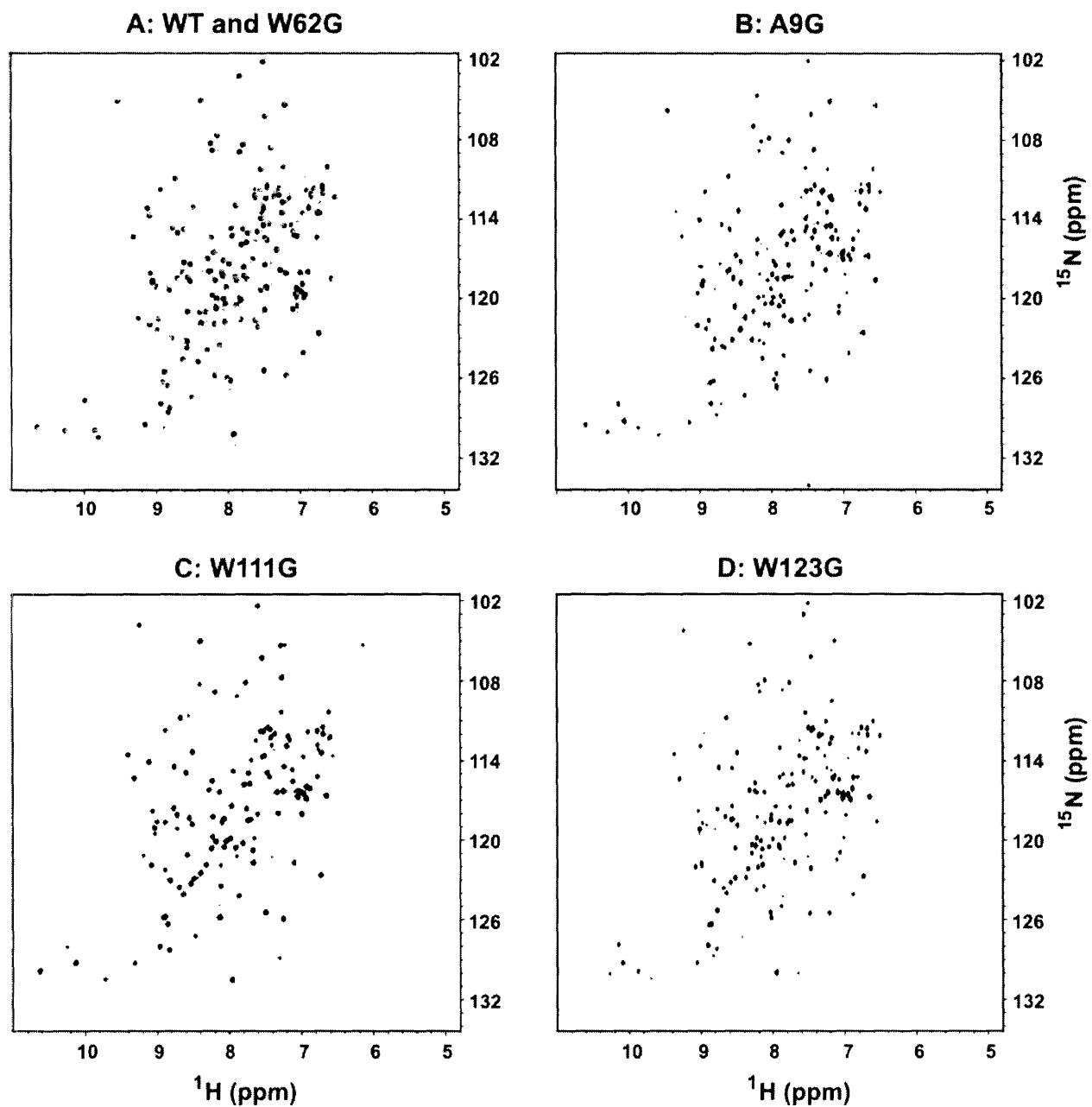
Figure 33: ^{1}H,^{15}N HSQC spectra of native lysozyme and mutants A: WT (gray) and W62G (black), B, C, D A9G, W111G and W123G as indicated. The spectra were recorded in unbuffered solution at pH 3-4.

4.3.4 Assignment of reduced mutants

The ^{1}H,^{15}N HSQC spectra of the reduced and methylated single point mutants in water at pH 2 do not vary significantly from that of WT-S^{ME} in water. An example is shown in Figure 34. The spectra of WT-S^{ME} and W62G-S^{ME} overlay very well.
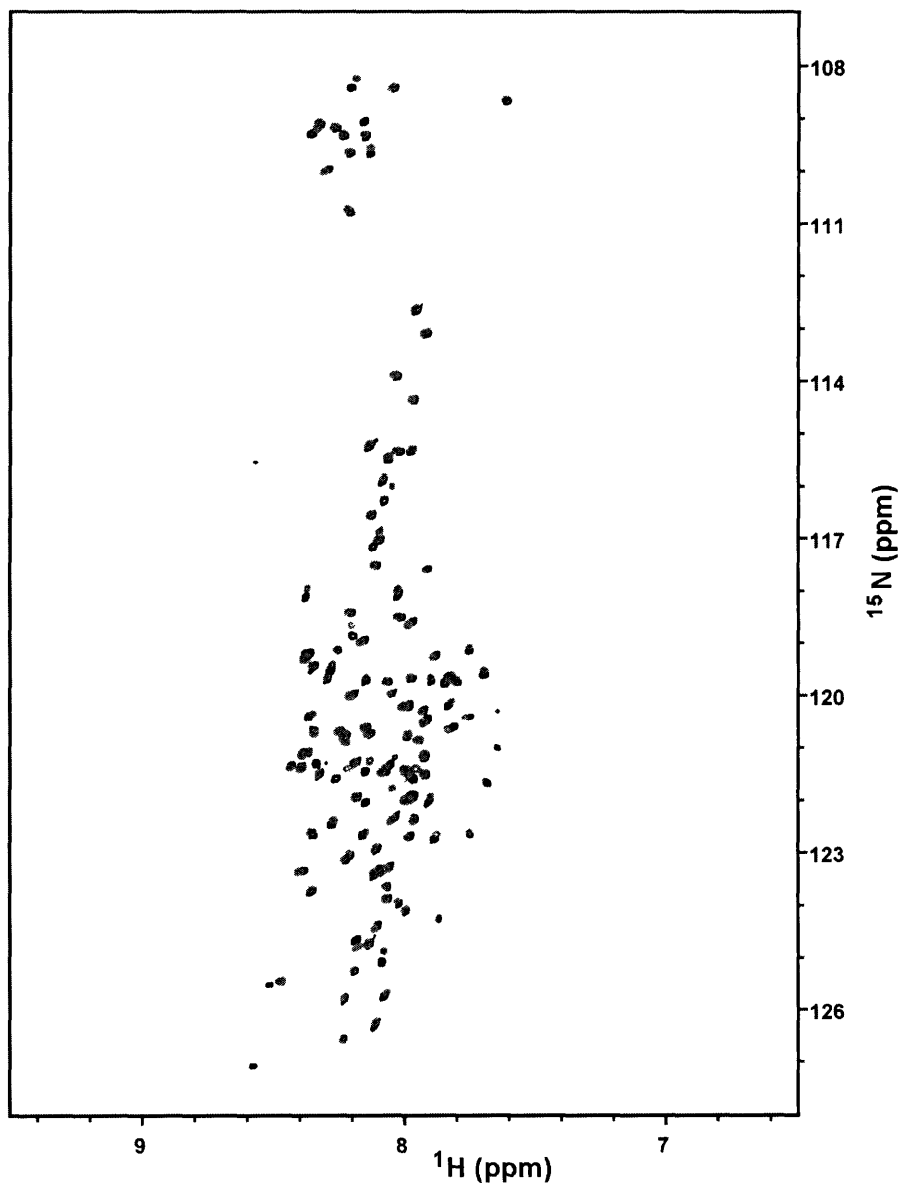
Figure 34: Overlaid ^{1}H,^{15}N HSQC spectra of WT-S^{ME} and W62G-S^{ME} in water at pH 2. In gray is the spectrum of WT-S^{ME} black signals arise from W62G-S^{ME}. The differences between the mutant and the WT are discussed in chapter 4.3.5.

A: A9G-S^{ME}

B: W62Y-S^{ME}

C: W111G-S^{ME}
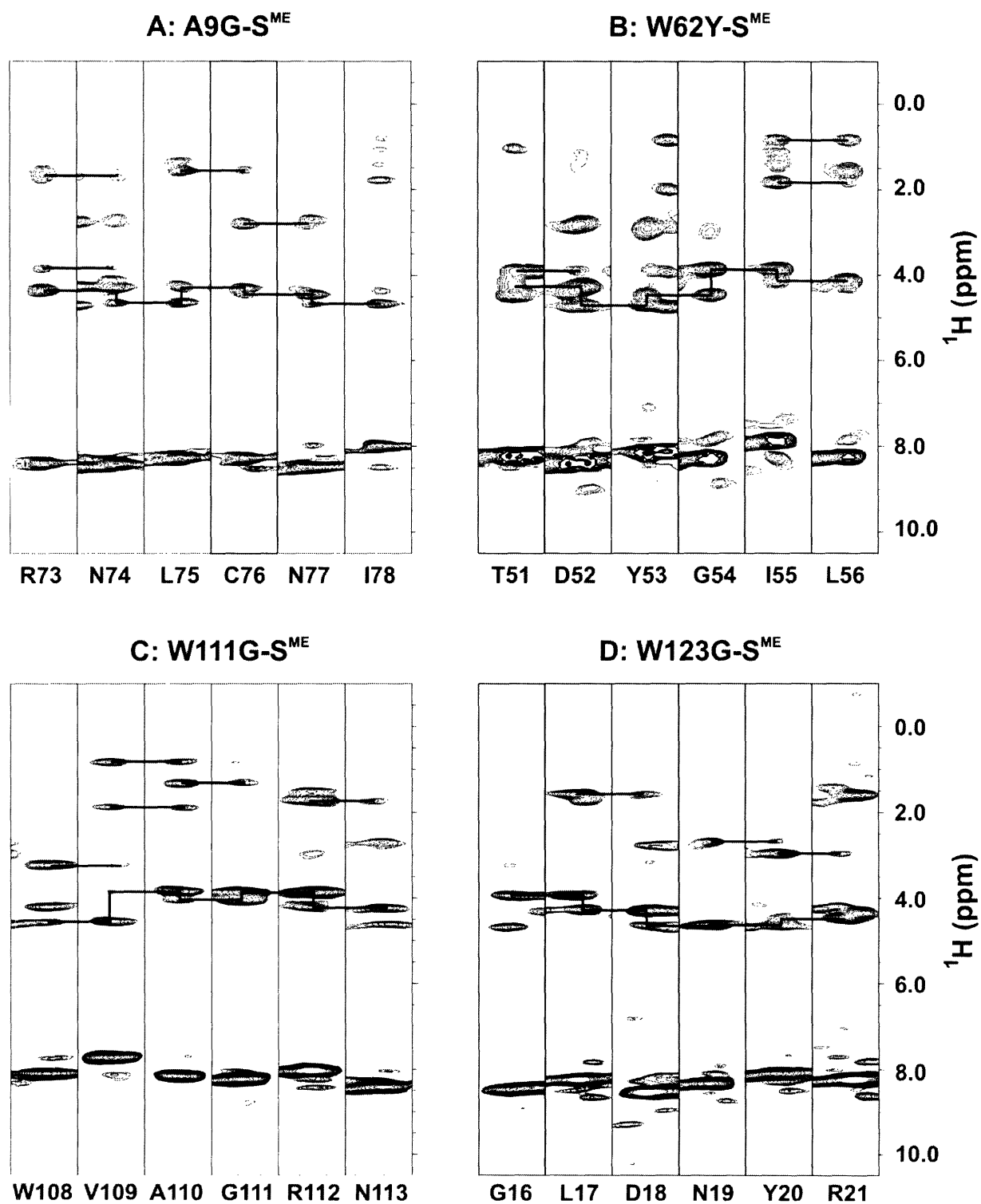
D: W123G-S^{ME}

Figure 35: ω1,ω3 strip plots from 3D NOESY-HSQC spectra of A) A9G-S^{ME}, B) W62Y-S^{ME}, C) W111G-S^{ME} and D) W123G-S^{ME} assignment. Residues 73-78, 51-56, 108-113 and 16-21 are shown in A: A9G-S^{ME}, B: W62Y-S^{ME}, C: W111G-S^{ME} and D: W123G-S^{ME}, respectively.

Starting out from the similarity of the mutant and WT ^{1}H,^{15}N HSQC spectra, assignment of the reduced lysozyme mutants was carried out using 3D NOESY-HSQC measurements. Figure 35 shows selected strip plots used for the assignment of A9G, W62Y, W111G and W123G.

Despite the small chemical shift resolution in the ^{1}H dimension of the spectra which is characteristic for unfolded proteins, the ^{15}N resolution is better and in particular, glycine, threonine and serine residues are well separated by their ^{15}N chemical shifts from the resonances of all the other amino acids, facilitating their assignment. In addition, in contrast to spectra of native proteins, where chemical shifts are substantially influenced by tertiary interactions, H$^\alpha$ and H$^\beta$ and sometimes even H$^\gamma$ resonances in unfolded proteins resonate at their random coil chemical shifts. This allows the identification of the respective amino acid in a so-called "NOESY strip": isoleucine residues for example, have characteristic signals between 0.5 and 2 ppm (Figure 35, panel A I78 and panel B I55). Besides the easy identification of a number of amino acids from their characteristic random coil chemical shifts the most helpful feature for the assignment of the lysozyme mutants were H^{N}H$^\alpha$ NOEs: in each strip H^{N}H$^\alpha$(i,i) and H^{N}H$^\alpha$(i,i-1) are present, as well as a number of (i,i) and (i,i-1) NOEs to the side chains.

Using the described approach 89% of the backbone amides could be assigned in A9G (112/126), 84% in W62G (106/126), 88% in W62Y (111/126), 85% in W111G (107/126) and 89% in W123G (112/126). The assignment rates obtained here are very similar to the assignment rate of 90% (114/126) that was obtained in WT-S^{ME} (Grimshaw). Mutated residues could be identified in A9G, W62G, W111G and W123G, but not in W62Y. This could be due either to peak overlap, or more probably, due to fast relaxation. Fast relaxation is also observed in WT-S^{ME}, where W62 is the weakest resonance of the spectrum.

4.3.5 Mutations do not affect residual secondary structure in unfolded lysozyme

The effect of the single amino acid replacements on the H$^\alpha$ chemical shifts is shown for all mutants in Figure 36. The data are presented as correlation plots between H$^\alpha$ chemical shifts in the single point mutants with those in WT-S^{ME}.
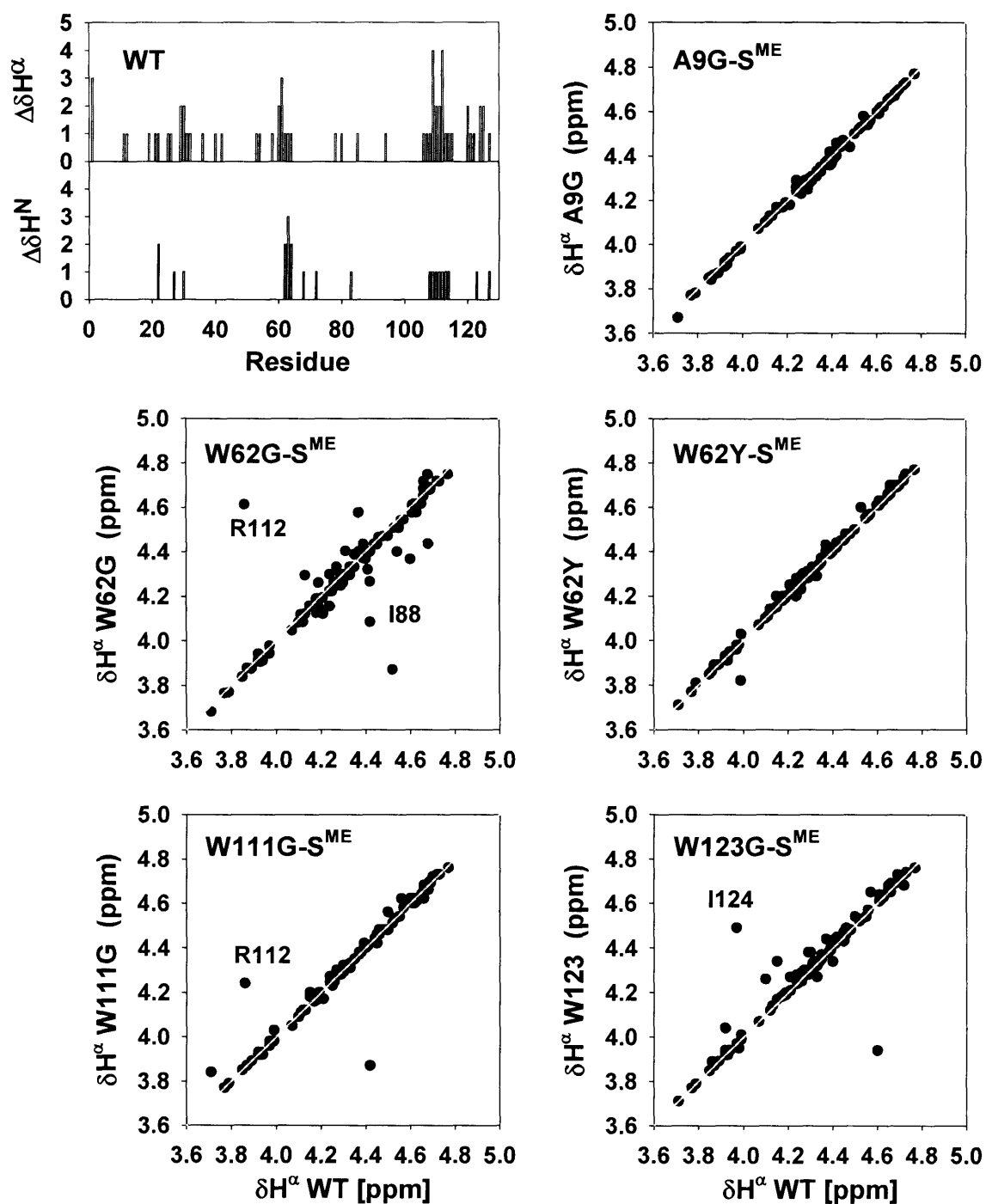
Figure 36: Correlation plots of the H^α chemical shifts in A9G-S^{ME}, W62G-S^{ME}, W62Y-S^{ME}, W111G-S^{ME} and W123G-S^{ME} with the H^α chemical shifts measured in WT-S^{ME}. The normalized H^α and H^N chemical shifts of WT-S^{ME} are shown again for better comparison. Outliers that do not correlate well are indicated in black, mutated residues are indicated in gray.

The conservative mutations W62Y and A9G result in chemical shifts virtually identical to those of the WT. The largest deviations were found for W62G, with wide-ranging effects, in particular on R112 and I88 which are distant from the point of mutation at position W62. W111G and W123G show predominantly local deviations, including the mutation site and its immediate neighbors in the sequence. Overall, however, deviations from WT H^{α} chemical shifts are small for all the mutants, even for the regions around the chosen mutation sites, with correlation coefficients ranging between 0.996 (A9G) and 0.849 (W62G). Deviations from WT H^{N} chemical shifts are even smaller, with a range between 0.938 (W123G) and 0.996 (W62Y). The local regions of residual secondary structure are therefore not disturbed by the mutations.

4.3.6 Residual tertiary contacts in unfolded lysozyme

Figure 37 shows the heteronuclear ^{15}N transverse relaxation rates (R_2^{exp}) of WT-S^{ME} and the five mutants: A9G-S^{ME}, W62Y-S^{ME}, W62G-S^{ME}, W111G-S^{ME} and W123G-S^{ME}. As described above, WT-S^{ME} displays six clusters centered at residues 1) L8, 2) C30, 3) S60, 4) S85, 5) W111, and 6) W123. While the hydrophobicity profile of the mutants is nearly unchanged in the mutants, the clusters based on R_2 relaxation measurements are partially diminished. The results of Gaussian fitting for the data of WT-S^{ME} and for the mutants which differ dramatically from either WT-S^{ME} or the segmental motion model (R_2^{rc}, Equation5) are summarized in Table 9.

	WT-S^{ME}			W111G-S^{ME}			W123G-S^{ME}		
	$x_{cluster}$	$R_{cluster}$	$\lambda_{cluster}$	$x_{cluster}$	$R_{cluster}$	$\lambda_{cluster}$	$x_{cluster}$	$R_{cluster}$	$\lambda_{cluster}$
1	8	1.30	2	10	0.81	3	10	1.64	1
2	30	4.35	4	29	1.57	7	32	1.09	11
3	60	5.16	5	60	3.31	4	60	3.45	4
4	85	1.00	10	71	0.87	24	81	0.69	7
5	111	3.80	3	117	0.53	9	111	2.54	3
6	123	3.00	3	125	2.07	1	125	0.57	4

Table 9: Results of Gaussian fitting for W111G-S^{ME} and W123G-S^{ME} using Equation 6: A cluster is characterized by its center ($x_{cluster}$), its amplitude ($R_{cluster}$) and its width ($\lambda_{cluster}$). Final fitting parameters used for the segmental motion dependent part are: $\lambda_0 = 7$ and $R_{int} = 0.2s^{-1}$. Fitting results for WT-S^{ME} are shown again for comparison.

The most drastic changes compared to WT-S^{ME} are found in W62G-S^{ME}: Replacement of Trp62 by glycine essentially abolishes not only its own cluster 3, but also clusters 1 through 4, and diminishes clusters 5 and 6. This result indicates that long-range interactions must be present in WT-S^{ME}. These are abolished by the non-conservative mutation. Thus, the tryptophan at position 62 is important for the stabilization of these long-range interactions.

In order to test whether this property of Trp62 is unique to tryptophan or may be more generally attributable to aromatic amino acids, the effect on the relaxation properties of unfolded lysozyme of replacing Trp62 by tyrosine was studied: The W62Y mutation has a minor effect on the relaxation properties of WT-S^{ME} compared to W62G-S^{ME}. The results of the Gaussian fitting of WT-S^{ME} are shown in the graph displaying the experimental relaxation rates of W62Y-S^{ME}. Deviations from the random coil are reasonably well described by the fit except for the region around W28 where minor deviations are found. The result presented here suggests that aromatic residues play an important role in the stabilization of long-range interactions in the unfolded states of lysozyme.

A9G-S^{ME} is almost identical to WT-S^{ME}. There is not even an effect on its own cluster, cluster 1. The hydrophobic clusters remain of virtually identical size and position, indicating that Ala9 is not important for stabilizing the hydrophobic core.

In contrast, the results of the replacement of tryptophan residues at positions 111 and 123 are very significant. When Trp111 (cluster 5) is replaced by glycine, the cluster where the mutation is located disappears entirely, and cluster 6 also decreases significantly in intensity. Loss is also observed in the intensity of cluster 2 (surrounding Trp28) and there is a small decrease in the intensity of cluster 3 (surrounding Trp62 and Trp63).

Similarly, when Trp123 at the C-terminus (cluster 6) is changed to a glycine, the cluster around the mutation site essentially disappears, and the intensities of clusters 2, 3, and 5 are lessened. The effect of W123G replacement is less significant than the one of W111G, as seen by the complete loss of cluster 5 in W111G-S^{ME} but not in W123G-S^{ME}. The largest changes in R_2 distribution as compared to WT-S^{ME} are observed with W62G-S^{ME}. W62G-S^{ME} appears to be complementary to W111G-S^{ME} and W123G-S^{ME}: while the W62G mutation causes dramatic disruption of clusters 1-4, and to a lesser extent of clusters 5-6, the W111G and W123G replacements cause more disruption in clusters 2, 5, and 6 and more subtle changes in cluster 3.
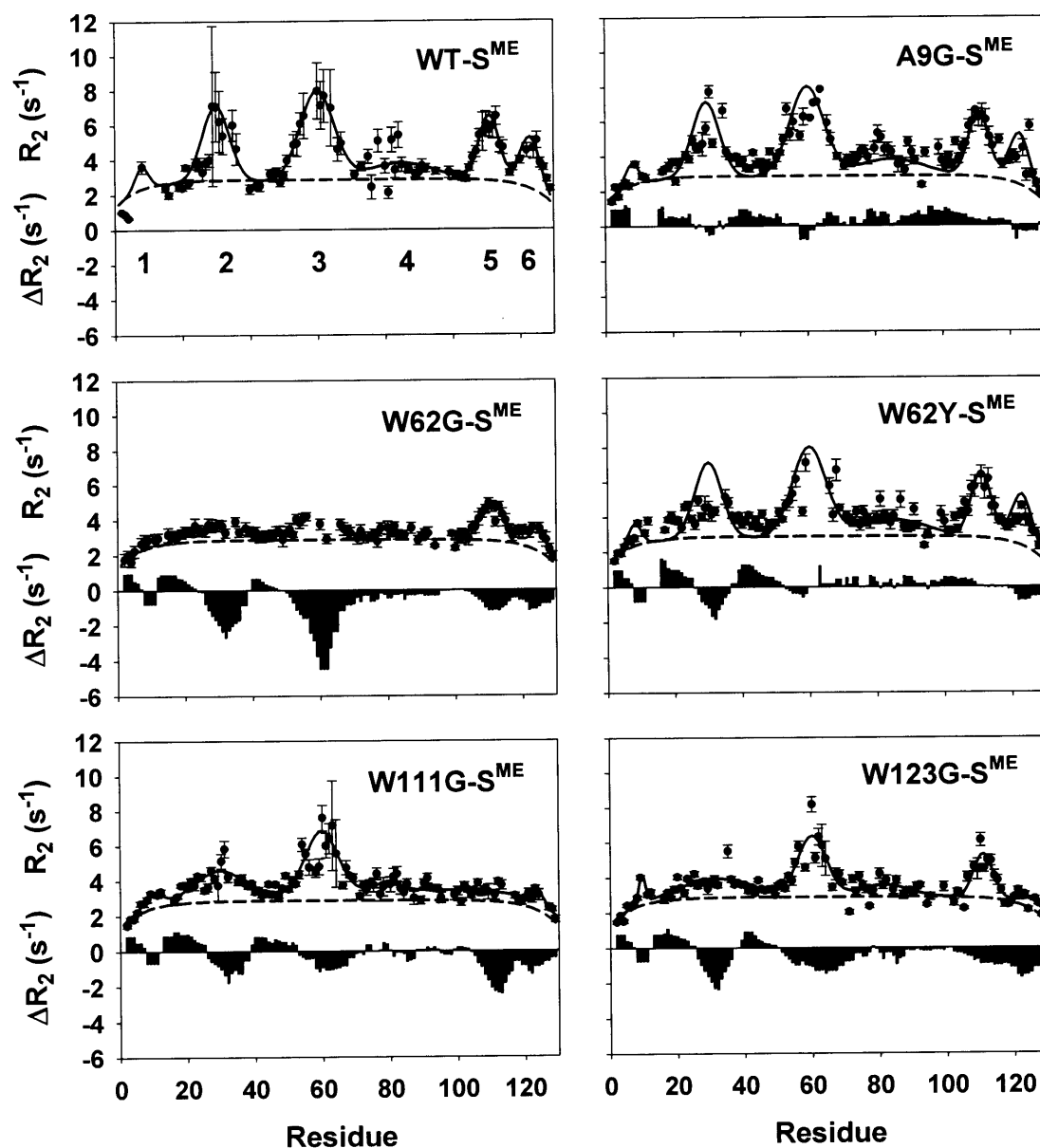
Figure 37: ^{15}N R_2 relaxation rates for WT-S^{ME} and A9G-S^{ME}, W62G-S^{ME}, W62Y-S^{ME}, W111G-S^{ME} and W123G-S^{ME} are shown. The experimental rates are shown as a scatter plot and the rates fitted by a model of segmental motion expected for a random coil with R_2^{rc} (Equation 5, $R_{int}=0.2s^{-1}$ and $\lambda_0=7$) are shown as dashed lines; black lines show Gaussian fits (Equation 6) for WT-S^{ME} (in panels of WT-S^{ME}, A9G-S^{ME}, W62Y-S^{ME}), W111G-S^{ME} and W123G-S^{ME}. Deviations of mutant R_2 relaxation rates from WT-S^{ME} relaxation rates (ΔR_2) are shown as bars. ΔR_2 was determined as (mutant-$S^{ME}(R_2)$-WT-$S^{ME}(R_2)$) and running average taken over 7 residues. Hydrophobic clusters are indicated by numbers 1 through 6 in the panel of WT-S^{ME}.

4.3.7 Compactness

The decrease of R_2 values upon single point mutations indicates that changes in the conformational ensembles of lysozyme have occurred. The diminishment of clusters far away in sequence from the respective mutation site indicates the presence of long-range interactions. The changes either reflect a shift in distribution from compact states where specific interactions are present to compact states where interactions are statistically distributed or more likely, a change from compact states to more extended states. In the second case, it is predicted that those point mutations that cause a change in relaxation profiles also cause a change in the compactness of the average unfolded state. A loss of conformational restriction in one or more of the hydrophobic clusters would increase the overall propensity for extended states in the ensemble. This hypothesis was directly tested by measuring the averaged hydrodynamic radii R_h, using NMR diffusion measurements of the different reduced and methylated proteins. Figure 38 shows the diffusion measurements for WT-S^{ME} and W62G-S^{ME}.
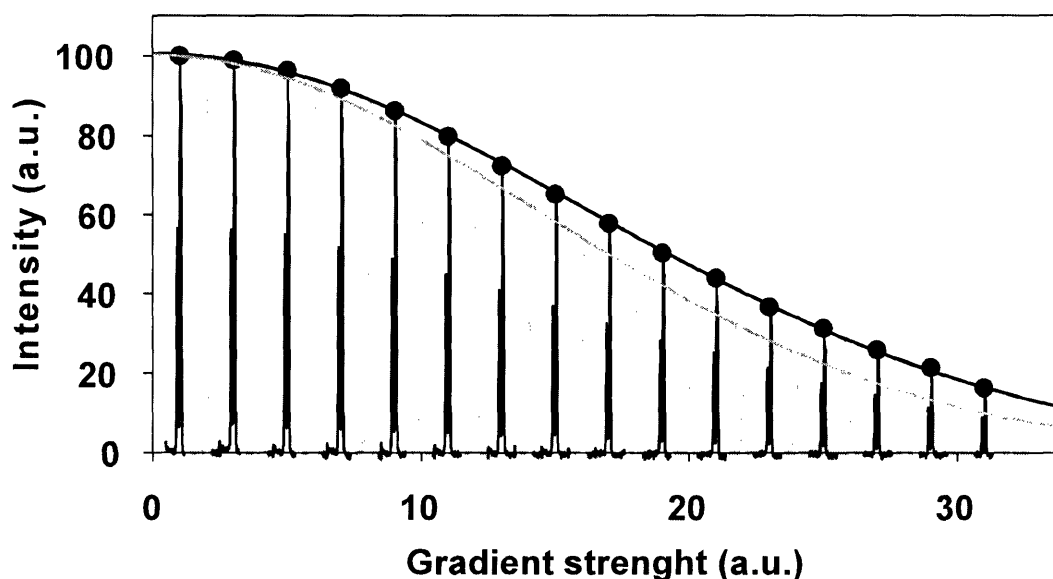
Figure 38: Diffusion experiments used for the determination of the hydrodynamic radii (R_h). NMR signal intensity (I) as a function of gradient strength (g) is shown for WT-S^{ME} (black) and W62G-S^{ME} (gray); the decay of the signal was fitted to a function of the form I=A*exp(-(Dg)^2). R_h values were calculated by comparison of protein decay rates (D) with decay rates of dioxane, as described in Methods.

The signal of WT-S^{ME} decays considerably faster than the signal of W62G-S^{ME}, indicating that WT-S^{ME} diffuses faster than W62G-S^{ME}. Summarizing the results of the other mutants (not shown) the following is observed: WT-S^{ME} and W123G-S^{ME} diffuse similarly fast, W111G-S^{ME} diffuses at intermediate velocity and W62G-S^{ME} diffuses the slowest. The R$_h$ values calculated from the diffusion rates (see methods) are 26.9±0.6Å, 25.4±1.2Å, 29.3±1.3Å and 32.1±0.6Å for WT-S^{ME}, W123G-S^{ME}, W111G-S^{ME}, and W62G-S^{ME}, respectively.

All R$_h$ values are considerably larger (at least 25%) than the R$_h$ of folded lysozyme (R$_h$=20.5Å) (Wilkins et al., 1999), indicating the extended nature of the reduced and methylated states in water at pH 2. However, the overall average shape of the ensemble of conformers markedly varies with the position of single point mutations in the different hydrophobic clusters. While a similar hydrodynamic radius was measured for WT-S^{ME} and W123G-S^{ME}, the hydrodynamic radius of W111G-S^{ME} is larger (R$_h$=29.3±1.3Å) and the W62G-S^{ME} mutant has the largest dimensions (R$_h$=32.1±0.6Å). This observation is visualized in Figure 39. The conformations shown were calculated using unbiased molecular dynamics simulations and randomly chosen to enable the visualization of the dimensions corresponding to a certain R$_h$; interactions present in the respective conformations shown are random and have no significance. As one can see, the R$_h$ of 32.1Å that was measured for W62G-S^{ME} corresponds to a very extended, almost linear conformation, consistent with the essentially random coil behavior observed for the relaxation rates of this mutant (Figure 37). The radii observed for the other mutants correspond to rather compact states, where tertiary contacts are clearly present. These states are much closer to the native state with R$_h$ of 20.5 Å than the very extended conformation with R$_h$ of 32.1 Å.
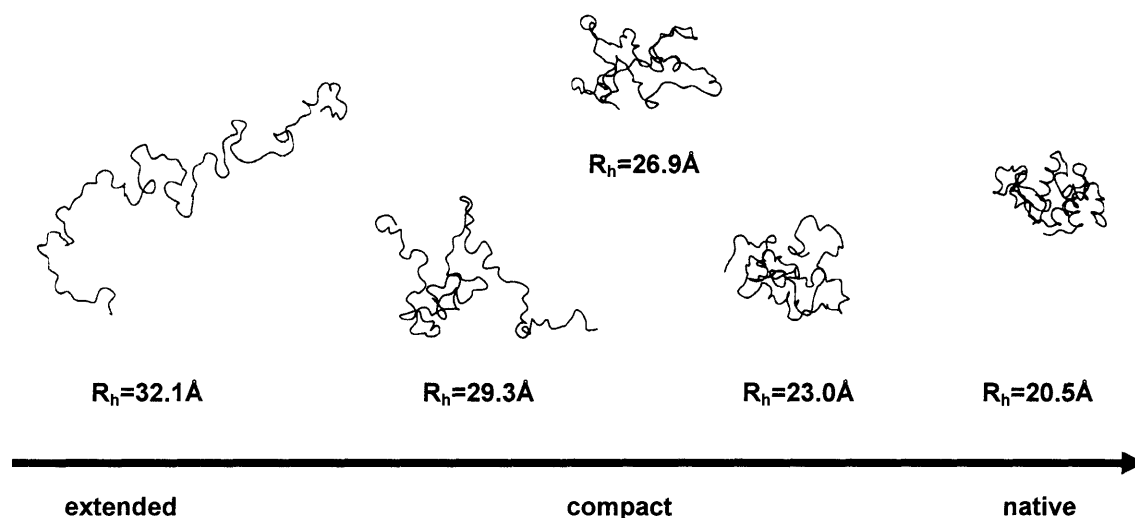
Figure 39: Randomly chosen conformations for a given R_h with the lysozyme amino acid sequence in order of their hydrodynamic radius. Conformations with increasing R_h with emphasis on experimental R_h values, are shown. The radii of different conformations were calculated using the program Hydropro (de la Torre et al., 2000). The diagram of the native structure (R_h=20.5Å) was prepared using the coordinates of the 193L data set of the pdb (Vaney et al., 1996).

4.3.8 Residual structure model from these data

The structures of individual conformers in the ensemble of inter-converting, fluctuating conformations in unfolded proteins cannot be measured experimentally but the average properties of this ensemble can be determined using NMR spectroscopy. The comparison between experimental NMR parameters and those expected for a true random coil show that there are six areas with strong deviations from random coil behavior in WT-S^{ME}. The observed hydrodynamic radius of 26.9±0.6Å indicates that the average conformations in WT-S^{ME} are rather compact (Figure 39), supporting the presence of extensive tertiary contacts in the unfolded state. These contacts are both native-like and non-native like (as discussed below). A graphic summarizing the different types of conformations depicting non-native and native interactions and for comparison with the conformation in the native state is shown in Figure 40. The color coding in the native state is according to the deviations from random coil behavior modeled by Gaussian fitting (Equation 6) in WT-S^{ME}, and kept for the description of possible non-native conformations.

A: Native state

B: Extended state

C: Compact states

I. Native-like

II. Native-like non-native-like

III. Non-native-like
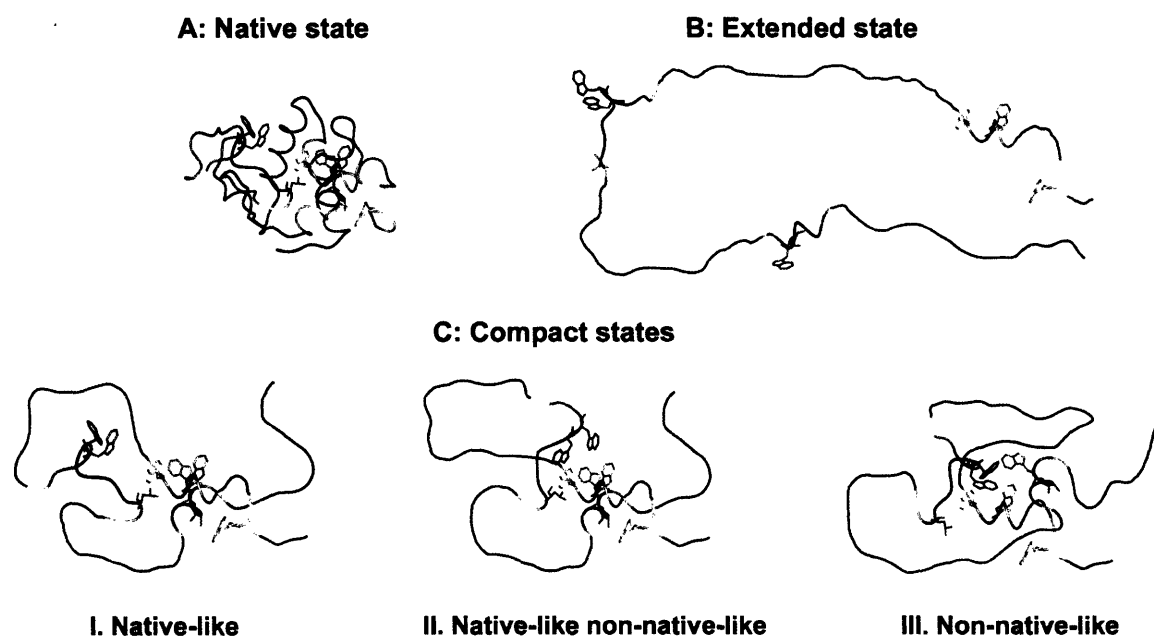
Figure 40: Illustrations of possible conformations in the native (A) extended unfolded (B) and compact unfolded states (C). The colors correspond to a mapping of the amount of restriction in conformational space based on R_2^{clus} (Equation 6) onto the native state structure (pdb:193L (Vaney et al., 1996)), where red indicates 0 % deviation, white 50 % deviation and blue 100 % deviation from R_2^{rc}. The largest R_2^{clus} of $5.16s^{-1}$ (centre of cluster 3) was taken as 100 % deviation. Fragments of the native state structure of highest deviation (corresponding to the blue areas in the structure) were taken and manually combined to visualize the unfolded state ensemble consisting of fully extended states (B), and native-like, native- and non-native-like, and non-native like unfolded compact states, shown in panels I, II and III in C, respectively.

The following conclusions can be drawn. Non-native contacts between cluster 3 (around Trp62 and Trp63, part of the β-domain in the native structure) and the clusters 1, 2, and 4 (part of the α-domain in the native structure) must be present in WT-S^{ME} since deviations from random coil relaxation behavior deplete in the W62G mutant (Figure 40C, panels II and III). Breakage of these long-range interactions results in a dramatic increase of the measured R_h from WT-S^{ME} to W62G-S^{ME} of 12% reflecting a shift towards very extended conformations (as depicted in Figure 40B). The presence of native-like contacts (Figure 40C, panel II) is suggested by the effect of the W123G mutation. This mutation simultaneously diminishes clusters 6 and 2, indicating native-like contacts between Trp123 and sites at or around Trp28. The importance of native-like

112

contacts is consolidated by the central role of the Trp111 interactions in WT-S^{ME}. Disruption of the contacts between Trp111 and the cluster surrounding Trp28 by the W111G mutation results in a large increase in R$_h$ of about 9%. Tryptophan residues or aromatic residues in general apparently play a special role in the stabilization of contacts in the compact states of unfolded lysozyme. Comparing the effects of replacing Trp62 with glycine or with tyrosine, there are little changes in W62Y-S^{ME} relaxation behaviour compared to that of WT-S^{ME}. In contrast, W62G-S^{ME} shows the strongest deviations from WT-S^{ME}. Aromatic residues therefore play a significant role in organizing nearby residues into arrangements that have restricted motions compared to a completely unfolded polypeptide chain.

4.3.9 Implications for lysozyme folding/disulfide bond formation

As shown here, interacting hydrophobic clusters in unfolded proteins might prevent unfolded proteins from aggregation by burying hydrophobic areas. Furthermore, the presence of residual structure in reduced unfolded proteins could guide the earliest steps of folding. In the case of lysozyme, it is indeed very likely that long-range interactions between the hydrophobic clusters seen here might help to form the correct disulfide bonds during oxidative folding of proteins. Results from oxidative refolding studies (van den Berg et al., 1999) showed that three out of the four disulfide bridges of native HEWL are formed quickly during refolding. Three different three-disulfide intermediates exist. The C30-C115 disulfide bridge is the only disulfide bridge always formed during the fast folding step at the beginning of the folding pathway. The results presented here suggest that hydrophobic clusters (cluster 2 and cluster 5 and 6) build a core that enables the fast formation of this particular disulfide bond.

4.3.10 Implications for amyloid formation

The results presented here for the model system for protein folding, hen lysozyme, have important implications for our understanding of the molecular basis of misfolding diseases. Native structures of the amyloidogenic variants of human lysozyme (>60% sequence identity with hen lysozyme) do not differ considerably from the WT structure but dramatically vary in the rate of amyloid formation (Chamberlain et al., 2001). Amyloids of hen lysozyme are also formed from the native (Goda et al., 2000; Krebs et al., 2000) and the WT-S^{ME} polypeptide chain (Cao et al., 2004). Here, we show that non-conservative single point mutations in hen lysozyme, which drastically alter the ensemble of unfolded conformations, do not change the structure of the native

proteins as deduced from native ^{15}N-HSQC measurements (Figure 33). Here it is shown that point mutations induce a shift in the distribution of conformers from compact to more extended states to differing degrees depending on the location of the mutation with respect to the hydrophobic clusters present. Such changes are expected to result in differences in the rates of folding and misfolding. Extended structures will clearly have different propensities, for example, in the formation of disulfide bonds. Similarly, the time that is needed for an extended stretch of amino acids within the protein sequence to rearrange into another extended or into a more compact state will be affected by the change in compactness induced by the different mutations. It is proposed here that it is this change in time scale that may provide opportunity for point mutations to alter the probability of amyloid formation (Chiti et al., 2003; Chiti et al., 2002). Variants of human lysozyme are known that cause familial amyloidosis due to an I56T or D67H mutation. Both mutations are non-conservative mutations: in I56T hydrophobicity is decreased drastically, whereas in D67H charges are reversed. Both mutations are flanking Trp62 to the left and right within cluster 3, suggesting that a change in distribution of conformers present in the unfolded state will occur and allow the formation of amyloids.

4.4 Conclusion

In this part of the thesis, the unfolded state of lysozyme has been investigated using NMR spectroscopic methods: Six hydrophobic clusters that display residual structure in the unfolded state of lysozyme were identified by chemical shift deviations and deviations from random coil relaxation rates. Non-conservative single point mutations (A9G, W62G, W62Y, W111G and W123G) within the main clusters of hydrophobic residues perturbed the stability of these clusters and the dynamics within the whole protein to different degrees as could be shown by relaxation data. Long-range interactions are therefore present between the hydrophobic clusters that are stabilized by hydrophobic and most likely aromatic interactions. Using NMR diffusion measurements it was shown that the disruption of the interactions between hydrophobic clusters dramatically alters the overall compactness of the unfolded state: single point mutations can turn a compact unfolded state into an extended state.

Detailed analysis of the interactions in this unfolded state of lysozyme (WT-S^{ME}) revealed that native and non-native interactions are present in the conformational ensemble. While interactions between parts of the protein that become the α-domain in folded lysozyme are native-like, non-

native interactions are found between parts of the α and the β-domain. Comparison of these results with results from oxidative refolding suggests that native-like interactions within the α–domain of lysozyme guide the formation of the correct disulfide bridges, particularly the formation of the 30-115 disulfide bridge.

These results also suggest the relevance of unfolded states in amyloid formation, particularly in familial amyloidosis: while the single point mutants investigated here do not influence the structure of the native state, the structural ensemble in the unfolded state is severely altered by the mutations.

A new method for the identification of long-range interactions in unfolded proteins based on a combination of mutagenesis and NMR measurements was developed. In particular, R_2 relaxation rates and diffusion constants were shown to be valuable tools for the identification of non-local contacts in non-native proteins. Using the methodology provided here long-range interactions in unfolded states of a large number of proteins can be identified and particularly the unfolded states of amyloidogenic proteins could be analyzed, to test the above hypothesis.

Chapter 5

Expression of bovine α-lactalbumin

5 Expression of bovine α-lactalbumin

5.1 Introduction

The availability of uniform isotope (^{15}N and/or ^{13}C) labeled protein in large quantities is a prerequisite for detailed NMR studies in solution. Typically, 0.5ml of a 1 mM sample are required. The most efficient way to introduce isotopic labels into proteins is the heterologous overexpression of the target protein in *Escherichia coli* (*E. coli*) grown in a so-called minimal medium (M9) containing ^{15}NH$_4$Cl and/or ^{13}C$_6$-glucose as the sole nitrogen and/or carbon source. The goal of this chapter was to establish conditions for the expression and subsequent purification of bovine α-lactalbumin (BLA) without additional methionine at the N-terminus. This heterologously overexpressed BLA has the same stability as the wild type protein from milk. Heterologous expression of α-lactalbumins in *E.coli* has been known for more than a decade (Wang et al., 1989), and overexpression using different constructs has been widely used, despite the fact that the protein forms inclusion bodies in *E.coli* and has to be subsequently refolded. The first approaches to α-lactalbumin expression involved fusion proteins that were subsequently cleaved to yield pure protein. Nowadays the protein is mostly expressed with an additional N-terminal methionine residue (Grobler et al., 1994), despite the fact that the N-terminal methionine remains at the protein. The stability of α-lactalbumin with the additional methionine (M-BLA) is severely altered: its presence lowers the melting point of the protein by about 16.5°C (33.9 → 16.5°C) in the absence of stabilizing Ca^{2+}-ions and about 8°C (64.2 → 55.9°C) in the presence of Ca^{2+}-ions (Ishikawa et al., 1998). This destabilization is not only found in BLA, but also in human α-lactalbumin (HLA). HLA is significantly destabilized against GdnHCl induced denaturation (Chaudhuri et al., 2000). Similarily we found that human α-lactalbumin is severely destabilized by the presence of the N-terminal methionine: whereas non recombinant human lactalbumin (HLA) is folded in 4M urea in the presence of two equivalents Ca^{2+}, recombinant methionine containing human α-lactalbumin (M-HLA) does not fold in the presence of 4M urea as indicated by NMR.

Therefore the expression of BLA without methionine at the N-terminus of the protein is described in this chapter.

5.2 Materials and Methods

5.2.1 Materials

All chemicals were purchased from either Roth or Sigma if not otherwise indicated.

5.2.2 Plasmid for expression of His-tagged bovine α-lactalbumin

A synthetic gene encoding for bovine α-lactalbumin (Hurley and Schuler, 1987; Wang et al., 1989) with an N-terminal polyhistidine sequence and a trypsin cleavage site was ordered in the expression vector pET11a from Entelechon. The vector is called pHISBLA, the potentially expressed protein is called HISBLA.

5.2.3 Expression of ^{15}N labeled bovine α-lactalbumin

The pHISBLA plasmid was transformed into competent *E.coli* BL21(DE3)-Gold cells (stratagene) and positive clones were selected from a 0.1mg/ml ampicilin containing luria broth (LB) agar plate. A single colony was used to inoculate a 50 ml (0.1mg/ml ampicilin) LB culture, and the bacteria were cultured over night at 37°C (OD_{600}=1.8-2). Cells were pelleted by centrifugation at 6000g for 10min (at 4°C) and washed two times using 15ml LB medium. The cell pellet was than resuspended in 10ml of the final growth medium (either LB or M9) and used to inoculate a 1l culture (either LB or M9) containing 0.1mg/ml ampicilin. The bacteria were grown at 37°C (180rpm) and induced at OD_{600}=0.7 using 0.5g/l IPTG (isopropylthiogalactoside) for 3h. Cells were collected by centrifugation at 6000g for 10min.

M9-medium for the expression of ^{15}N labeled protein:

42 mM Na$_2$HPO$_4$	18.5 mM ^{15}N- NH$_4$Cl (1g/l)	Trace metals:	
22 mM KH$_2$PO$_4$	0.1 mM CaCl$_2$	1 μM Cu^{2+}	1 μM Ni^{2+}
8.5 mM NaCl	2 mM MgSO$_4$	1 μM Zn^{2+}	1 μM MoO$_4{}^{2-}$
pH 7.4	50 mM Glucose	1 μM Mn^{2+}	1 μM Cu^{2+}
	5 mg/l Thiaminchloride	1 μM Co^{2+}	
	30 μM FeCl$_3$		

5.2.4 Purification of HISBLA

All steps of the purification except the centrifugation were carried out at room temperature. The cell pellet was resuspended in 20ml Sol.#1 and lysed by three times sonification for 1min (50% power, 50% strokes) on ice. The cell lysate was centrifuged (10000g, 30min, 4°C) and the insoluble fraction was washed again with 15ml Sol.#1 followed by 20ml Sol.#2 (centrifugation each time for 30min, at 4°C, 10000g). The resulting pellet (inclusion bodies) was solubilized using 10ml of Sol.#3; after centrifugation, the supernatant was diluted 1:1 with IMAC (immobilized-metal affinity chromatography) I buffer and loaded onto a freshly prepared 5ml Ni-NTA (nickel-nitrilotriacetic acid) column equilibrated with 10 column volumes (CV) IMAC I buffer. The column was washed with two CV IMAC I and three CV IMAC II buffer. Elution of the protein was performed stepwise (three CV each) by increasing imidazole concentration (using mixtures of IMAC I and IMAC III) as follows:

Step I:	20mM imidazole
Step II:	50mM imidazole
Step III:	100mM imidazole
Step IV:	250mM imidazole
Step V:	500mM imidazole

Sol.#1- Sonication buffer:	Sol.#2- Washing buffer:	Sol#3- Solubilization buffer:
50mM Tris/HCl	20mM Tris/HCl	50mM Tris/HCl
25% sucrose	1% Triton X-100	300mM NaCl
1mM EDTA	1mM EDTA	8M urea
pH 8	pH 8	20mM β-mercaptoethanol
		pH 8.5

IMAC I- Ni-NTA column mobile phase:	IMAC II- Ni-NTA column mobile phase - wash buffer:	IMAC II- Ni-NTA column mobile phase - elution buffer:
50mM Tris	50mM Tris	50mM Tris
300mM NaCl	4M urea	4M urea
4M urea	500mM NaCl	300mM NaCl
5mM β-mercaptoethanol	5mM β-mercaptoethanol	5mM β-mercaptoethanol
pH 8.5	pH 8.5	500mM Imidazol
		pH 8.5

5.2.5 Refolding of HISBLA

Fractions containing HISBLA were analyzed by SDS-PAGE and pure fractions were pooled. Protein refolding and formation of the four native disulfides bridges was done by stepwise dialysis (1:300 v/v ratio, 12h each step) against the following buffers:

1) in buffer X + 4M urea containing 2.5mM reduced and 0.5mM oxidized glutathione,

2) buffer X containing 2.5mM reduced and 0.5mM oxidized glutathione and

3) buffer X.

Protein concentration in the dialysis chamber was between 0.6 and 0.9 mM. Protein refolding was confirmed using analytical HPLC.

Buffer X:

   10 mM Tris

   50 mM NaCl

   1 mM CaCl$_2$

   pH 8.5

5.2.6 Optimization of cleavage conditions

HISBLA contains two N-terminal trypsin cleavage sites. One of the cleavage sites coincides with a lysC cleavage site. Cleavage conditions were optimized using samples containing 200µg HISBLA in 170µl buffer X by variation of temperature (4°C - 37 °C) and protease:HISBLA ratio (1:20-1:500). Trypsin (Promega, sequencing grade modified) as well as lysC (Roche, sequencing grade) were used for cleavage optimization. The reaction was stopped either by denaturation using gel-buffer (Laemmli, 1970) or by the addition of 0.2mM 4-2(aminoethyl) benzene sulfonyl fluoride hydrochloride (AEBSF, purchased from AppliChem), to allow storage before further treatment.

The success of the cleavage procedure was monitored using SDS-PAGE, HPLC chromatography and MALDI-TOF.

5.2.7 Ionexchange chromatography

The solution of the cleavage reaction was diluted 1:1 with water and applied onto a Q-sepharose column equilibrated with 25mM phosphate buffer pH 6.5. Protein was eluted by a linear gradient ranging from 25mM to 300mM NaCl.

5.2.8 Analytic

5.2.8.1 Determination of protein concentration using UV measurements

Protein concentrations were determined by UV-measurements at 280nm using a Cary UV-specrometer (Varian) equipped with 500µl quartz cuvette (1cm pathlength). The following extinction coefficients were used: $\varepsilon_{BLA}=29210M^{-1}cm^{-1}$ (Sugai et al., 1973), the extinction coefficient of the Tag alone was estimated as $\varepsilon=0$ from the amino acid sequence using the ExPASy ProtParm tool (http://www.expasy.org/tools/protparam.html), therefore $\varepsilon_{HISBLA}=29210M^{-1}cm^{-1}$ was used.

5.2.8.2 Mass spectrometry

MALDI-TOF (matrix-assisted laser desorption/ionization time of flight) mass spectrometry was used to analyze the masses of BLA, HISBLA and the cleavage products using a VG TofSpec mass spectrometeter (Fisons).

5.2.8.3 HPLC

The analysis of successful refolding, with the right formation of disulfides, was performed by HPLC. A L6200 pump (Merck Hitachi) and L3000 photo diode array detector (Merck Hitachi) and a reverse phase RP4 4.6x250mm column were used. The column was equilibrated with water and BLA was eluted over a linear gradient from 0-100% Acetonitril (AcN) within 50min. All solvents contained 0.1% trifluoroacetic acid (TFA).

5.2.8.4 SDS-PAGE

SDS-PAGE (sodium dodecylsulfate polyacrylamide gel electrophoresis) (Laemmli, 1970) was performed to analyze protein expression, purification and cleavage using a Whatmann Biometra minigel apparatus and a Biometra Power Pack 25.

5.3 Results and Discussion

5.3.1 Design of a plasmid suitable for the expression of ^{15}N labeled α-lactalbumin

The aim of this work is to express of WT-BLA, without methionine at the N-terminus. There are two general possibilities to achieve this if the protein is expressed in *E.coli*: 1) Expression of the protein with the addition of the N-terminal methionyl residue and subsequent BrCN cleavage and 2) expression of a larger construct, followed by an enzymatic proteolytic digest. As BLA contains a methinonine (M90) in its sequence, the BrCN cleavage cannot be performed on the WT-protein since BrCN cleaves behind every Met residue. Figure 41 shows the amino acid sequence chosen here. A polyhistidine sequence (His-tag) containing six histidines is followed by a sequence that contains two trypsin cleavage sites. One cleavage site (directly at the N-terminus of the protein (after position -1)) coincides with a LysC cleavage site. The sequence from position -9 to position -1 is identical to the last nine aminoacids of a fusion protein that had been used by Wang et al. (Wang et al., 1989) for expression and subsequently successful cleavage of BLA using Trypsin.

Trypsin
Cleavage site   Trypsin + LysC
                 cleavage site
His-Tag

MGHHHHHHTEGPGRGPKEQLTKCEVFRELKD
LKGYGGVSLPEWVCTTFHTSGYDTQAIVQNN
DSTEYGLFQINNKIWCKDDQNPHSSNICNIS
CDKFLDDDLTDDIMCVKKILDKVGINYWLAH
KALCSEKLDQWLCEKL
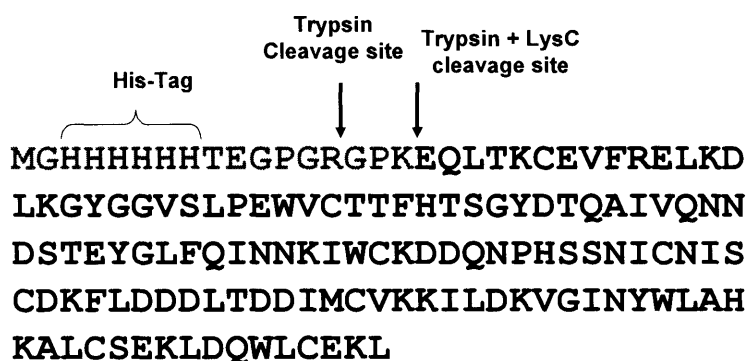
Figure 41: Amino acid sequence of the synthetic gene for the expression of His-tagged BLA (HISBLA).

A synthetic gene coding for the sequence shown in Figure 41, was optimized for the use in *E.coli*. It was purchased from Entelechon in a pET11a vector with NdeI cleavage site at the N-terminus of the construct and BamH1 cleavage site at the C-terminus (optimized DNA sequence in the supplementary material). The vector is called pHISBLA, the encoded protein HISBLA.

5.3.2 Expression and purification of HISBLA

The pHISBLA vector was transformed into competent *E.coli* BL21(DE3)-Gold cells and HISBLA was expressed as described in the materials and methods section. Figure 42 shows a SDS-PAGE analysis of the expression and initial purification of the protein: the presence of a large protein band with an apparent mass slightly bigger than 14.4kDa in lane 2 indicates the successful expression of the construct, which is 16.076kDa (calculated from the sequence).
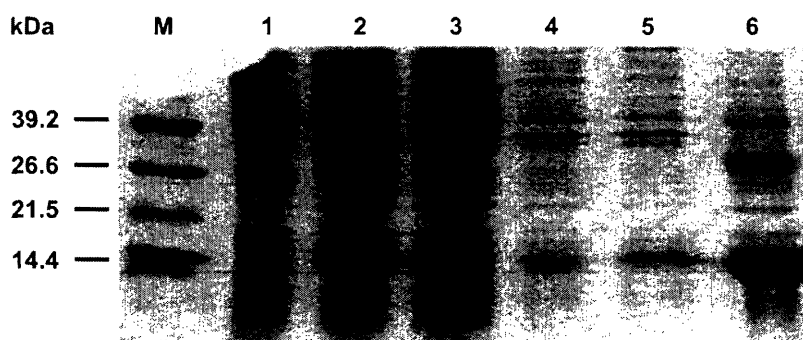
Figure 42: SDS-PAGE (15%) analysis of the expression of HISBLA and purification of inclusion bodies. M: protein molecular weight (MW) marker, 1: lysate of transformed bacteria before induction, 2: lysate of transformed bacteria 3h after induction with IPTG, 3: supernatant after ultrasonication, 4 and 5: supernatant after washing steps (4 with sonication buffer, 5 with triton-X containing washing buffer), 6: solubilized inclusion bodies.

α-LA has four disulfide bridges. These disulfide bridges are not correctly oxidized and folded in *E.coli*. Therefore, the protein is insoluble in the cell and accumulates as inclusion bodies. Consequently, it is possible to purify the protein coarsely, by washing the insoluble inclusion bodies after lysis of the cells. Lanes 3-5 show the supernatants of lysis and several washing steps: the concentration of the expression product is very low. Solubilized inclusion bodies (lane 6) contain mainly the target protein, which is applied to a Ni-NTA column for further purification. Despite the high concentration of denaturant and reducing agents necessary to solubilize the inclusion bodies, the protein binds well to the Ni-NTA column, as can be seen by the absence of relevant amounts HISBLA in the flow through of the Ni-NTA column (Figure 43, lane 2).
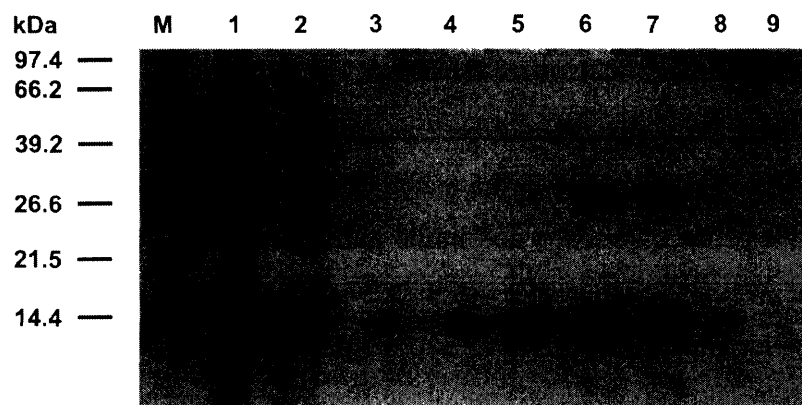
Figure 43: SDS-PAGE (15%) for the analysis of eluted fractions from a Ni-NTA column using imidazole. M: protein MW marker, 1: solubilized inclusion bodies, 2: proteins in the flow through, 3 proteins in washing step (mobile phase), 4: washing step (mobile phase, high salt), 5: elution with 20mM imidazole, 6: elution with 50mM imidazole, 7: elution with 100mM imidazole, 8: elution with 250mM imidazole, 9: elution with 500mM imidazole.

HISBLA elutes between 20mM and 250mM imidazole concentrations (lanes 5-8), the main fractions elute at 50mM to 100mM imidazole (lanes 5 and 6). The column, however, fails to completely purify the protein, an additional protein (between 27 and 39kDa) coelutes. The yield of the almost pure HISBLA after the Ni-NTA column is 20mg/l for ^{15}N labeled protein and 27.5mg/l for LB medium determined by UV measurements.

5.3.3 Refolding of HISBLA

Oxidative refolding of the protein is performed in three steps, starting out under moderate denaturant conditions (4M urea) and a glutathione redox shuffling system, ending without any denaturing and redox agent. Refolding is almost quantitative as judged from analytic HPLC measurements (Figure 44) and 1D NMR measurements. The coeluted protein (Figure 43, lanes 6 and 7) however does not prevent successful folding.
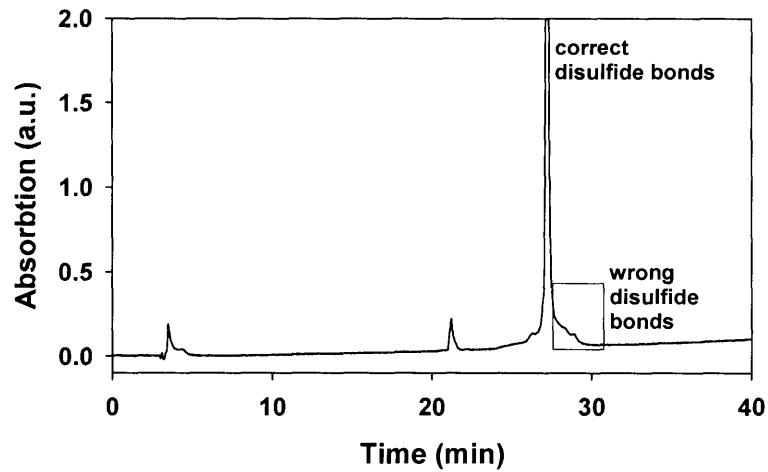
Figure 44: HPLC-chromatogram after refolding (see methods). Correctly refolded protein, i.e. the right disulfide bridges formed, elutes after 27.2min (indicated as right disulfides); protein without properly formed disulfide bridges elutes after 28.9min (indicated as wrong disulfides).

5.3.4 Cleavage of HISBLA and purification of BLA

The enzymatic cleavage of refolded HISBLA was optimized by a variation in temperature, protease concentration and buffer conditions. Two proteases, trypsin and lysC that are known to cleave behind lysine (K) residues were tried. Figure 45 shows a representative SDS-PAGE analysis of a cleavage reaction using 1:100 trypsin:protein ratio at 37°C.
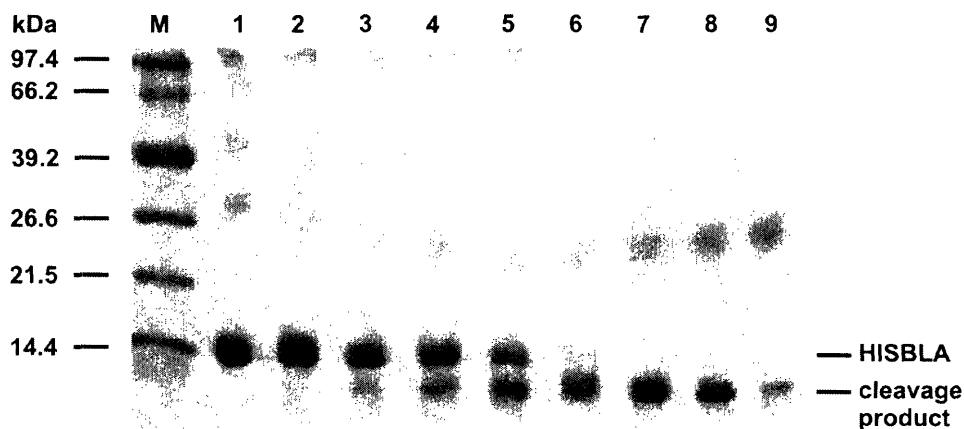
Figure 45: SDS-PAGE (18%) for the analysis of trypsin cleavage reaction at 37°C; trypsin:HISBLA ratio was 1:100. M: protein MW marker, 1: HISBLA after refolding, before cleavage 2: 1 min, 3: 5 min, 4: 10 min, 5: 20 min, 6: 1 h, 7: 2 h, 8: 3 h, 9: 4.5 h.

The cleavage reaction starts after five minutes (lane 3), two bands (smaller than 21.5 kDa) can be observed 10min after the initiation of the cleavage reaction (lane 4). The higher molecular weight band corresponds to HISBLA, while the lower molecular weight bond corresponds to the cleavage product. The digestion of HISBLA is completed after 2h (lane 7), and degradation is observed after longer incubation times (lanes 8 and 9). Analysis of the cleaved product using mass spectrometry reveals that the cleavage product is a mixture of the wanted BLA and a construct three amino acids longer (BLA+3, Figure 46). BLA+3 results from cleavage at the first trypsin cleavage site only. The difference in molecular weight between the two proteins (BLA and BLA+3) is small (0.282kDa) so that a clear separation based on SDS-PAGE is not possible; further analysis of cleavage conditions is therefore mainly based on HPLC and mass spectrometry. Time dependent analysis (by HPLC and mass spectrometry) of the cleavage reaction reveals that the two trypsin cleavage sites are cleaved sequentially.

$$
\begin{array}{c}
\text{T} \qquad \text{T + L} \\
\diagdown \qquad \diagdown \qquad \text{BLA} \\
\text{MGHHHHHHTEGPGRGPKEQLT} \ldots \\
\text{-17} \qquad \text{-10} \qquad \text{-3} \quad \text{1} \\
\text{BLA+3}
\end{array}
$$
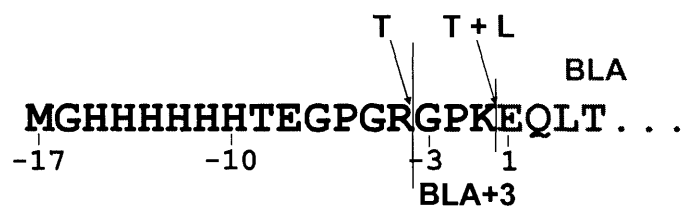
Figure 46: Sequence of the cleavable His-tag. T=trypsin cleavage site, L=lysC cleavage site. The start of the BLA sequence is shown in gray, BLA+3 indicates the construct that is obtained if cleavage only happens before the -3 position.

Table 10 summarizes representative conditions employed during the optimization of the cleavage reaction for both enzymes. Successful cleavage is not possible using the enzyme lysC. The protein is not cleaved under native conditions, or completely digested into small peptides under mildly destabilizing conditions (5% AcN or 4mM EDTA, Table 10).

protease	Protease: HISBLA (w/w)	T (°C)	Time (h)	HISBLA : BLA+3 : BLA	Yield BLA (%)
lysC	1:100	RT	6	1:0:0	0
	1:100	37	6	1:0:0	0
	1:100	RT	1 ***	0:0:1	< 10
	1:500	RT	1 ***	0:0:1	< 10
	1:100	RT	1 ****	0:0:0	0
	1:500	RT	1 ****	0:0:1	< 10
trypsin	1:100	37	2	0:1:1	< 10
	1:100	37	3	0:1:2	< 10
	1:100	37	3 *	0:1:2	< 10
	1:100	30	4	0:4:1	> 10
	1:50	RT	4	0:1:1	> 10
	1:50	30	4	0:1:5	> 30
	1:20	RT	4 **	0:1:1	< 10
	1:20	30	4 **	0:4:1	> 10

Table 10: Cleavage conditions, and rough estimates of ratios of uncleaved protein (HISBLA): cleavage products (BLA+3, BLA). Yields of BLA are relative to the amount of refolded HISBLA. Estimates were made based on SDS-PAGE (differentiation between HISBLA and cleavage products) and HPLC and mass spectrometry (differentiation between BLA and BLA+3). Yields of BLA are relative to the amount of refolded HISBLA. * cleavage in the presence of Ni-NTA resin, **visible aggregation of some component in the cleavage reaction, *** addition of 5 % acetonitril to the cleavage reaction, ****addition of 4 mM EDTA to the cleavage reaction. Finally used cleavage conditions are shaded in gray.

The final conditions used for the digest of ^{15}N labeled HISBLA are 1:50 trypsin:protein ratio , 4h digest at 30°C in refolding buffer (buffer X, see materials and methods). HISBLA is almost quantitatively cleaved under these conditions; the cleavage product consists mainly of BLA, little BLA+3 is present. Further purification of the cleavage products is performed using ion exchange chromatography as shown in Figure 47. BLA has a theoretical pI of 4.8 (http://us.expasy.org/cgi-bin/protparam), whereas BLA+3 has a theoretical pI of 4.94. At pH 6.5 for the ion exchange chromatography it is thus predicted that BLA binds to the column, whereas the slightly more basic BLA+3 does not bind to the column. As expected, pure BLA elutes (in two fractions that do not differ from each other spectroscopically) from the column as determined by mass

spectrometry. The pure fractions were pooled, the yield of ^{15}N BLA of the whole purification procedure was 7mg/l medium.
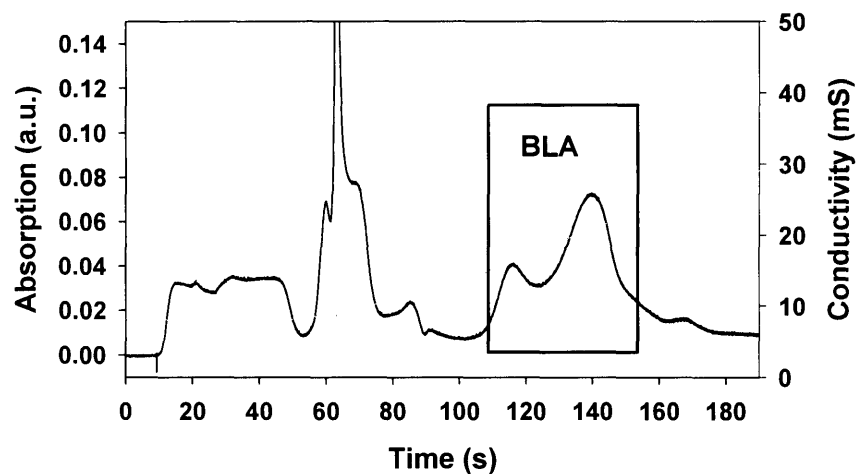
Figure 47: FPLC chromatogram of the purification of ^{15}N labeled BLA after cleavage reaction. Protein was eluted from a Q-sepharose column by increasing the salt concentration. Black line indicates the absorption at 280nm, gray line indicates the conductivity. Fractions that contain pure BLA are indicated by a black box.

5.3.5 Spectroscopic characterization

Mass spectrometric characterization of the purified BLA revealed a molecular weight of 14246.75Da. This corresponds to the calculated weight (14186Da) plus a K^+ and a Na^+ ion (Figure 48). The weight of the purified ^{15}N labeled BLA is 14347.87Da. This matches the calculated weight (14186 (unlabeled protein)+162(^{15}N label)=14348) of ^{15}N BLA (data not shown).
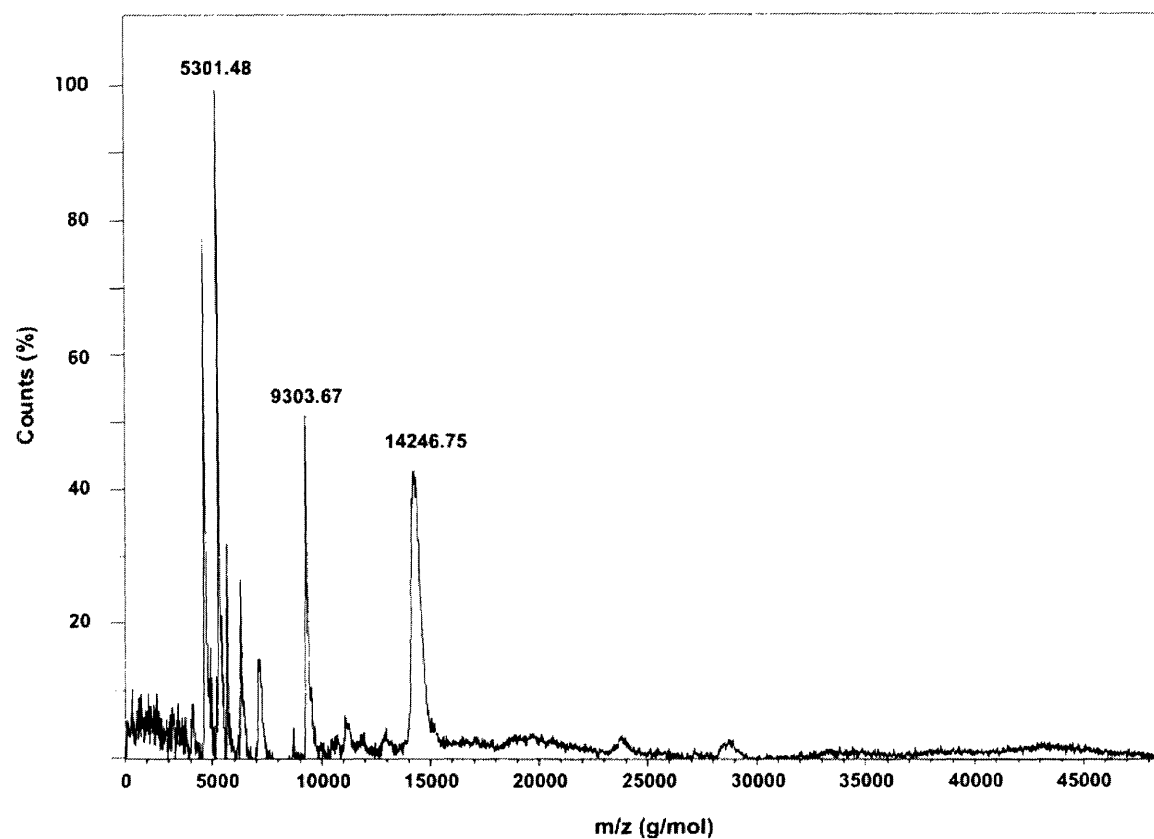
Figure 49: MALDI-toff spectrum of purified unlabeled BLA. The molar mass of 14246.75 Da (g/mol) corresponds to the molar weigth of BLA (14186) plus one K^+ and one Na^+ ion.

NMR spectroscopic characterization using $^1H^{15}N$ heteronuclear single quantum correlation (HSQC) spectra confirmed that the protein is folded under native conditions (100mM Tris, 10mM Ca^{2+}, pH 7.0).

131
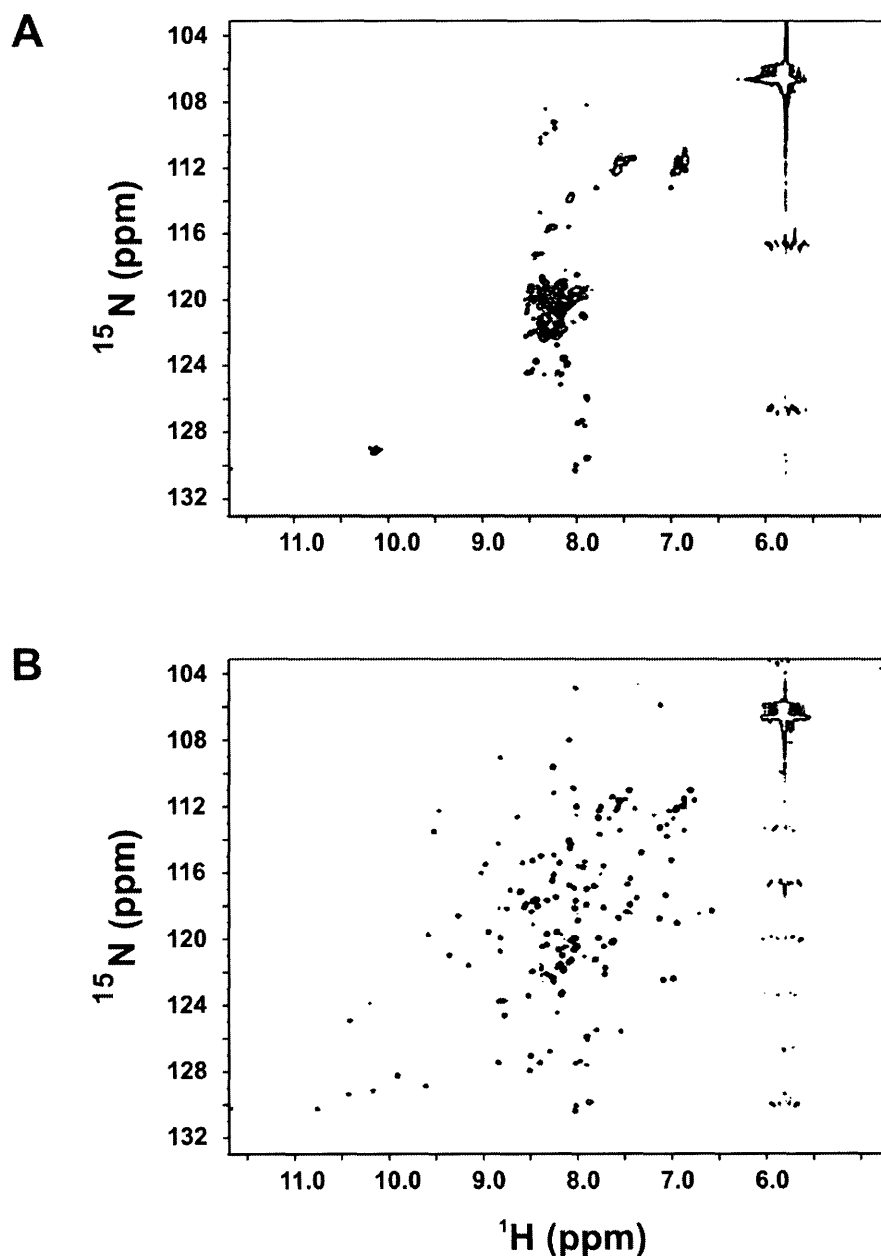
Figure 50: ^{1}H^{15}N-HSQC spectra of BLA in 4M urea at pH=7, 35°C A) in the absence of Ca^{2+}, unfolded protein and B) in the presence of two equivalent Ca^{2+} ions, folded protein.

Furthermore it could be shown (Figure 50) that the heterologous expressed BLA has the same stability as the WT protein isolated from milk: The protein is unfolded in 4M urea in the absence of Ca^{2+}-ions, as can be seen by the low chemical shift dispersion of the HSQC spectrum in Figure 50A. In the presence of two equivalents of Ca^{2+}-ions the expressed BLA is folded, as reflected by the large chemical shift dispersion of the HSQC spectrum in Figure 50B. The high stability of

132

this expressed BLA is in contrast to the typically expressed methionine containing M-BLA. M-BLA is unfolded in 4M urea even in the presence of high concentrations of Ca^{2+}-ions.

5.4 Conclusion

In this chapter, the heterologous expression, purification and spectroscopic characterisation of ^{15}N labeled bovine α-lactalbumin (BLA) is described. In contrast to the stability of M-BLA (see Introduction) the stability of the expressed BLA is comparable to the protein isolated from milk. The goal to set up an expression system that allows the expression of large amounts of BLA with similar stability as the isolated WT protein could be achieved. The overexpressed BLA has 100% sequence identity compared to the WT-BLA isolated from cows milk.

BLA is expressed with a His-tag (HISBLA) in inclusion bodies, which allows purification of the HISBLA by washing and subsequent affinity chromatography. A very efficient refolding scheme using a redox shuffling system and decreasing denaturant concentration stepwise was introduced to refold the purified HISBLA. Refolding yields for the refolding from inclusion bodies are almost quantitative; overall 20mg/l medium ^{15}N labeled HISBLA could be purified and refolded. Cleavage of HISBLA to yield BLA was achieved by sequential cleavage of the two trypsin cleavage sites prior to the N-terminus of BLA. A mixture of both constructs is obtained during the cleavage reaction. This mixture can be separated by ion exchange chromatography. The yield of the cleavage reaction, including subsequent purification is 35%. 7mg/l medium BLA are obtained.

Using the protocol it is not only possible to obtain BLA resembling the protein isolated from cows milk, but it is also possible to isotope label the protein. Using isotope labeled protein, the unfolded state of BLA can be structurally characterized using NMR spectroscopy as shown in Chapter 6. The protein furthermore could be used to characterize folding intermediates of the BLA folding pathway following e.g. the approach chosen in Chapter 7.

Chapter 6

The unfolded state of bovine α-lactalbumin and comparison of the unfolded state with unfolded states of homologous proteins

6 The unfolded state of bovine α-lactalbumin and comparison of the unfolded state with unfolded states of homologous proteins

6.1 Introduction

The goal of this chapter is to investigate the unfolded state of bovine α-lactalbumin (BLA) at atomic resolution using NMR spectroscopy. In the first part of this chapter, the assignment and analysis of the chemical shifts of unfolded BLA are discussed. In the second part of the chapter, the unfolded states of the structurally related proteins lysozyme (HEWL) and human α-lactalbumin (HLA) are compared with the unfolded state of BLA.

HEWL and both HLA and BLA are used as model systems to study protein folding (Dobson et al., 1994; Forge et al., 1999; McKenzie and White, 1991). The native structures of the proteins are very well superimposable (Figure 12, backbone RMSD=1.27Å for residues 1-116 in HEWL, corresponding to residues 1-112 in α-LA). While their functions in the native state are well understood, their unfolded states are less well characterized, particularly the unfolded state of BLA and HLA. Differences in the oxidative folding of HEWL and the α-lactalbumins (α-LAs) have been found. However, those differences are not yet completely understood (Chang, 2004; Chang and Li, 2002; Li and Chang, 2004; van den Berg et al., 1999) (Chapter 1.4.2.7).

One key towards the understanding of oxidative refolding is the investigation of the unfolded state of proteins in the absence of disulfide bridges; the starting point of protein folding. This state is called unbranched from here on. While unbranched HEWL (HEWL-S^{ME}) and unbranched HLA (All-Ala-HLA) have already been studied to some extent using NMR spectroscopy and heteronuclear resonance assignments are available (Klein-Seetharaman et al., 2002; Redfield et al., 1999; Schwalbe et al., 1997; Wirmer et al., 2004), unbranched BLA (BLA-S^{ME}) has not been studied yet.

In this chapter, unbranched BLA (BLA-S^{ME}) is prepared by reduction and methylation of its four disulfide bridges (BLA-S^{ME}). BLA-S^{ME} is unfolded at pH 2 in 8M urea and heteronuclear NMR assignments were obtained. Residual secondary structure was identified and compared with the residual structure and motional restrictions of the unbranched states of HLA and HEWL.

Christina Redifield provided ^{15}N relaxation rates of unbranched HLA making the comparison possible.

136

6.2 Materials and Methods

6.2.1 NMR samples of BLA-S^{ME}

^{15}N and ^{13}C,^{15}N labeled bovine α-lactalbumin (BLA) were prepared as described in Chapter 5. The ^{13}C,^{15}N labeled sample was kindly prepared by Holger Berk, following the approach developed in Chapter 5.

Methylation of BLA to yield BLA-S^{ME} was performed following the procedure of Heinrikson (Heinrikson, 1971): 7mg BLA were dissolved in 8ml denaturing buffer (6M GdnHCl, 25% v/v acetonitrile (AcN), 250mM Tris, 5mM EDTA, pH 8.5) and the solution was purged for one minute with N$_2$. 3.5µl β-mercaptoethanol (~12 fold molar excess of cysteine residues) were added and the sample was kept for 1h at 50°C, to reduce all the disulfide bridges. The sample was cooled down to 37°C before 150µl 0.6M methyl-4-nitrobenzene sulfonate (MNBS) in AcN (at least 1.5 fold excess of the total amount of –SH groups in the sample) were added. After 2h at 37°C, the sample was dialyzed against water at pH 2 (5 times against 5l, 12h each step). The samples were subsequently lyophilized and dissolved in NMR buffer:

The BLA-S^{ME} NMR samples contained 0.5mM (^{15}N) and 1.5mM (^{13}C,^{15}N) BLA-S^{ME} in 8M urea, pH 2, 90% H$_2$O, 10% D$_2$O. The spectra of the protein were concentration independent.

6.2.2 NMR assignment measurements

All NMR measurements were performed at 20°C on Bruker spectrometers as summarized in Table 11.

Spectrometer	probe head
DRX600	TXI HCN xyzgrad
DMX600	TXI HCN cryoprobe
AV800	TXI HCN zgrad
AV900	TXI HCN xyzgrad

Table 11: Spectrometers and probe heads used for the assignment experiments.

Table 12 lists all the experiments used for the assignment of BLA-S^{ME} and the most important acquisition parameters. Pulsed field gradient versions of all these experiments using sensitivity enhancement for the back transfer (Schleucher et al., 1994) were used. Carrier positions were as follows: ^{1}H 4.7ppm, ^{13}C^{ali} 39ppm, ^{13}CO 172ppm and ^{15}N 117.5ppm and the recycle delay was set

to 1.5s. Carbon pulses were implemented as Gaussian pulse cascades (Emsley and Bodenhausen, 1990); in particular Q3 and Q5 cascades (Emsley and Bodenhausen, 1992) were applied.

experiment	t_3 points (n) SW	t_2 points (n) SW	t_1 points (n) SW	sample	field (MHz)
HSQC		2048 (^{1}H) 8389	256 (^{15}N) 1946	^{13}C,^{15}N	800
NOESY-HSQC (Marion et al., 1989a)	2048 (^{1}H) 12626	100 (^{15}N) 2007	180 (^{1}H) 12626	^{15}N	900
TOCSY-HSQC (Marion et al., 1989b)	2048 (^{1}H) 8389	100 (^{15}N) 1520	200 (^{1}H) 7198	^{15}N	600
HNCA (Grzesiek and Bax, 1992)	2048 (^{1}H) 12626	94 (^{15}N) 2007	100 (^{13}C) 4074	^{15}N	900
HNCACB (Wittekind and Mueller, 1993)	2048 (^{1}H) 8389	96 (^{15}N) 1429	70 (^{13}C) 9054	^{13}C,^{15}N	600
CBCACONH(Muhandiram and Kay, 1994)	2048 (^{1}H) 8389	68 (^{15}N) 1429	104 (^{13}C) 9054	^{13}C,^{15}N	600
HN(CA)CO (Clubb et al., 1992)	2048 (^{1}H) 11161	88 (^{15}N) 1946	180 (^{13}C) 2012	^{13}C,^{15}N	800
HNCO (Grzesiek and Bax, 1992)	2048 (^{1}H) 11161	88 (^{15}N) 1946	125 (^{13}C) 2012	^{13}C,^{15}N	800

Table 12: Acquisition parameters used for the assignment of BLA-S^{ME}. t_3, t_2 and t_1 indicate the dimensions in the order of their appearance in the pulse sequence. t_3 is the direct dimension (with the exception of the 2D HSQC where t2 is the direct dimension), while t_2 and t_1 are indirect dimensions. Points indicate the number of points measured during acquisition of the respective dimension, (n) indicates the recorded nucleus, and SW indicates the spectral width of the dimension. The sample column indicates which sample was being used for the experiment, ^{15}N denotes the ^{15}N labeled sample, ^{13}C,^{15}N denotes the ^{13}C,^{15}N labeled sample; field indicates the field strength used for the respective experiment.

Data processing was done using the program Xwinnmr version 3.5 (Bruker, Karlsruhe), linear prediction, as implemented in Xwinnmr, was used to extend the data in the ^{15}N dimension and all datasets were zero-filled to yield final matrix sizes of 256* 256* 2048 complex points. Data analysis was done using the program CARA (Keller, in preparation), downloaded from www.nmr.ch.

^{1}H chemical shifts were referenced directly to internal (2,2,3,3-d4) trimethyl-3-propionic acid, sodium salt (TMSP) while ^{13}C and ^{15}N chemical shifts were referenced indirectly calculating the true frequency using the approach of Wishart et al. (Wishart et al., 1995b).

6.2.3 Chemical shift deviations

Random coil chemical shifts were extracted from NMR view using the "Wishart peptide" scale (Wishart et al., 1995a) including sequence specific corrections (Schwarzinger et al., 2001) and used to determine deviations from random coil chemical shifts. The chemical shift deviations were calculated according to $\Delta\delta=\delta^{exp}-\delta^{rc}$, with δ^{exp}=experimental chemical shifts, δ^{rc}=random coil chemical shifts according to Wishart (Wishart et al., 1995a).

δ^{exp} for BLA-S^{ME} were used from the assignment described in this chapter, δ^{exp} for unbranched HLA (All-Ala-HLA) were provided by C. Redfield (Redfield et al., 1999), and δ^{exp} HEWL-S^{ME} were taken from previous work (Schwalbe et al., 1997).

6.2.4 R$_2$ relaxation rates

R$_2$ relaxation rates of BLA-S^{ME} were determined as described in chapter 3.2.3. Relaxation rates of HEWL-S^{ME} were taken from earlier work in the group (Schwalbe et al., 1997). R$_2$ relaxation rates of All-Ala-HLA were kindly provided from C. Redfield (unpublished). All R$_2$ relaxation rates compared here were measured at 20°C at a proton frequency of 600MHz. The analysis of the relaxation rates was done as described in chapter 4.

6.2.5 Hydrophobicity predictions

Hydrophobicity predictions were performed as described in chapter 3.2.4 using the following sequences from the swiss prot database: P00709 LCA_HUMAN, P00711 LCA_BOVIN and P00698 LYC_CHICK, each leaving out the first 19 N-terminal amino acids that belong to the precursor peptide.

6.2.6 Sequence alignment

Sequence alignment was performed using the BLAST database algorithm based on the publication of Altschul et al. (Altschul et al., 1997). The same sequences were used as for the hydrophobicity predictions.

6.3 Assignment of BLA-S^{ME} and chemical shift analysis: Results and Discussion

6.3.1 Backbone assignment

6.3.1.1 ^{1}H,^{15}N HSQC

While a majority of proteins are unfolded in 8M urea at pH 2 at 20°C, the conformation of BLA with intact disulfide bridges in various urea concentrations is more complicated, and harsher conditions have to be applied to unfold BLA: 8M GdnHCl at pH 2 and 50°C (Wijesinha-Bettoni et al., 2001). In contrast, standard denaturing conditions can be applied, for the investigation of BLA in the absence of disulfide bridges, as discussed below. This unbranched state of BLA (BLA-S^{ME}) is used as a model for BLA before the start of oxidative folding.
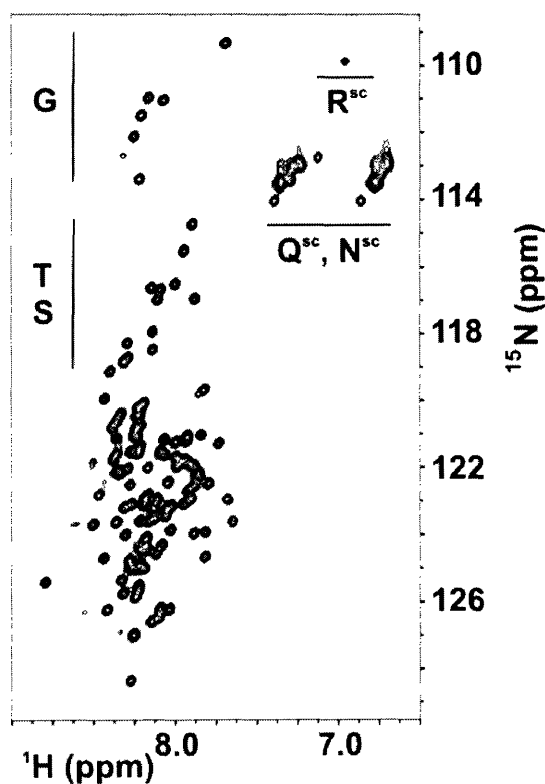
Figure 51: ^{1}H,^{15}N correlation spectrum of BLA-S^{ME} in 8M urea, pH 2 and 20°C. The spectrum was recorded at 800 MHz. The regions where glycine (G), threonine (T) and serine (S) residues as well as the

region where NH$_2$ groups of arginine (R^{sc}), glutamine and asparagine side chains (Q^{sc},N^{sc}) are found are indicated.

For the investigation of the unbranched state of BLA, the four native disulfide bridges of the protein have been reduced and subsequently methylated (BLA-S^{ME}, see methods). Figure 51 shows a ^{1}H,^{15}N correlation spectrum of BLA-S^{ME}. 105 signals in the backbone region can be observed. This fits reasonably to the number of 120 theoretically observable amide backbone signals expected for a monomeric protein taking peak overlap into account. The spectrum shows the signature of an unfolded protein at standard unfolding conditions (8M urea at pH 2 and 20°C): All signals are sharp and have similar intensities, reflecting similar motional properties of all residues in the protein. The observed chemical shift dispersion is small, particularly in the ^{1}H dimension: signals in unfolded proteins appear near their random coil chemical shifts (Chapter 1.2.1), due to the lack of pronounced secondary structure. Only glycine, threonine and serine residues are as well resolved as in native proteins due to their distinct ^{15}N random coil chemical shifts. Despite the small chemical shift dispersion, the signals in the spectrum of BLA-S^{ME} (Figure 51) are reasonably well resolved allowing for the assignment of the protein.

6.3.1.2 Resolved spin systems

The first step in assigning a protein is the picking of peaks that belong to one spin system. As the resolution of the HSQC (Figure 51) is in some parts limited, it proves to be more complicated than in well resolved protein spectra to assign a particular peak to one spin system. Often only the HN-C plane is used for peak picking, here the N-C plane was also used to dissolve ambiguities as shown in Figure 52 in strips of a three dimensional (3D) CBCA(CO)NH experiment (Chapter 1). Four peaks, numbered 1 through 4, can be seen in the middle of an HN-C strip of the CBCA(CO)NH experiment for residue D63 (Figure 52A). Figure 52B shows the N-C strip of the same experiment. Only two of the peaks resonate exactly at the N chemical shift chosen, peaks 1 and 4. These peaks can therefore be assigned to the spin system of D63. Peaks number 2 and 3 resonate at the frequency of D64 as can be seen in Figure 52C and D.
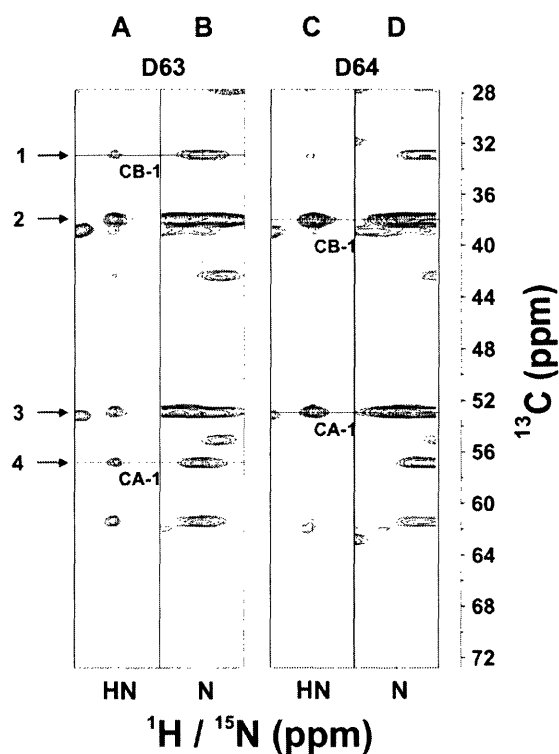
Figure 52: Strip plots of a three dimensional CBCA(CO)NH of residues D63 and D64. A) strip plots showing the HN-C plane of residue D63, B) strip plot showing the N-C plane of residue D63. C, D) strip plots of HN-C, N-C plane of D64. CA-1 and CB-1 indicate the C_α and C_β signals from the preceding residue. 1-4 indicate the signals discussed in the text.

6.3.1.3 Sequential assignment

Sequential assignment of BLA-S^{ME} was performed as described in chapter 1. The starting points of the assignment were glycine, threonine and alanine residues. These each have a very unique sets of random coil chemical shifts. Figure 53 shows so-called "strip plots" that illustrate the sequential assignment of residues K58 to residue N66 using HNCACB and HN(CA)CO spectra. C_α and C_β chemical shifts in BLA-S^{ME} are very similar to the random coil chemical shifts. This often prevents an unambiguous assignment. In contrast, the chemical shift dispersion for each type of amino acid is larger in the carbonyls. The combination of C_α and C_β chemical shifts on the one hand and carbonyl chemical shifts on the other hand was therefore crucial for the successful assignment. Additional verification of the assignment was done using TOCSY-HSQC (Marion et al., 1989b) and NOESY-HSQC (Marion et al., 1989a) spectra.
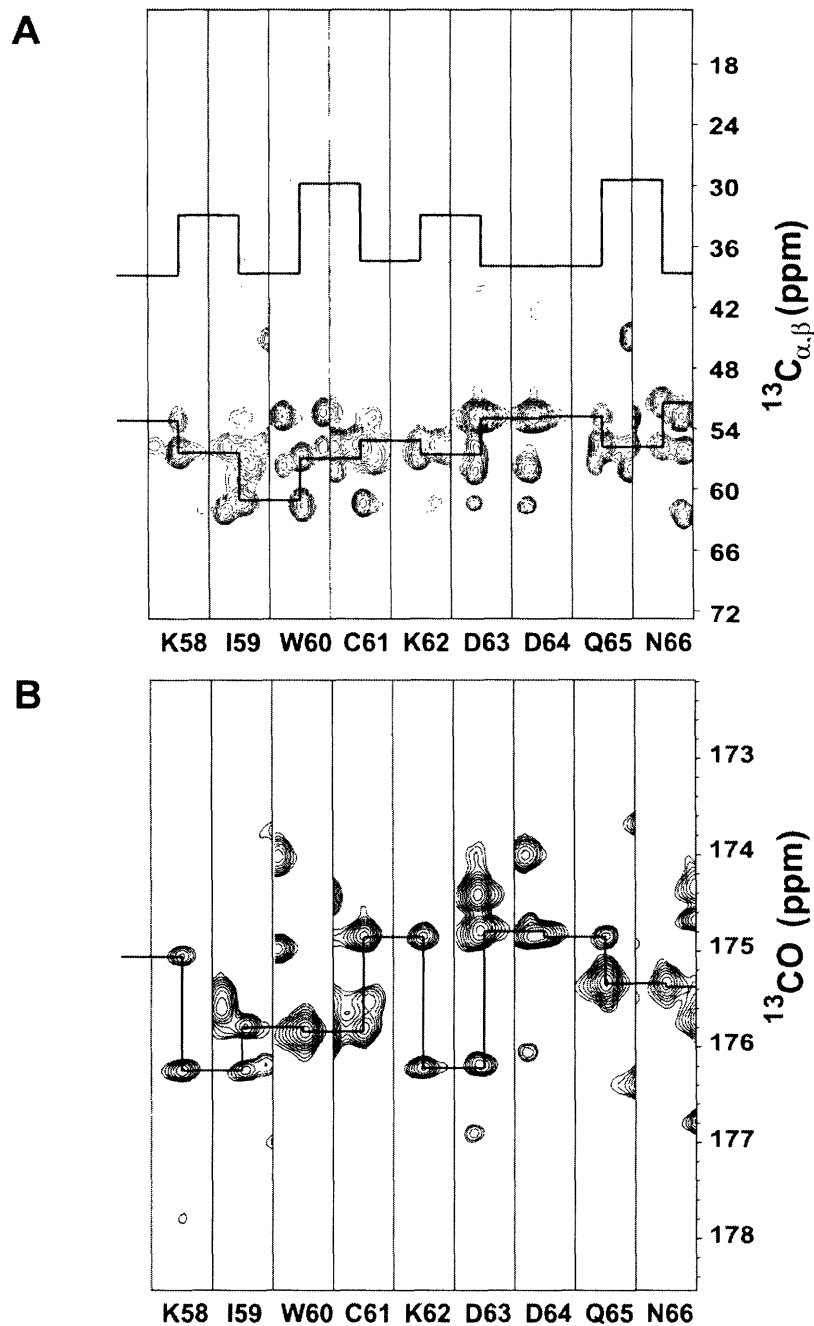
Figure 53: Strip plots of residue K58 to residue N66. A) Strip plot extracted from an HNCACB spectrum. B) Strip plots extracted from an HN(CA)CO spectrum of BLA-S^{ME}. Each strip in the HNCACB shows resonances for the C_α and the C_β of the respective residue and the preceding residue, while each strip in the HN(CA)CO shows resonances for the CO carbon of the respective residue and the preceding one. Horizontal lines connect strips of a certain residue between the i and the i-1 strip, while vertical lines connect the i peak and the i-1 peak within one strip.

6.3.1.4 Assignment success

Figure 54 shows a ^{1}H,^{15}N heteronuclear single quantum (HSQC) spectrum of BLA-S^{ME} with the assigned residues indicated.
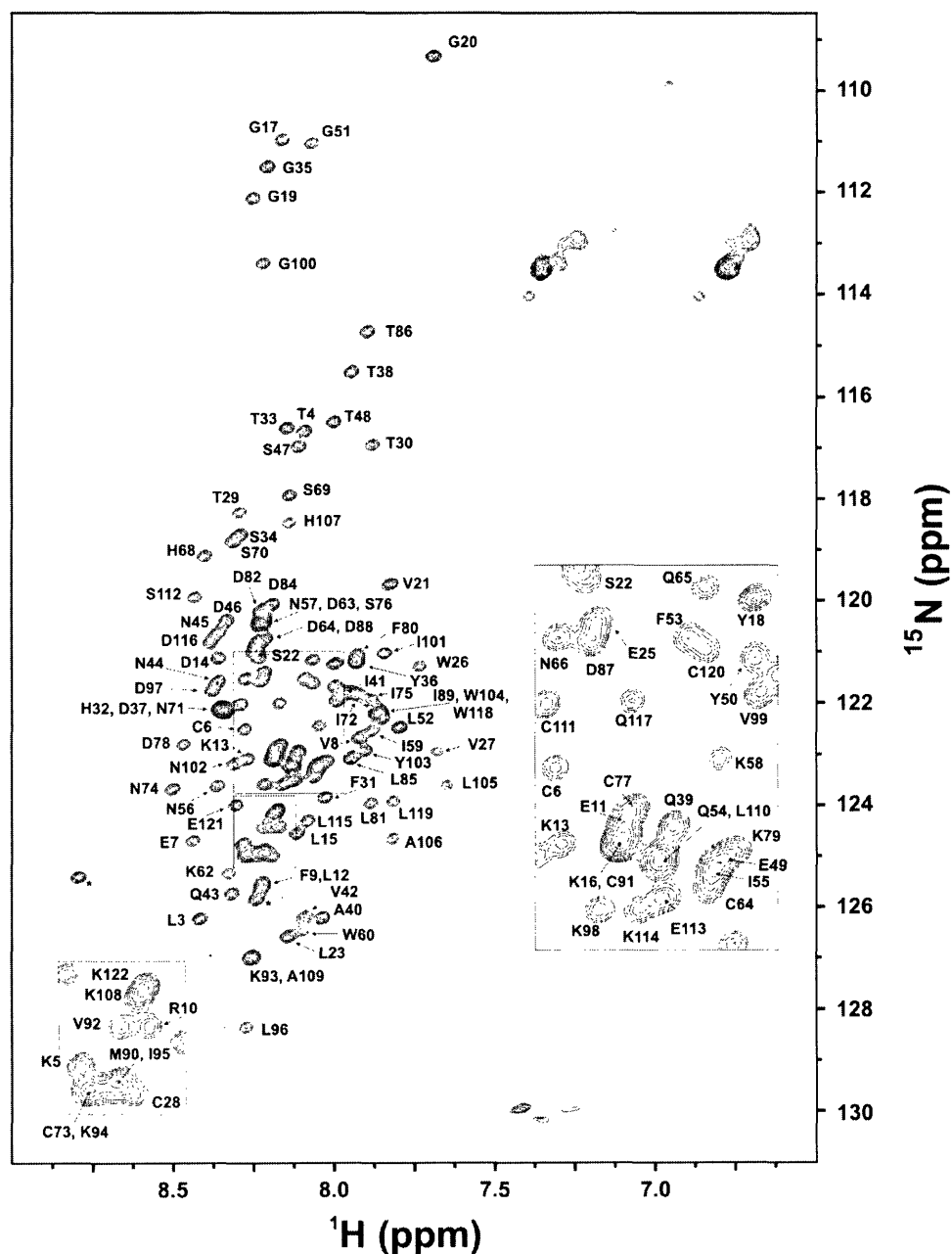
Figure 54: ^{1}H,^{15}N HSQC spectrum of BLA-S^{ME} at 800 MHz. Signals arising from the protein backbone are indicated by the residue type and number. Enlargements of two regions are shown due to limited space for indicating the respective residue number.

144

117 out of potentially 120 assignable backbone residues could be assigned using $^{13}C,^{15}N$ labeled BLA-S^{ME}. $^{13}C,^{15}N$ labeled BLA-S^{ME} had to be used for the assignment since only 48% of the protein could be assigned based on NOESY-HSQC and TOCSY-HSQC spectra (chapter 1) using only ^{15}N labeled protein (not discussed further here, Figure 55).
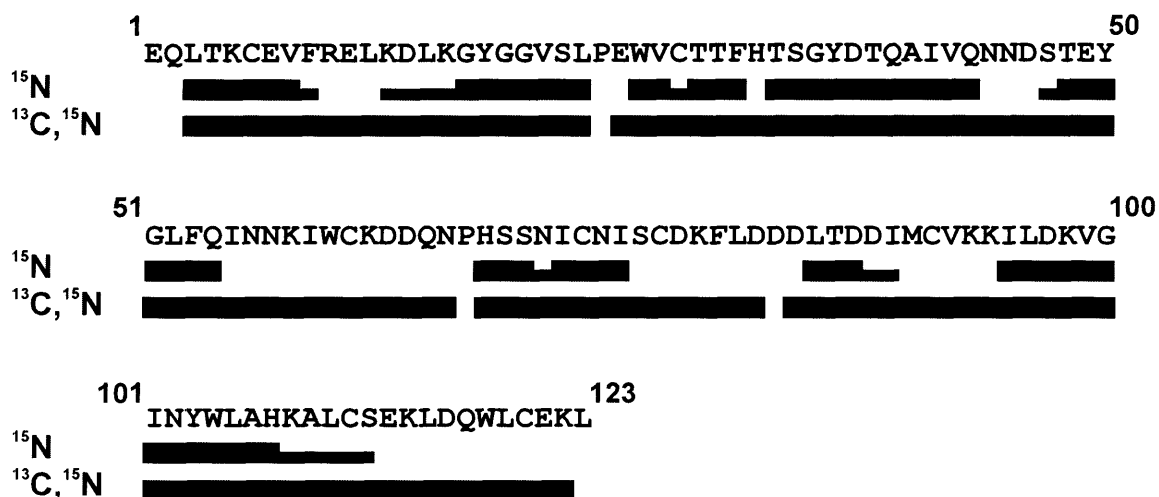
Figure 55: Success of assignment using ^{15}N labeled and/or $^{13}C,^{15}N$ labeled protein. Full bars indicate definite assignment, while half bars indicate tentative assignment only.

6.3.2 Residual secondary structure monitored by chemical shift deviations from random coil chemical shifts

Residual secondary structure in unfolded proteins is reflected in the deviations of experimental chemical shifts from the chemical shifts expected in a completely unstructured protein. There are different definitions of random coil chemical shifts: chemical shifts are either derived statistically from databases using unstructured regions of proteins (Wishart and Sykes, 1994), or experimentally derived by measuring the shifts of small unstructured peptides (Wishart et al., 1995a). Here, experimentally derived random coil chemical shifts were used (Wishart et al., 1995a), as statistically derived random coil chemical shifts do not sample the complete conformational space and should therefore not be regarded as true random coil chemical shifts (Wishart and Nip, 1998).
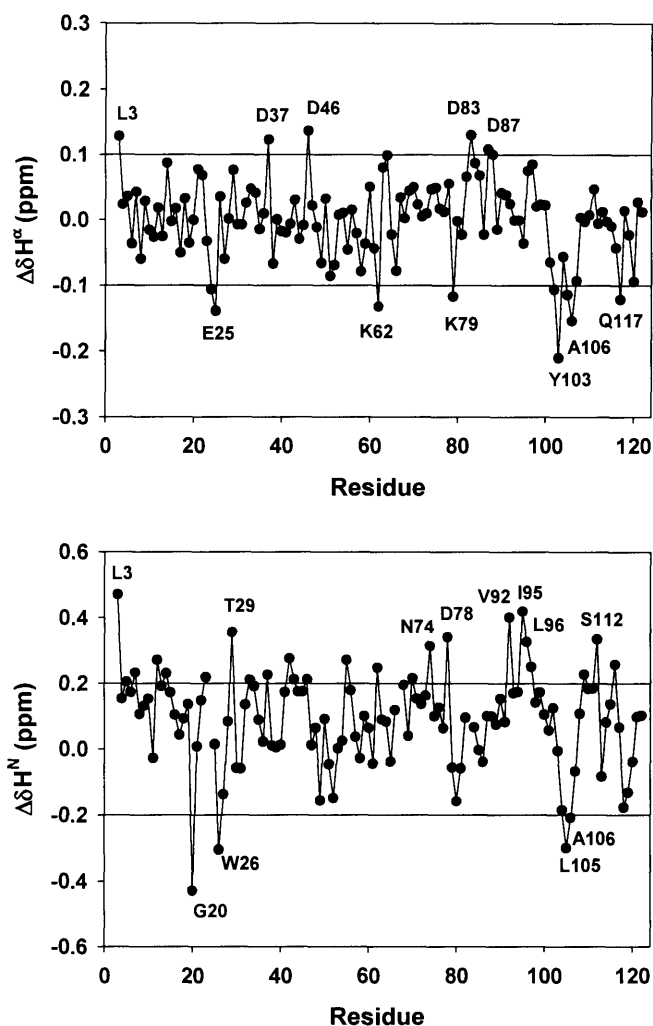
Figure 56: Deviations of H^α and H^N ($\Delta\delta H^\alpha$ and $\Delta\delta H^N$) chemical shifts in BLA-S^{ME} (δ^{ex}) from random coil chemical shifts (δ^{rc}, $\Delta\delta = \delta^{ex} - \delta^{rc}$). Black horizontal lines indicate the limits of the random coil regions (Wishart and Sykes, 1994; Wishart et al., 1992). Residues that have chemical shifts significantly deviating from random coil chemical shifts are labeled.

Figure 56 shows the deviation of H^α and H^N chemical shifts from random coil chemical shifts (Wishart et al., 1995a). Negative deviations of H^α and H^N chemical shifts from the expected random coil chemical shifts indicate residual α-helical structure. Significant negative deviations in both chemical shifts indicating residual α-helical structure are found in the region around W26 (H^N of W26, H^α of E25) and around L105 (H^α of Y103, H^N of L105 and H^α and H^N of A106). H^α chemical shifts alone furthermore also predict α-helical structure for K62 and K79, while H^N

chemical shifts predict α-helicity for G20. Significant positive deviations indicating β-sheet structure are found for the region around I95 (H^N of V92, I95 and L96) and for S112, N74, T29 (all H^N) and L3 (H^N and H^α). A number of aspartate residues also show positive deviations (D37, D46, D78, D83, D87). It is likely that these deviations are not due to residual β-sheet structure but rather due to differences in the chemical structure of the aspartate residue. Measurements were performed at pH 2, where aspartates are protonated. In contrast, unprotonated aspartate residues were used for the determination of random coil chemical shifts (pKa=4.4 (Stryer, 1988), Figure 57).
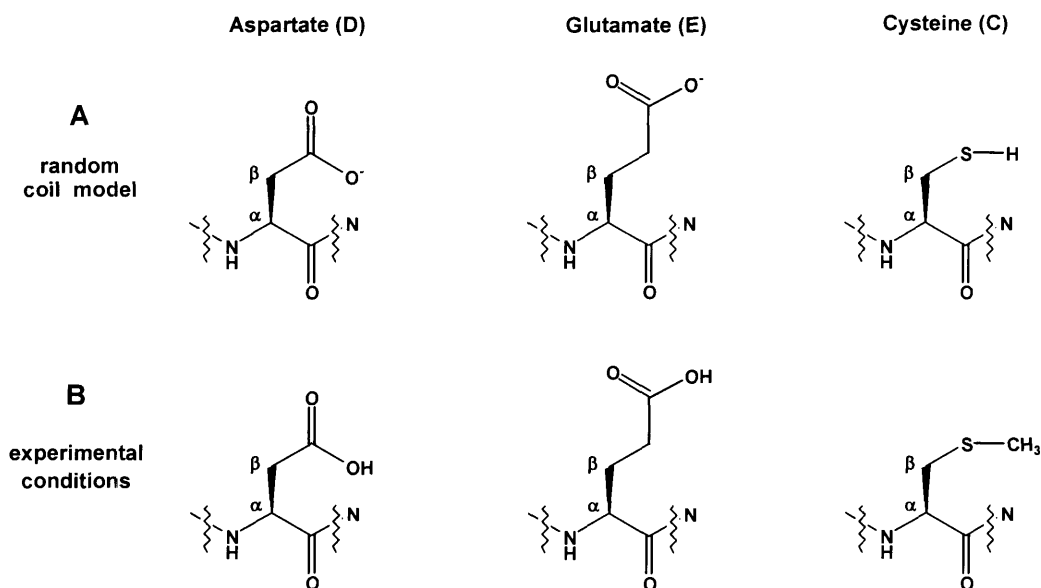
Figure 57: Chemical structures of aspartate, glutamate and cysteine residues within a protein. A) Structures of amino acid residues used to determine the chemical shifts in the random coil model at pH 5. B) Structures under experimental conditions at pH 2. α and β indicate the positions of C_α, H^α and C_β, discussed in the text and in Figure 56 and Figure 58. The pKa value of the side chains of glutamate and aspartate residues is approximately 4.4 (Stryer, 1988), justifying the chemical structures proposed for the respective pH values.

Amino acid specific deviations from random coil chemical shifts are even more pronounced for C_α and C_β chemical shifts (Figure 58). The largest deviations are observed for the reduced and methylated cysteine residues (Cys-S^{ME}): deviations of C_α chemical shifts range from -2.3 to -3.0ppm, while deviations of the C_β chemical shifts range from 9.1 to 9.3ppm. These large

deviations are not due to residual secondary structure in BLA-S^{ME} but due to differences in the chemical structure of the cysteine residues used for the determination of random coil chemical shifts (Wishart et al., 1995a), and the reduced and methylated variant of the cysteine residues used here (Figure 57).

As observed in the analysis of H$^\alpha$ chemical shift deviations, significant perturbations of C$_\alpha$ and C$_\beta$ chemical shifts are observed in aspartate residues and differently to the analysis of H$^\alpha$ chemical shift deviations also in glutamate residues. C$_\alpha$ chemical shift deviations of aspartates and glutamates range from -0.7 to -1.5ppm, C$_\beta$ chemical shift deviations range from -2.7 to -3.5ppm for aspartate residues and from -0.9 to -1.3ppm for glutamate. The deviations are again not due to residual structure but rather due to the different protonation state of the carboxy group of the amino acid side chain (Figure 57). Deviations of the C$_\beta$ chemical shifts of aspartates are larger than deviations of the C$_\beta$ chemical shifts of the glutamate due to the closer proximity of the aspartate C$_\beta$ to the carboxy group.
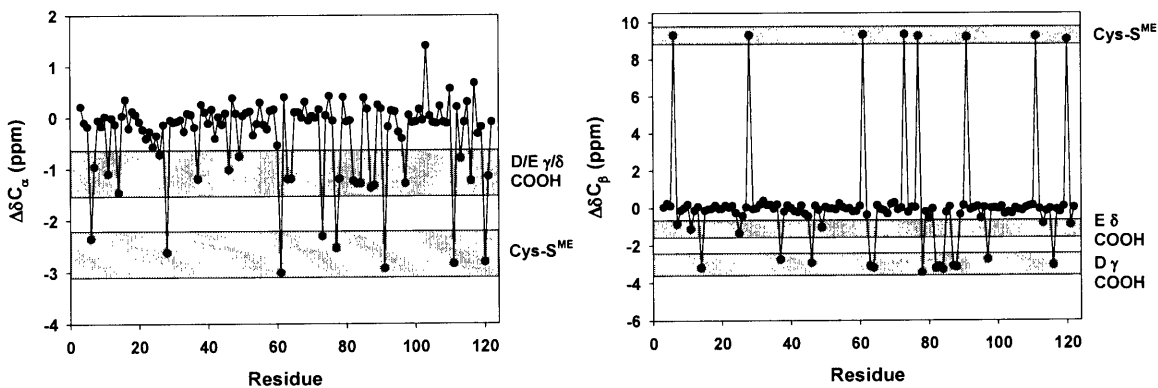
Figure 58: Deviations of C$_\alpha$ and C$_\beta$ ($\Delta\delta$C$_\alpha$ and $\Delta\delta$C$_\beta$) chemical shifts in BLA-S^{ME} δ^{ex} from expected random coil chemical shifts (δ^{rc}, $\Delta\delta=\delta^{ex}-\delta^{rc}$). Gray shaded areas indicate areas where only deviations of aspartates (D), glutamates (E) and methylated cysteines (Cys-S^{ME}) are found.

For an accurate analysis of residual secondary structure in BLA-S^{ME} monitored by carbon chemical shifts, all aspartate, glutamate and cysteine residues were taken out (Figure 59). No secondary structure is indicated by the C$_\beta$ chemical shifts, while only two residues deviate from the area of random coil chemical shifts monitored by the C$_\alpha$ chemical shifts: Y103 and Q117.
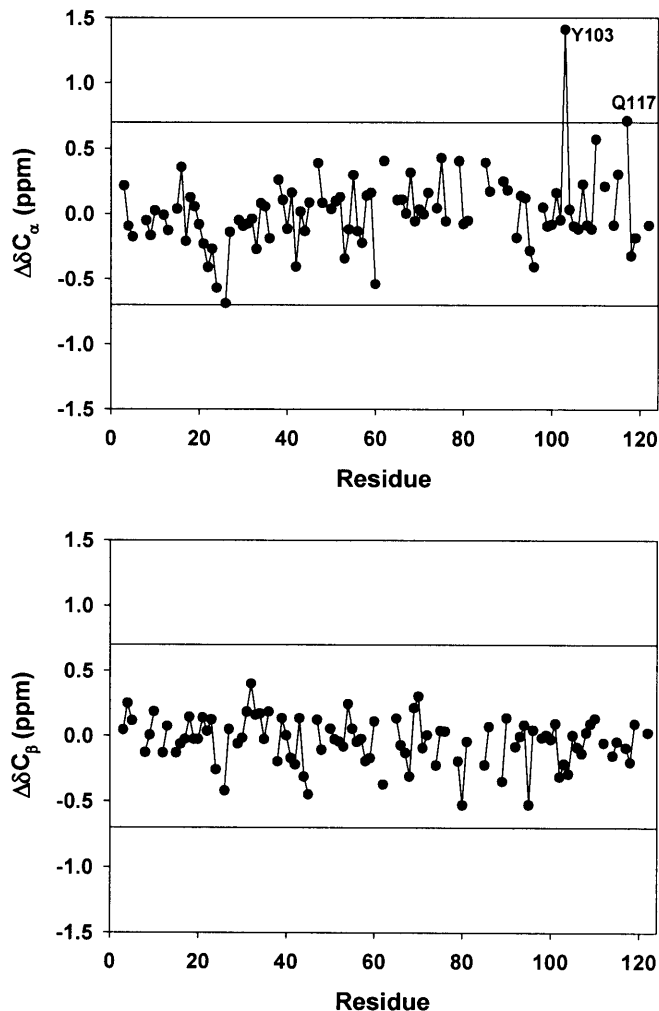
Figure 59: Deviations of C_α and C_β ($\Delta\delta C_\alpha$ and $\Delta\delta C_\beta$) chemical shifts in BLA-S^{ME} (δ^{ex}) from the expected random coil chemical shifts (δ^{rc}, $\Delta\delta=\delta^{ex}-\delta^{rc}$). Deviations of aspartate, glutamate and cysteine residues are not shown. Black vertical lines at 0.7 and -0.7ppm indicate borders used for chemical shift indexing (Wishart and Sykes, 1994; Wishart et al., 1992), residues within these borders are in a random coil state. Residues that deviate from the random coil are indicated.

6.3.3 Residual α-helical structure indicates a possible folding nucleus

BLA-S^{ME} is largely unfolded, as indicated by the few deviations from random coil chemical shifts that are observed. Three regions that exhibit pronounced residual α-helical structure, indicated by deviations of at least two types of chemical shifts, could be identified: around W26 (H$^\alpha$ and H^N chemical shift deviations), around Y103 and L105 (H$^\alpha$, H^N and C$_\alpha$) and at Q117 (H$^\alpha$ and C$_\alpha$) (Figure 56 and Figure 59). Aromatic residues are found in all of these regions (W26, Y103, W104 and W118). Furthermore, all these regions belong to the part of the protein that becomes the α-domain in the folded state of the protein. All three regions are part of helices found in the native state of the protein: the region around W26 and L105 and A106 belong to α-helices in folded BLA, while Y103 and the region around Q117 are part of 3_{10} helices in the folded protein (see also Table 13).

The presence of residual helical structure in regions that become helical in native BLA suggests that these regions are nuclei for the folding of the BLA. This assumption is supported by the close proximity of the respective folding nuclei in the native 3D structure (Figure 60). All three regions are in the α-domain of folded BLA.
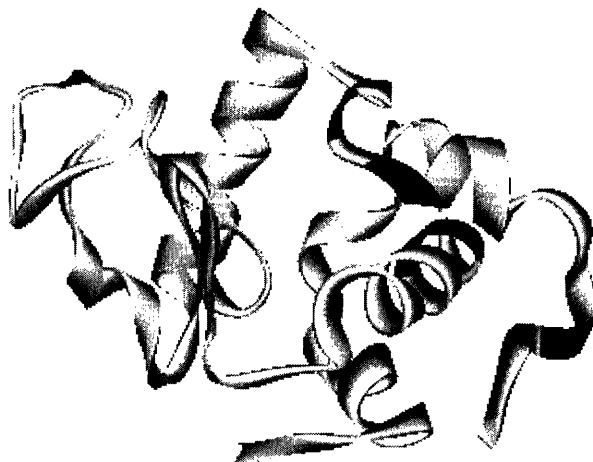
Figure 60: Graphical representation of the backbone structure of native BLA (pdb: 1HFZ, (Pike et al., 1996)). Residues that show large α-propensities (E25, W26, Y103, L105, A106 and Q117) in the unfolded state are colored in black. The picture was prepared using the program ViewerLite version 5.0 (Accelrys).

α-helix	³10-helix	β-sheet
5-11	18-20	41-43 *
23-34	77-80 *	48-50 *
86-98	101-103	55-56 *
105-110	115-118	

Table 13: Secondary structure elements in folded BLA (structure: 1HFZ from the pdb database (Pike et al., 1996)). Secondary structure is calculated and described according to an implementation of the method of Kabsch and Sander (Kabsch and Sander, 1983). Asterisks (*) indicate secondary structure elements in the β-domain of the protein, all other structure elements belong to the α-domain. Shaded secondary structure elements are discussed in the text.

6.4 Comparison of unbranched states of bovine and human α-lactalbumin and hen lysozyme: Results and Discussion

Comparison of BLA-S^{ME} with unbranched HLA and HEWL requires the absence of disulfide bridges in the respective proteins. Unbranched HEWL (HEWL-S^{ME}) was prepared by reduction and methylation of disulfide bridges as in BLA and has already been investigated in chapter 4 of this thesis. Unbranched HLA was obtained by mutation of all cysteines to alanines (All-Ala-HLA) (Redfield et al., 1999). BLA-S^{ME} and All-Ala-HLA are unfolded in 8M urea at pH 2 (see above for BLA-S^{ME}, (Redfield et al., 1999)) while HEWL-S^{ME} is unfolded at pH 2 in 8M urea as well as in water (pH 2 up to pH 6).

6.4.1 Sequence homology and hydrophobicity predictions

Sequence alignments of human and bovine α-LA and HEWL are shown in Figure 61. The sequence identity between HEWL and both α-LAs is approximately 40%, while the sequence similarity is almost 60%. Sequence identity between HLA and BLA is 75% whereas the similarity is 88%. In addition, the eight cysteine residues that form four disulfide bridges in the native state of the proteins align.

```
              10        20        30        40        50        60
HEWL: KVFGRCELAAAMKRHGLDNYRGYSLGNWVCAAKFESNFNTQATNRNTDGSTDYGILQINSRWWCN 65
      +CE+    +K    L   Y G SL  WVC      S ++TQA   +N D ST+YG+ QIN++ WC
BLA : EQLTKCEVFRELK--DLKGYGGVSLPEWVCTTFHTSGYDTQAIVQNND-STEYGLFQINNKIWCK 62
      +Q TKCE+ +  LK--D+ GYGG++LPE +CT FHTSGYDTQAIV+NN+-STEYGLFQI+NK+WCK
HLA : KQFTKCELSQLLK--DIDGYGGIALPELICTMFHTSGYDTQAIVENNE-STEYGLFQISNKLWCK 62
              10        20        30        40        50        60


              70        80        90       100       110       120
HEWL: DGRTPGSRNLCNIPCSALLSSDITASVNCAKKIVSDGNGMNAWVAWRNRCKGTDV-QAWIRGCRL 129
      D + P S N+CNI C   L  D+T  + C KKI+ D  G+N W+A + C  + + Q W+     L
BLA : DDQNPHSSNICNISCDKFLDDDLTDDIMCVKKIL-DKVGINYWLAHKALCS-EKLDQ-WLC-EKL 123
      Q P S NIC+ISCDKFLDDD+TDDIMC KKIL-D  GI+YWLAHKALC+-EKL+Q-WLC-EKL
HLA : SSQVPQSRNICDISCDKFLDDDITDDIMCAKKIL-DIKGIDYWLAHKALCT-EKLEQ-WLC-EKL 123
              70        80        90       100       110       120
```

Figure 61: Sequence alignment of HEWL, BLA and HLA. Alignment of HEWL and BLA results in 111 residues overlap (5-115) and indicates 42% sequence identity (47/111) and 58% sequence similarity (65/111). Similarly 39.1% sequence identity (45/115) and 59.1% sequence similarity (68/115) are found for the alignment of HEWL and HLA (not shown). All residues overlap in the alignment of BLA and HLA: 75% sequence identity (93/123) and 88% sequence similarity (109/123) are found.

Hydrophobic regions (regions that are prone to form residual secondary structure) also correspond to each other in sequence alignment, although there are differences if hydrophobicities are observed more closely.
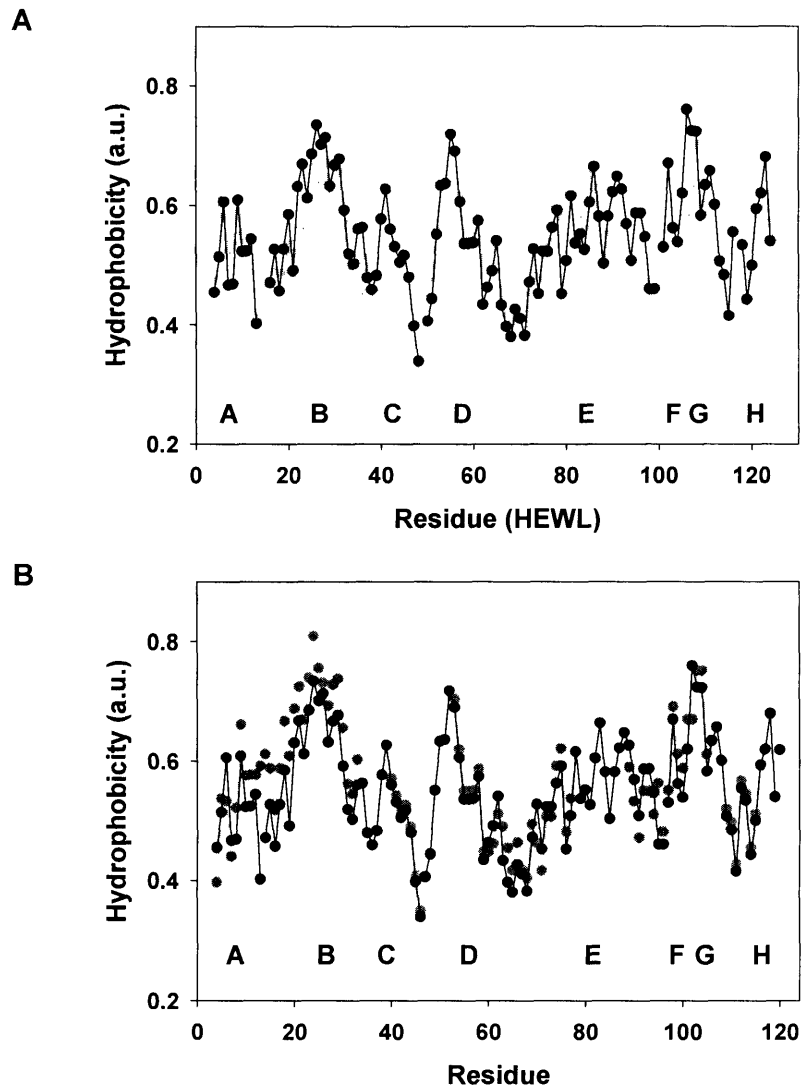
Figure 62: Hydrophobicity predictions using the scale of Abraham and Leo (Abraham and Leo, 1987). A) Hydrophobicities of HEWL (light gray triangles) and BLA (black circles). Residue numbering is according to the HEWL sequence and BLA was aligned according to Figure 61. B) Hydrophobicities of BLA (black circles) and HLA (dark gray circles). Areas of increased hydrophobicity in HEWL are indicated by numbers (1 through 6) while in BLA and HLA they are indicated by letters (A through G), area F is indicated not due to increased hydrophobicities but for further discussion in the following sections. The C to A mutations in HLA to obtain the unbranched All-Ala-HLA does not change the hydrophobicity profile ($R_2=0.9991$).

Figure 62 shows hydrophobicity predictions using the scale of Abraham and Leo (Abraham and Leo, 1987), for BLA, HLA and HEWL. Six regions of elevated hydrophobicity can be identified in HEWL, numbered 1-6 in Figure 62A. The hydrophobicity profiles of BLA and HLA (Figure 62B) are quite similar to the one of HEWL and almost identical with respect to each other, with the exception of area B where higher hydrophobicities are found in HLA. Hydrophobicities in BLA and HLA are on average slightly higher than in HEWL and regions of elevated hydrophobicity (compared to the basis level of the protein) are not as pronounced as in HEWL.

Eight areas of elevated hydrophobicity numbered A through G are indicated in Figure 62. Almost systematically higher hydrophobicities are found in area 2/B ($G16^{HEWL}$ and $S36^{HEWL}$, $D14^{BLA/HLA}$ and $S34^{BLA/HLA}$), and the most pronounced differences are found in the regions between $T40^{HEWL}$-$T47^{HEWL}$ ($T38^{BLA/HLA}$-$N45^{BLA/HLA}$): while increased hydrophobicity is present in BLA and HLA (area C), rather polar residues are found in the corresponding part of the sequence in HEWL. Hydrophobicities correspond closely for all three proteins in the region of area 3/D between residue $D52^{HEWL}$ and $R61^{HEWL}$ ($E49^{BLA/HLA}$ and $K58^{BLA/HLA}$; $R^2=0.988$ between HLA and HEWL).

6.4.2 Deviations from random coil chemical shifts

Figure 63 shows a comparison of H^α and H^N perturbations from the chemical shifts measured in small peptides (Wishart et al., 1995a) for BLA-S^{ME}, All-Ala-HLA in 8M urea, HEWL-S^{ME} in 8M urea and in water. The majority of residues in the proteins fall within the range of negligible perturbations from the expected random coil values. Significant negative deviations, indicative of residual helical structure are found in all proteins for H^N of $G20^{BLA/HLA}$ ($G22^{HEWL}$) and for H^N and H^α chemical shifts in the regions around $W104/L105^{BLA/HLA}$ ($W108/W111^{HEWL}$) and $W118/L119^{BLA/HLA}$ ($W123^{HEWL}$) indicating residual α-helical structure elements. Furthermore, significant deviations around W62/W63 are found in HEWL-S^{ME} both in water and in urea. These coincide with the small deviations found in BLA-S^{ME} at K62 (H^α). In contrast to the other proteins, additional chemical shift perturbations around W26 are observed in BLA-S^{ME}.

154
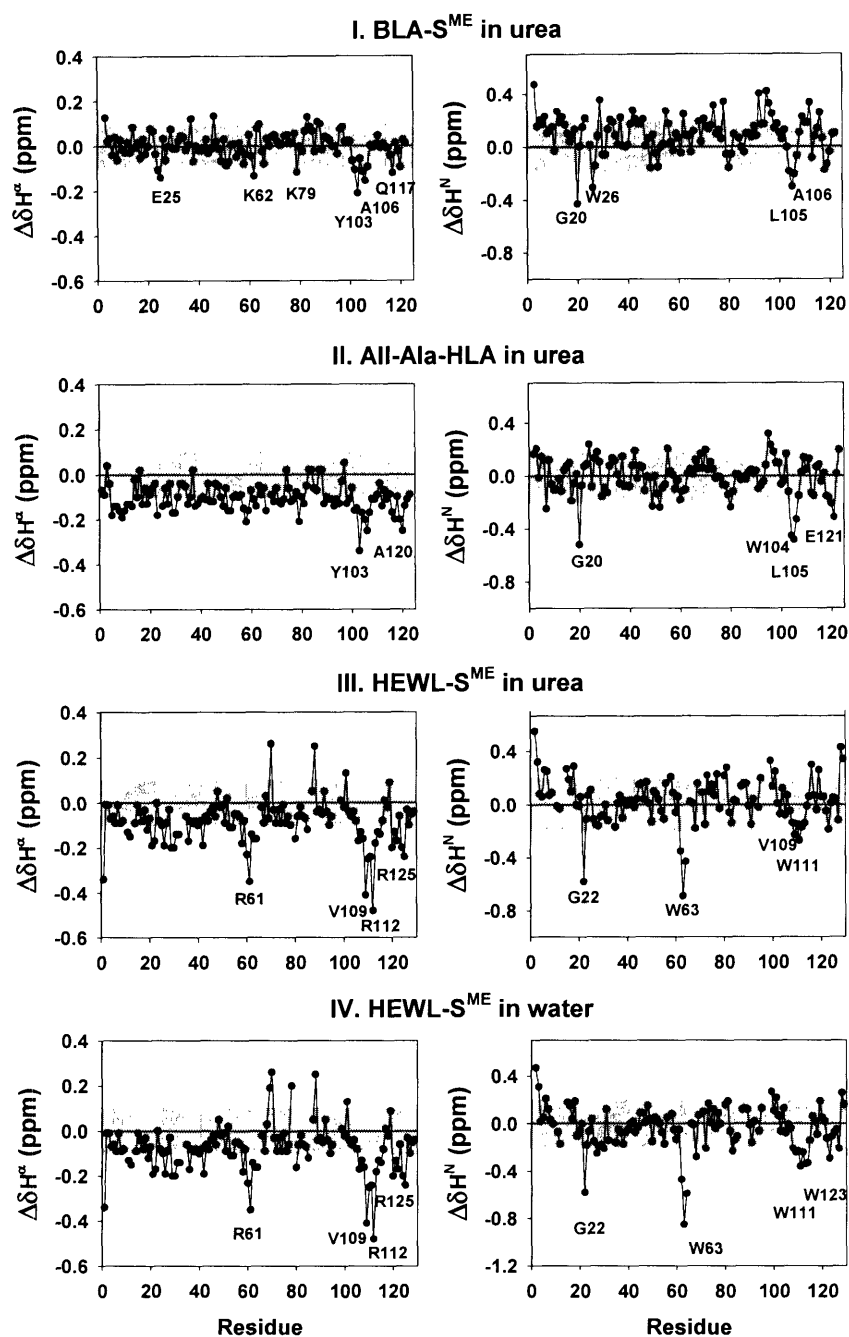
Figure 63: Deviations of H^α and H^N ($\Delta\delta H^\alpha$ and $\Delta\delta H^N$) experimental chemical shifts (δ^{ex}) from random coil chemical shifts (δ^{rc}, $\Delta\delta=\delta^{ex}-\delta^{rc}$) (Wishart et al., 1995a). I.) BLA-S^{ME} in urea, II.) All-Ala-α-LA in urea, III.) HEWL-S^{ME} in urea (reproduced from (Schwalbe et al., 1997)) IV.) HEWL-S^{ME} in water (reproduced from (Klein-Seetharaman et al., 2002)). Areas that show only small chemical shift perturbations are shaded in gray according to the areas used in chemical shift indexing (Wishart and Sykes, 1994; Wishart et al., 1992). Residues that show significant deviations are labeled.

155

6.4.3 R_2 relaxation rates

The R_2 relaxation rates measured in BLA-S^{ME}, All-Ala-HLA, and HEWL-S^{ME} are shown in Figure 64. The deviations from the random coil relaxation rates (R_2^{rc}) differ considerably for the three proteins and can be identified by Gaussian fitting as described in chapter 4 (Table 14): eight clusters of deviations from R_2^{rc} are found in BLA-S^{ME}, while only five clusters were identified in All-Ala-HLA, six clusters had been identified in HEWL-S^{ME} (see chapter 4).
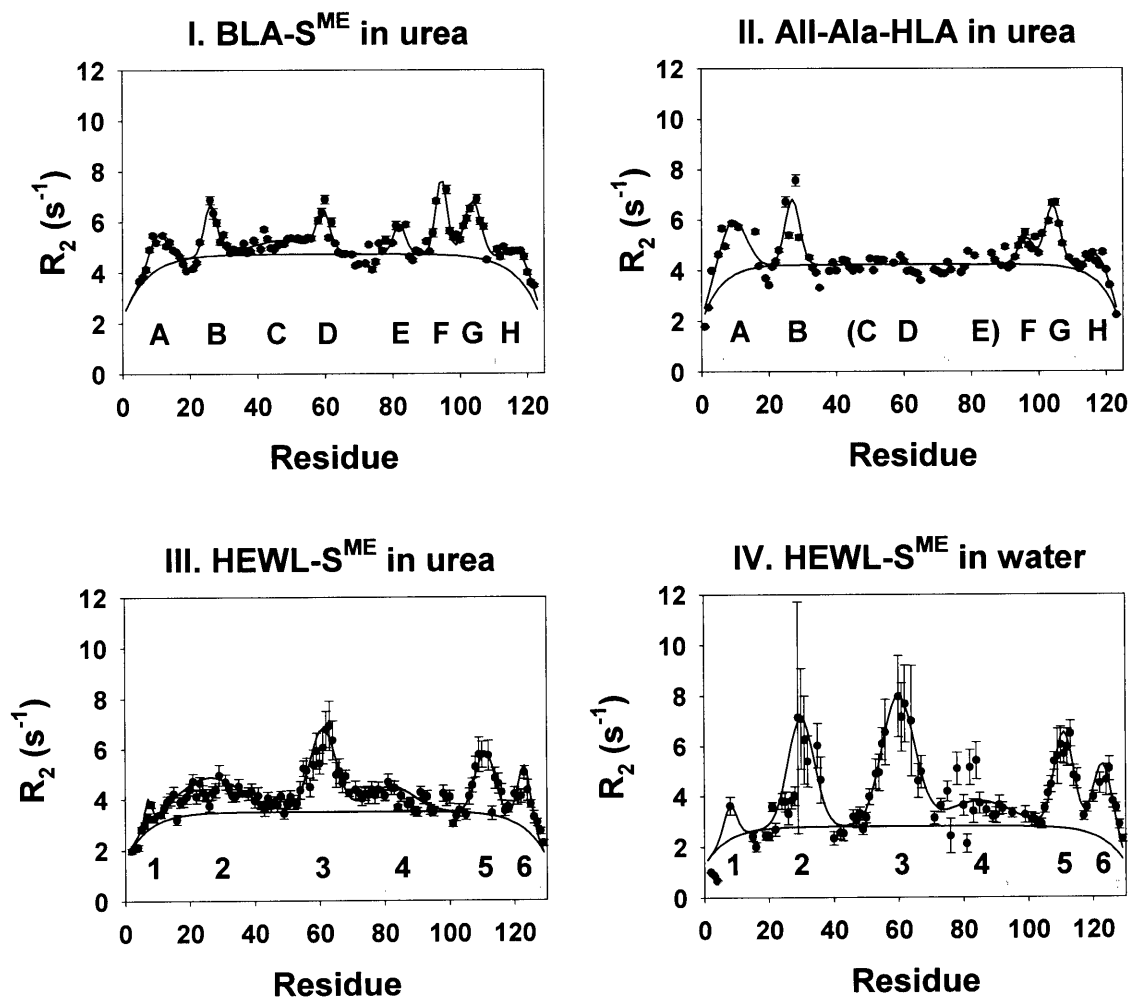
Figure 64: ^{15}N transverse relaxation rates (R_2) as function of residue for I.) BLA-SME in urea, II.) All-Ala-HLA in urea, III.) HEWL-S^{ME} in urea (reproduced from (Schwalbe et al., 1997)) and IV.) HEWL-S^{ME} in water (reproduced from (Klein-Seetharaman et al., 2002). The relaxation rates expected for a random coil (R_2^{rc}) and the Gaussian fit (Equation 6, Table 14) of experimentally determined relaxation rates are shown as black lines. Gray circles indicate the position of methylated cysteines (in I., III. and IV.) or cysteine to alanine mutations (in II.).

156

Gaussian clusters in BLA-S^{ME} center on (A) E11, (B) W26, (C) S47, (D) W60, (E) D82, (F) L96, (G) W104 and (H) Q117. The highest relaxation rates are observed in clusters (B), (D), (F) and (G): R_2(W26)=6.86s^{-1}, R_2(W60)=6.87s^{-1}, R_2(L96)=7.26s^{-1} and R_2(L105)=6.88s^{-1}. The five Gaussian clusters in All-Ala-HLA are centered around (A) S9, (B) I27, (F) L96, (G) W104, and (H) W118. The highest relaxation rates are measured in cluster (B) with R_2(A28)=7.56s^{-1} (C28 in wild type HLA). High R_2 rates are also found in cluster (A) R_2(S9)=5.86s^{-1} and in cluster (G) R_2(W104)=6.65s^{-1}, and R_2(L105)=6.67s^{-1}.

Cluster	BLA-S^{ME}			All-Ala-HLA		
	$x_{cluster}$	$R_{cluster}$	$\lambda_{cluster}$	$x_{cluster}$	$R_{cluster}$	$\lambda_{cluster}$
A	11	1.36	2.78	9	2.16	4.00
B	26	2.01	1.85	27	2.60	2.17
C	47	0.55	8.72	-	-	-
D	60	1.55	1.94	-	-	-
E	82	1.20	1.80	-	-	-
F	95 (95)	2.99 (1.06)	1.72 (3.44)	96	1.07	2.27
G	104	2.23	2.48	104	2.39	2.26
H	117	1.09	3.85	118	1.04	3.08

Table 14: Fitting results of Gaussian fitting for BLA-S^{ME} and All-Ala-HLA. Fitting was performed using Equation 6. $x_{cluster}$=center amino acid of the cluster (residue number), $R_{cluster}$=amplitude of the cluster in s^{-1}, $\lambda_{cluster}$=width of the cluster in units of residues. Fitting results for HEWL-S^{ME} are found in Chapter 4 and in (Klein-Seetharaman et al., 2002).

The clusters identified in the α-domain of HEWL-S^{ME} (1, 2, 5, and 6, see chapter 4), have their equivalent in clusters A, B, G and H in BLA-S^{ME} and in All-Ala-HLA. Furthermore, an additional cluster, cluster F, is present in both α-LAs.

Striking differences are found in the β-domain: Cluster 3 in HEWL-S^{ME} has its counterpart in BLA-S^{ME} (cluster D) but not in All-Ala-HLA. Moreover a very flat cluster around S47 (cluster C) and a cluster around residue D82 (cluster E) can be identified in BLA-S^{ME} but neither in All-Ala-HLA nor in HEWL-S^{ME}.

6.4.4 Comparison

For further discussion of the similarities and dissimilarities of the unfolded states of the three proteins, the results presented above are summarized in Table 15. Motional restrictions as monitored by R_2 relaxation rates coincide in all three proteins with hydrophobic areas, justifying

the name hydrophobic cluster. Residual secondary structure as monitored by chemical shift deviations only occurs within the hydrophobic clusters except for $G20^{HLA/BLA}/G22^{HEWL}$ (see above). Interestingly, residual secondary structure is not found in all of the hydrophobic clusters identified by R_2 relaxation rates, but only in approximately half of them. It is thus not the necessarily residual secondary structure that leads to decreased flexibility in the unfolded state. Similarly, aromatic residues and specifically tryptophan residues occur within some of the hydrophobic clusters. Interestingly while at least one aromatic residue is present in each of the major clusters in HEWL, this is not the case in the two α-LAs. Aromatic residues are absent in clusters A (All-Ala-HLA), B (All-Ala-HLA, BLA-S^{ME}) and F (BLA-S^{ME}), all of which are significant clusters. Furthermore, an aromatic residue is present at $W60^{HLA}$, however no cluster is present in area D of All-Ala-HLA, as monitored by R_2 relaxation rates. Even more striking is the fact, that the major difference between All-Ala-HLA and BLA-S^{ME} in area D is a L^{HLA} to I^{BLA} replacement at residue 59.

	Cluster 1/A				Cluster 2/B				Cluster C				Cluster 3/D			
	hyd	sec	R_2	ar	hyd	sec	R_2	ar	hyd	sec	R_2	ar	hyd	sec	R_2	ar
HLA	▨															W60
BLA	▨			F9		E25 W26		W26			▨			K62		W60
HEWL	▨		▨					W28						R61 W63		W62 W63

	Cluster 4/E				Cluster F				Cluster 5/G				Cluster 6/H			
	hyd	sec	R_2	ar	hyd	sec	R_2	ar	hyd	sec	R_2	ar	hyd	sec	R_2	ar
HLA					▨		▨			Y103 W104 L105		Y103 W104		A102 E121	▨	W118
BLA		K79	▨		▨					Y103 L105 A106		Y103 W104		Q117	▨	W118
HEWL			▨							V109 W111 R112		W108 W111		W123 R125		W123

Table 15: Comparison of the properties of the unfolded state of the three proteins. HLA=All-Ala-HLA, BLA=BLA-S^{ME}, HEWL=HEWL-S^{ME}, hyd=increased hydrophobicity, sec=residual secondary structure from chemical shift deviations, R_2=increased R_2 relaxation rates, ar=aromatic residues. Striped areas indicate moderate increases in hydrophobicity/R_2 relaxation rates, while gray areas depict significant increases in hydrophobicity/R_2 relaxation rates.

6.4.5 The unfolded state of proteins is determined by the primary sequence

The data presented here show that the unfolded states of the three proteins differ considerably from each other and from random coil behavior, with respect to motional restrictions and residual secondary structure. As the native structures of the three proteins are very similar, the structural properties of the unfolded state of the proteins are determined by the primary amino acid sequence and thus independent from the native structure. One property of the primary amino acid sequence is the hydrophobicity: Areas that exhibit motional restrictions in the unfolded state (Figure 64) often coincide with hydrophobic areas in the protein sequence (Table 15) and consequently also with areas with large volumes. However, this coincidence is by chance. From the data presented in this thesis it is clear that the properties of the unfolded state are neither predetermined by the hydrophobicity (see above and chapters 3 and 4) or volumes (see chapter 3) of the respective amino acids in the primary sequence, as proposed earlier (Schwarzinger et al., 2002). Rather specific interactions between amino acids and even long-range interactions lead to deviations from the expected random coil behavior.

It could be shown here that even the smallest differences in the sequence such as a conservative single point mutation of an isoleucine residue to a leucine dramatically changes the motional restrictions in the unfolded state. The combination of amino acid I59W60 in BLA-S^{ME} imposes motional restrictions on the region of cluster D, whereas no cluster is present in All-Ala-HLA where the combination L59W60 is found.

6.4.6 Relevance of hydrophobic clusters

The unfolded states of closely related proteins differ considerably. Those differences in reduced unfolded states could have some implications for the oxidative refolding of the respective protein. Figure 65 shows the deviations of fitted R$_2$ relaxation rates in the unfolded state from the expected random coil relaxation rates mapped onto the native structures of BLA, HLA and HEWL, respectively. Clusters in BLA-S^{ME} are found all over the protein, while clusters in All-Ala-HLA are only located in the α-domain of the protein. Similarly to BLA-S^{ME}, clusters in HEWL-S^{ME} are found in both domains of the proteins (see also chapter 4). Comparison of the regions exhibiting deviations of R$_2$ relaxation rates from R$_2$^{rc} in HEWL-S^{ME} in 8M urea at pH 2 (Schwalbe et al., 1997) (panel IV) and in water at pH 2 (Klein-Seetharaman et al., 2002; Wirmer et al., 2004) (panel III) reveals that those regions overlap under both conditions. Furthermore, the

extent of secondary structure, as determined from chemical shift perturbances, is nearly identical (Figure 63). The general features of the denatured states are therefore similar under refolding (pH 2) and under denaturing (pH 2, 8M urea) conditions. Residual structure in the unfolded state under any conditions can therefore be considered as the starting point of oxidative refolding under native conditions.
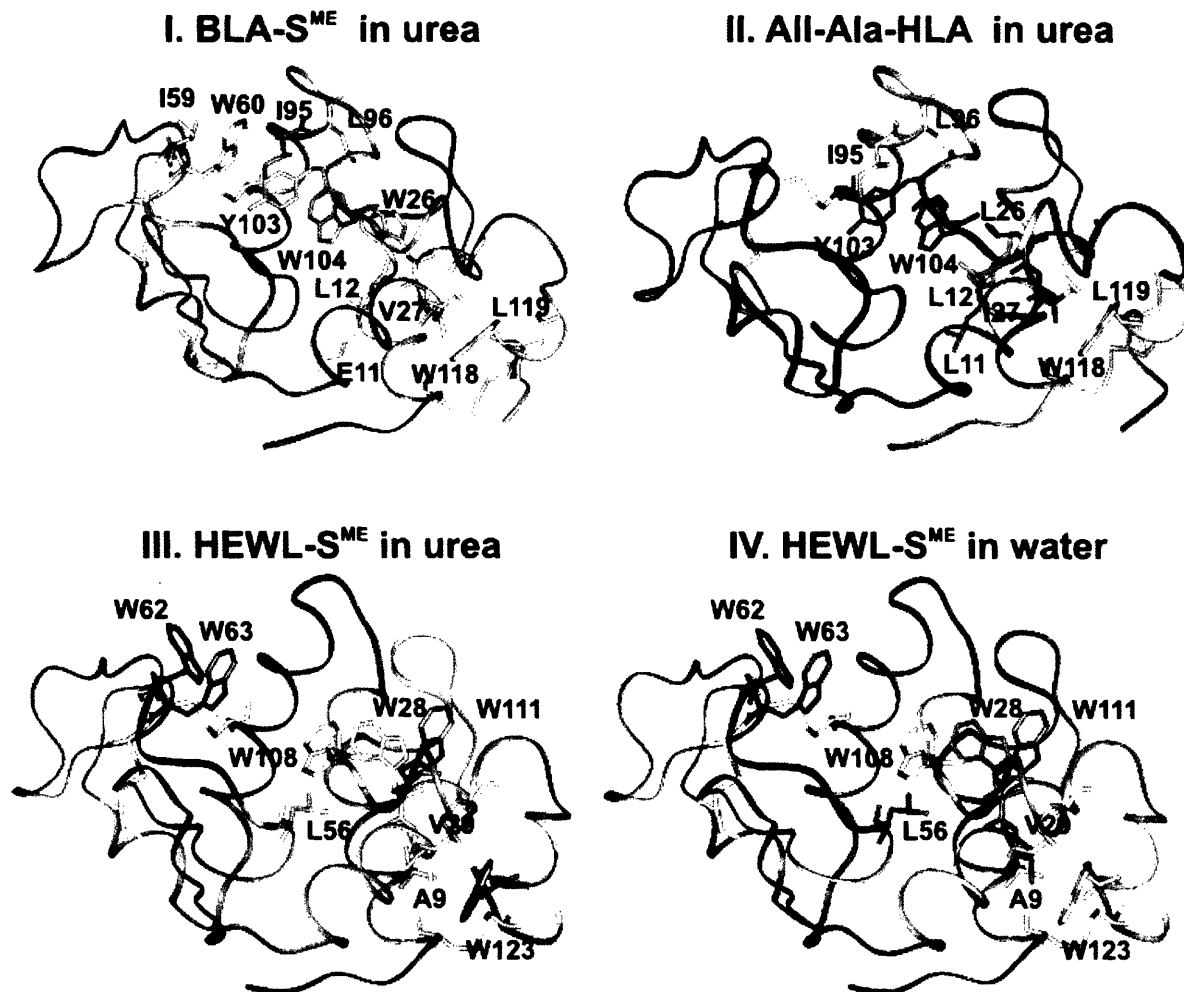
Figure 65: Deviations of fitted R_2 relaxation rates from the relaxation rates expected in a random coil mapped onto the native structure of I.) BLA (pdb-file: 1HFZ (Pike et al., 1996)), II.) HLA (pdb-file: 1HFZ (Pike et al., 1996)), III.) and IV.) HEWL (pdb-file: 193L (Vaney et al., 1996)). No deviations are indicated in red, while 100% deviations are colored in blue, 100% deviation is defined independently for each of the proteins/conditions. In the case of BLA-S^{ME} the highest deviation was scaled according to cluster G. The picture was prepared using the program InsightII (msi).

Analysis of the hydrophobic clusters in HEWL-S^{ME} by a combination of non-conservative single point mutations and NMR techniques (chapter 4) revealed long-range interactions between the clusters. These long-range interactions are stabilized by cluster 3, as well as by interactions between the clusters in the α-domain of the protein. As discussed in chapter 4, the interactions found within the α-domain could account for the formation of the 30-115 disulfide bridge during oxidative folding, while interactions of the β-domain cluster 3 with clusters in the α-domain could be important for the build up of domain-domain interactions, and structuring of the β-domain.

Similarly, motional restrictions and residual secondary structure in both unfolded α-LAs in regions within the α-domain could account for the formation of the 28-111 disulfide bridge during oxidative folding of these proteins (Figure 65).

The situation is different with respect to the regions of α-LA that are in the β-domain of the folded protein: there are no motional restrictions in the unfolded state around W60 in All-Ala-HLA, while there is cluster D in BLA-S^{ME}. As protein folding of BLA and HLA is alike (Chang, 2004; Chang and Li, 2002; Peng and Kim, 1994; Peng et al., 1995; Wu et al., 1995b) it can be assumed that cluster D in BLA-S^{ME} has no influence on the folding of BLA. Consequently, there are no long-range interactions between the β- and the α-domain in the unfolded state of α-LA. This is different to HEWL-S^{ME}, where cluster 3 mediates long-range interactions between the two domains. Possibly the guiding role of cluster 3 in the oxidative refolding of HEWL is replaced in BLA by the Ca^{2+} binding site that connects the two domains in α-LA. The importance of the Ca^{2+} binding site in oxidative folding has been proven by the fact that oxidative refolding is faster and a lot more efficient in the presence of Ca^{2+}-ions than in its absence (Chang, 2004; Chang and Li, 2002).

This hypothesis is further supported by the observed sequence conservation: cluster 3 in HEWL-S^{ME} is stabilized by aromatic interactions in all known and investigated lysozyme variants by either a YW or a WW pair (McKenzie and White, 1991; Nitta and Sugai, 1989). The corresponding position in α-LA contains, in contrast, an XW or XF pair where X is no aromatic residue, but an aliphatic residue with X=L, I, M, G, N and D (McKenzie and White, 1991; Nitta and Sugai, 1989). It can be argued that the conservation of aromatic residues at position 62 in lysozymes is not due to folding properties but due to its catalytic function and its importance in substrate recognition. However, it has been shown that the bacteriolytic efficiency of lysozyme

mutants containing a non-aromatic hydrophobic amino acid at position 62 is equal or even increased compared to the wild type (Maenaka et al., 1994). It is therefore concluded here, that while the presence of a WW or YW pair in lysozyme is necessary to promote protein folding, it is not necessary in α-lactalbumin due to the presence of the Ca^{2+}-binding site.

6.5 Conclusion

The assignment of methylated bovine α-lactalbumin (BLA-S^{ME}) has been described here. It was possible to assign 97% of the protein using $^{13}C,^{15}N$ labeled protein. ^1H and ^{13}C backbone chemical shifts were used to analyze the degree of residual secondary structure present in the unfolded protein by comparison to random coil chemical shifts derived from small peptides (Wishart et al., 1995a). Residual α-helical structure and motional restrictions are found in parts of the α-domain of BLA that become helical in the folded state, indicating the presence of a folding nucleus.

A comparison of the properties of the unfolded state of BLA-S^{ME} with All-Ala-HLA and HEWL-S^{ME} showed that unfolded states of proteins differ considerably for proteins that possess very similar structures in the native state. The properties of the unfolded state are solely determined by the primary structure of the protein and even the smallest conservative mutations can considerably change the structural ensemble present in the unfolded state.

A closer inspection of the properties of the three unfolded proteins in the light of oxidative protein folding revealed the following: Residual structure and long-range interactions in unfolded states of proteins can be nuclei for protein folding. This property of residual structure can be replaced by ion-binding sites.

Chapter 7

Monitoring of an intermediate on the Ca^{2+} induced folding pathway of bovine a-lactalbumin using timeresolved photo-CIDNP

7 Monitoring of an intermediate on the Ca^{2+} induced folding pathway of bovine α-lactalbumin using time-resolved photo-CIDNP NMR

7.1 Introduction

The aim of this chapter was to develop a method to study laser triggered kinetics by photo-CIDNP NMR and to apply this method to study the Ca^{2+} dependent refolding kinetics of bovine α-lactalbumin (BLA).

The refolding of BLA by tenfold dilution from GdnHCl has been investigated in the groups of Dobson and Hore. Time-resolved NMR and photo-CIDNP NMR studies revealed a close similarity between a folding intermediate populated after a rapid collapse during the slow (>120s) refolding pathway (Forge et al., 1999; Maeda et al., 2000) and the molten globule state (MG, chapter 1.4.2.2) of BLA. Here, a different approach for the study of refolding of BLA was chosen, which allows the detection of folding intermediates at high time-resolution (200ms) at site-specific resolution. Ca^{2+} ions have little effect on the structure of the native state of BLA (Wijesinha-Bettoni et al., 2001). However, Ca^{2+} stabilizes the protein against denaturants such as GdnHCl and urea, high temperature or pressure (Dolgikh et al., 1981). The refolding of BLA can therefore be induced at a constant urea concentration by the release of Ca^{2+}-ions (Kühn and Schwalbe, 2000) from the photolabile ion chelator DiMethoxy-Nitrophen (DMN) (Figure 66) (Ellis-Davies and Kaplan, 1988).
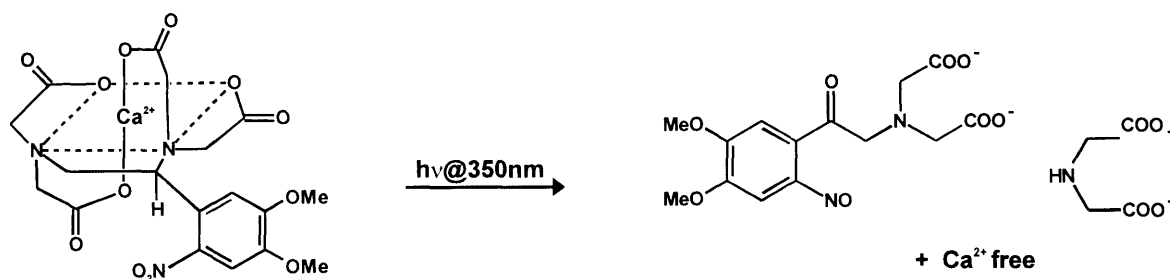
Figure 66: Structure of DMN (Ellis-Davies and Kaplan, 1988) and its photo products upon irradiation at 350nm. $K_D=10^{-9}M$ for Ca^{2+} which decreases to $K_D=10^{-3}M$ upon irradiation at 350nm. The metal is released within 10μs and the quantum yield of the photolysis is 0.18. Effective photolysis is additionally favored since intact DMN absorbs a lot stronger than its photo products ($\varepsilon^{\lambda=350}$(DMN) > $\varepsilon^{\lambda=350}$(photo-prod)).

High temporal resolution can be achieved using this setup. The time resolution of the experiment is only limited by the irradiation power and the concentration of ions that have to be released to initiate folding. The ion concentration in turn depends on the protein concentration. Kinetics of the Ca^{2+} induced folding monitored by stopped-flow fluorescence, agree remarkably well: Folding proceeds via an intermediate with rate constants of $k_1=11.4s^{-1}$, $k_2=1.3s^{-1}$ and $k_1=8.9s^{-1}$, $k_2=1.0s^{-1}$, respectively. Refolding of methyl-groups monitored by 1D proton NMR in contrast follows slower single-exponential kinetics $(0.16s^{-1} \leq k \leq 0.33s^{-1})$ (Kühn and Schwalbe, 2000). It was concluded, that the formation of secondary structure precedes the acquisition of the exact position of methyl groups.

A

$$U \xrightarrow{k_1} MG \xrightarrow{k_2} F$$

dilution from GdnHCl ▶

B

$$U \xrightarrow{k_1} I \xrightarrow{k_2} F$$

$+ Ca^{2+}$ in Urea ▶
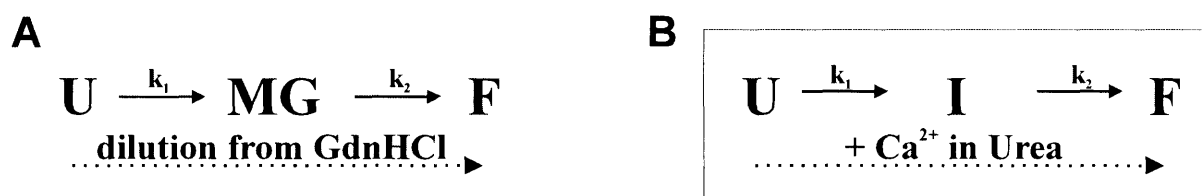
Figure 67: Schematic drawing of BLA refolding A) by dilution from 6M GdnHCl (Forge et al., 1999) and B) by the addition of Ca^{2+}-ions at constant urea concentration (Kühn and Schwalbe, 2000). U, MG, I and F indicate the unfolded, molten globule, intermediate and folded state of BLA.

The nature of the folding intermediate populated in the Ca^{2+}-induced refolding of BLA in 4M urea is yet unknown. It could be either molten globule like, as in the refolding by dilution from GdnHCl, or could have a distinct structure. Previous results (Kühn and Schwalbe, 2000) showed that the folding intermediate can not be picked up in methyl-group kinetics, it should however be visible in the kinetics of aromatic residues, since the rearrangement of the aromatic core is known to be one driving force of protein folding. The aromatic region of protein NMR spectra on the other hand (Figure 74) is significantly less resolved than the methyl-region, even in the folded state of a protein. Therefore, it is not possible to follow the kinetics of single aromatic side-chains using conventional 1D proton NMR spectroscopy. Here the changes in the accessibility of aromatic amino acids significant for the rearrangement of the hydrophobic core during refolding of α-lactalbumin from the unfolded state were characterized. This could be achieved by developing a new method. The refolding is photo-initiated by ion release from photo-labile chelators and changes in the accessibility are detected using photo-CIDNP NMR during the course of refolding. Combination of ion release from DMN with photo-CIDNP combines the

advantages from photochemical folding initiation (Kühn and Schwalbe, 2000) and photo-CIDNP-NMR signal detection (Bargon et al., 1967; Hore and Broadhurst, 1993; Kaptein, 1982; Maeda et al., 2000). (i) The dead-time of the experiment is only governed by the signal-to-noise of the NMR experiment, mixing inhomogeneities are avoided and (ii) resolution and intensities of aromatic resonances (such as Trp, Tyr and His) that are not within the core of the protein but exposed to solvent and therefore in van-der-Waals contact with CIDNP dyes, are enhanced.

7.2 Materials and Methods

7.2.1 Protein preparation

All experiments have been performed on Ca^{2+}-free bovine α-lactalbumin (BLA). Since commercially available bovine α-lactalbumin (Fluka) contains traces of Ca^{2+}, the protein was depleted from Ca^{2+}-ions by ultrafiltration. The protein was dissolved in 50ml desalting buffer containing 100mM EDTA and 100mM Tris, pH 7 to chelate Ca^{2+}-ions. Subsequently the protein was washed two times with 50ml Tris-buffer (100mM Tris, pH 7) and four times with 50ml water using amicon YM10000 membranes. The protein was lyophilized and stored at -20°C until usage.

7.2.2 Synthesis of DM-Nitophen

General procedures:

All solvents were dried before use following general drying procedures and all reactions were performed under argon using dried glassware. As DMN and it precursors are light sensitive, all steps were carried out in the dark, using brown flasks.

7.2.2.1 Synthesis of 2-Nitro-4,5-dimethoxystyrene (2)

100ml of a 1M solution of sodium hexamethyl disilazid in THF were added under argon to cooled down (0°C) suspension of methyl-triphenylphosphonium bromide (37.5g 0.1mol) in 100ml dry THF. The mixture was stirred for 1h at 0°C before 100ml of a solution of 15.9g (75mmol) 6-Nitrovertaldehyd in 150ml THF was added drop wise. Stirring continued over night at room temperature before the reaction was quenched by the addition of 200ml of saturated NH_4Cl solution. The crude product was extracted using CH_2Cl_2, the organic phases were combined and tried using $MgSO_4$ and concentrated. Further purification was done using flash

chromatography on a silica column using n-hexane/ethylacetat 3:1 as mobile phase. The product was the first compound that eluted from the column. 2-Nitro-4,5-dimethoxystyrene is a yellow solid which melts at 61°C. The yield of the synthesis was 41%.

^{1}H-NMR: ^{1}H-NMR (CDCl$_3$): 7.60 (s, 1 H), 7.32 (dd, J=11.0, 0.8 Hz, 1 H), 6.99 (s, 1 H), 5.64 (dd, J=17.2, 0.7 Hz, 1 H), 5.34 (dd, J=11.0, 0.8 Hz, 1 H), 4.00 (s, 3 H), 3.96 (s, 3 H)

R$_f$: 0.46 (n-Hexan/Essigester 3:1)

7.2.2.2 Synthesis of 1-(2-Nitro-4,5-dimethoxyphenyl)ethan-1,2-diyl-1,2-dibromide (3)

1.4ml bromine (27mmol) in 20ml CH$_2$Cl$_2$ were added drop wise (within 10min.) to a solution of 24mmol 2-Nitro-4,5-dimethoxystyrene in 80ml CH$_2$Cl$_2$ at 0°C. Stirring was continued at room temperature for 4h before the reaction was quenched using 100ml of a saturated Na$_2$S$_2$O$_3$ solution. The crude product was extracted using CH$_2$Cl$_2$, the organic phases were combined and tried using MgSO$_4$ and concentrated. Further purification was done using flash chromatography on a silica column using toluene/ethylacetate 9:1 as mobile phase. The product is very light sensitive brown oil; the yield of the synthesis was 85%.

^{1}H-NMR (CDCl$_3$): 7.54 (s, 1 H), 7.11 (s, 1 H), 6.21 (dd, J=5.8, 9.8 Hz, 1 H), 4.05 – 3.97 (m, 2 H), 4.01 (s, 3 H), 3.96 (s, 3 H)

R$_f$: 0.72 (Toluol/Essigester 9:1)

7.2.2.3 Synthesis of 1-(2-Nitro-4,5-dimethoxyphenyl)-N,N,N',N'-tetrakis[(ethoxycarbonyl)-methyl]-1,2-ethanediamin (4)

1-(2-Nitro-4,5-dimethoxyphenyl)ethan-1,2-diyl-1,2-dibromide, (7g, 19mmol), imminodiacetic acid diethyl ester (7.9g (42mmol) and 0.1mmol Na$_2$CO$_3$ (0.1g) in 40ml acetonitril (AcN) were refluxed over night. The reaction mixture was distributed between 60ml water and 60ml CH$_2$Cl$_2$ and the water phase was extracted four times using 20ml CH$_2$Cl$_2$. The solvent and any residual water and AcN were carefully removed using high vacuum and the crude product was further purified using flash chromatography on a silica column using hexane/ethylacetat 7:3 as mobile phase. The product is the 2nd compound that elutes from the columng, the first compound eluting from the column is the educt which can be used again. The product is a yellow light sensitive oil, and yield is 15%.

^{1}H-NMR (CDCl$_3$): 7.70 (s, 1 H), 7.43 (s, 1 H), 5.12 (dd, J=7.1, 8.9 Hz, 1 H), 4.14 – 4.03 (m, 8 H), 3.95 (s, 1 H), 3.87 (s, 3 H), 3.77 – 3.42 (m, 8 H), 3.14 (dd, J=7.1, 13.8 Hz, 1 H), 2.89 (dd, J=8.9, 13.8 Hz, 1 H), 1.25 – 1.14 (m, 12 H).

7.2.2.4 Synthesis of 1-(2-Nitro-4,5-dimethoxyphenyl)- N,N,N',N'-tetrakis[acetacid]-1,2-ethanediamin (DM-Nitrophen, DMN)

0.7g (1.2mmol) of 1-(2-Nitro-4,5-dimethoxyphenyl)-N,N,N',N'-tetrakis [(ethoxycarbonyl)methyl]-1,2-ethanediamin in a solution of 2.5 ml 2M NaOH, 2.5 ml pyridine and 1ml ethanol was stirred for 16h at room temperature. The progress of the reaction was monitored by TLC (methanol). The reaction product was extracted three times with 5ml diethyl ether. This was than reextracted using 6ml deionized water. The pH of the combined water phases was adjusted to pH 7 using HCl. To achieve complete removal of pyridine, the solution was concentrated four times to half the volume in the vacuum and subsequently diluted with deionized water again. The solution was lyophilized and DMN was gained as yellow solid.

The obtained DMN is not 100 % pure, but contains substantial amounts of NaCl. The yield of the reaction therefore was determined by NMR (by comparison of signal intensities to signal intensities of a known amount of the reference compound TMSP) and by UV spectroscopy using the following extinction coefficient ε=4.33*10^{3}M^{-1}cm^{-1} at 350nm (Kaplan and Ellis-Davies, 1988). The yield of the reaction was 79%, which leads to an apparent molecular weight for DMN of 1200g.

^{1}H-NMR (D$_2$O, pH 7.0): 7,6 (1H, s), 6,82 (1H, s), 4,92 (1H, CH-CH$_2$, dd, J=11,4; 2,2 Hz), 3,98 (3H, OMe, s), 3,86 (3H, OMe, s), 3,18 (1H, CH-CH$_2$, dd, J=13,6; 11,4 Hz), 3,39; 3,29; 3,04; 2,60 (8H, NCH$_2$COO, d, J=16,3 Hz), 2,55 (1H, CH-CH$_2$, dd, J=13,6; 2,2 Hz)

7.2.3 Stopped-flow fluorescence

Refolding kinetics were recorded using an Applied Photophysics π*-180 stopped flow spectrometer at 35°C (α^{exc}=289nm, α^{em}>320nm). The refolding was initiated by mixing 0.1mM BLA against 0.2mM CaCl$_2$, both in buffer (2M urea, 50mM Tris/HCl at pH 7), yielding an end concentration of 50μM BLA. A baseline was recorded by mixing BLA against buffer to exclude dilution effects. The dead time of the experiments was 1ms.

Data analysis by fitting of the experimental data to mono- and biexponential decays was performed using the program Origin. Simulation of folding kinetics was done using the Maple software package.

7.2.4 Lasercoupling

Materials:

Argon-ion lasers (spectra physics: model 2085 emitting at 350nm (beam-diameter 1.9nm, power 7.25W) and model 2017 emitting at 488 and 515nm (beam-diameter 1.4nm, power 6.25W)),

electronical shutters (Uniblitz®, Vincent Associates),

dichroic beamsplitter (CVI Laseroptics),

reflecting mirror (CVI Laseroptics),

multimode silica/silica fiber (1500μM core, Ceram Optecs) held in a fiber holder,

lens (CVI Laseroptics),

cone shaped shigemi tip (shigemi),

beam block (thorlabs),

power meter (Soliton, Laser- und Meßtechnik).

All optical elements were mounted on a breadboard (Thorlabs) using post holders (Thorlabs).

Method:

For the implementation of the laser coupling all elements of the coupling optics were prepositioned as shown in Figure 69. Using a piece of paper attached to a post holder, the focus of the UV-Laser was determined by sliding the paper back and forth between the fiber and the lens at low laser power level (100mW). This visualizes the position of the beam as fluorescence and yields to burning a hole into the paper at the maximum focus. Further optimization of coupling of the UV-laser was achieved using a power meter (Soliton, Laser- und Meßtechnik) and adjusting the screws at the fiber holder. Coarse focusing of the VIS-laser was achieved by variation of the position of the reflecting mirror, again a piece of paper was used to visualize the laser beam (without burning a hole). Fine adjustment of the VIS-laser was achieved by turning the screws of the reflecting mirror while monitoring the output power.

7.2.5 NMR measurements

All NMR measurements were done at a 750MHz homebuilt NMR spectrometer (Magnex magnet) at 35°C. Sample conditions were as follows:

Static photo-CIDNP: all samples: 0.5mM BLA, 0.2mM FMN in D_2O; U: 8M urea, pH*2; F: 50mM cacodylat, 10mM $CaCl_2$, pH* 6.7 (uncorrected pH, measured in D_2O). MG: pH*2.

Kinetic photo-CIDNP: 0.1mM BLA, 0.5mM $CaCl_2$, 0.5mM DMN, 0.05mM FMN, 2M urea, 50mM cacodylat buffer in D_2O, pH* 6,7.

Laser-titration: 0.25mM bovine α-BLA, 0.5mM $CaCl_2$, 0.5mM DMN, 4M urea, 100mM Tris buffer in D_2O, pH* 6,7.

Samples containing DMN were prepared under red light to prevent photo-degradation. All samples were 170μl only.

7.3 Results and Discussion

7.3.1 Synthesis of DM-Nitrophen

The synthesis of DM-Nitrophen was performed similarly to the approaches described in the literature (Ellis-Davies and Kaplan, 1988; Kühn and Schwalbe, 2000) with modifications, as illustrated in Figure 68. DM-Nitrophen is synthesized starting from Nitrovertraldehyd (1). The first step is a Wittig reaction by which the styrene derivative (2) is synthesized. This is subsequently brominated to yield (3). The key step in the synthesis is the nucleophilic substitution reaction by which the nitrophenester (4) is obtained. In contrast to the pressure (5000psi) catalyzed method described by Ellis-Davies and Kaplan (Ellis-Davies and Kaplan, 1988) the reaction was carried out catalyzed by a base. After the thorough purification of the Ester (4), cleavage yields DM-Nitrophen (5).
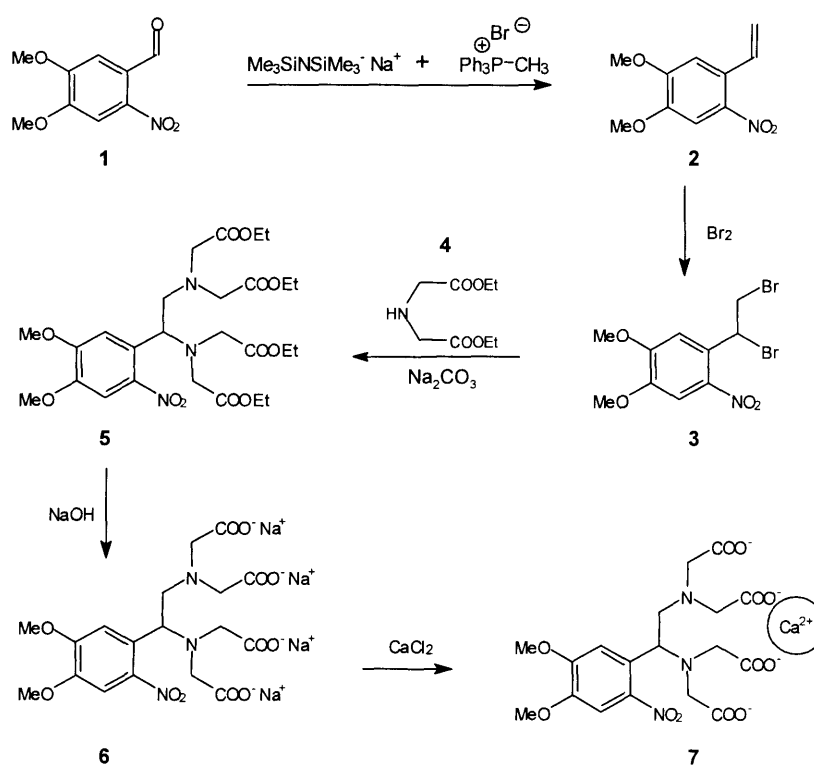
Figure 68: Synthesis of DM-Nitrophen (DMN). 1: 6-Nitrovertraldehyd, 2: 2-Nitro-4,5-dimethoxystyrene, 3: 1-(2-Nitro-4,5-dimethoxyphenyl)ethan-1,2-diyl-1,2-dibromide, 4: 1-(2-Nitro-4,5-dimethoxyphenyl)-N,N,N',N'-tetrakis[(ethoxycarbonyl)methyl]-1,2-ethanediamin, 5: DM-Nitrophen.

7.3.2 Laser setup

7.3.2.1 Laser coupling

For the combination of folding initiation by Ca^{2+} release from DMN (350nm) and photo-CIDNP (488 and 515nm) detection, light beams from two argon-ion lasers emitting at 350nm (UV-laser) and at 488 and 515nm (VIS-laser) were coupled into an NMR tube inside a 750MHz homebuilt spectrometer. Coupling of one laser into an NMR tube in the spectrometer via quartz fibers had been described in the literature (Kühn and Schwalbe, 2000; Scheffler et al., 1985), whereas coupling of two high power lasers in contrast has not been described before. The following requirements for both lasers had to be fulfilled:

→ The dead time of the kinetic NMR experiment depends on effective ion release from DMN, which in turn depends on the power input in the NMR sample. Therefore the efficiency of the coupling for the UV-laser has to be maximized and at best be as good as in a setup where a UV-laser alone is used (78% at 7W).

→ Induction of photo-CIDNP using the VIS-laser in contrast does not require the maximum power output (7W), but can be successfully performed at 50% power level (3W). This allows a coupling efficiency as low as 50% (at 6W).

The aim of the laser coupling is to find a setup for the combination of both lasers into one quartz fiber without reducing the coupling efficiency of the UV-laser, compared to the use of the UV-laser alone. This could be achieved using the following sheme (Figure 69).
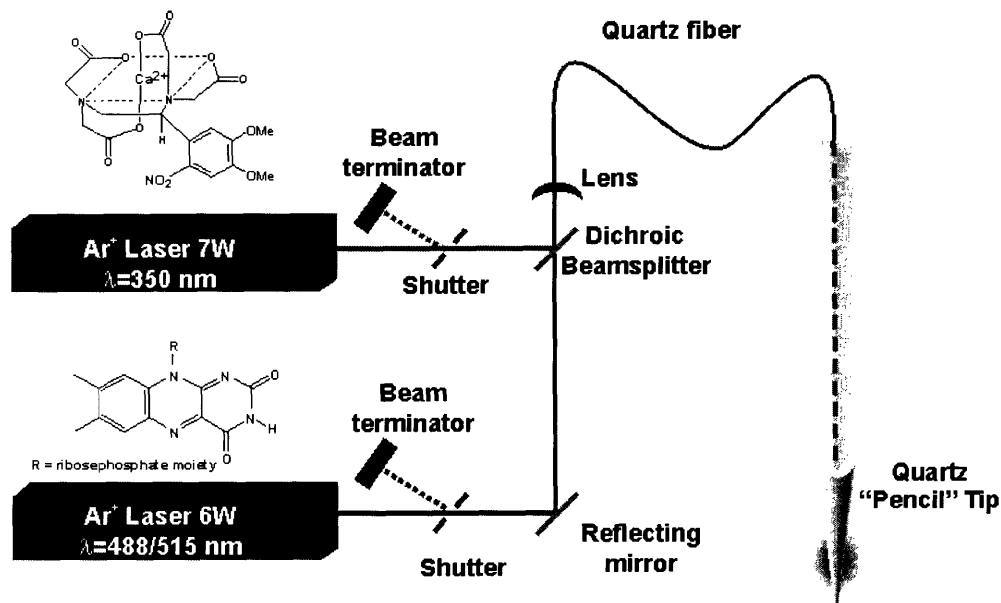
Figure 69: Setup used to couple the light of two high power lasers into an NMR spectrometer. Both laser beams were guided through electronical shutters that were controlled by the spectrometer, to allow synchronization of laser irradiation for ion release and photo-CIDNP initiation and NMR signal detection. Combination of the two laser beams was performed using a reflecting mirror and a dichroic beamsplitter. A lens focused the combined beam on a multimode silica/silica fiber (held in a fiber holder). Final coupling of the light into the NMR tube inside the spectrometer was achieved by connecting the quartz fiber into a pencil shaped shigemi tip. This allows best distribution of light in the sample.

The challenging part of this setup was the choice of coupling optics, to combine two laser beams into one beam, while all other elements were chosen based on previous experience in the group (Kühn and Schwalbe, 2000). Combination of both beams was achieved using a dichroic beamsplitter. Dichroic beamsplitters can be used as beamcombiner: while one side of the dichroic beamsplitter reflects a certain wavelength, the other side of the dichroic beamsplitter transmits a different wavelength. However, transmission yield of dichroic beamsplitter is never 100%. A LWP (long wavelength pass) dichroic beamsplitter was chosen here, to achieve maximum yield for the reflected UV-laser, while allowing loss during the transmission of the longer wavelength of the VIS-laser.

The coupling efficiency for both lasers is shown in Table 16. It was possible to use the UV-laser in a combined setup (77% light output) without reducing the coupling efficiency compared to a direct setup (78% light output). Coupling yield is 66 % for VIS laser in the combined setup, while

it is 81% if the VIS-laser is used alone. The reduction of the coupling efficiency in the combined setup for the VIS-laser is a) due to expected loss of light power during the transmission through the dichroic beamsplitter and b) due to imperfections in the fiber coupling, which is optimized for the UV-laser.

The aim of obtaining a coupling efficiency of the UV-laser in the combined setup (77%) comparable to the coupling efficiency alone (78%), while the coupling efficiency of the VIS-laser does not drop below 50% could be achieved.

	UV-laser		VIS-laser	
	alone	combined	alone	combined
Fiber output	78%	77%	81%	66%
Fiber input	86%	86%	88%	78%
Loss fiber	9%	10.5%	8%	15.5%
Loss lens/lens +beamsplitter	14%	14%	12%	22%

Table 16: Yield and loss of laser coupling in % of primary output from the respective laser.

7.3.2.2 Timeresolution: Timescales of ion release from DM-Nitrophen

Ion release from DMN within the NMR spectrometer can be monitored by the fraction of BLA that is folded after a specified laser pulse. Figure 70 shows the result of such a laser-titration experiment using a protein concentration of 0.25mM, 0.5mM Ca^{2+} and 0.5mM DMN and laser pulses of 100ms each. Ion release during the first laser pulses is almost linear. Approximately 70% of BLA is folded after three laser pulses have been applied to the sample. 70% folded protein is enough protein to detect simple kinetics by subsequent measurement of 1D NMR spectra without repeating irradiation at a later time and taking difference measurements. This is a huge increase of time resolution compared to earlier measurements where several irradiations of 1s each were needed to release enough Ca^{2+}-ions to fold the protein, due to the fact that a less powerful laser was used (2W, 50% coupling efficiency (Kühn and Schwalbe, 2000), compared to 7W, 77% coupling efficiency described here). This time resolution will allow to characterize folding intermediates that are populated during the 1s of refolding, e.g. the folding intermediate of BLA populated 300ms after folding initiation (Kühn and Schwalbe, 2000).

Power (J)

Number of 100ms irradiations
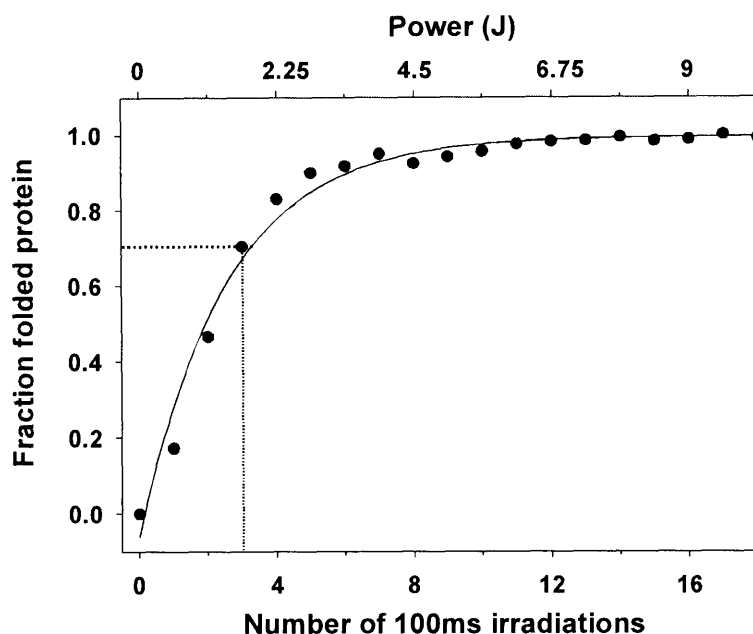
Figure 70: Fraction of folded protein as a function of laser irradiations. Each laser pulse is 100ms at 350nm (7W). Sample conditions were 0.25mM BLA, 0.5mM Ca^{2+}, 0.5mM DMN in 4M urea, 100mM Tris, pH 7, 35°C, D_2O.

7.3.2.3 Optimized refolding conditions for the combination of folding initiation by ion release and photo-CIDNP folding detection

Earlier refolding experiments of BLA by Ca^{2+} release from DMN monitoring methyl-kinetics, have been performed in 4M urea, 100mM Tris at pH 7 (Kühn and Schwalbe, 2000). Changes of these conditions had to be made to allow the combination of ion release with photo-CIDNP signal detection. The following aspects had to be considered and are discussed in detail below:

(i)   photo-CIDNP is quenched by tertiary amines,

(ii)   ion-release from DMN has to be possible in the presence of the photo-CIDNP dye,

(iii)   almost complete folding (>95%) of BLA at the end of refolding.

(i) Quenching of the photo-CIDNP effect by secondary and tertiary amines due to quenching of triplet state of flavines has been reported in the literature (Broadhurst et al., 1991; Kaptein, 1982). Tris as well as DMN are tertiary amines. To reduce the quenching effect, Tris-buffer was replaced by cacodylic acid: cacodylic acid does not interact with Ca^{2+}, and has no NMR signals

175

in the aromatic region. DMN however is the only photo labile Ca^{2+} chelator known that is suitable for the proposed experiment, and could therefore not be replaced by anything else. Quantification of the quenching effect using a mixture of Tyr and Trp (each 0.25mM) and the CIDNP sensitizer flavin-mononucleotide (FMN, 0.05mM) (Broadhurst et al., 1991; Kaptein, 1982) in the presence and absence of DMN (0.5mM), revealed that ~ 75% of CIDNP-intensity is attained even in the presence of DMN.

(ii) The most commonly used dye for photo-CIDNP NMR is FMN (Broadhurst et al., 1991; Kaptein, 1982). FMN, however, absorbs more strongly at 350nm than the photolabile Ca^{2+}-chelator DMN (Figure 71). Initiation of protein folding by ion release from DMN is prevented in the presence of equimolar amounts of FMN, even at very long overall irradiation times (5s compared to 200ms without FMN). This is presumably not only due to stronger absorption of FMN, but also due to a radical reaction that occurs if FMN and DMN are excited at the same time. It was shown, that folding initiation was only possible using FMN:DMN ratios as low as 1:10 (data not shown).
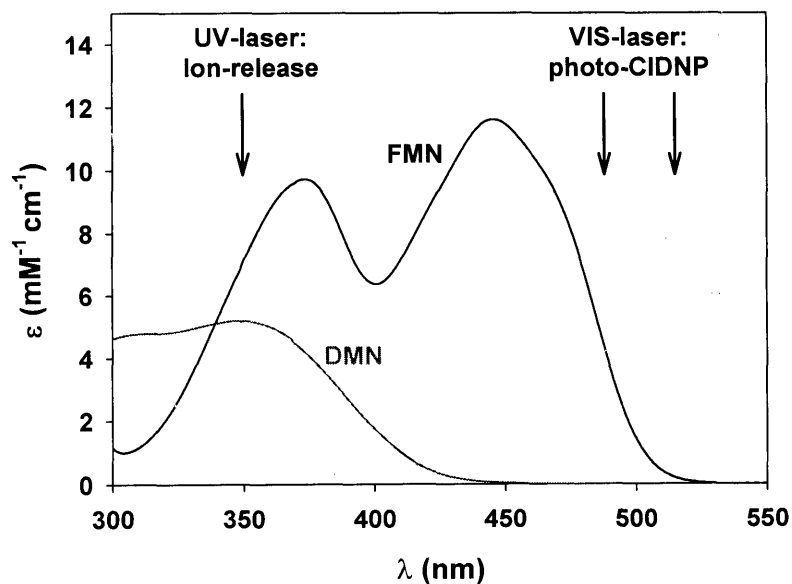
Figure 71: UV absorption spectra of DMN (dark gray), FMN (black) and Thionin (light gray). Arrows indicate wavelenghts used to initiate folding by ion release (350nm) and photo-CIDNP (488 and 515nm).

Thionine is the only known photo-CIDNP dye that does not absorb at 350nm (Glazkov and Rudakov, 1994). However, displacement of the well established FMN with Thionin is not possible. Thionin modifies BLA during the photo-CIDNP experiment. This modification in turn leads to unfolding of the folded protein (data not shown).

For the above mentioned reasons a ratio of 1:10 FMN:DMN is used for the combination of photo-CIDNP with ion release. Using 0.5mM Ca^{2+} and 0.5mM DMN, 0.2mM Ca^{2+} can be released by 200ms irradiation at 350nm to fold 0.1mM BLA.

(iii) At the conditions chosen to monitor the kinetics of methyl group refolding (4 M urea) (Kühn and Schwalbe, 2000), a minor fraction of the protein is still unfolded in the presence of Ca^{2+}-ions. This fraction can be neglected in the kinetic experiment using conventional ^1H NMR. However, this minor fraction cannot be neglected in photo-CIDNP NMR. The accessibility in the unfolded state of any protein is a lot higher than in the folded state of a protein. Even a small fraction (<20%) of unfolded protein dominates the CIDNP spectra of any folded protein. To shift the equilibrium towards a more folded state (~95% at the end of folding) only 2M urea are used in the kinetic photo-CIDNP experiments.

Refolding conditions can be summarized as follows:

0.1mM BLA, 0.5mM Ca^{2+}, 0.5mM DMN, 0.05mM FMN, 50mM cacodylat, 2M urea at pH* 6.7 in D_2O, 35°C (pH* uncorrected meter reading in D_2O, corresponds to pH 7 in H_2O).

7.3.3 Stopped-flow fluorescence kinetics of Ca^{2+} induced refolding of BLA

Changes in secondary and tertiary structure of proteins can be monitored by fluorescence spectroscopy: Aromatic residues, mainly tryptophans absorb light between 275nm and 297nm and emit it at higher wavelengths after relaxation processes have taken place. These relaxation processes depend on the environment of the fluorescing residue and can therefore be used to monitor kinetics. BLA has four tryptophan residues: Trp 26, Trp 60, Trp 104 and Trp 118; they are well distributed over the sequence and cover both, the alpha domain (W26, W104 and W118) and the beta domain (W60) of the protein (structure of BLA is shown in Figure 77).

Combination of fluorescence detection with stopped-flow technology where folding is initiated by rapid mixing of two solutions in the stopped-flow detection chamber, allows to detect folding

177

kinetics with dead times as low as 1ms. A fluorescence trace of the Ca^{2+}-induced refolding of BLA in 2M urea is shown in Figure 72. A comparison with dilution of the protein in the absence of Ca^{2+} reveals that no reaction takes place during the dead time of the experiment. The kinetics can be best described using double exponential decay with rate constants of $k_1=6.02\pm0.06s^{-1}$ and $k_2=1.29\pm0.02s^{-1}$ ($R^2=0.99993$ compared to $R^2=0.99698$ if fitted using monoexponential kinetics). These rate constants compare well with results from previous stopped-flow fluorescence measurements at 4M urea ($k_1=11.4s^{-1}$, $k_2=1.3s^{-1}$ (Kühn and Schwalbe, 2000)) indicating that the folding process is not significantly changed by the different denaturant concentration, and that folding proceeds via a folding intermediate.
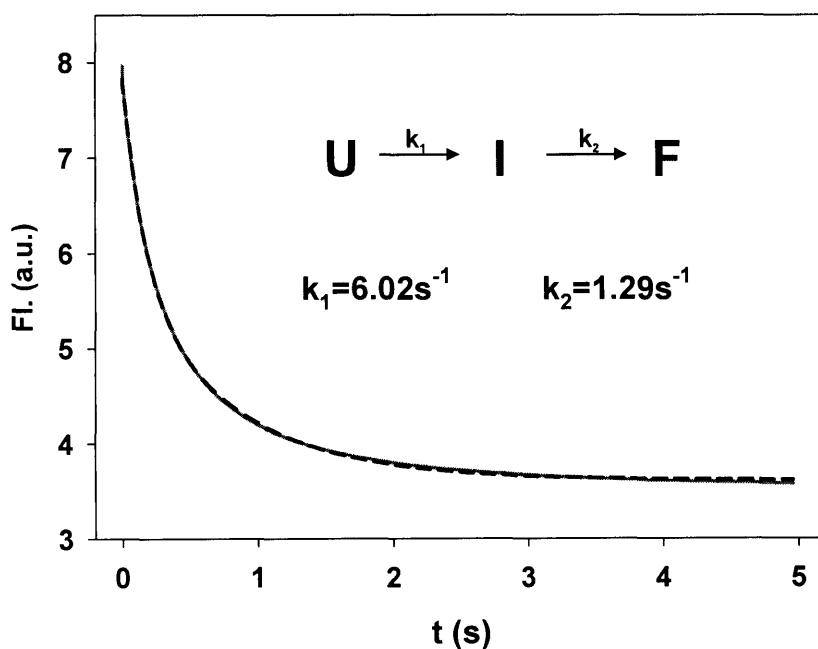
Figure 72: Stopped-flow fluorescence data ($\alpha^{exc}=289nm$) for the refolding of BLA in 2M urea upon the addition of two equivalents Ca^{2+}(Fl=fluorescence, gray line=experimental data, fitting results shown as black dots). Data are averages from 10 measurements, urea background was subtracted from the kinetic trace, fluorescence remains unchanged in blank dilution experiments in the absence of Ca^{2+}. Fitting to double exponential decay yields $k_1=6.02\pm0.06s^{-1}$, $A_1=2.06\pm0.02$ and $k_2=1.29\pm0.02s^{-1}$, $A_2=2.15\pm0.02$, $R^2=0.99993$ and $\chi^2=0.00047$.

Using the rates constants from the kinetic analysis a time-concentration profile of unfolded (U), intermediate (I) and folded (F) states of BLA can be derived (Figure 73). A folding intermediate is populated to 70% at τ=350ms. As the fluorescence of BLA is an overall property of both the β- and the α-domain it is not possible to deduct any specific information about the nature of this folding intermediate from the data. Time resolved photo-CIDNP NMR as shown in Chapter 7.3.5 can be used to characterize this folding intermediate.
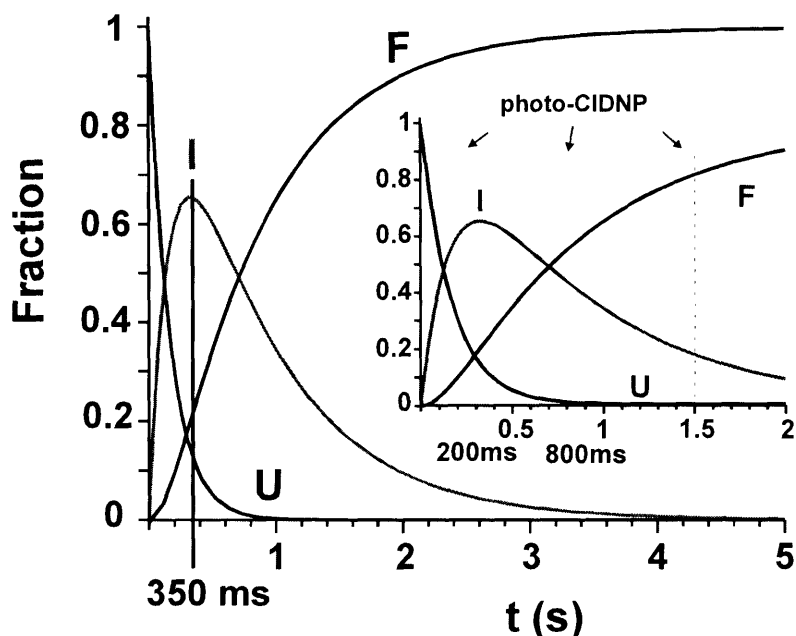
Figure 73: Time concentration profile of unfolded (U), folded (F), and intermediate (I) as calculated from k_1 and k_2 (Figure 72) using the software package Maple. Time points at which kinetic photo-CIDNP NMR spectra were recorded are indicated in the figure insert.

7.3.4 Static photo-CIDNP

NMR spectra of the three states of BLA are shown in Figure 74. Large chemical shift dispersion can be observed in the folded state (F) of BLA. In contrast, the chemical shift dispersion is small in the unfolded (U) and in the MG of BLA. While the signals are sharp in the unfolded protein, broad signals are observed in the MG due to slow conformational reorganization of the protein. Differentiation of the three states of the protein and every other protein can therefore be done based on these characteristics. On the other hand, the aromatic region of proteins are very

crowded and the identification of signals that belong to single amino-acid side chains is not possible in a conventional 1D NMR spectrum (Figure 74A). Photo-CIDNP (Hore and Broadhurst, 1993; Kaptein, 1982) spectra (Figure 74B) in contrast, are in general well resolved, due to the fact, that only side-chains that are accessible to the dye in solution are present in the spectrum. A second advantage of photo-CIDNP NMR is that signals are enhanced compared to conventional NMR spectra), furthermore, direct identification of amino-acid side chains is possible: while absorptive signals are observed for Trp and His residues, Tyr signals are emissive (Figure 74).

A

F

MG

U

7.8 7.4 7.0 6.6

¹H (ppm)

B

Trp 118

Trp 118

F

His 68

Tyr 18

MG

4 Trp

U

4 Tyr

8.0   7.5   7.0   6.5   6.0

¹H (ppm)
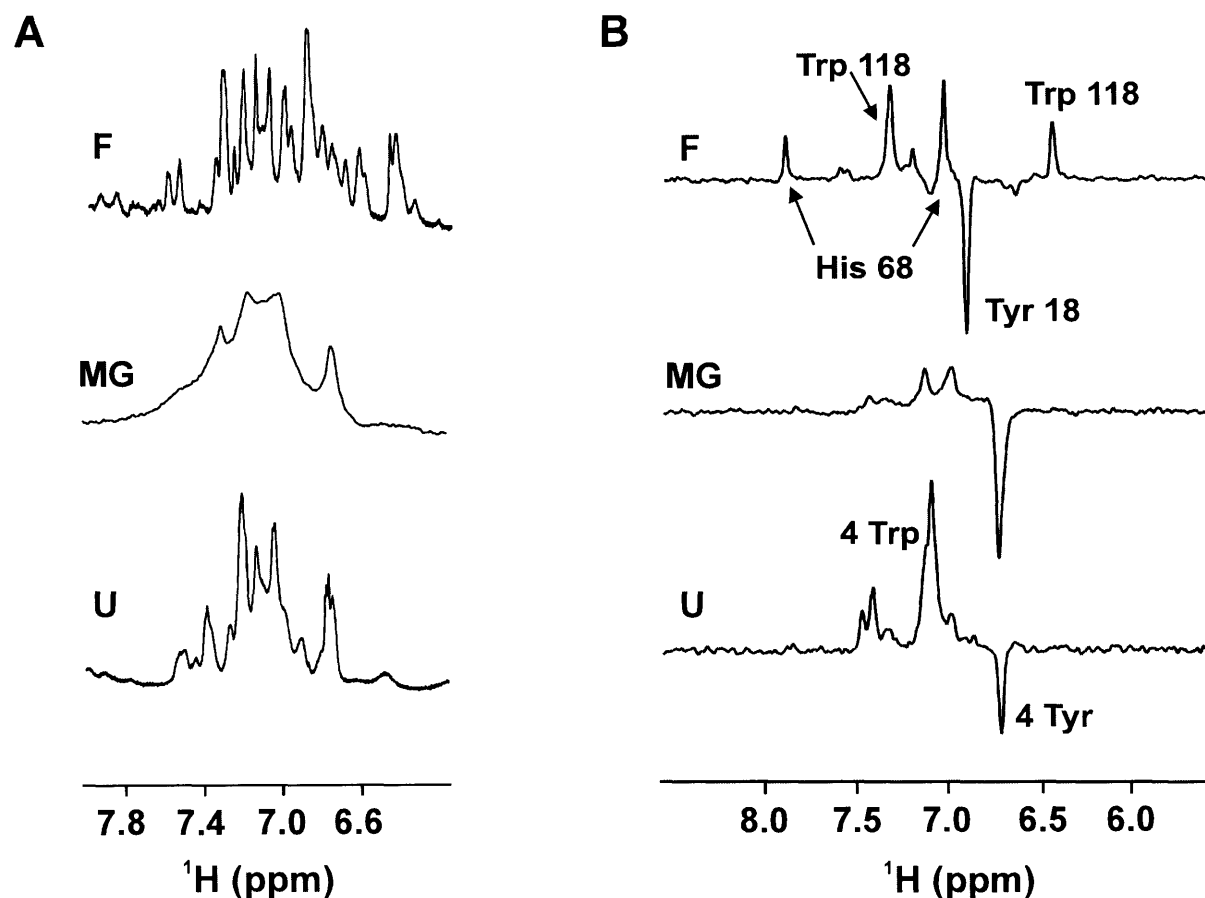
Figure 74: A) 1D proton NMR and A) 1D photo-CIDNP NMR spectra of the aromatic region of the three states of α-lactalbumin. (U, MG, F denotes unfolded, molten globule and folded, respectively). NS=128 for the NMR spectra shown in a), while NS=1 in the photo-CIDNP spectra in b). Sample conditions: 0.5mM BLA, 0.2mM FMN in D_2O F: 50mM cacodylat, 10mM $CaCl_2$, pH* 6.7 (uncorrected pH, measured in D_2O); U: 8M urea, pH* 2; MG: pH* 2.

Photo-CIDNP spectra of the three states of BLA are shown in Figure 74B. In agreement with previous measurements in the literature (Maeda et al., 2000), only Trp 118, Tyr 18 and His 68 contribute to the CIDNP spectrum in folded BLA. While Trp118 and His68 have positive signals, the signal of Tyr 18 is emissive (Figure 74). These three side-chains are the most accessible (Lee and Richards, 1971) within a given amino acid type (Trp, Tyr and His) as can be seen from Table 17. It becomes furthermore obvious, that photo-CIDNP is not the same for all amino acid side chains, but that Trp > Tyr >> His (Hore and Broadhurst, 1993; Kaptein, 1982). His 32 has a normalized accessibility of 0.458 and is not present in the spectrum. In contrast, Trp 118 with an accessibility of 0.134 can be seen.

Trp	SA		Tyr	SA		His	SA	
	abs	norm		Abs	norm		Abs	norm
Trp 26	1.854	0.008	Tyr 18	67.663	0.302	His 32	85.699	0.458
Trp 60	12.203	0.048	Tyr 36	15.783	0.069	His 68	147.549	0.818
Trp 104	12.166	0.05	Tyr 50	23.19	0.111	His 107	17.712	0.098
Trp 118	32.623	0.134	Tyr 103	51.319	0.228			

Table 17: Solvent accessibilities (SA) of Trp-, Tyr- and His-residues in folded BLA (pdb: 1HFZ, (Pike et al., 1996)) using the approach of Lee and Richards (Lee and Richards, 1971). Absolute (in Å^2) as well as normalized values (to the surface of the amino-acid) are shown.

The same can be seen in the CIDNP spectrum of unfolded BLA (Figure 74): Accessibilities of histidines, tyrosins and tryptophans in any unfolded protein are expected to be approximately the same, however only the more CIDNP active tryptophan and tyrosine residues give signals at their respective random coil chemical shifts. While the spectra of the folded and the unfolded BLA exhibit sharp lines, broader line widths are characteristic for the MG of BLA, reflecting the ensemble of slow inter-converting conformers. A distinctive fingerprint for each of the three states can therefore be observed using photo-CIDNP spectroscopy.

7.3.5 Time resolved photo-CIDNP

The pulse scheme used for the time-resolved photo-CIDNP studies is shown in Figure 75. Ca^{2+}-ions are released within the NMR tube in the spectrometer by irradiation at 350nm (^{UV}L). Subsequently all magnetization is destroyed by means of four trim pulses and a gradient before new magnetization of accessible aromatic side chains is generated by photo-CIDNP using the

VIS-laser (^{VIS}L). Detection of the NMR signal is done using a jump-return echo water suppression scheme. After a recycle delay (RD), photo-CIDNP is created again.

initiation   **detection**
photo-CIDNP   **photo-CIDNP**
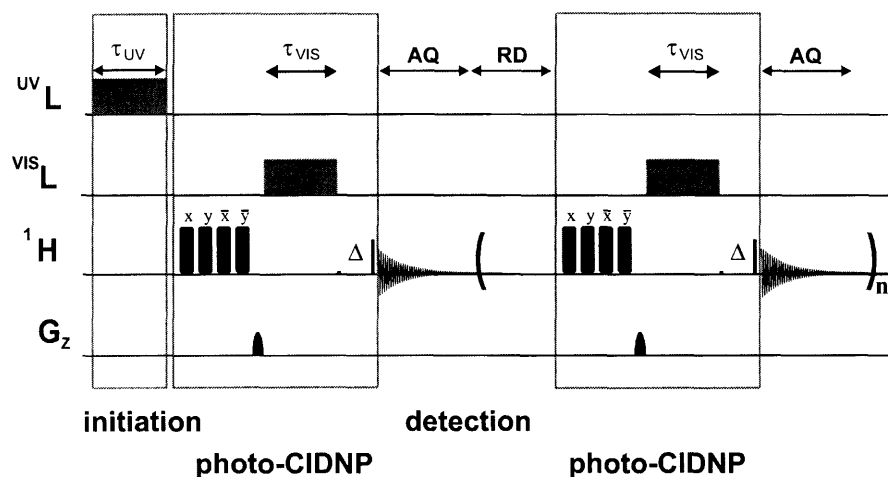
Figure 75: Pulse sequence used for time-resolved photo-CIDNP measurements. ^{UV}L and ^{VIS}L represent the UV- and the VIS-lasers, respectively. ^{1}H denotes pulses on protons, G_z denotes z gradients. Wide bars indicate trim pulses (500ms at high proton power), while thin bars indicate the use of a jump-return pulse scheme (τ_{JR}=140μs) for detection and water suppression. τ_{UV} =200ms, τ_{VIS}=100ms, Δ=5ms, AQ=56ms, RD=440ms or 1140ms, $\tau_{gradient}$=1ms (50G/cm), n=1-4. Static photo-CIDNP spectra were collected omitting the UV-laser, using τ'_{laser}=200ms, AQ=224ms and RD=2s, while the rest of the parameters remained the same.

The results of the time-resolved photo-CIDNP experiments of the ion induced refolding of BLA are shown in Figure 76. Before initiation of refolding (labeled with start), the protein is mainly unfolded. In the first kinetic experiment 200ms after folding initiation, native signals of Trp 118 (absorptive signals by δ=6.4 and 7.42ppm) and Tyr 18 (emissive signal by δ=6.95ppm) are observed. Those signals do not significantly change in intensity at later time points. Folding is finished at 1500ms after folding initiation, as the spectrum does not change afterwards. This is in good agreement with the data from stopped-flow fluorescence. A sharp emissive signal by δ=7.05ppm arising from one of the three additional Tyr residues in BLA is present after 200ms and disappears between 800ms and 1500ms. This signal can be assigned to the folding intermediate observed also in the kinetic analysis of the stopped-flow fluorescence experiments. The chemical shift of the signal does not correspond to any Tyr signal in either the native, the unfolded or the molten globule state of BLA (Figure 74B) and additional NMR data of the native

state (Alexandrescu et al., 1992)). It can therefore be assigned to an intermediate with non-native environment.
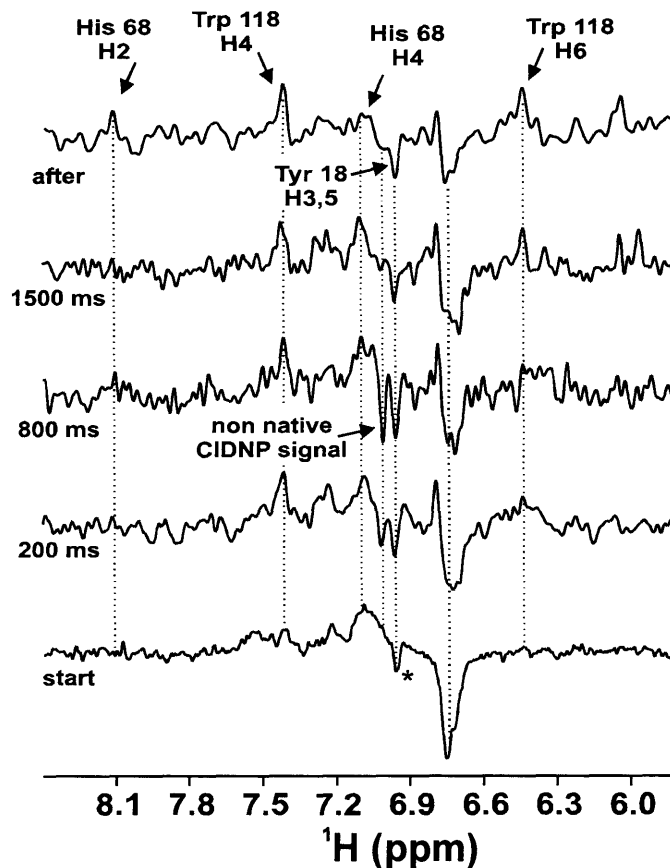
Figure 76: Time resolved photo-CIDNP. Spectra shown here, are averaged over eight samples. The first and the last trace are photo-CIDNP spectra prior and after the kinetic experiments, respectively; * denotes a native signal present at the start of the experiment.

7.3.6 Discussion

A folding intermediate populated between 200ms and 800ms after the initiation of the Ca^{2+} induced refolding pathway of BLA has been identified using stopped-flow fluorescence and time-resolved photo-CIDNP experiments. Time resolved photo-CIDNP experiments furthermore reveal that one tyrosine in a non-native environment is accessible in this folding intermediate, indicated by a sharp negative signal. It is important to note that the emissive Tyr signal seen in the intermediate does not arise from a Tyr residue in a native environment. This implies that at least part of the polypeptide chain samples non-native conformations during protein refolding. In

183

addition, this emissive signal observed for the folding intermediate is sharp and has therefore no characteristics of a molten globule.

The assignment of this emissive signal by δ=7.05ppm to one of the three remaining tyrosines, Tyr 36, 50, and 103 cannot be ascertained. However, the following conclusion can be drawn: It is unlikely that the signal arises from Tyr 36 or Tyr 103, because both Tyr are part of hydrophobic clusters in the folded state one is formed between Tyr 36 and Trp 118 and a second is formed between Tyr103, Trp 60 and Trp 104. Tyr 36 and Tyr 103 are both part of the α-domain in α-lactalbumin. In contrast, Tyr 50 is in the β-sheet of the β-domain of α-lactalbumin. The assignment of the non-native signal to Tyr 50 would imply that the hydrophobic core in the α-domain of BLA is locked in its native-like environment while parts of the β-domain, maybe those involved in β-sheet formation, would undergo slower structural reorganization.
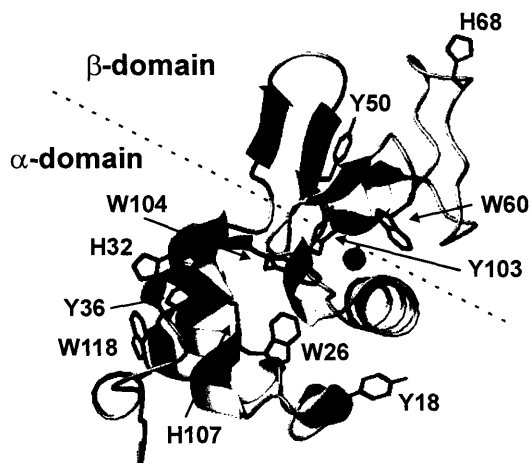
Figure 77: Ribbon display of the X-ray structure of BLA (pdb: 1HFZ, (Pike et al., 1996)). Tyrosine- (blue), tryptophan- (black) and histidine- (green) residues are indicated. The figure was prepared using the program MOLMOL (Konradi et al., 1996).

A comparison of the results from the ion induced refolding of BLA presented here with results from refolding of BLA by dilution from GdnHCl (Figure 67A) suggest that different pathways can be populated during refolding, leading to the folded state of a protein via different folding

intermediates. Here a distinct intermediate is populated, while the folding intermediate in the dilution induced pathway resembles the MG.

7.3.7 Outlook

Using time-resolved photo-CIDNP NMR a folding intermediate of BLA was detected on the Ca^{2+} induced refolding pathway at constant denaturant. This folding intermediate has no characteristics of a MG. Further characterization of this folding intermediate could be achieved by monitoring the changes of the backbone amides during the refolding process by time-resolved ^1H,^{15}N-HSQCs using heterologoulsy expressed BLA as shown in Chapter 5.

The laser setup established in this work has two advantages compared to systems that had been available before: (i) higher laser power in the UV regime allows either to trigger reactions in one shot, without taking difference measurements or to go to higher polymer concentrations. (ii) Combination with photo-CIDNP NMR simplifies spectra and increases the signal-to-noise ratio.

The availability of various photo-labile compounds such as photo-protected ATP, ADP and AMP (Geissler et al., 2003), photo-protected decylubiquinol (Hansen et al., 2000), caged H^+ (Barth and Corrie, 2002) and photo-protected nucleotides (Pitsch et al., 1999) combined with the two advantages mentioned above, allows to investigate more complicated systems not accessible before:

RNA folding is difficult to monitor due to intrinsically lower (compared to proteins) signal-to-noise ratio, due to relaxation. Using the new setup, RNA folding becomes accessible to time-resolved NMR, since higher RNA concentrations as before and photo-CIDNP can be used to increase resolution and signal-to-noise to monitor conformational changes of purine bases. Conformational changes of larger proteins or even membrane proteins could be monitored without the need of labeling and assigning the protein, using photo-CIDNP.

7.4 Conclusion

A new method for the investigation of laser triggered kinetics using photo-CIDNP NMR was developed. Two lasers were coupled into an NMR spectrometer: one for initiation of folding by releasing ions from photo-labile chelators with dead times as low as 200ms, and one for induction of photo-CINDP NMR. The Ca^{2+}-ion induced folding of BLA was investigated using the method: The folding was initiated by ion release from the beforehand synthesized photolabile chelator DM-Nitrophen within 200ms by laser irradiation at 350nm. Photo-CIDNP was initiated by laser

irradiation at 488/515nm using FMN. A comparison of static photo-CIDNP spectra of the three states of BLA with spectra recorded during refolding revealed the presence of a non-native tyrosine in a folding intermediate populated 200ms after initiation of folding. This non-native signal can be assigned to Tyr 50, implying that the hydrophobic core in the α-domain of BLA is locked in its native-like environment while parts of the β-domain, maybe those involved in β-sheet formation, undergo slower structural reorganization.

Literature

Abraham, D.J. and Leo, A.J. (1987) Extension of the fragment method to calculate amino acid zwitterion and side chain partition coefficients. *Proteins*, **2**, 130-152.

Akke, M. and Palmer, A.G. (1996) Monitoring macromolecular motions on microsecond to millisecond time scales by R(1)rho-R(1) constant relaxation time NMR spectroscopy. *J Am Chem Soc*, **118**, 911-912.

Alexandrescu, A.T., Broadhurst, R.W., Wormald, C., Chyan, C.L., Baum, J. and Dobson, C.M. (1992) 1H-NMR assignments and local environments of aromatic residues in bovine, human and guinea pig variants of alpha-lactalbumin. *Eur J Biochem*, **210**, 699-709.

Allerhand, A. and Hailstone, R.K. (1972) C-13 Fourier-Transform Nuclear Magnetic-Resonance. X. Effect of Molecular-Weight on C-13 Spin-Lattice Relaxation-Times of Polystyrene in Solution. *J Chem Phys*, **56**, 3718-3720.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, **25**, 3389-3402.

Anfinsen, C.B., Haber, E., Sela, M. and White, F.H., Jr. (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci U S A*, **47**, 1309-1314.

Appel, R.D., Bairoch, A. and Hochstrasser, D.F. (1994) A new generation of information retrieval tools for biologists: the example of the ExPASy WWW server. *Trends Biochem Sci*, **19**, 258-260.

Arcus, V.L., Vuilleumier, S., Freund, S.M., Bycroft, M. and Fersht, A.R. (1995) A comparison of the pH, urea, and temperature-denatured states of barnase by heteronuclear NMR: implications for the initiation of protein folding. *J Mol Biol*, **254**, 305-321.

Bachmann, A. and Kiefhaber, T. (2001) Apparent two-state tendamistat folding is a sequential process along a defined route. *J Mol Biol*, **306**, 375-386.

Balbach, J., Forge, V., Lau, W.S., van Nuland, N.A., Brew, K. and Dobson, C.M. (1996) Protein folding monitored at individual residues during a two-dimensional NMR experiment. *Science*, **274**, 1161-1163.

Balbach, J., Forge, V., van Nuland, N.A., Winder, S.L., Hore, P.J. and Dobson, C.M. (1995) Following protein folding in real time using NMR spectroscopy. *Nat Struct Biol*, **2**, 865-870.

Baldwin, R. (1993) Pulsed H/D-exchange studies of folding intermediates. *Curr Opin Struct Biol*, **3**, 84-91.

Baldwin, R.L. and Rose, G.D. (1999) Is protein folding hierarchic? II. Folding intermediates and transition states. *Trends Biochem Sci*, **24**, 77-83.

Banci, L., Bertini, I., Savellini, G.G., Romagnoli, A., Turano, P., Cremonini, M.A., Luchinat, C. and Gray, H.B. (1997) Pseudocontact shifts as constraints for energy minimization and molecular dynamics calculations on solution structures of paramagnetic metalloproteins. *Proteins*, **29**, 68-76.

Barbieri, R., Bertini, I., Lee, Y.M., Luchinat, C. and Velders, A.H. (2002) Structure-independent cross-validation between residual dipolar couplings originating from internal and external orienting media. *J Biomol NMR*, **22**, 365-368.

Bargon, J., Fischer, H. and Johnsen, U. (1967) Kernresonanz-Emissionslinien während rascher Radikalreaktionen. I. Aufnahmeverfahren und Beispiele. *Zeitschrift für Naturforschung Part A-Astrophysik, Physik und Physikalische Chemie*, **A 22**, 1551-1555.

187

Bartels, C., Xia, T.H., Billeter, M., Guntert, P. and Wüthrich, K. (1995) The Program Xeasy for Computer-Supported NMR Spectral-Analysis of Biological Macromolecules. *J Biomol NMR*, **6**, 1-10.

Barth, A. and Corrie, J.E. (2002) Characterization of a new caged proton capable of inducing large pH jumps. *Biophys J*, **83**, 2864-2871.

Bentrop, D., Bertini, I., Iacoviello, R., Luchinat, C., Niikura, Y., Piccioli, M., Presenti, C. and Rosato, A. (1999) Structural and dynamical properties of a partially unfolded Fe4S4 protein: role of the cofactor in protein folding. *Biochemistry*, **38**, 4669-4680.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Res*, **28**, 235-242.

Bevington, P.R. (1969) In *Data reduction and error analysis for the physical sciences*. McGraw-Hill, New York, pp. 195-203.

Bhavesh, N.S., Chatterjee, A., Panchal, S.C. and Hosur, R.V. (2003) Application of HN(C)N to rapid estimation of 1J(N-C(alpha)) coupling constants correlated to psi torsion angles in proteins: implication to structural genomics. *Biochem Biophys Res Commun*, **311**, 678-684.

Billeter, M., Neri, D., Otting, G., Qian, Y.Q. and Wüthrich, K. (1992) Precise vicinal coupling constants 3JHN alpha in proteins from nonlinear fits of J-modulated [15N,1H]-COSY experiments. *J Biomol NMR*, **2**, 257-274.

Blake, C.C., Fenn, R.H., North, A.C., Phillips, D.C. and Poljak, R.J. (1962) Structure of lysozyme. A Fourier map of the electron density at 6 angstrom resolution obtained by x-ray diffraction. *Nature*, **196**, 1173-1176.

Blake, C.C., Koenig, D.F., Mair, G.A., North, A.C., Phillips, D.C. and Sarma, V.R. (1965) Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Angstrom resolution. *Nature*, **206**, 757-761.

Blanco, F.J., Serrano, L. and Forman-Kay, J.D. (1998) High populations of non-native structures in the denatured state are compatible with the formation of the native folded state. *J Mol Biol*, **284**, 1153-1164.

Blum, A.D., Smallcombe, S.H. and Baldwin, R.L. (1978) Nuclear magnetic resonance evidence for a structural intermediate at an early stage in the refolding of ribonuclease A. *J Mol Biol*, **118**, 305-316.

Bozzi, M., Bianchi, M., Sciandra, F., Paci, M., Giardina, B., Brancaccio, A. and Cicero, D.O. (2003) Structural characterization by NMR of the natively unfolded extracellular domain of beta-dystroglycan: towards the identification of the binding epitope for alpha-dystroglycan. *Biochemistry*, **42**, 13717-13724.

Broadhurst, R.W., Dobson, C.M., Hore, P.J., Radford, S.E. and Rees, M.L. (1991) A photochemically induced dynamic nuclear polarization study of denatured states of lysozyme. *Biochemistry*, **30**, 405-412.

Bronstein, I.N. and Semendjajew, K.A. (1989) In *Taschenbuch der Mathematik*. Verlag Harri Deutsch, Thun & Frankfurt, pp. 24-27.

Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comp Chem*, **4**, 187-217.

Brutscher, B., Bruschweiler, R. and Ernst, R.R. (1997) Backbone dynamics and structural characterization of the partially folded A state of ubiquitin by 1H, 13C, and 15N nuclear magnetic resonance spectroscopy. *Biochemistry*, **36**, 13043-13053.

Bryngelson, J.D., Onuchic, J.N., Socci, N.D. and Wolynes, P.G. (1995) Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins*, **21**, 167-195.

Buck, M., Radford, S.E. and Dobson, C.M. (1993) A partially folded state of hen egg white lysozyme in trifluoroethanol: structural characterization and implications for protein folding. *Biochemistry*, **32**, 669-678.

Buck, M., Radford, S.E. and Dobson, C.M. (1994) Amide hydrogen exchange in a highly denatured state. Hen egg-white lysozyme in urea. *J Mol Biol*, **237**, 247-254.

Buck, M., Schwalbe, H. and Dobson, C.M. (1995) Characterization of conformational preferences in a partly folded protein by heteronuclear NMR spectroscopy: assignment and secondary structure analysis of hen egg-white lysozyme in trifluoroethanol. *Biochemistry*, **34**, 13219-13232.

Buck, M., Schwalbe, H. and Dobson, C.M. (1996) Main-chain dynamics of a partially folded protein: 15N NMR relaxation measurements of hen egg white lysozyme denatured in trifluoroethanol. *J Mol Biol*, **257**, 669-683.

Buevich, A.V., Shinde, U.P., Inouye, M. and Baum, J. (2001) Backbone dynamics of the natively unfolded pro-peptide of subtilisin by heteronuclear NMR relaxation studies. *J Biomol NMR*, **20**, 233-249.

Bundi, A. and Wüthrich, K. (1979) 1H-NMR parameters of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH. *Biopolymers*, **18**, 285-297.

Bussell, R., Jr. and Eliezer, D. (2001) Residual structure and dynamics in Parkinson's disease-associated mutants of alpha-synuclein. *J Biol Chem*, **276**, 45996-46003.

Bystrov, V.P. (1976) Spin-spin coupling and the conformational states of peptide systems. *Prog NMR Spectrosc*, **10**, 44-81.

Cao, A., Hu, D. and Lai, L. (2004) Formation of amyloid fibrils from fully reduced hen egg white lysozyme. *Protein Sci*, **13**, 319-324.

Chamberlain, A.K., Receveur, V., Spencer, A., Redfield, C. and Dobson, C.M. (2001) Characterization of the structure and dynamics of amyloidogenic variants of human lysozyme by NMR spectroscopy. *Protein Sci*, **10**, 2525-2530.

Chandra, N., Brew, K. and Acharya, K.R. (1998) Structural evidence for the presence of a secondary calcium binding site in human alpha-lactalbumin. *Biochemistry*, **37**, 4767-4772.

Chang, J.Y. (2004) Evidence for the underlying cause of diversity of the disulfide folding pathway. *Biochemistry*, **43**, 4522-4529.

Chang, J.Y. and Li, L. (2002) Pathway of oxidative folding of alpha-lactalbumin: a model for illustrating the diversity of disulfide folding pathways. *Biochemistry*, **41**, 8405-8413.

Chaudhuri, T.K., Arai, M., Terada, T.P., Ikura, T. and Kuwajima, K. (2000) Equilibrium and kinetic studies on folding of the authentic and recombinant forms of human alpha-lactalbumin by circular dichroism spectroscopy. *Biochemistry*, **39**, 15643-15651.

Chiti, F., Stefani, M., Taddei, N., Ramponi, G. and Dobson, C.M. (2003) Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature*, **424**, 805-808.

Chiti, F., Taddei, N., Baroni, F., Capanni, C., Stefani, M., Ramponi, G. and Dobson, C.M. (2002) Kinetic partitioning of protein folding and aggregation. *Nat Struct Biol*, **9**, 137-143.

Chothia, C. (1992) Proteins. One thousand families for the molecular biologist. *Nature*, **357**, 543-544.

Choy, W.Y., Shortle, D. and Kay, L.E. (2003) Side chain dynamics in unfolded protein states: an NMR based 2H spin relaxation study of delta131delta. *J Am Chem Soc*, **125**, 1748-1758.

Clubb, R.T., Thanbal, V. and Wagner, G. (1992) A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue 1HN, 15N, and 13C' chemical shifts in 15N,13C-labelled proteins. *J Magn Reson*, **97**, 213-217.

Collins, E.S., Hirai, K., Tachibana, H., S., S. and Schwalbe, H. (submitted) Characterization of the Slow Disulfide Bond Dynamics in Non-native States of Wild Type Lysozyme and its Disulfide Deletion Mutants by NMR.

189

Collins, E.S., Whittaker, S.B., Tozawa, K., MacDonald, C., Boetzel, R., Penfold, C.N., Reilly, A., Clayden, N.J., Osborne, M.J., Hemmings, A.M., Kleanthous, C., James, R. and Moore, G.R. (2002) Structural dynamics of the membrane translocation domain of colicin E9 and its interaction with TolB. *J Mol Biol*, **318**, 787-804.

Cornilescu, G., Delaglio, F. and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR*, **13**, 289-302.

Cornilescu, G., Marquardt, J.L., Ottiger, M. and Bax, A. (1998) Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J Am Chem Soc*, **120**, 6836-6837.

Crowhurst, K.A. and Forman-Kay, J.D. (2003) Aromatic and methyl NOEs highlight hydrophobic clustering in the unfolded state of an SH3 domain. *Biochemistry*, **42**, 8687-8695.

Daughdrill, G.W., Hanely, L.J. and Dahlquist, F.W. (1998) The C-terminal half of the anti-sigma factor FlgM contains a dynamic equilibrium solution structure favoring helical conformations. *Biochemistry*, **37**, 1076-1082.

de la Torre, J.G., Huertas, M.L. and Carrasco, B. (2000) Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys J*, **78**, 719-730.

Delaglio, F., Torchia, D.A. and Bax, A. (1991) Measurement of 15N-13C J couplings in staphylococcal nuclease. *J Biomol NMR*, **1**, 439-446.

Derix, N.M., Wechselberger, R.W., van der Horst, M.A., Hellingwerf, K.J., Boelens, R., Kaptein, R. and van Nuland, N.A. (2003) Lack of negative charge in the E46Q mutant of photoactive yellow protein prevents partial unfolding of the blue-shifted intermediate. *Biochemistry*, **42**, 14501-14506.

Ding, K. and Gronenborn, A.M. (2004) Protein Backbone 1H(N)-13Calpha and 15N-13Calpha residual dipolar and J couplings: new constraints for NMR structure determination. *J Am Chem Soc*, **126**, 6232-6233.

Dobson, C.M. (2003) Protein folding and misfolding. *Nature*, **426**, 884-890.

Dobson, C.M. (2004) Principles of protein folding, misfolding and aggregation. *Semin Cell Dev Biol*, **15**, 3-16.

Dobson, C.M., Evans, P.A. and Radford, S.E. (1994) Understanding how proteins fold: the lysozyme story so far. *Trends Biochem Sci*, **19**, 31-37.

Dolgikh, D.A., Gilmanshin, R.I., Brazhnikov, E.V., Bychkova, V.E., Semisotnov, G.V., Venyaminov, S. and Ptitsyn, O.B. (1981) Alpha-Lactalbumin: compact state with fluctuating tertiary structure? *FEBS Lett*, **136**, 311-315.

Edison, A.S., Markley, J.L. and Weinhold, F. (1994a) Calculations of one-, two- and three-bond nuclear spin-spin couplings in a model peptide and correlations with experimental data. *J Biomol NMR*, **4**, 519-542.

Edison, A.S., Weinhold, F., Westler, W.M. and Markley, J.L. (1994b) Estimates of phi and psi torsion angles in proteins from one-, two- and three-bond nuclear spin-spin couplings: application to staphylococcal nuclease. *J Biomol NMR*, **4**, 543-551.

Ellis-Davies, G.C.R. and Kaplan, J.H. (1988) A new class of photolabile chelators for the rapid release of divalent cations: generation of caged calcium and caged magnesium. *J Org Chem*, **53**, 1966-1969.

Emsley, L. and Bodenhausen, G. (1990) Gaussian Pulse Cascades - New Analytical Functions for Rectangular Selective Inversion and in-Phase Excitation in NMR. *Chem Phys Lett*, **165**, 469-476.

Emsley, L. and Bodenhausen, G. (1992) Optimization of shaped selective pulses for NMR using a quaternion description of their overall propagators. *J Magn Reson*, **97**, 135-148.

Englander, S.W. and Mayne, L. (1992) Protein folding studied using hydrogen-exchange labeling and two-dimensional NMR. *Annu Rev Biophys Biomol Struct*, **21**, 243-265.

Evans, P.A., Topping, K.D., Woolfson, D.N. and Dobson, C.M. (1991) Hydrophobic clustering in nonnative states of a protein: Interpretation of chemical shifts in NMR spectra of denatured states of lysozyme. *Proteins: Struct Funct Genet*, **9**, 248-266.

Fieber, W., Kristjansdottir, S. and Poulsen, F.M. (2004) Short-range, long-range and transition state interactions in the denatured state of ACBP from residual dipolar couplings. *J Mol Biol*, **339**, 1191-1199.

Fiebig, K.M., Rice, L.M., Pollock, E. and Brunger, A.T. (1999) Folding intermediates of SNARE complex assembly. *Nat Struct Biol*, **6**, 117-123.

Fiebig, K.M., Schwalbe, H., Buck, M., Smith, L.J. and Dobson, C.M. (1996) Towards a description of the conformations of denatured states of proteins. Comparison of a random coil model with NMR measurements. *J Chem Phys*, **100**, 2661-2666.

Flory, P.J. (1953) *Principles of polymer chemistry*. Cornell University Press, Ithaca, N. Y.

Flory, P.J. (1969) *Statistical mechanics of chain molecules*. Wiley, New York.

Forge, V., Wijesinha, R.T., Balbach, J., Brew, K., Robinson, C.V., Redfield, C. and Dobson, C.M. (1999) Rapid collapse and slow structural reorganisation during the refolding of bovine alpha-lactalbumin. *J Mol Biol*, **288**, 673-688.

Frank, M.K., Clore, G.M. and Gronenborn, A.M. (1995) Structural and dynamic characterization of the urea denatured state of the immunoglobulin binding domain of streptococcal protein G by multidimensional heteronuclear NMR spectroscopy. *Protein Sci*, **4**, 2605-2615.

Franz, K.J., Nitz, M. and Imperiali, B. (2003) Lanthanide-binding tags as versatile protein coexpression probes. *Chembiochem*, **4**, 265-271.

Frieden, C., Hoeltzli, S.D. and Ropson, I.J. (1993) NMR and protein folding: equilibrium and stopped-flow studies. *Protein Sci*, **2**, 2007-2014.

Friel, C.T., Beddard, G.S. and Radford, S.E. (2004) Switching two-state to three-state kinetics in the helical protein Im9 via the optimisation of stabilising non-native interactions by design. *J Mol Biol*, **342**, 261-273.

Gaemers, S. and Bax, A. (2001) Morphology of three lyotropic liquid crystalline biological NMR media studied by translational diffusion anisotropy. *J Am Chem Soc*, **123**, 12343-12352.

Garcia, P., Serrano, L., Durand, D., Rico, M. and Bruix, M. (2001) NMR and SAXS characterization of the denatured state of the chemotactic protein CheY: implications for protein folding initiation. *Protein Sci*, **10**, 1100-1112.

Gast, K., Damaschun, H., Eckert, K., Schulze-Forster, K., Maurer, H.R., Muller-Frohne, M., Zirwer, D., Czarnecki, J. and Damaschun, G. (1995) Prothymosin alpha: a biologically active protein with random coil conformation. *Biochemistry*, **34**, 13211-13218.

Geissler, D., Kresse, W., Wiesner, B., Bendig, J., Kettenmann, H. and Hagen, V. (2003) DMACM-caged adenosine nucleotides: ultrafast phototriggers for ATP, ADP, and AMP activated by long-wavelength irradiation. *Chembiochem*, **4**, 162-170.

Geyer, M., Munte, C.E., Schorr, J., Kellner, R. and Kalbitzer, H.R. (1999) Structure of the anchor-domain of myristoylated and non-myristoylated HIV-1 Nef protein. *J Mol Biol*, **289**, 123-138.

Gillespie, J.R. and Shortle, D. (1997a) Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels. *J Mol Biol*, **268**, 158-169.

Gillespie, J.R. and Shortle, D. (1997b) Characterization of long-range structure in the denatured state of staphylococcal nuclease. II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures. *J Mol Biol*, **268**, 170-184.

Glazkov, U.V. and Rudakov, I.A. (1994) Chemically induced dynamic nuclear polarization in the dye-quinone-solvent system. *Vestn Beloruss Gos Univ*, **2**, 13-16.

Goda, S., Takano, K., Yamagata, Y., Nagata, R., Akutsu, H., Maki, S., Namba, K. and Yutani, K. (2000) Amyloid protofilament formation of hen egg lysozyme in highly concentrated ethanol solution. *Protein Sci*, **9**, 369-375.

Green, T.B., Ganesh, O., Perry, K., Smith, L., Phylip, L.H., Logan, T.M., Hagen, S.J., Dunn, B.M. and Edison, A.S. (2004) IA3, an aspartic proteinase inhibitor from Saccharomyces cerevisiae, is intrinsically unstructured in solution. *Biochemistry*, **43**, 4071-4081.

Grimshaw, S.B. (1999) PhD thesis. University of Oxford, UK.

Grobler, J.A., Wang, M., Pike, A.C. and Brew, K. (1994) Study by mutagenesis of the roles of two aromatic clusters of alpha-lactalbumin in aspects of its action in the lactose synthase system. *J Biol Chem*, **269**, 5106-5114.

Grzesiek, S. and Bax, A. (1992) Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *J Magn Reson*, **96**, 432-440.

Hansen, K.C., Schultz, B.E., Wang, G. and Chan, S.I. (2000) Reaction of Escherichia coli cytochrome bo(3) and mitochondrial cytochrome bc(1) with a photoreleasable decylubiquinol. *Biochim Biophys Acta*, **1456**, 121-137.

Harata, K., Abe, Y. and Muraki, M. (1999) Crystallographic evaluation of internal motion of human alpha-lactalbumin refined by full-matrix least-squares method. *J Mol Biol*, **287**, 347-358.

Hartl, F.U. and Hayer-Hartl, M. (2002) Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science*, **295**, 1852-1858.

Heikkinen, S., Permi, P. and Kipeläinen, I. (2001) Methods for the measurement of 1JNC alpha and 2JNC alpha from a simplified 2D 13C alpha-coupled 15N SE-HSQC spectrum. *J Magn Reson*, **148**, 53-60.

Heinrikson, R.L. (1971) The selective S-methylation of sulfhydryl groups in proteins and peptides with methyl-p-nitrobenzenesulfonate. *J Biol Chem*, **246**, 4090-4096.

Hennig, M., Bermel, W., Spencer, A., Dobson, C.M., Smith, L.J. and Schwalbe, H. (1999) Side-chain conformations in an unfolded protein: chi1 distributions in denatured hen lysozyme determined by heteronuclear 13C, 15N NMR spectroscopy. *J Mol Biol*, **288**, 705-723.

Hill, R.L. and Brew, K. (1975) Lactose synthetase. *Adv Enzymol Relat Areas Mol Biol*, **43**, 411-490.

Hoeltzli, S.D. and Frieden, C. (1995) Stopped-flow NMR spectroscopy: real-time unfolding studies of 6-19F-tryptophan-labeled Escherichia coli dihydrofolate reductase. *Proc Natl Acad Sci U S A*, **92**, 9318-9322.

Holm, L. and Sander, C. (1998) Touring protein fold space with Dali/FSSP. *Nucleic Acids Res*, **26**, 316-319.

Hore, P.J. and Broadhurst, R.W. (1993) Photo-CIDNP of biopolymers. *Prog NMR Spectrosc*, **25**, 345-402.

Hore, P.J., Winder, S.L., Roberts, C.H. and Dobson, C.M. (1997) Stopped-flow photo-CIDNP observation of protein folding. *J Am Chem Soc*, **119**, 5049-5050.

Hu, J.S. and Bax, A. (1996) Measurement of three-bond C-13-C-13 J couplings between carbonyl and carbonyl/carboxyl carbons in isotopically enriched proteins. *J Am Chem Soc*, **118**, 8170-8171.

Hu, J.S. and Bax, A. (1997) Determination of Phi and Chi1 Angles in Proteins from 13C-13C Three-Bond J Couplings Measured by Three-Dimensional Heteronuclear NMR. How Planar Is the Peptide Bond? *J Am Chem Soc*, **119**, 6360-6368.

Hu, J.S. and Bax, A. (1998) Measurement of three-bond, C-13 '-C-13(beta) J couplings in human ubiquitin by a triple resonance, E. COSY-type NMR technique. *J Biomol NMR*, **11**, 199-203.

Hubbell, W.L., McHaourab, H.S., Altenbach, C. and Lietzow, M.A. (1996) Watching proteins move using site-directed spin labeling. *Structure*, **4**, 779-783.

Hurley, W.L. and Schuler, L.A. (1987) Molecular cloning and nucleotide sequence of a bovine alpha-lactalbumin cDNA. *Gene*, **61**, 119-122.

Ishikawa, N., Chiba, T., Chen, L.T., Shimizu, A., Ikeguchi, M. and Sugai, S. (1998) Remarkable destabilization of recombinant alpha-lactalbumin by an extraneous N-terminal methionyl residue. *Protein Eng*, **11**, 333-335.

Jaroniec, C.P., MacPhee, C.E., Astrof, N.S., Dobson, C.M. and Griffin, R.G. (2002) Molecular conformation of a peptide fragment of transthyretin in an amyloid fibril. *Proc Natl Acad Sci U S A*, **99**, 16748-16753.

Jaroniec, C.P., MacPhee, C.E., Bajaj, V.S., McMahon, M.T., Dobson, C.M. and Griffin, R.G. (2004) High-resolution molecular structure of a peptide in an amyloid fibril determined by magic angle spinning NMR spectroscopy. *Proc Natl Acad Sci U S A*, **101**, 711-716.

Jenson, J., Goldstein, G. and Breslow, E. (1980) Physical-chemical properties of ubiquitin. *Biochim Biophys Acta*, **624**, 378-385.

Jimenez, J.L., Guijarro, J.I., Orlova, E., Zurdo, J., Dobson, C.M., Sunde, M. and Saibil, H.R. (1999) Cryo-electron microscopy structure of an SH3 amyloid fibril and model of the molecular packing. *Embo J*, **18**, 815-821.

Jones, J.A., Wilkins, D.K., Smith, L.J. and Dobson, C.M. (1997) Characterisation of protein unfolding by NMR diffusion measurements. *J Biomol NMR*, **10**, 199-203.

Juranic, N. and Macura, S. (2001) Correlations among 1JNC' and h3JNC' coupling constants in the hydrogen-bonding network of human ubiquitin. *J Am Chem Soc*, **123**, 4099-4100.

Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577-2637.

Kaplan, J.H. and Ellis-Davies, G.C. (1988) Photolabile chelators for the rapid photorelease of divalent cations. *Proc Natl Acad Sci U S A*, **85**, 6571-6575.

Kaptein, R. (1971) Simple Rules for Chemically Induced Dynamic Nuclear Polarization. *J Chem Soc Chem Commun*, **14**, 732-733.

Kaptein, R. (1982) Photo-CIDNP studies of proteins. *Biol Magn Reson*, **4**, 145-191.

Karplus, M. (1963) Vicinal Proton Coupling in Nuclear Magnetic Resonance. *J Am Chem Soc*, **85**, 2870-2871.

Katou, H., Kanno, T., Hoshino, M., Hagihara, Y., Tanaka, H., Kawai, T., Hasegawa, K., Naiki, H. and Goto, Y. (2002) The role of disulfide bond in the amyloidogenic state of beta(2)-microglobulin studied by heteronuclear NMR. *Protein Sci*, **11**, 2218-2229.

Kautz, R.A. and Fox, R.O. (1993) NMR analysis of staphylococcal nuclease thermal quench refolding kinetics. *Protein Sci*, **2**, 851-858.

Kay, L.E., Ikura, M., Tschudin, R. and Bax, A. (1990) Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J Magn Reson*, **89**, 496-514.

Kay, L.E., Torchia, D.A. and Bax, A. (1989) Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry*, **28**, 8972-8979.

Keller, R. (in preparation) Optimizing the process of nuclear magnetic resonance spectrum analysis and computer aided resonance assignment.

Kiefhaber, T. (1995) Kinetic traps in lysozyme folding. *Proc Natl Acad Sci U S A*, **92**, 9029-9033.

Kim, P.S. and Baldwin, R.L. (1982) Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu Rev Biochem*, **51**, 459-489.

Kim, P.S. and Baldwin, R.L. (1990) Intermediates in the folding reactions of small proteins. *Annu Rev Biochem*, **59**, 631-660.

Klein-Seetharaman, J., Oikawa, M., Grimshaw, S.B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L.J., Dobson, C.M. and Schwalbe, H. (2002) Long-range interactions within a nonnative protein. *Science*, **295**, 1719-1722.

Koide, S., Dyson, H.J. and Wright, P.E. (1993) Characterization of a folding intermediate of apoplastocyanin trapped by proline isomerization. *Biochemistry*, **32**, 12299-12310.

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

Kosen, P.A., Scheek, R.M., Naderi, H., Basus, V.J., Manogaran, S., Schmidt, P.G., Oppenheimer, N.J. and Kuntz, I.D. (1986) Two-dimensional 1H NMR of three spin-labeled derivatives of bovine pancreatic trypsin inhibitor. *Biochemistry*, **25**, 2356-2364.

Krebs, M.R., Wilkins, D.K., Chung, E.W., Pitkeathly, M.C., Chamberlain, A.K., Zurdo, J., Robinson, C.V. and Dobson, C.M. (2000) Formation and seeding of amyloid fibrils from wild-type hen lysozyme and a peptide fragment from the beta-domain. *J Mol Biol*, **300**, 541-549.

Kuboniwa, H., Grzesiek, S., Delaglio, F. and Bax, A. (1994) Measurement of HN-H alpha J couplings in calcium-free calmodulin using new 2D and 3D water-flip-back methods. *J Biomol NMR*, **4**, 871-878.

Kühn, T. and Schwalbe, H. (2000) Monitoring the kinetics of ion-dependent protein folding by time-resolved NMR spectroscopy at atomic resolution. *J Am Chem Soc*, **122**, 6169-6174.

Kuwajima, K. (1996) The molten globule state of alpha-lactalbumin. *Faseb J*, **10**, 102-109.

Kuwata, K., Li, H., Yamada, H., Batt, C.A., Goto, Y. and Akasaka, K. (2001) High pressure NMR reveals a variety of fluctuating conformers in beta-lactoglobulin. *J Mol Biol*, **305**, 1073-1083.

Kyte, J. and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol*, **157**, 105-132.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680-685.

Lassalle, M.W., Li, H., Yamada, H., Akasaka, K. and Redfield, C. (2003) Pressure-induced unfolding of the molten globule of all-Ala alpha-lactalbumin. *Protein Sci*, **12**, 66-72.

Lazaridis, T. and Karplus, M. (1999) Effective energy function for proteins in solution. *Proteins*, **35**, 133-152.

Lee, B. and Richards, F.M. (1971) The interpretation of protein structures: estimation of static accessibility. *J Mol Biol*, **55**, 379-400.

Lee, C.L. and Atassi, M.Z. (1973) Conformation and immunochemistry of methylated and carboxymethyulated derivatives of lysozyme. *Biochemistry*, **12**, 2690-2695.

Lee, J., Kim, S.Y., Joo, K. and Kim, I. (2004a) Prediction of protein tertiary structure using PROFESY, a novel method based on fragment assembly and conformational space annealing. *Proteins*, **56**, 704-714.

Lee, S., Cho, M.K., Jung, J.W., Kim, J.H. and Lee, W. (2004b) Exploring protein fold space by secondary structure prediction using data distribution method on grid platform. *Bioinformatics*, advanced online publication.

Lenkinski, R.E., Chen, D.M., Glickson, J.D. and Goldstein, G. (1977) Nuclear magnetic resonance studies of the denaturation of ubiquitin. *Biochim Biophys Acta*, **494**, 126-130.

Levinthal, C. (1969) Mossbauer spectroscopy in biological systems: Proocedings of a meeting held at Allerton House. In DeBrunner, J.T.P. and Munck, E. (eds.). University of Illinois Press, Monticello, Illinois, pp. 22-24.

Levitt, M. (1976) A simplified representation of protein conformations for rapid simulation of protein folding. *J Mol Biol*, **104**, 59-107.

Li, L. and Chang, J.Y. (2004) Two-state folding of lysozyme versus multiple-state folding of alpha-lactalbumin illustrated by the technique of disulfide scrambling. *Protein J*, **23**, 3-10.

Lietzow, M.A., Jamin, M., Jane Dyson, H.J. and Wright, P.E. (2002) Mapping long-range contacts in a highly unfolded protein. *J Mol Biol*, **322**, 655-662.

Linding, R., Jensen, L.J., Diella, F., Bork, P., Gibson, T.J. and Russell, R.B. (2003) Protein disorder prediction: implications for structural proteomics. *Structure (Camb)*, **11**, 1453-1459.

Lindorff-Larsen, K., Kristjansdottir, S., Teilum, K., Fieber, W., Dobson, C.M., Poulsen, F.M. and Vendruscolo, M. (2004) Determination of an ensemble of structures representing the denatured state of the bovine acyl-coenzyme A binding protein. *J Am Chem Soc*, **126**, 3291-3299.

Lipari, G. and Szabo, A. (1982a) Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .1. Theory and Range of Validity. *J Am Chem Soc*, **104**, 4546-4559.

Lipari, G. and Szabo, A. (1982b) Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .2. Analysis of Experimental Results. *J Am Chem Soc*, **104**, 4559-4570.

Lisse, T., Bartels, D., Kalbitzer, H.R. and Jaenicke, R. (1996) The recombinant dehydrin-like desiccation stress protein from the resurrection plant Craterostigma plantagineum displays no defined three-dimensional structure in its native state. *Biol Chem*, **377**, 555-561.

Liu, A., Riek, R., Wider, G., von Schroetter, C., Zahn, R. and Wüthrich, K. (2000) NMR experiments for resonance assignments of 13C, 15N doubly-labeled flexible polypeptides: application to the human prion protein hPrP(23-230). *J Biomol NMR*, **16**, 127-138.

Logan, T.M., Theriault, Y. and Fesik, S.W. (1994) Structural characterization of the FK506 binding protein unfolded in urea and guanidine hydrochloride. *J Mol Biol*, **236**, 637-648.

Löhr, F., Blümel, M., Schmidt, J.M. and Rüterjans, H. (1997) Application of H(N)CA,CO-E.COSY experiments for calibrating the phi angular dependences of vicinal couplings J(C'(i-1),H alpha(i)), J(C'(i-1),C-i(beta)) and J(C'(i-1),C'(i)) in proteins. *J Biomol NMR*, **10**, 107-118.

Loll, P.J. and Lattman, E.E. (1989) The crystal structure of the ternary complex of staphylococcal nuclease, Ca2+, and the inhibitor pdTp, refined at 1.65 A. *Proteins*, **5**, 183-201.

Louhivuori, M., Paakkonen, K., Fredriksson, K., Permi, P., Lounila, J. and Annila, A. (2003) On the origin of residual dipolar couplings from denatured proteins. *J Am Chem Soc*, **125**, 15647-15650.

Ludvigsen, S., Andersen, K.V. and Poulsen, F.M. (1991) Accurate measurements of coupling constants from two-dimensional nuclear magnetic resonance spectra of proteins and determination of phi-angles. *J Mol Biol*, **217**, 731-736.

Maeda, K., Lyon, C.E., Lopez, J.J., Cemazar, M., Dobson, C.M. and Hore, P.J. (2000) Improved photo-CIDNP methods for studying protein structure and folding. *J Biomol NMR*, **16**, 235-244.

Maenaka, K., Kawai, G., Watanabe, K., Sunada, F. and Kumagai, I. (1994) Functional and structural role of a tryptophan generally observed in protein-carbohydrate interaction. TRP-62 of hen egg white lysozyme. *J Biol Chem*, **269**, 7070-7075.

Marion, D., Driscoll, P.C., Kay, L.E., Wingfield, P.T., Bax, A., Gronenborn, A.M. and Clore, G.M. (1989a) Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H-15N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta. *Biochemistry*, **28**, 6150-6156.

Marion, D., Kay, L.E., Sparks, S.W., Torchia, D.A. and Bax, A. (1989b) Three-dimensional heteronuclear NMR of nitrogen-15 labeled proteins. *J Am Chem Soc*, **111**, 1515-1517.

Matthews, B.W. (1995) Studies on protein stability with T4 lysozyme. *Adv Protein Chem*, **46**, 249-278.

Mayor, U., Grossmann, J.G., Foster, N.W., Freund, S.M. and Fersht, A.R. (2003) The denatured state of Engrailed Homeodomain under denaturing and native conditions. *J Mol Biol*, **333**, 977-991.

McKenzie, H.A. and White, F.H., Jr. (1991) Lysozyme and alpha-lactalbumin: structure, function, and interrelationships. *Adv Protein Chem*, **41**, 173-315.

Merutka, G., Dyson, H.J. and Wright, P.E. (1995) 'Random coil' 1H chemical shifts obtained as a function of temperature and trifluoroethanol concentration for the peptide series GGXGG. *J Biomol NMR*, **5**, 14-24.

Mierke, D.F., Grdadolnik, S.G. and Kessler, H. (1992) Use of one-bond C alpha - H alpha coupling constants as restraints in MD simulations. *J Am Chem Soc*, **114**, 8283-8284.

Miranker, A., Radford, S.E., Karplus, M. and Dobson, C.M. (1991) Demonstration by NMR of folding domains in lysozyme. *Nature*, **349**, 633-636.

Mizushima, T., Hirao, T., Yoshida, Y., Lee, S.J., Chiba, T., Iwai, K., Yamaguchi, Y., Kato, K., Tsukihara, T. and Tanaka, K. (2004) Structural basis of sugar-recognizing ubiquitin ligase. *Nat Struct Mol Biol*, **11**, 365-370.

Mogridge, J., Legault, P., Li, J., Van Oene, M.D., Kay, L.E. and Greenblatt, J. (1998) Independent ligand-induced folding of the RNA-binding domain and two functionally distinct antitermination regions in the phage lambda N protein. *Mol Cell*, **1**, 265-275.

Muhandiram, D.R. and Kay, L.E. (1994) Gradient-Enhanced Triple-Resonance Three-Dimensional NMR Experiments with Improved Sensitivity. *J Magn Reson B*, **103**, 203-216.

Nash, D.P. and Jonas, J. (1997a) Structure of pressure-assisted cold denatured lysozyme and comparison with lysozyme folding intermediates. *Biochemistry*, **36**, 14375-14383.

Nash, D.P. and Jonas, J. (1997b) Structure of the pressure-assisted cold denatured state of ubiquitin. *Biochem Biophys Res Commun*, **238**, 289-291.

Neri, D., Billeter, M., Wider, G. and Wüthrich, K. (1992) NMR determination of residual structure in a urea-denatured protein, the 434-repressor. *Science*, **257**, 1559-1563.

Neri, D., Otting, G. and Wüthrich, K. (1990) New nuclear magnetic resonance experiment for measurements of the vicinal coupling constants 3JHN alpha in proteins. *J Am Chem Soc*, **112**, 3663-3665.

Nitta, K. and Sugai, S. (1989) The evolution of lysozyme and alpha-lactalbumin. *Eur J Biochem*, **182**, 111-118.

Nitz, M., Franz, K.J., Maglathlin, R.L. and Imperiali, B. (2003) A powerful combinatorial screen to identify high-affinity terbium(III)-binding peptides. *Chembiochem*, **4**, 272-276.

Nitz, M., Sherawat, M., Franz, K.J., Peisach, E., Allen, K.N. and Imperiali, B. (2004) Structural origin of the high affinity of a chemically evolved lanthanide-binding peptide. *Angew Chem Int Ed Engl*, **43**, 3682-3685.

Nordstrand, K., Ponstingl, H., Holmgren, A. and Otting, G. (1996) Resonance assignment and structural analysis of acid denatured E. coli [U-15N]-glutaredoxin 3: use of 3D 15N-HSQC-(TOCSY-NOESY)-15N-HSQC. *Eur Biophys J*, **24**, 179-184.

Onuchic, J.N., Wolynes, P.G., Luthey-Schulten, Z. and Socci, N.D. (1995) Toward an outline of the topography of a realistic protein-folding funnel. *Proc Natl Acad Sci U S A*, **92**, 3626-3630.

Palmer, A.G. and Case, D.A. (1992) Molecular-Dynamics Analysis of Nmr Relaxation in a Zinc-Finger Peptide. *J Am Chem Soc*, **114**, 9059-9067.

Pardi, A., Billeter, M. and Wüthrich, K. (1984) Calibration of the angular dependence of the amide proton-C alpha proton coupling constants, 3JHN alpha, in a globular protein. Use of 3JHN alpha for identification of helical secondary structure. *J Mol Biol*, **180**, 741-751.

Peng, Z.Y. and Kim, P.S. (1994) A protein dissection study of a molten globule. *Biochemistry*, **33**, 2136-2141.

Peng, Z.Y., Wu, L.C. and Kim, P.S. (1995) Local structural preferences in the alpha-lactalbumin molten globule. *Biochemistry*, **34**, 3248-3252.

Pepys, M.B., Hawkins, P.N., Booth, D.R., Vigushin, D.M., Tennent, G.A., Soutar, A.K., Totty, N., Nguyen, O., Blake, C.C., Terry, C.J., Feest, T.G., Zalin, A.M. and Hsuan, J.J. (1993) Human lysozyme gene mutations cause hereditary systemic amyloidosis. *Nature*, **362**, 553-557.

Permi, P. and Annila, A. (2000) Transverse relaxation optimised spin-state selective NMR experiments for measurement of residual dipolar couplings. *J Biomol NMR*, **16**, 221-227.

Permi, P. and Annila, A. (2004) Coherence transfer in proteins. *Prog NMR Spectrosc*, **44**, 97-137.

Permi, P., Rosevear, P.R. and Annila, A. (2000) A set of HNCO-based experiments for measurement of residual dipolar couplings in 15N, 13C, (2H)-labeled proteins. *J Biomol NMR*, **17**, 43-54.

Permyakov, E.A. and Berliner, L.J. (2000) alpha-Lactalbumin: structure and function. *FEBS Lett*, **473**, 269-274.

Peti, W., Hennig, M., Smith, L.J. and Schwalbe, H. (2000) NMR spectroscopic investigation of psi torsion angle distribution in unfolded ubiquitin from analysis of 3J(C alpha, C alpha) coupling costants and cross-correlated gamma C HNN,C alpha H alpha relaxation rates. *J Am Chem Soc*, **122**, 12017-12018.

Peti, W., Smith, L.J., Redfield, C. and Schwalbe, H. (2001) Chemical shifts in denatured proteins: resonance assignments for denatured ubiquitin and comparisons with other denatured proteins. *J Biomol NMR*, **19**, 153-165.

Petkova, A.T., Buntkowsky, G., Dyda, F., Leapman, R.D., Yau, W.M. and Tycko, R. (2004) Solid state NMR reveals a pH-dependent antiparallel beta-sheet registry in fibrils formed by a beta-amyloid peptide. *J Mol Biol*, **335**, 247-260.

Petkova, A.T., Ishii, Y., Balbach, J.J., Antzutkin, O.N., Leapman, R.D., Delaglio, F. and Tycko, R. (2002) A structural model for Alzheimer's beta -amyloid fibrils based on experimental constraints from solid state NMR. *Proc Natl Acad Sci U S A*, **99**, 16742-16747.

Pike, A.C., Brew, K. and Acharya, K.R. (1996) Crystal structures of guinea-pig, goat and bovine alpha-lactalbumin highlight the enhanced conformational flexibility of regions that are significant for its action in lactose synthase. *Structure*, **4**, 691-703.

Pitsch, S., Weiss, P.A., Wu, X., Ackermann, D. and Honegger, T. (1999) Fast and Reliable Automated Synthesis of RNA and Partially 2'-O-protected Precursors ('Caged RNA') Based on Two Novel, Orthogonal 2'-O-Protecting Groups. *Helv. Chim. Acta*, **82**, 1753-1761.

Plaxco, K.W., J. Morton, C.J., Grimshaw, S.B., Jones, J.J., Pitkeathly, M., Campbell, I.D. and Dobson, C.M. (1997) The effects of guanidine hydrochloride on the 'random coil' conformations and NMR chemical shifts of the peptide series GGXGG. *J Biomol NMR*, **10**, 221-230.

Radford, S.E., Buck, M., Topping, K.D., Dobson, C.M. and Evans, P.A. (1992a) Hydrogen exchange in native and denatured states of hen egg-white lysozyme. *Proteins*, **14**, 237-248.

Radford, S.E., Dobson, C.M. and Evans, P.A. (1992b) The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature*, **358**, 302-307.

Ramachandran, G.N. and Sasisekharan, V. (1968) Conformation of polypeptides and proteins. *Adv Protein Chem*, **23**, 283-438.

Ramakrishnan, B. and Qasba, P.K. (2001) Crystal structure of lactose synthase reveals a large conformational change in its catalytic component, the beta1,4-galactosyltransferase-I. *J Mol Biol*, **310**, 205-218.

Ramakrishnan, B., Shah, P.S. and Qasba, P.K. (2001) alpha-Lactalbumin (LA) stimulates milk beta-1,4-galactosyltransferase I (beta 4Gal-T1) to transfer glucose from UDP-glucose to N-acetylglucosamine. Crystal structure of beta 4Gal-T1 x LA complex with UDP-Glc. *J Biol Chem*, **276**, 37665-37671.

Ramboarina, S. and Redfield, C. (2003) Structural characterisation of the human alpha-lactalbumin molten globule at high temperature. *J Mol Biol*, **330**, 1177-1188.

Rechensteiner, M. (1988) *Ubiquitin*. Plenum Press, New York.

Redfield, C., Schulman, B.A., Milhollen, M.A., Kim, P.S. and Dobson, C.M. (1999) Alpha-lactalbumin forms a compact molten globule in the absence of disulfide bonds. *Nat Struct Biol*, **6**, 948-952.

Robertson, A.D. and Baldwin, R.L. (1991) Hydrogen exchange in thermally denatured ribonuclease A. *Biochemistry*, **30**, 9907-9914.

Roder, H., Wagner, G. and Wüthrich, K. (1985) Individual amide proton exchange rates in thermally unfolded basic pancreatic trypsin inhibitor. *Biochemistry*, **24**, 7407-7411.

Roder, H. and Wuthrich, K. (1986) Protein folding kinetics by combined use of rapid mixing techniques and NMR observation of individual amide protons. *Proteins*, **1**, 34-42.

Rohl, C.A. and Baldwin, R.L. (1997) Comparison of NH exchange and circular dichroism as techniques for measuring the parameters of the helix-coil transition in peptides. *Biochemistry*, **36**, 8435-8442.

Ropson, I.J. and Frieden, C. (1992) Dynamic NMR spectral analysis and protein folding: identification of a highly populated folding intermediate of rat intestinal fatty acid-binding protein by 19F NMR. *Proc Natl Acad Sci U S A*, **89**, 7222-7226.

Rost, B., Yachdav, G. and Liu, J. (2004) The PredictProtein server. *Nucleic Acids Res*, **32**, W321-326.

Saab-Rincon, G., Gualfetti, P.J. and Matthews, C.R. (1996) Mutagenic and thermodynamic analyses of residual structure in the alpha subunit of tryptophan synthase. *Biochemistry*, **35**, 1988-1994.

Sari, N., Alexander, P., Bryan, P.N. and Orban, J. (2000) Structure and dynamics of an acid-denatured protein G mutant. *Biochemistry*, **39**, 965-977.

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

Scheffler, J.E., Cottrell, C.E. and Berliner, L.J. (1985) An inexpensive, versatile sample illuminator for photo-CIDNP on any NMR spectrometer. *J Magn Reson*, **63**, 199-201.

Schleucher, J., Schwendinger, M., Sattler, M., Schmidt, P., Schedletzky, O., Glaser, S.J., Sorensen, O.W. and Griesinger, C. (1994) A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed field gradients. *J Biomol NMR*, **4**, 301-306.

Schlörb, C., Katrin Ackermann, K., Wirmer, J., Wöhnert, J. and Schwalbe, H. (in preparation) Heterologous expression of hen egg white lysozyme in E. coli for isotope enrichment and 1H,13C,15N chemical shift assignment for an non-native state of lysozyme.

Schmidt, J.M., Blümel, M., Löhr, F. and Rüterjans, H. (1999) Self-consistent 3J coupling analysis for the joint calibration of Karplus coefficients and evaluation of torsion angles. *J Biomol NMR*, **14**, 1-12.

Schmidt, P.G. and Kuntz, I.D. (1984) Distance measurements in spin-labeled lysozyme. *Biochemistry*, **23**, 4261-4266.

Schulman, B.A., Kim, P.S., Dobson, C.M. and Redfield, C. (1997) A residue-specific NMR view of the non-cooperative unfolding of a molten globule. *Nat Struct Biol*, **4**, 630-634.

Schulman, B.A., Redfield, C., Peng, Z.Y., Dobson, C.M. and Kim, P.S. (1995) Different subdomains are most protected from hydrogen exchange in the molten globule and native states of human alpha-lactalbumin. *J Mol Biol*, **253**, 651-657.

Schwalbe, H., Fiebig, K.M., Buck, M., Jones, J.A., Grimshaw, S.B., Spencer, A., Glaser, S.J., Smith, L.J. and Dobson, C.M. (1997) Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea. *Biochemistry*, **36**, 8977-8991.

Schwalbe, H., Grimshaw, S.B., Spencer, A., Buck, M., Boyd, J., Dobson, C.M., Redfield, C. and Smith, L.J. (2001) A refined solution structure of hen lysozyme determined using residual dipolar coupling data. *Protein Sci*, **10**, 677-688.

Schwarzinger, S., Kroon, G.J., Foss, T.R., Chung, J., Wright, P.E. and Dyson, H.J. (2001) Sequence-dependent correction of random coil NMR chemical shifts. *J Am Chem Soc*, **123**, 2970-2978.

Schwarzinger, S., Wright, P.E. and Dyson, H.J. (2002) Molecular hinges in protein folding: the urea-denatured state of apomyoglobin. *Biochemistry*, **41**, 12681-12686.

Serrano, L. (1995) Comparison between the phi distribution of the amino acids in the protein database and NMR data indicates that amino acids have various phi propensities in the random coil conformation. *J Mol Biol*, **254**, 322-333.

Shaka, A.J., Barker, P.B. and Freeman, R. (1985) Computer-optimized decoupling scheme for wideband applications and low-level operation. *J Magn Reson*, **64**, 547-552.

Shortle, D. and Abeygunawardana, C. (1993) NMR analysis of the residual structure in the denatured state of an unusual mutant of staphylococcal nuclease. *Structure*, **1**, 121-134.

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

Skolnick, J., Kolinski, A., Brooks, C.L., 3rd, Godzik, A. and Rey, A. (1993) A method for predicting protein structure from sequence. *Curr Biol*, **3**, 414-423.

Smith, L.J., Bolin, K.A., Schwalbe, H., MacArthur, M.W., Thornton, J.M. and Dobson, C.M. (1996a) Analysis of main chain torsion angles in proteins: prediction of NMR coupling constants for native and random coil conformations. *J Mol Biol*, **255**, 494-506.

Smith, L.J., Fiebig, K.M., Schwalbe, H. and Dobson, C.M. (1996b) The concept of a random coil. Residual structure in peptides and denatured proteins. *Fold Des*, **1**, R95-106.

Spraul, M., Hoffmann, M. and Schwalbe, H. (1996) NMR-Meßzelle und Verfahren zur schnellen Mischung mindestens zweier Reaktionsfluide in der NMR-Meßzelle. German Patent, DE 19548977 C 1.

Stone, M.J., Chandrasekhar, K., Holmgren, A., Wright, P.E. and Dyson, H.J. (1993) Comparison of backbone and tryptophan side-chain dynamics of reduced and oxidized Escherichia coli thioredoxin using 15N NMR relaxation measurements. *Biochemistry*, **32**, 426-435.

Stone, M.J., Fairbrother, W.J., Palmer, A.G., 3rd, Reizer, J., Saier, M.H., Jr. and Wright, P.E. (1992) Backbone dynamics of the Bacillus subtilis glucose permease IIA domain determined from 15N NMR relaxation measurements. *Biochemistry*, **31**, 4394-4406.

Stryer, L. (1988) Part 1 Molecular design of life. In *Biochemistry*, p. 21.

Sugai, S., Yashiro, H. and Nitta, K. (1973) Equilibrium and kinetics of the unfolding of alpha-lactalbumin by guanidine hydrochloride. *Biochim Biophys Acta*, **328**, 35-41.

Sunde, M. and Blake, C. (1997) The structure of amyloid fibrils by electron microscopy and X-ray diffraction. *Adv Protein Chem*, **50**, 123-159.

Swindells, M.B., MacArthur, M.W. and Thornton, J.M. (1995) Intrinsic phi, psi propensities of amino acids, derived from the coil regions of known structures. *Nat Struct Biol*, **2**, 596-603.

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

Tjandra, N. and Bax, A. (1997b) Measurement of dipolar contributions to 1JCH splittings from magnetic-field dependence of J modulation in two-dimensional NMR spectra. *J Magn Reson*, **124**, 512-515.

Tjandra, N., Grzesiek, S. and Bax, A. (1996) Magnetic field dependence of nitrogen-proton J splittings in 15N-enriched human ubiquitin resulting from relaxation interference and residual dipolar coupling. *J Am Chem Soc*, **118**, 6264-6272.

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

Tompa, P. (2002) Intrinsically unstructured proteins. *Trends Biochem Sci*, **27**, 527-533.

Tompa, P. (2003) Intrinsically unstructured proteins evolve by repeat expansion. *Bioessays*, **25**, 847-855.

Tycko, R., Blanco, F.J. and Ishii, Y. (2000) Alignment of biopolymers in strained gels: A new way to create detectable dipole-dipole coupling in high-resolution biomolecular NMR. *J Am Chem Soc*, **122**, 9340-9341.

Ueda, T., Nakashima, A., Hashimoto, Y., Miki, T., Yamada, H. and Imoto, T. (1994) Formation of alpha-helix 88-98 is essential in the establishment of higher-order structure from reduced lysozyme. *J Mol Biol*, **235**, 1312-1317.

Ueda, T., Ohkuri, T. and Imoto, T. (1996) Identification of the peptide region that folds native conformation in the early stage of the renaturation of reduced lysozyme. *Biochem Biophys Res Commun*, **228**, 203-208.

Ueda, T., Yamada, H., Aoki, H. and Imoto, T. (1990) Effect of chemical modifications of tryptophan residues on the folding of reduced hen egg-white lysozyme. *J Biochem (Tokyo)*, **108**, 886-892.

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

Uversky, V.N., Gillespie, J.R. and Fink, A.L. (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins*, **41**, 415-427.

van den Berg, B., Chung, E.W., Robinson, C.V., Mateo, P.L. and Dobson, C.M. (1999) The oxidative refolding of hen lysozyme and its catalysis by protein disulfide isomerase. *Embo J*, **18**, 4794-4803.

Vaney, M.C., Maignan, S., RiesKautt, M. and Ducruix, A. (1996) High-resolution structure (1.33 angstrom) of a HEW lysozyme tetragonal crystal grown in the APCF apparatus. Data and structural comparison wiht a crystal grown under microgravity from SpaceHab-01 mission. *Acta Cryst*, **D52**, 505-517.

Vuister, G.W., Delaglio, F. and Bax, A. (1993) The use of 1JC alpha H alpha coupling constants as a probe for protein backbone conformation. *J Biomol NMR*, **3**, 67-80.

Wagner, G. (1993) NMR Relaxation and protein mobility. *Curr Opin Struct Biol*, **3**, 748-754.

Wang, A.C. and Bax, A. (1996) Determination of the backbone dihedral angles phi in human ubiquitin from reparametrized empirical Karplus equations. *J Am Chem Soc*, **118**, 2483-2494.

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

Wang, M., Scott, W.A., Rao, K.R., Udey, J., Conner, G.E. and Brew, K. (1989) Recombinant bovine alpha-lactalbumin obtained by limited proteolysis of a fusion protein expressed at high levels in Escherichia coli. *J Biol Chem*, **264**, 21116-21121.

Wenter, P., Fürtig, B., Hainard, A., Schwalbe, H. and Pitsch, S. (submitted) Kinetic investigation of photoinduced RNA refolding by realtime NMR spectroscopy.

Wienk, H.L., Martinez, M.M., Yalloway, G.N., Schmidt, J.M., Perez, C., Ruterjans, H. and Lohr, F. (2003) Simultaneous measurement of protein one-bond and two-bond nitrogen-carbon coupling constants using an internally referenced quantitative J-correlated [(15)N,(1)H]-TROSY-HNC experiment. *J Biomol NMR*, **25**, 133-145.

Wijesinha-Bettoni, R., Dobson, C.M. and Redfield, C. (2001) Comparison of the structural and dynamical properties of holo and apo bovine alpha-lactalbumin by NMR spectroscopy. *J Mol Biol*, **307**, 885-898.

Wildegger, G. and Kiefhaber, T. (1997) Three-state model for lysozyme folding: triangular folding mechanism with an energetically trapped intermediate. *J Mol Biol*, **270**, 294-304.

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

Wilkinson, K.D. and Mayer, A.N. (1986) Alcohol-induced conformational changes of ubiquitin. *Arch Biochem Biophys*, **250**, 390-399.

Wirmer, J., Kühn, T. and Schwalbe, H. (2001) Millisecond time resolved photo-CIDNP NMR reveals a non-native folding intermediate on the ion-induced refolding pathway of bovine alpha-lactalbumin. *Angew Chem Int Ed Engl*, **40**, 4248-4251.

Wirmer, J., Schlörb, C., Klein-Seetharaman, J., Hirano, R., Ueda, T., Imoto, T. and Schwalbe, H. (2004) Modulation of compactness and long-range interactions of unfolded lysozyme by single point mutations. *Angew Chem Int Ed Engl*, **43**, 5780-5785.

Wirmer, J., Schlörb, C. and Schwalbe, H. (2005) Conformation and Dynamics of Nonnative States of Proteins studied by NMR Spectroscopy. In Buchner, J. and Kiefhaber, T. (eds.), *Protein Folding Handbook*. WILEY-VCH, Weinheim.

Wirmer, J. and Schwalbe, H. (2002) Angular dependence of 1J(Ni,Calphai) and 2J(Ni,Calpha(i-1)) coupling constants measured in J-modulated HSQCs. *J Biomol NMR*, **23**, 47-55.

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

Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L. and Sykes, B.D. (1995b) 1H, 13C and 15N chemical shift referencing in biomolecular NMR. *J Biomol NMR*, **6**, 135-140.

Wishart, D.S. and Nip, A.M. (1998) Protein chemical shift analysis: a practical guide. *Biochem Cell Biol*, **76**, 153-163.

Wishart, D.S. and Sykes, B.D. (1994) The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. *J Biomol NMR*, **4**, 171-180.

Wishart, D.S., Sykes, B.D. and Richards, F.M. (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry*, **31**, 1647-1651.

Wishart, D.S., Sykes, B.D. and Richards, F.M. (1993) Improved synthetic methods for the selective deuteration of aromatic amino acids: applications of selective protonation towards the identification of protein folding intermediates through nuclear magnetic resonance. *Biochim Biophys Acta*, **1164**, 36-46.

Wittekind, M. and Mueller, L. (1993) HNCACB, a high-sensitivity 3d NMR experiment to correlate amide-proton and nitrogen resonances with the alpha-carbon and beta-carbon resonances in proteins. *J Magn Reson B*, **101**, 201-205.

Wohnert, J., Franz, K.J., Nitz, M., Imperiali, B. and Schwalbe, H. (2003) Protein alignment by a coexpressed lanthanide-binding tag for the measurement of residual dipolar couplings. *J Am Chem Soc*, **125**, 13338-13339.

Wolf, Y.I., Grishin, N.V. and Koonin, E.V. (2000) Estimating the number of protein folds and families from complete genome data. *J Mol Biol*, **299**, 897-905.

Wong, K.B., Clarke, J., Bond, C.J., Neira, J.L., Freund, S.M., Fersht, A.R. and Daggett, V. (2000) Towards a complete description of the structural and dynamic properties of the denatured state of barnase and the role of residual structure in folding. *J Mol Biol*, **296**, 1257-1282.

Wong, K.B., Freund, S.M. and Fersht, A.R. (1996) Cold denaturation of barstar: 1H, 15N and 13C NMR assignment and characterisation of residual structure. *J Mol Biol*, **259**, 805-818.

Wu, D.H., Chen, A.D. and Johnson, C.S. (1995a) An Improved Diffusion-Ordered Spectroscopy Experiment Incorporating Bipolar-Gradient Pulses. *J Magn Reson A*, **115**, 260-264.

Wu, L.C., Peng, Z.Y. and Kim, P.S. (1995b) Bipartite structure of the alpha-lactalbumin molten globule. *Nat Struct Biol*, **2**, 281-286.

Yao, J., Chung, J., Eliezer, D., Wright, P.E. and Dyson, H.J. (2001) NMR structural and dynamic characterization of the acid-unfolded state of apomyoglobin provides insights into the early events in protein folding. *Biochemistry*, **40**, 3561-3571.

Yao, J., Dyson, H.J. and Wright, P.E. (1997) Chemical shift dispersion and secondary structure prediction in unfolded and partly folded proteins. *FEBS Letters*, **419**, 285-289.

Yi, Q., Scalley-Kim, M.L., Alm, E.J. and Baker, D. (2000) NMR characterization of residual structure in the denatured state of protein L. *J Mol Biol*, **299**, 1341-1351.

Zeev-Ben-Mordehai, T., Rydberg, E.H., Solomon, A., Toker, L., Auld, V.J., Silman, I., Botti, S. and Sussman, J.L. (2003) The intracellular domain of the Drosophila cholinesterase-like neural adhesion protein, gliotactin, is natively unfolded. *Proteins*, **53**, 758-767.

Zhang, H., Neal, S. and Wishart, D.S. (2003) RefDB: a database of uniformly referenced protein chemical shifts. *J Biomol NMR*, **25**, 173-195.

Zimmerman, S.B. and Trach, S.O. (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. *J Mol Biol*, **222**, 599-620.

Supplementary material

S2 Chapter 2

S2.1 Pulse sequence for the measurement of $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants

;jwnca, pulse sequence for the measurement of J(N,CA)

;J.Wirmer & H.Schwalbe, J. Biomol. NMR, 23, 47-57 (2002)

```
#include <Avance.incl>            d0
#include <Grad.incl>              DELTA
#include <Delay.incl>             (p22 ph1):f3
                                  50u UNBLKGRAD
"p2=p1*2"                         p16:gp1*EA
"p22=p21*2"                       d16
"d0=3u"                           (p1 ph1) (p21 ph4):f3
"d4=1s/(cnst2*4)"                 d24
"d14=d4-p17"                      (p2 ph1) (p22 ph1):f3
"d11=30m"                         d24
"d13=3u"                          (p1 ph2) (p21 ph5):f3
"DELTA=p16+d16+50u-p14*2-p2-d0*2" p17:gp3
"DELTA1=d13+p16+d16+4u"           d14
"d31=(p14-p2)/2"                  (p2 ph1) (p22 ph1):f3
"d30=(p24-p22)/2"                 p17:gp3
"l3=(td1/2)"                      d14
                                  (p1 ph1)
1 ze                              DELTA1
  d11 pl13:f3                     (p2 ph1)
2 d1 do:f3                        d13
  6m                              p16:gp2
3 d11                             d16 pl13:f3
  18m                             4u  BLKGRAD
4 (p1 ph1)                        go=2 ph31 cpd3:f3
  d4 pl3:f3                       d1 do:f3 mc #0 to 4
  (p2 ph1) (p22 ph6):f3                F1EA(igrad EA & ip5*2, id0 &
  d4 pl0:f2                             ip3*2 & ip6*2 & ip31*2)
  p28 ph1                         exit
  d13
  (p1 ph2) (p21 ph3):f3
  d23*0.5                         ph1=0
  (p24:sp10 ph1):f2 (d30 p22 ph1):f3  ph2=1
  d23*0.5                         ph3=0 2
  d0                              ph4=0 0 2 2
  (p14:sp5 ph1):f2                ph5=1 1 3 3
  (p2 ph1):f1                     ph6=0
  (p14:sp3 ph1):f2                ph7=0 0 2 2
                                  ph31=0 2 2 0
```

;pl1 : f1 channel - power level for pulse (default)

;pl2 : f3 channel - power level for pulse (default)

;pl13: f3 channel - power level for CPD/BB decoupling

;p1 : f1 channel - 90 degree high power pulse

;p2 : f1 channel - 180 degree high power pulse

;p21 : f3 channel - 90 degree high power pulse

```
;p22 : f3 channel - 180 degree high power pulse
;p16: homospoil/gradient pulse
;p17: short homospoil/gradient pulse
;p28: f1 channel - trim pulse
;d0 : incremented delay (2D)                    [3 usec]
;d1 : relaxation delay; 1-5 * T1
;d4 : 1/(4J)XH
;d11: delay for disk I/O                        [30 msec]
;d13: short delay                               [3 usec]
;d16: delay for homospoil/gradient recovery
;d23: delay for measurement of J(N,Ca)
;d24: 1/(4J)XH for XH
;cnst2: = J(XH)
;l3: loop for phase sensitive 2D using  E/A method : l3 = td1/2
;in0: 1/(2 * SW(X)) = DW(X)
;nd0: 2
;NS: 1 * n
;DS: >= 16
;td1: number of experiments
;MC2: echo-antiecho
;cpd3: decoupling according to sequence defined by cpdprg2
;pcpd3: f3 channel - 90 degree pulse for decoupling sequence

;for z-only gradients:
;gpz1: 40%
;gpz2: -8.1%
;gpz3: 80%
;gpz4: 30%
```

S2. 2 Values of $^1J(N_i,C_{\alpha i})$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants

	native ubiquitin		unfolded ubiquitin	
	$^1J(N_i,C_{\alpha i})$ (Hz)	$^2J(N_i,C_{\alpha(i-1)})$ (Hz)	$^1J(N_i,C_{\alpha i})$ (Hz)	$^2J(N_i,C_{\alpha(i-1)})$ (Hz)
Q2	11.4±0.1	8.8±0.1	11.1±0.1	9.1±0.1
I3	11.9±0.1	9.1±0.2		
F4	12.3±0.1	8.4±0.1	10.9±0.1	7.8±0.1
V5	10.8±0.1	8.8±0.2	10.7±0.1	8.3±0.2
K6	10.7±0.1	7.9±0.1	10.8±0.1	7.8±0.1
E16	11.00.3	8.5±0.4	11.0±0.1	8.2±0.1
V17	12.1±0.1	8.5±0.1	10.8±0.1	8.2±0.1
E18	12.3±0.1	8.2±0.1	11.1±0.1	7.8±0.1
E24			10.7±0.2	7.9±0.2
N25	10.3±0.2	5.6±0.1	11.2±0.1	8.0±0.1
K27	9.7±0.2	6.2±0.2		
A28	9.6±0.1	5.6±0.1	11.3±0.1	7.9±0.1
K29	9.8±0.2	6.2±0.1	11.1±0.1	8.4±0.1
I30	9.9±0.2	6.0±0.1	10.7±0.1	8.1±0.1
Q31	9.9±0.2	6.1±0.1	11.0±0.1	7.6±0.1
D32	10.0±0.1	6.0±0.1	11.1±0.1	8.2±0.1
K33	10.0±0.1	6.2±0.1	10.7±0.1	8.0±0.1
E34	9.8±0.1	6.3±0.1	11.1±0.1	7.7±0.1
D39	10.5±0.1	6.4±0.1	11.1±0.4	8.3±0.1
Q40	10.2±0.1	6.0±0.1	11.1±0.1	7.7±0.1
Q41	11.5±0.1	6.2±0.1		
R42	10.6±0.1	9.1±0.1	10.6±0.1	8.2±0.2
L43	10.9±0.1	8.6±0.1	10.6±0.1	8.1±0.2
I44	11.0±0.1	9.2±0.1	10.6±0.1	8.1±0.1
F45	10.8±0.1	8.1±0.2	10.9±0.1	7.5±0.1
A46			10.7±0.1	7.9±0.1
Q49	11.0±0.1	8.9±0.1	11.0±0.1	8.0±0.1
L50	11.2±0.1	8.7±0.1	10.8±0.1	8.0±0.1
E51	11.3±0.1	8.8±0.2		
D52	11.0±0.2	9.6±0.2	11.1±0.1	8.1±0.1
Y59	10.6±0.1	6.0±0.1		
N60	10.3±0.1	6.0±0.1	11.1±0.1	7.7±0.1
I61	10.9±0.1	6.7±0.1	11.0±0.2	6.9±0.2
Q62	12.3±0.1	8.7±0.1	10.9±0.1	7.8±0.1
K63	11.8±0.1	8.4±0.1	10.9±0.1	8.0±0.1
E64	9.9±0.4	8.9±0.4	11.1±0.1	7.8±0.1
H68	11.7±0.1	8.7±0.1	11.0±0.1	8.2±0.1
L69	10.4±0.1	8.0±0.1	10.9±0.1	8.1±0.1
V70	11.3±0.1	8.5±0.1		
L71	11.3±0.1	8.0±0.1	10.8±0.1	7.7±0.1
R72			10.9±0.1	8.2±0.1
L73	10.7±0.1	9.0±0.2		
R74			11.0±0.1	8.2±0.1

S3 Chapter 3

S3.1 Relaxation data in unfolded ubiquitin

	R_1 (s^{-1})	R_2 (s^{-2})	HetNOE (a.u.)
Q2	1.18 ±0.04	1.74 ±0.03	-1.00 ±0.04
I3		2.97 ±0.02	-0.10 ±0.04
F4	1.56 ±0.08	2.96 ±0.04	-0.28 ±0.04
V5			-0.10 ±0.04
K6	1.66 ±0.08	2.90 ±0.04	0.15 ±0.04
T7			0.10 ±0.04
L8	1.63 ±0.08	3.38 ±0.05	0.06 ±0.04
T9	1.51 ±0.10	6.37 ±0.11	0.08 ±0.04
G10	1.52 ±0.08	2.86 ±0.04	-0.10 ±0.04
K11	1.60 ±0.10	3.13 ±0.05	0.10 ±0.04
T12	1.53 ±0.09	3.81 ±0.05	0.06 ±0.04
I13	1.67 ±0.09	3.38 ±0.06	0.18 ±0.04
T14	1.58 ±0.09	4.11 ±0.06	-0.02 ±0.04
L15	1.61 ±0.10	3.22 ±0.05	0.15 ±0.04
E16	1.57 ±0.08	3.49 ±0.05	0.10 ±0.04
V17	1.60 ±0.10	3.56 ±0.06	-0.04 ±0.04
E18	1.61 ±0.09	3.55 ±0.05	0.11 ±0.04
S20	1.54 ±0.08	3.70 ±0.05	0.12 ±0.04
D21	1.57 ±0.08	3.53 ±0.05	0.05 ±0.04
T22	1.51 ±0.08	3.41 ±0.05	0.15 ±0.04
I23	1.60 ±0.09	3.46 ±0.05	0.09 ±0.04
E24	1.58 ±0.08	3.47 ±0.05	-0.11 ±0.04
N25	1.55 ±0.09	3.55 ±0.05	-0.07 ±0.04
V26	1.46 ±0.05	3.46 ±0.03	0.17 ±0.04
K27	1.57 ±0.04	3.34 ±0.03	0.08 ±0.04
A28	1.59 ±0.06	3.53 ±0.04	-0.06 ±0.04
K29	1.48 ±0.07	3.53 ±0.05	-0.08 ±0.04
I30	1.59 ±0.09	3.46 ±0.05	-0.08 ±0.04
Q31	1.71 ±0.06	3.59 ±0.03	0.15 ±0.04
D32	1.56 ±0.08	3.33 ±0.04	-0.04 ±0.04
K33	1.59 ±0.09	3.38 ±0.05	0.14 ±0.04
E34	1.48 ±0.08	3.49 ±0.05	
G35	1.55 ±0.03	2.94 ±0.04	-0.04 ±0.04
I36	1.51 ±0.10	3.95 ±0.07	0.19 ±0.04
D39	1.67 ±0.10	4.40 ±0.07	0.22 ±0.04
Q40	1.50 ±0.08	3.87 ±0.06	0.12 ±0.04
Q41	1.59 ±0.10	3.54 ±0.05	0.01 ±0.04
R42	1.61 ±0.10	3.85 ±0.06	0.19 ±0.04
L43	1.53 ±0.09	3.69 ±0.06	0.07 ±0.04
I44	1.58 ±0.11	3.60 ±0.06	-0.06 ±0.04
F45	1.64 ±0.09	3.55 ±0.05	0.16 ±0.04
A46	1.49 ±0.07	3.87 ±0.04	-0.06 ±0.04
G47	1.35 ±0.06	3.93 ±0.05	-0.26 ±0.04

	R_1 (s^{-1})	R_2 (s^{-2})	HetNOE (a.u.)
K48	1.52 ±0.07	3.16 ±0.04	-0.11 ±0.04
Q49	1.56 ±0.08	3.16 ±0.05	-0.04 ±0.04
L50	1.65 ±0.07	3.43 ±0.03	0.10 ±0.04
E51	1.60 ±0.06	3.37 ±0.04	0.16 ±0.04
D52	1.58 ±0.09	3.32 ±0.06	-0.02 ±0.04
G53	1.50 ±0.08	2.74 ±0.05	0.02 ±0.04
R54	1.59 ±0.08	3.32 ±0.05	0.16 ±0.04
T55	1.65 ±0.07	3.37 ±0.04	-0.10 ±0.04
L56	1.57 ±0.04	3.34 ±0.03	0.08 ±0.04
S57	1.62 ±0.05	3.31 ±0.03	0.10 ±0.04
D58	1.58 ±0.06	3.37 ±0.04	0.17 ±0.04
Y59	1.46 ±0.05	3.46 ±0.03	0.17 ±0.04
N60	1.58 ±0.09	3.78 ±0.05	0.21 ±0.04
I61	1.60 ±0.09	3.50 ±0.05	0.14 ±0.04
Q62	1.66 ±0.06	3.53 ±0.03	0.15 ±0.04
K63	1.53 ±0.05		-0.10 ±0.04
E64		3.44 ±0.05	0.03 ±0.04
S65	1.53 ±0.07	3.41 ±0.04	0.14 ±0.04
T66	1.66 ±0.07	3.37 ±0.04	-0.10 ±0.04
L67	1.66 ±0.07	3.27 ±0.04	0.15 ±0.04
H68	1.58 ±0.08	3.29 ±0.04	-0.05 ±0.04
L69	1.58 ±0.07	2.84 ±0.03	-0.06 ±0.04
V70	1.55 ±0.08	2.83 ±0.04	0.06 ±0.04
L71	1.59 ±0.07	±0.06	0.05 ±0.04
R72	1.57 ±0.07	2.56 ±0.03	-0.06 ±0.04
R74	1.43 ±0.05	2.04 ±0.03	-0.57 ±0.04
G75	1.12 ±0.04	1.37 ±0.02	-0.67 ±0.04
G76	0.77 ±0.03	0.93 ±0.02	-1.96 ±0.04

S4 Chapter 4

S4.1 Assignment of A9G-S^{ME} and W62Y-S^{ME}

Residue	A9G-S^{ME}			W62Y-S^{ME}		
	H^N (ppm)	N (ppm)	H$^\alpha$ (ppm)	H^N (ppm)	N (ppm)	H$^\alpha$ (ppm)
1			3.97			3.98
2	8.48	123.4	4.11	8.47	123.3	4.11
3	8.60	126.3	4.59	8.60	126.1	4.61
4	8.34	111.6	3.87	8.34	111.5	3.89
5	8.24	121.3	4.29	8.27	121.3	4.30
6	8.51	122.1	4.46	8.52	121.8	4.46
7	8.59	124.3		8.53	124.0	4.34
8			4.29	8.19	124.0	4.25
9	8.41	110.3	3.88	8.25	125.1	4.24
10						
11	8.18	123.0	4.19	8.17	123.0	4.19
12	8.14	119.6	4.33	8.11	119.3	4.33
13						
14						4.28
15			4.67	8.63	121.4	4.72
16	8.45	110.0	3.91	8.48	110.0	3.92
17	8.23	122.3	4.25	8.21	122.1	4.26
18	8.53	120.1	4.61	8.52	119.9	4.63
19	8.29	119.8	4.62	8.29	119.6	4.62
20	8.04	121.1	4.44	8.04	121.1	4.48
21	8.24	124.1	4.16	8.22	124.0	4.20
22	7.74	109.5	3.78	7.73	109.4	3.81
23	7.93	120.5	4.55	7.93	120.4	4.55
24	8.24	118.3	4.36	8.24	118.2	4.39
25	8.20	124.6	4.26	8.21	124.5	4.28
26	8.17	109.2	3.77	8.17	109.1	3.77
27	8.14	119.3	4.65	8.14	119.1	4.65
28	8.05	122.2	4.62	8.05	122.1	4.63
29	7.83	122.3	3.90	7.82	122.3	3.92
30	8.11	123.4	4.33	8.11	123.3	4.35
31	8.36	127.3	4.17	8.36	127.3	4.18
32	8.08	123.0	4.17	8.09	123.1	4.20
33						
34						
35	8.25	122.0	4.29	8.26	122.0	4.29
36	8.25	117.2	4.30	8.25	117.2	4.30
37	8.31	120.7	4.66	8.33	120.7	4.66
38	8.12	120.9	4.53	8.13	120.9	4.60
39	8.31	120.7	4.66	8.33	120.7	4.70
40	8.09	115.1	4.24	8.10	115.0	4.25
41	8.30	122.6	4.24	8.31	122.6	4.29
42	8.25	125.4		8.26	125.3	
43	8.07	113.3	4.29	8.08	113.3	4.29

Residue	A9G-S^{ME}			W62Y-S^{ME}		
	H^N (ppm)	N (ppm)	H$^\alpha$ (ppm)	H^N (ppm)	N (ppm)	H$^\alpha$ (ppm)
44	8.35	121.5	4.66	8.36	121.4	4.69
45	8.31	122.0	4.28	8.32	122.0	4.29
46	8.47	120.1	4.72	8.48	120.0	4.73
47	8.15	114.7	4.29	8.16	114.6	4.31
48	8.48	121.4	4.69	8.47	121.3	4.70
49	8.36	110.1	3.94	8.37	109.9	3.95
50	8.14	116.1	4.46	8.15	116.0	4.48
51	8.19	116.3	4.26	8.19	116.2	4.23
52	8.35	121.5	4.66	8.36	121.4	4.70
53	8.11	121.4	4.44	8.11	121.5	4.45
54	8.25	110.3	3.85	8.26	110.2	3.85
55	7.82	120.3	4.13	7.82	120.2	4.14
56	8.21	125.6	4.25	8.21	125.6	4.30
57			4.29			4.28
58	8.10	122.2	3.98	8.10	122.1	4.03
59	8.31	122.0	4.66	8.39	122.1	4.69
60	8.17	116.7	4.29	8.19	116.9	4.20
61	8.16	122.5	3.99	8.22	122.5	3.82
62	7.77	121.1	4.52			
63	7.37	121.1	4.50			
64	7.71	121.4	4.42			
65						
66	8.33	119.2	4.62	8.34	119.4	4.62
67	8.30	109.0	3.86	8.32	108.9	3.89
68	7.93	120.5	4.37	7.97	120.3	4.40
69	8.21	117.7	4.58	8.23	117.7	
70						
71	8.44	110.9	3.91	8.43	110.7	3.93
72	8.08	116.1	4.38	8.10	116.0	4.40
73	8.39	123.1	4.29	8.40	123.1	4.33
74	8.40	120.3	4.65	8.40	120.1	4.66
75	8.27	123.3	4.31	8.28	123.3	4.33
76	8.27	120.3	4.45	8.28	120.4	4.48
77	8.48	121.8	4.67	8.49	121.8	4.70
78	7.99	123.4	4.36	8.01	123.4	4.43
79						
80	8.48	121.4	4.39	8.49	121.0	4.42
81	8.49	118.6	4.41	8.50	118.8	4.42
82	8.17	126.4	4.30	8.20	126.4	4.30
83	7.92	120.8	4.28	7.96	120.8	4.30
84	8.00	122.6	4.28	8.03	122.7	4.30
85	8.19	116.6	4.35	8.21	116.6	4.37
86						
87	8.45	122.1	4.69	8.45	122.1	4.69
88	8.10	122.2	4.46	8.14	122.1	4.42
89	8.25	116.0	4.31	8.26	115.9	4.32
90	8.23	126.9	4.26	8.24	127.0	4.28

Residue	A9G-S^{ME}			W62Y-S^{ME}		
	H^N (ppm)	N (ppm)	H$^\alpha$ (ppm)	H^N (ppm)	N (ppm)	H$^\alpha$ (ppm)
91	8.14	118.7	4.43	8.15	118.7	4.44
92	8.04	121.9	4.17	8.05	121.7	4.18
93			4.70			4.70
94	8.24	121.3	4.45	8.27	121.3	4.45
95	8.35	126.5	4.23	8.36	126.4	4.26
96						
97						
98						
99	8.31	125.9	4.13	8.32	125.9	4.14
100	8.40	120.3	4.45	8.43	120.4	4.45
101	8.56	122.1	4.77	8.56	122.0	4.77
102	8.39	110.0	3.92	8.39	109.9	3.92
103	8.33	119.2	4.68	8.34	119.0	4.70
104	8.45	110.0	3.91	8.45	109.8	3.93
105	8.19	120.6	4.37	8.19	120.5	4.40
106	8.38	119.9	4.56	8.38	119.8	4.57
107	8.19	124.4	4.18	8.20	124.2	4.20
108	8.02	120.5	4.50	8.03	120.4	4.50
109	7.79	121.7	3.67	7.77	121.8	3.71
110	7.99	125.0	4.07	7.99	124.9	4.07
111	7.89	120.0	4.40	7.88	119.8	4.44
112	7.98	121.4	3.84	7.96	121.3	3.86
113	8.04	118.4	4.54	8.04	118.2	4.56
114	7.87	121.2	4.18	7.88	121.1	4.25
115	8.17	120.7	4.39	8.17	120.7	4.42
116	8.34	123.8	4.24	8.36	123.8	4.26
117	8.34	110.4	3.97	8.33	110.3	3.96
118	8.04	113.9	4.31	8.04	113.8	4.29
119	8.51	122.1	4.73	8.52	122.0	4.75
120	8.07	121.6	3.90	8.08	121.6	3.93
121	8.34	123.8	4.18	8.33	123.7	4.21
122	8.11	124.8	4.17	8.13	124.8	4.15
123	7.93	120.5	4.60	7.97	120.3	4.61
124	7.88	123.4	3.97	7.88	123.3	3.98
125	8.15	124.5	4.10	8.15	124.6	4.10
126	8.27	110.1	3.93	8.27	110.0	3.91
127	8.10	120.5	4.47	8.10	120.3	4.45
128	8.48	124.5	4.28	8.48	124.4	4.28
129	8.33	125.2	4.30	8.32	125.4	4.30

S4.2 Assignment of W62G-S^{ME}, W111G-S^{ME} and W123G-S^{ME}

Residue	W62G-S^{ME}		W111G-S^{ME}			W123G-SME		
	H^N (ppm)	H^α (ppm)	H^N (ppm)	N (ppm)	H^α (ppm)	H^N (ppm)	N (ppm)	H^α (ppm)
1								3.95
2	8.50	4.12	8.50	123.3	4.11	8.50	123.2	
3	8.61	4.58	8.62	126.1	4.60	8.62	126.1	4.61
4	8.35	3.87	8.06	112.8	3.89	8.37	111.4	3.89
5	8.29	4.26	8.30	121.3	4.28	8.30	121.2	4.28
6		--	8.54	121.8	4.46	8.53	121.8	4.46
7	8.53	4.33	8.55	124.0		8.55	124.0	4.34
8	8.20	4.23	8.22	124.0		8.20	123.9	4.25
9	8.24	4.16	8.26	125.1	4.24	8.26	125.1	4.24
10		--						
11	8.17	4.15	8.19	123.0	4.17	8.19	123.0	4.20
12	8.11	4.33	8.13	119.3		8.13	119.3	4.34
13		--						
14		--						
15		--	8.62	120.5				4.68
16	8.50	3.91	8.42	110.6	3.93	8.45	110.7	3.94
17	8.21	4.26	8.23	122.1	4.27	8.23	122.1	4.29
18	8.52	4.61	8.53	119.9	4.62	8.49	120.1	4.64
19	8.30	4.61	8.31	119.6	4.62	8.31	119.6	4.63
20	8.06	4.47	8.07	121.1	4.48	8.07	121.2	4.48
21	8.24	4.15	8.24	124.0	4.18	8.23	124.0	4.17
22	7.75	3.77	7.75	109.3	3.79	7.75	109.3	3.79
23	7.94	4.51	7.96	120.4	4.54	8.06	120.4	4.54
24	8.25	4.37	8.26	118.2	4.40	8.26	118.2	4.39
25		--	8.22	124.5	4.27	8.21	124.5	4.28
26	8.18	3.77	8.20	109.1	3.77	8.19	109.1	3.77
27	8.15	4.61	8.17	119.1	4.27	8.16	119.2	4.66
28	8.06	4.58	8.07	122.1	4.61	8.06	122.2	4.63
29	7.83	3.91	7.85	122.3	3.92	7.83	122.4	3.94
30	8.12	4.33	8.14	123.4	4.35	8.14	123.3	4.35
31	8.37	4.19	8.39	127.3	4.19	8.39	127.3	4.19
32	8.10	4.13	8.11	123.1	4.19	8.11	123.0	4.19
33		--						
34		--						
35	8.29	4.26	8.28	121.9	4.29	8.28	122.0	4.29
36	8.26	4.30	8.28	117.2	4.30	8.28	117.2	4.30
37		--	8.35	120.7	4.62	8.45	120.4	4.65
38	8.14	4.51	8.15	120.9	4.53	8.14	120.9	4.53
39	8.34	4.65	8.35	120.7	4.68	8.35	120.6	4.69
40	8.10	4.22	8.12	115.1	4.23	8.11	115.0	4.27
41	8.32	4.26	8.33	122.6	4.25	8.33	122.6	4.26
42	8.27	4.29	8.29	125.4		8.28	125.4	
43	8.09	4.25	8.11	113.3	4.29	8.10	113.3	4.29
44	8.39	4.72	8.38	121.4	4.66	8.38	121.5	4.68

| Residue | W62G-S^{ME} | | W111G-S^{ME} | | | W123G-SME | | |
	H^N (ppm)	H^α (ppm)	H^N (ppm)	N (ppm)	H^α (ppm)	H^N (ppm)	N (ppm)	H^α (ppm)
45	8.32	4.26	8.34	121.9	4.29	8.34	121.9	4.30
46	8.49	4.72	8.50	120.1	4.73	8.43	120.1	4.73
47	8.17	4.30	8.19	114.6	4.29	8.18	114.6	4.29
48	8.49	4.68	8.49	121.3	4.70	8.49	121.3	4.70
49	8.40	3.91	8.38	110.0	3.92	8.38	109.9	3.94
50	8.16	4.46	8.17	116.0	4.47	8.17	116.0	4.46
51	8.20	4.26	8.21	116.1	4.26	8.21	116.2	4.28
52	8.37	4.65	8.38	121.4	4.66	8.41	121.3	4.69
53	8.13	4.43	8.13	121.4	4.44	8.13	121.4	4.45
54	8.27	3.84	8.28	110.2	3.85	8.27	110.2	3.85
55	7.84	4.08	7.85	120.2	4.12	7.84	120.2	4.12
56	8.22	4.29	8.23	125.6	4.30	8.23	125.5	4.30
57		--			4.28			
58		--	8.13	122.1	4.03	8.09	122.1	4.01
59	8.41	4.68	8.34	121.9	4.68	8.34	121.9	4.69
60	8.22	4.30	8.19	116.7	4.24	8.19	116.6	4.25
61		--	8.19	122.4	3.98	8.19	122.4	3.99
62	8.37	3.87	7.79	120.9	4.51	7.97	120.8	4.52
63		--	8.59	125.8	4.48	7.58	121.3	
64		--	7.73	121.3	4.42	7.73	121.3	
65		--						
66	8.34	4.61	8.36	119.1	4.60	8.34	119.4	4.62
67	8.35	3.88	8.33	108.9	3.87	8.33	108.9	3.87
68	7.97	4.58	7.96	120.4	4.37	7.98	120.5	
69	8.23	4.40	8.25	117.7		8.25	117.7	
70		--			4.66			
71	8.44	3.91	8.45	110.7	3.93	8.40	110.5	3.94
72	8.12	4.40	8.11	116.0	4.38	8.11	116.0	4.38
73	8.42	4.30	8.42	123.1		8.42	123.0	4.33
74		--	8.42	119.9	4.64	8.41	119.9	4.68
75	8.30	4.30	8.31	123.2	4.33	8.30	123.2	4.32
76	8.29	4.46	8.29	120.3	4.48	8.52	120.0	4.49
77	8.53	4.44	8.54	121.8	4.68	8.53	121.8	4.69
78	8.08	4.40	8.03	123.4	4.38	8.03	123.3	4.44?
79		--						
80	8.51	4.43	8.51	121.1	4.40	8.49	121.0	4.40
81	8.52	4.32	8.53	118.8	4.41	8.52	118.7	4.42
82	8.22	4.29	8.22	126.4	4.32	8.22	126.4	4.30
83	7.97	4.26	7.98	120.8	4.28	8.00	120.7	4.30
84	8.04	4.33	8.06	122.7	4.30	8.06	122.7	4.30
85	8.22	4.39	8.24	116.6	4.35	8.24	116.6	4.37
86		--						
87	8.47	4.69	8.49	122.3	4.69	8.48	122.2	4.73
88	8.13	4.08				8.14	122.1	4.45
89	8.27	4.40	8.30	115.9	4.32	8.29	115.9	4.32
90	8.25	4.30	8.27	127.0	4.28	8.26	127.0	4.28

Residue	W62G-S^{ME}		W111G-S^{ME}			W123G-SME		
	H^{N} (ppm)	H$^{\alpha}$ (ppm)	H^{N} (ppm)	N (ppm)	H$^{\alpha}$ (ppm)	H^{N} (ppm)	N (ppm)	H$^{\alpha}$ (ppm)
91	8.16	4.26	8.18	118.8		8.17	118.7	
92	8.06	4.15	8.07	121.8	4.16	8.07	121.8	4.18
93		--			4.72			4.71
94	8.27	4.43	8.41	121.4	4.42	8.29	121.4	4.47
95	8.37	4.24	8.39	126.5	4.25	8.38	126.5	4.25
96		--						
97		--						
98		--						
99	8.33	4.12	8.36	126.1	4.12	8.35	126.0	4.14
100		--	8.46	120.5	4.46	7.79	120.9	4.43
101	8.57	4.75	8.61	122.1	4.76	8.60	122.3	4.76
102	8.37	3.94	8.42	109.9	3.92	8.42	109.9	3.94
103	8.35	4.75	8.36	119.1	4.69	8.36	119.1	4.69
104	8.46	3.91	8.49	109.8	3.92	8.47	109.6	3.92
105	8.20	4.37	8.20	120.2	4.40	8.29	120.3	4.34
106	8.39	4.54	8.42	119.9	4.58	8.04	119.9	4.65
107	8.20	4.26	8.17	124.3	4.20	8.21	124.3	4.19
108	8.04	4.47	8.11	120.8	4.56	8.20	120.3	4.54
109	7.78	3.68	7.75	122.9	3.84	7.76	121.9	3.71
110	8.01	4.05	8.18	127.1	4.05	8.03	125.1	4.07
111	7.89	4.40	8.25	108.0	3.87	7.90	119.8	4.45
112	7.95	4.61	8.04	120.7	4.24	7.95	121.4	3.89
113	8.05	4.54	8.44	119.7	4.62	8.07	118.4	4.57
114	7.90	4.19				7.92	121.2	4.44
115	8.18	4.40				8.25	120.8	4.20
116	8.37	4.22			4.24	8.41	124.0	4.26
117	8.35	3.94	8.37	111.4	3.96	8.37	110.4	3.97
118	8.06	4.29	8.10	113.9	4.31	8.08	113.9	4.27
119	8.54	4.72	8.57	122.2	4.73	8.58	122.2	4.73
120	8.08	3.91	8.12	121.5	3.92	8.13	121.6	4.04
121	8.34	4.12	8.38	124.0	4.17	8.44	124.5	4.27
122	8.13	4.16	8.19	125.0	4.20	8.27	123.7	4.34
123	7.99	4.37	8.02	120.5	4.62	8.49	109.9	3.94
124	7.89	3.98	7.91	123.6	3.98	8.49	125.7	4.49
125	8.16	4.08			4.09	8.28	124.1	4.26
126	8.28	3.91	8.34	110.1	3.92	8.34	108.4	3.92
127	8.11	4.44	8.15	120.4	4.46	7.95	120.3	
128	8.50	4.26	8.52	124.4	4.29	8.57	124.5	4.38
129	8.33	4.29	8.36	125.5	4.29	8.39	125.6	4.38

S4.3 R_2 relaxation rates of WT-S^{ME*}, A9G-S^{ME} and W62Y-S^{ME}

Residue	R_2(WT-S^{ME}) (s^{-1})	R_2(A9G-S^{ME}) (s^{-1})	R_2(W62G-S^{ME}) (s^{-1})
1			
2	1.43±0.06	1.45±0.05	1.78±0.36
3		2.27±0.05	1.99±0.32
4	1.56±0.07	1.72±0.05	1.62±0.25
5	2.21±0.07	2.33±0.04	2.25±0.34
6	2.30±0.07	2.52±0.04	
7	2.46±0.08	2.33±0.05	2.67±0.36
8	2.90±0.08		2.85±0.35
9	3.37±0.11	3.57±0.07	2.86±0.29
10			
11	3.05±0.08	2.89±0.08	2.90±0.35
12	3.09±0.10	2.77±0.07	2.97±0.30
13			
14			
15			3.19±0.40
16			3.04±0.38
17	3.19±0.06	3.13±0.08	3.12±0.28
18	2.89±0.05	3.30±0.07	3.08±0.32
19	3.21±0.11	3.42±0.10	3.25±0.26
20	3.43±0.08	3.69±0.08	3.38±0.28
21	3.64±0.09	2.60±0.05	3.33±0.24
22	4.70±0.14	3.64±0.18	3.36±0.28
23	3.71±0.07		3.70±0.41
24	3.71±0.11	4.05±0.10	3.46±0.29
25	3.73±0.11	3.85±0.09	3.31±0.20
26	5.09±0.13	4.94±0.11	3.40±0.31
27	3.92±0.11	4.50±0.10	3.74±0.37
28	4.40±0.10	4.49±0.10	3.71±0.37
29	5.11±0.39	4.71±0.51	3.65±0.23
30	4.30±0.20	5.61±0.20	3.80±0.24
31	5.02±0.23	7.72±0.25	3.68±0.31
32	3.90±0.12	4.77±0.10	3.29±0.27
33			
34			
35	5.47±0.20	6.63±0.24	3.93±0.25
36	3.88±0.12	4.27±0.11	3.31±0.25
37	3.39±0.06	3.98±0.06	
38	3.34±0.08	3.92±0.08	3.54±0.35
39	3.39±0.06	3.98±0.06	3.38±0.29
40	3.42±0.08	3.61±0.07	3.24±0.34
41	3.48±0.10	3.79±0.09	3.30±0.32
42	3.98±0.08	3.32±0.07	3.06±0.33
43	3.08±0.09	3.30±0.08	3.07±0.32
44	3.83±0.07	4.20±0.06	3.14±0.29
45	3.47±0.09	3.53±0.08	3.07±0.27
46	3.12±0.08	3.64±0.09	3.16±0.31

Residue	R_2(WT-S^{ME}) (s^{-1})	R_2(A9G-S^{ME}) (s^{-1})	R_2(W62G-S^{ME}) (s^{-1})
47	3.08±0.08	3.21±0.08	3.29±0.37
48	3.28±0.08	3.59±0.08	3.31±0.37
49	3.00±0.08	4.26±0.09	2.87±0.37
50	3.44±0.08	3.44±0.07	3.26±0.39
51	3.68±0.09	3.91±0.08	3.47±0.37
52	3.83±0.07	4.20±0.06	3.19±0.15
53	4.33±0.11	4.97±0.12	4.05±0.28
54	4.18±0.15	6.71±0.19	3.84±0.30
55	4.94±0.16	5.26±0.15	4.04±0.34
56	5.50±0.25	5.97±0.27	4.21±0.24
57			
58	6.11±0.21	5.17±0.11	
59	5.90±0.27	6.26±0.38	
60	7.62±0.31		3.75±0.30
61	5.53±0.24	6.18±0.24	
62	6.25±0.36	7.05±0.37	2.92±0.27
63	7.61±0.99	7.12±1.99	
64		7.83±1.35	
65			
66	4.89±0.25	5.90±0.19	3.74±0.30
67	5.34±0.21	4.96±0.16	3.49±0.36
68	3.71±0.07		3.34±0.36
69	3.69±0.11	3.97±0.10	3.26±0.32
70			
71	2.66±0.10	3.44±0.10	2.94±0.33
72	3.64±0.12	3.81±0.10	3.36±0.36
73	3.46±0.13	3.64±0.11	2.99±0.31
74	3.56±0.13	3.91±0.12	
75	3.54±0.16	3.87±0.12	3.04±0.26
76	3.11±0.17	4.28±0.14	3.19±0.28
77	4.16±0.18	4.26±0.15	2.75±0.33
78	3.73±0.21	3.94±0.16	3.50±0.31
79			
80	3.91±0.17	4.43±0.16	3.54±0.38
81	4.32±0.20	5.31±0.20	3.24±0.23
82	4.02±0.20	5.06±0.17	3.63±0.24
83	3.33±0.18	4.37±0.17	3.17±0.27
84	3.95±0.19	4.40±0.17	
85	3.65±0.16	4.21±0.13	3.13±0.25
86			
87	3.71±0.16	3.51±0.14	3.52±0.31
88	3.91±0.08	3.65±0.08	
89	3.11±0.12	3.17±0.11	2.81±0.26
90	3.99±0.14	4.80±0.13	3.04±0.24
91	3.70±0.15	4.21±0.14	3.28±0.20
92	3.63±0.09	3.75±0.09	3.21±0.35
93			
94	2.21±0.07	2.33±0.04	2.49±0.08

Residue	R_2(WT-S^{ME}) (s^{-1})	R_2(A9G-S^{ME}) (s^{-1})	R_2(W62G-S^{ME}) (s^{-1})
95	3.96±0.16	4.18±0.12	
96			
97			
98			
99		4.48±0.21	3.20±0.14
100	2.95±0.15	3.74±0.14	2.37±0.12
101	2.57±0.13	3.40±0.11	3.03±0.33
102	2.73±0.11	3.96±0.11	3.05±0.43
103	3.02±0.11	3.49±0.09	3.05±0.36
104	3.55±0.08	4.80±0.11	2.92±0.40
105	3.54±0.13	3.82±0.11	3.51±0.35
106	3.98±0.14	4.44±0.13	3.54±0.33
107	4.18±0.15	4.68±0.12	4.03±0.38
108	4.54±0.17	5.68±0.15	4.36±0.26
109	5.64±0.49	5.96±0.45	4.59±0.26
110	5.14±0.28	6.59±0.26	4.92±0.29
111	5.32±0.22	6.18±0.20	4.73±0.31
112	5.33±0.30	6.29±0.29	3.85±0.27
113	5.32±0.26	6.04±0.21	4.75±0.24
114	4.46±0.23	5.09±0.21	4.29±0.33
115	3.61±0.22	4.55±0.20	3.94±0.34
116	3.63±0.21		3.44±0.27
117	3.07±0.13	4.91±0.13	3.05±0.29
118	3.18±0.12	3.52±0.10	3.28±0.34
119	3.53±0.10		3.25±0.31
120	3.36±0.09	3.72±0.09	3.26±0.40
121	3.93±0.16	4.04±0.10	3.34±0.20
122	3.39±0.15	3.84±0.13	3.22±0.28
123	3.71±0.07		3.50±0.29
124	3.60±0.24	4.41±0.22	
125	3.76±0.20	2.85±0.04	3.46±0.30
126	3.04±0.17	5.69±0.17	3.00±0.34
127	2.17±0.12	2.91±0.08	2.82±0.40
128	2.05±0.11	2.39±0.09	2.29±0.28
129	2.36±0.11	1.91±0.08	1.90±0.27

* Grimshaw, 1999

S4.4 R_2 relaxation rates of W62Y-S^{ME}, W111G-S^{ME} and W123G-S^{ME}

Residue	R_2(W62Y-S^{ME}) (s^{-1})	R_2(W111G-S^{ME}) (s^{-1})	R_2(W123G-S^{ME}) (s^{-1})
1			
2	1.51±0.14	1.45±0.10	1.50±0.09
3	1.98±0.13	1.86±0.09	1.86±0.09
4	1.95±0.15	1.84±0.11	1.54±0.10
5	2.34±0.13	2.22±0.12	2.43±0.09
6	2.65±0.16	2.73±0.14	2.33±0.09
7	2.65±0.19	2.74±0.17	2.50±0.12
8	2.80±0.17	2.96±0.15	2.87±0.11
9	3.66±0.17	3.29±0.13	4.02±0.12
10			
11	3.11±0.17	3.23±0.15	3.04±0.12
12	3.81±0.22	3.40±0.15	3.19±0.15
13			
14			
15			
16		2.89±0.23	2.76±0.19
17	3.27±0.14	3.12±0.10	3.15±0.10
18	3.86±0.14	3.74±0.09	2.87±0.09
19	4.07±0.25	3.76±0.19	3.34±0.18
20	3.84±0.21	3.58±0.15	4.03±0.12
21	3.80±0.22	3.98±0.17	3.41±0.14
22	3.39±0.29	3.73±0.25	3.12±0.19
23	4.59±0.19	4.29±0.12	3.93±0.11
24	4.65±0.26	4.23±0.21	3.70±0.17
25	3.89±0.21	3.47±0.16	4.21±0.14
26	3.57±0.26	3.65±0.23	3.54±0.16
27	4.94±0.29	4.59±0.22	3.72±0.16
28	4.02±0.23	4.11±0.20	3.78±0.14
29	4.54±0.66	3.73±0.82	3.38±0.31
30	4.54±0.37	5.13±0.43	3.73±0.25
31	4.13±0.36	5.83±0.47	3.91±0.20
32	4.34±0.25	4.21±0.27	3.58±0.15
33			
34			
35	5.15±0.41	3.94±0.17	5.49±0.36
36	4.81±0.28	4.16±0.22	3.89±0.16
37	4.08±0.16	3.87±0.12	3.58±0.10
38	3.82±0.19	3.70±0.15	3.68±0.11
39	4.08±0.16	3.87±0.12	3.58±0.10
40	3.58±0.21	3.38±0.16	3.57±0.12
41	3.77±0.22	3.46±0.19	3.24±0.13
42	4.10±0.17	3.19±0.13	3.43±0.10
43	3.49±0.21	3.30±0.17	3.09±0.13
44	3.85±0.14	3.82±0.11	3.84±0.10
45	3.39±0.21	3.23±0.17	3.33±0.14
46	4.16±0.23	3.79±0.14	3.26±0.14

Residue	R_2(W62Y-S^{ME}) (s^{-1})	R_2(W111G-S^{ME}) (s^{-1})	R_2(W123G-S^{ME}) (s^{-1})
47	3.25±0.21	3.18±0.16	3.30±0.14
48	3.61±0.21	3.57±0.15	3.33±0.14
49	3.51±0.19	4.30±0.16	2.98±0.13
50	3.66±0.18	3.52±0.14	3.55±0.12
51	4.26±0.24	4.05±0.16	3.67±0.14
52	3.85±0.14	3.82±0.11	3.36±0.15
53	4.66±0.29	4.35±0.22	4.09±0.14
54	4.88±0.34	6.07±0.38	3.80±0.25
55	5.30±0.37	5.51±0.30	4.83±0.24
56	6.16±0.45	4.77±0.45	5.77±0.29
57			
58	4.29±0.26	4.50±0.31	4.56±0.20
59	7.09±0.49	4.80±0.46	
60		7.58±0.74	8.15±0.44
61		6.00±0.70	5.10±0.29
62		6.40±0.86	6.30±0.48
63		7.14±2.57	5.81±1.21
64		5.56±1.93	5.06±0.97
65			
66	5.76±0.44	3.74±0.22	3.44±0.20
67	4.10±0.37	4.74±0.40	4.18±0.34
68	6.67±0.61	4.29±0.12	4.16±0.24
69	4.50±0.19	4.15±0.21	3.69±0.14
70			
71	4.06±0.24	3.11±0.17	2.02±0.12
72	3.85±0.26	3.72±0.28	3.90±0.21
73	3.69±0.27	3.48±0.32	3.19±0.21
74	4.26±0.30	3.85±0.19	4.05±0.23
75	3.50±0.31	3.74±0.38	3.50±0.24
76	4.26±0.34	4.38±0.34	3.33±0.24
77	3.76±0.32	3.21±0.37	2.33±0.09
78	3.74±0.42	3.64±0.53	3.25±0.22
79			
80	3.93±0.36	3.78±0.42	4.19±0.29
81	5.00±0.44	4.20±0.49	3.53±0.22
82	3.98±0.34	4.36±0.41	3.81±0.29
83	3.97±0.38	3.56±0.34	
84	3.63±0.37	3.31±0.44	3.28±0.25
85	4.03±0.32	3.78±0.30	3.84±0.21
86			
87	4.96±0.47	2.99±0.37	2.98±0.23
88	3.40±0.22		3.22±0.13
89	3.34±0.27	2.87±0.22	3.22±0.17
90	3.42±0.27	3.83±0.23	3.03±0.18
91	4.40±0.34	4.12±0.28	3.48±0.20
92	3.50±0.23	3.58±0.19	3.54±0.13
93			
94	2.34±0.13	3.54±0.20	2.43±0.09

Residue	R_2(W62Y-S^{ME}) (s^{-1})	R_2(W111G-S^{ME}) (s^{-1})	R_2(W123G-S^{ME}) (s^{-1})
95	2.99±0.26	3.00±0.29	2.89±0.20
96			
97			
98			
99	4.16±0.43	3.17±0.30	3.45±0.27
100	3.70±0.34	3.38±0.25	
101	3.11±0.29	2.77±0.21	
102	3.11±0.24	3.68±0.20	2.54±0.17
103	3.96±0.25	3.45±0.16	3.15±0.17
104	3.25±0.24	3.13±0.15	2.99±0.14
105	4.13±0.28	3.34±0.16	2.22±0.09
106	4.78±0.33	3.85±0.19	4.00±0.28
107	4.28±0.29	3.29±0.20	
108	5.62±0.39	3.45±0.20	4.58±0.27
109	5.70±0.63	3.33±0.44	4.43±0.60
110			6.10±0.37
111	6.37±0.51	3.16±0.28	4.47±0.23
112	5.62±0.65	3.91±0.17	4.97±0.28
113	6.15±0.59	3.72±0.26	4.92±0.30
114	4.56±0.53		4.05±0.33
115	4.54±0.45		3.79±0.29
116	3.84±0.26		3.26±0.24
117	3.75±0.26	3.11±0.17	2.48±0.22
118	3.50±0.27	2.85±0.17	2.43±0.14
119	3.38±0.23	2.76±0.18	2.55±0.11
120	3.55±0.23	2.96±0.16	2.85±0.11
121	3.84±0.26	3.23±0.26	3.24±0.17
122	3.76±0.30	2.92±0.22	2.92±0.23
123	4.59±0.19	3.50±0.16	2.99±0.14
124	3.90±0.49	3.15±0.32	2.13±0.12
125	3.73±0.35		
126	3.14±0.32	4.30±0.31	2.74±0.13
127	3.05±0.23	2.45±0.19	2.48±0.09
128	2.28±0.22	2.34±0.14	1.82±0.09
129	2.70±0.20	1.77±0.18	2.21±0.14

S6 Chapter 6

S6.1 Assignment of BLA-S^{ME}

AA	Residue	H^N (ppm)	N (ppm)	H$^\alpha$ (ppm)	C$_\alpha$ (ppm)	C$_\beta$ (ppm)	CO (ppm)
Q	2						
L	3	8.63	126.4	4.47	55.3	42.4	177.45
T	4	8.30	116.8	4.37	61.7	70.0	174.49
K	5	8.49	124.9	4.36	56.0	33.2	176.31
C	6	8.49	122.7	4.51	55.8	37.3	175.04
E	7	8.65	124.9	4.39	55.6	29.0	175.47
V	8	8.14	122.9	4.06	62.1	32.8	175.62
F	9	8.43	125.8	4.65	57.5	39.6	175.55
R	10	8.38	124.6	4.32	56.0	31.1	175.66
E	11	8.39	123.1	4.32	55.5	28.8	175.84
L	12	8.43	125.7	4.36	55.1	42.3	177.24
K	13	8.48	123.3	4.29	56.1	33.2	176.17
D	14	8.57	121.3	4.73	52.7	37.9	174.92
L	15	8.33	124.7	4.34	55.1	42.3	177.20
K	16	8.39	123.3	4.34	56.6	33.0	176.96
G	17	8.37	111.1	3.91	44.9		173.69
Y	18	8.21	121.4	4.58	58.0	38.9	176.40
G	19	8.47	112.3	3.92	45.2		174.36
G	20	7.90	109.5	3.96	45.0		173.77
V	21	8.04	119.8	4.20	62.0	33.0	176.08
S	22	8.46	121.1	4.54	57.9	63.8	174.01
L	23	8.36	126.7	4.60	52.8	41.8	175.00
P	24			4.31	62.7	31.8	176.84
E	25	8.43	121.6	4.21	56.3	28.6	175.71
W	26	7.95	121.5	4.70	56.8	29.2	175.94
V	27	7.89	123.1	4.06	62.1	32.9	175.42
C	28	8.40	125.2	4.55	55.6	37.3	175.38
T	29	8.51	118.4	4.43	61.8	69.7	174.43
T	30	8.09	117.1	4.34	61.7	69.8	173.95
F	31	8.24	124.0	4.61	57.6	39.8	175.31
H	32	8.55	122.3	4.76	55.0	29.4	174.24
T	33	8.36	116.8	4.40	61.5	70.0	174.51
S	34	8.50	118.9	4.51	58.4	64.0	174.82
G	35	8.42	111.7	3.95	45.2		173.67
Y	36	8.14	121.4	4.56	57.7	39.0	175.59
D	37	8.57	122.3	4.76	53.0	38.4	175.19
T	38	8.16	115.7	4.28	62.1	69.6	174.60
Q	39	8.32	123.1	4.34	55.8	29.5	175.51
A	40	8.25	126.4	4.30	52.4	19.1	177.39
I	41	8.17	122.0	4.15	61.3	38.6	176.22
V	42	8.31	126.4	4.11	61.8	32.7	175.90
Q	43	8.53	125.9	4.37	55.7	29.5	175.48
N	44	8.57	121.8	4.71	53.0	38.6	174.95

AA	Residue	H^N (ppm)	N (ppm)	H$^\alpha$ (ppm)	C$_\alpha$ (ppm)	C$_\beta$ (ppm)	CO (ppm)
N	45	8.57	120.8	4.73	53.2	38.5	175.12
D	46	8.55	120.6	4.78	53.2	38.2	175.29
S	47	8.32	117.2	4.49	58.7	63.9	174.94
T	48	8.21	116.7	4.34	61.9	69.7	174.52
E	49	8.26	123.4	4.28	55.8	28.9	175.56
Y	50	8.21	121.8	4.58	57.9	38.9	176.27
G	51	8.28	111.2	3.87	45.2		173.75
L	52	8.01	122.6	4.27	55.2	42.4	176.98
F	53	8.30	121.7	4.63	57.4	39.5	175.48
Q	54	8.34	123.4	4.35	55.6	29.6	175.53
I	55	8.27	123.5	4.12	61.4	38.9	175.77
N	56	8.58	123.8	4.76	53.0	38.9	174.86
N	57	8.44	120.6	4.72	52.9	38.9	175.06
K	58	8.26	122.6	4.24	56.3	32.9	176.27
I	59	8.10	122.7	4.13	61.3	38.6	175.80
W	60	8.31	126.6	4.71	57.0	29.7	175.85
C	61	8.28	123.6	4.51	55.2	37.3	174.85
K	62	8.54	125.5	4.19	56.6	32.7	176.23
D	63	8.43	120.6	4.72	53.0	38.0	174.79
D	64	8.42	120.9	4.74	53.0	37.9	174.84
Q	65	8.28	121.3	4.32	55.8	29.5	175.35
N	66	8.49	121.7	4.92	51.4	38.6	175.34
P	67			4.45	63.3	32.0	176.73
H	68	8.61	119.3	4.73	55.3	28.7	174.49
S	69	8.35	118.1	4.51	58.2	64.0	174.45
S	70	8.53	119.0	4.52	58.3	64.1	174.14
N	71	8.55	122.3	4.76	53.1	38.8	175.04
I	72	8.14	122.1	4.18	61.3	38.8	175.87
C	73	8.48	125.1	4.56	55.9	37.3	174.64
N	74	8.71	123.8	4.79	53.1	38.7	175.01
I	75	8.10	122.1	4.22	61.5	38.8	176.15
S	76	8.44	120.6	4.49	58.2	63.8	174.43
C	77	8.38	123.0	4.56	55.7	37.3	174.83
D	78	8.68	123.0	4.70	53.0	37.7	174.98
K	79	8.23	123.3	4.20	56.6	32.9	176.14
F	80	8.14	121.2	4.62	57.6	39.1	175.72
L	81	8.10	124.1	4.32	55.0	42.4	176.89
D	82	8.44	120.4	4.71	53.0	37.9	174.91
D	83			4.77	52.9	38.0	174.75
D	84	8.41	120.2	4.73	52.9	37.8	174.98
L	85	8.16	123.3	4.41	55.5	42.2	177.56
T	86	8.11	114.9	4.33	62.0	69.9	174.33
D	87	8.44	121.7	4.75	52.8	38.0	174.69
D	88	8.44	120.9	4.74	52.9	38.0	174.86
I	89	8.07	122.4	4.16	61.4	38.5	176.00
M	90	8.43	125.1	4.52	55.6	33.0	175.83

AA	Residue	H^N (ppm)	N (ppm)	H^α (ppm)	C_α (ppm)	C_β (ppm)	CO (ppm)
C	91	8.40	123.3	4.59	55.3	37.2	174.84
V	92	8.43	124.6	4.14	62.0	32.8	175.85
K	93	8.46	127.1	4.32	56.3	33.1	176.24
K	94	8.46	125.1	4.32	56.3	33.2	176.24
I	95	8.42	125.0	4.13	60.8	38.3	176.00
L	96	8.49	128.5	4.42	54.7	42.4	176.76
D	97	8.59	121.9	4.73	52.9	38.4	174.87
K	98	8.43	123.8	4.34	56.3	33.1	176.27
V	99	8.20	122.1	4.14	62.1	32.9	176.51
G	100	8.43	113.6	3.98	45.0		173.90
I	101	8.06	121.2	4.11	61.3	38.9	175.83
N	102	8.52	123.4	4.63	53.1	38.6	175.28
Y	103	8.11	123.1	4.34	59.3	38.6	175.85
W	104	8.06	122.5	4.60	57.5	29.3	176.18
L	105	7.86	123.8	4.23	55.0	42.4	176.86
A	106	8.03	124.8	4.17	52.4	19.0	177.50
H	107	8.35	118.6	4.64	55.2	28.9	174.24
K	108	8.40	124.3	4.32	56.1	33.1	176.05
A	109	8.47	127.2	4.32	52.4	19.2	177.56
L	110	8.34	123.4	4.35	55.7	42.5	177.30
C	111	8.50	122.2	4.60	55.4	37.3	175.17
S	112	8.64	120.1	4.46	58.5	63.7	174.43
E	113	8.34	123.7	4.36	55.8	29.1	175.85
K	114	8.37	123.7	4.32	56.1	32.9	176.46
L	115	8.30	124.5	4.33	55.4	42.4	177.13
D	116	8.60	121.0	4.60	53.0	38.1	175.17
Q	117	8.39	122.2	4.22	56.4	29.3	175.60
W	118	8.07	122.4	4.67	57.2	29.4	176.03
L	119	8.03	124.1	4.32	54.9	42.5	176.91
C	120	8.28	121.8	4.46	55.4	37.1	175.00
E	121	8.52	124.2	4.38	55.5	29.0	175.47
K	122	8.39	124.3	4.33	56.1	33.1	176.25
L	123						

S6.1 R_2 relaxation rates of WT-S^{ME*}, All-Ala-HLA** and BLA-S^{ME} in urea

Residue	R_2 (WT-S^{ME} in urea) (s^{-1})	R_2 (All-Ala-HLA in urea) (s^{-1})	R_2 (BLA-S^{ME} in urea) (s^{-1})
1		1.79±0.01	
2	1.98±0.10	2.54±0.02	
3	2.07±0.11	4.00±0.03	
4	2.11±0.11		
5	2.82±0.11	4.63±0.07	3.66±0.05
6	3.00±0.12	5.67±0.09	3.84±0.07
7	3.29±0.17	4.95±0.08	4.12±0.09
8	3.80±0.12		4.90±0.08
9	3.24±0.13	5.86±0.08	5.47±0.08
10			5.18±0.07
11	3.40±0.11	5.72±0.09	
12	3.65±0.17		5.47±0.08
13	3.91±0.22		5.04±0.08
14	4.01±0.24		5.20±0.08
15	4.21±0.30		4.89±0.06
16	3.20±0.20	5.54±0.09	4.81±0.03
17	3.88±0.20	4.17±0.04	4.63±0.06
18	3.94±0.23		4.33±0.05
19	4.09±0.26	3.70±0.04	4.07±0.05
20	4.39±0.24	3.41±0.03	
21	4.70±0.30	4.14±0.04	4.16±0.07
22	4.10±0.25	4.36±0.06	4.39±0.05
23	4.64±0.25	4.80±0.07	5.20±0.07
24	4.27±0.29		
25	4.16±0.25	6.71±0.21	
26	3.74±0.22	5.39±0.11	6.86±0.16
27	4.26±0.26		6.34±0.21
28	4.36±0.28	7.56±0.23	5.97±0.14
29	4.95±0.41	5.31±0.11	5.20±0.11
30			5.50±0.11
31	4.68±0.29		5.07±0.09
32	4.43±0.19	4.52±0.08	4.77±0.03
33	4.15±0.25	4.10±0.05	4.91±0.07
34	4.11±0.19	3.91±0.04	4.85±0.05
35	4.51±0.29	3.31±0.03	4.86±0.05
36	4.23±0.14		5.15±0.04
37			4.77±0.03
38	4.22±0.25	3.97±0.04	4.82±0.06
39	4.30±0.31	4.32±0.04	5.26±0.06
40	3.93±0.26	3.99±0.04	
41	4.14±0.33		4.91±0.06
42	3.77±0.16	4.41±0.07	5.70±0.07
43	3.60±0.27	4.39±0.07	5.33±0.07
44	4.04±0.25	4.13±0.05	4.95±0.06
45	3.74±0.25	3.99±0.04	4.92±0.06

Residue	R_2 (WT-S^{ME} in urea) (s^{-1})	R_2 (All-Ala-HLA in urea) (s^{-1})	R_2 (BLA-S^{ME} in urea) (s^{-1})
46	3.87±0.24	4.08±0.04	5.06±0.06
47	3.77±0.26	4.04±0.04	
48	4.11±0.21		5.12±0.06
49	3.43±0.20		5.33±0.04
50	3.81±0.19	4.47±0.05	5.36±0.06
51	4.08±0.27	4.00±0.04	
52	3.88±0.28	4.43±0.04	5.35±0.07
53	3.84±0.19	4.41±0.06	5.31±0.07
54	4.35±0.35	4.41±0.06	
55	5.22±0.31		5.36±0.07
56	5.13±0.51		5.34±0.09
57	4.47±0.37	4.29±0.05	
58	5.36±0.61		6.04±0.11
59	5.92±0.73	4.59±0.06	6.37±0.16
60	5.41±0.42	4.36±0.06	6.87±0.17
61	6.05±0.66	3.97±0.04	5.36±0.07
62	6.78±0.72	3.99±0.05	5.97±0.17
63	6.92±0.99	3.92±0.04	5.15±0.04
64	6.34±0.78	3.87±0.04	4.78±0.04
65	4.96±0.49	3.61±0.04	4.71±0.07
66	5.11±0.49		4.72±0.07
67	4.64±0.40		
68	4.90±0.25		4.70±0.07
69	4.24±0.14	4.02±0.05	4.25±0.05
70		3.97±0.05	4.32±0.05
71	4.36±0.36	3.87±0.04	
72	4.24±0.25	3.86±0.04	4.34±0.07
73	4.02±0.29	4.32±0.04	5.09±0.05
74	4.19±0.25	4.01±0.04	4.10±0.09
75	4.16±0.34		4.40±0.08
76	4.37±0.31		5.15±0.04
77	4.14±0.33	3.91±0.04	4.84±0.10
78	4.36±0.31	4.12±0.04	5.27±0.15
79		4.75±0.04	
80	4.13±0.20		5.15±0.04
81	4.67±0.31	4.57±0.05	5.84±0.17
82	4.44±0.28		5.74±0.07
83	4.43±0.27		
84	3.67±0.09		5.87±0.09
85	4.12±0.22		4.61±0.06
86		4.67±0.06	4.46±0.06
87	3.77±0.23	4.40±0.05	4.85±0.04
88	3.85±0.20		4.78±0.04
89	3.53±0.16	4.17±0.05	
90	3.47±0.17	4.92±0.08	5.22±0.06
91	4.25±0.24	4.08±0.05	4.81±0.03
92	4.01±0.23	4.18±0.05	5.54±0.16

Residue	R_2 (WT-S^{ME} in urea) (s^{-1})	R_2 (All-Ala-HLA in urea) (s^{-1})	R_2 (BLA-S^{ME} in urea) (s^{-1})
93	4.08±0.25	4.52±0.07	6.80±0.09
94	3.57±0.19	4.96±0.10	
95	3.50±0.19	5.19±0.12	
96		5.50±0.13	7.26±0.17
97		4.98±0.08	5.63±0.11
98	4.21±0.31	4.83±0.07	5.36±0.09
99	4.03±0.31	5.31±0.09	5.47±0.09
100	4.07±0.26	4.67±0.07	5.28±0.10
101	3.02±0.12	5.43±0.08	5.82±0.10
102	3.33±0.21		6.11±0.11
103	3.55±0.25	5.94±0.08	6.50±0.13
104		6.65±0.11	
105	3.37±0.21	6.67±0.13	6.88±0.18
106	4.09±0.32	5.82±0.08	6.01±0.13
107	4.53±0.28	5.05±0.08	5.80±0.12
108	5.26±0.42		4.48±0.07
109	5.76±0.69	4.48±0.06	
110	5.76±0.52		
111		4.31±0.06	4.90±0.10
112	5.72±0.61	4.05±0.05	4.59±0.11
113	3.43±0.27	4.17±0.05	5.00±0.08
114	4.81±0.42	4.60±0.07	4.81±0.10
115	4.60±0.40	4.45±0.06	4.77±0.08
116	4.25±0.32	4.69±0.07	
117	3.53±0.22	4.33±0.06	4.83±0.10
118	3.68±0.23	4.17±0.05	4.84±0.04
119		4.71±0.07	4.59±0.13
120	4.17±0.28	4.01±0.04	3.98±0.08
121	4.04±0.29	3.41±0.04	3.58±0.08
122	4.10±0.27		3.46±0.06
123	5.03±0.25	2.22±0.02	
124	4.33±0.42		
125	3.78±0.15		
126	3.27±0.25		
127	3.06±0.15		
128	2.72±0.13		
129	2.25±0.11		

* Schwalbe et al., 1997

** kindly provided from Christina Redfield

Curriculum Vitae   Julia Wirmer-Bartoschek

Personal data   Born: 8/7/1975 in Frankfurt, Germany; married to Dr. Stefan Bartoschek; nationality: german

Address   Center for Biological Magnetic Resonance
                        Goethe University Frankfurt
                        Marie-Curie-Str. 11   Phone: ++49 69 7982 9934
                        D-60439 Frankfurt   Fax: ++49 69 7982 9515
                        Germany   Email: wirmer@nmr.uni-frankfurt.de

Education

Aug. 2000 - present   **Massachussetts Institute of Technology**   **Cambridge, USA**
                        PhD
Oct. 1995 – Jul. 2000   **Goethe University Frankfurt**   **Frankfurt, Germany**
                        Diploma in Chemistry (grade 1.2)
Aug. 1986 – Jun. 1995   **Schule am Ried**   **Frankfurt, Germany**
                        School, finished with "Abitur" (grade 1.1)

Research Experience

Oct. 2001 – present   **Goethe University Frankfurt**   **Frankfurt, Germany**
                        Continuation of PhD thesis in the group of Professor Dr. H. Schwalbe
Aug. 2000 – Sep. 2001   **Massachussetts Institute of Technology**   **Cambridge, USA**
                        PhD thesis in the group of Professor Dr. H. Schwalbe

                        Title: "Investigation of the kinetics of protein folding and the ensemble of conformations in non-native states of proteins by liquid NMR spectroscopy."
Nov. 1999 – Jul. 2000   **Massachussetts Institute of Technology**   **Cambridge, USA**
                        Diploma thesis in the group of Professor Dr. H. Schwalbe

                        Title: "NMR spectroscopic investigations of protein folding and synthesis of caged-calcium compounds." (grade 1.0)

Teaching Experience

May & Oct. 2002,   **Goethe University Frankfurt**   **Frankfurt, Germany**
July & Nov. 2003   Teaching assistant for the laboratory course "Organic chemistry for chemists", supervision of practical work (20 students).
Dec. 2001 – Jan. 2002, **Goethe University Frankfurt**   **Frankfurt, Germany**
Jan. 2003   Teaching assistant for the course "Biophysical chemistry for biochemists", Responsible for the part "Kinetics": lectures and supervision of practical work (35 students).

Awards

2002	R.-R.-Ernst Prize of the German Chemical Society (GdCh)
Jun. 2001 - May 2003	Fellowship from the Fonds of the Chemical Industry
Jan. 2001 – Jun. 2001	Rosenblith stipend from the Massachussetts Institute of Technology
June 1995	Prize for best "Abitur" in Chemistry