A NMR-based Study on the Characteristics of Nonnative Hen and Human Lysozyme and their Disease-Related Mutants

Dissertation

zur Erlangung des Doktorgrades

der Naturwissenschaften

Vorgelegt beim Fachbereich Biochemie, Chemie und Pharmazie

der Goethe Universität

Frankfurt am Main

von

Friederike Sziegat

(geb. Heinicke)

geboren in Heidelberg

Frankfurt am Main

2011

GOETHE

UNIVERSITÄT

FRANKFURT AM MAIN

Vom Fachbereich Biochemie, Chemie und Pharmazie der Goethe Universität Frankfurt am Main als Dissertation angenommen.

Dekan: Prof. Dr. Thomas Prisner

Gutachter: Prof. Dr. Harald Schwalbe

Prof. Dr. Joachim Engels

Datum der Disputation: 2012

„So eine Arbeit wird eigentlich nie fertig, man muss sie für fertig erklären, wenn man nach Zeit und Umständen das Möglichste getan hat"

(Johann Wolfgang von Goethe, Italienische Reise, 16.März 1787)

Table of Contents

Abbreviations

APS	ammonium persulfate
CM	carboxymethyl
CSA	hemical shift anisotropy
d	delay between NMR experiments
D_2O	deuterium oxide
ddH_2O	deionized destilled water
DEAE	diethylaminoethyl-
DNA	desoxyribonucleic acid
DTT	dithiothreitol
dMTSL	1-acetyl-2,2,5,5-tetramethyl-η^3-pyrroline-3-methylmethane-thiosulfonate
DOSY	diffusion ordered spectroscopy
E.coli	*Escherichia coli*
EDTA	ethylene diamine tetracetic acid
Fig.	figure
FPLC	fast protein liquid chromatography
GdnHCl	guanidinium hydrochloride
η	$h/2\pi$, $h= Planck$ constant
hewl	hen egg white lysoyzme
hLys	all-ala-variant of human lysozyme
D67H	D67H mutant of hLys-all-ala
I56T	I56T mutant of hLys-all-ala
HSQC	heteronuclear single quantum coherence
Hz	Hertz
IPTG	isopropyl β-D-thiogalactopyranoside
kDa	kilo dalton
LB	Luria Broth
M.	marker
M9	Minimal media for isotopic labelling
MD	molecular dynamics
MTSSL	1-oxyl-2,2,5,5-tetramethyl-η^3-pyrrolin-3-methyl methane thiosulfonate
MWCO	molecular weight cutoff

ns	number of scans
NMR	nuclear magnetic resonance
OD	optical density
PALES	Prediction of Alignment from structure
PCR	Polymerase Chain Reaction
Pfu	Pyrococcus furiosus
ppm	parts per million
PRE	paramagnetic relaxation enhancement
RDC	residual dipolar coupling
RNA	ribonucleic acid
rpm	rounds per minute
s	seconds
SAXS	small angle x-ray scattering
SDS-PAGE	sodium dodecylsulfat polyacrylamide gel electrophoresis
Tab.	table
TCA	trichloroacetic acid
TEMED	N,N,N',N'-tetramethylenediamine
TFA	trifluoroacetic acid
Tris	tris-(hydroxymethyl-)aminoethane
Triton X-100	Octylphenoxypolyethoxyethanol
W28G^{Ala}	W28G mutant of hewl
W62G^{Ala}	W62G mutant of hewl
W108G^{Ala}	W108G mutant of hewl
W111G^{Ala}	W111G mutant of hewl
W123G^{Ala}	W123G mutant of hewl
WT^{Ala}	all-ala wild type of hewl
w/v	weight per volume

List of Figures

List of Tables

1   Introduction

1.1   Protein misfolding, aggregation and non-native states

Proteins play one of the crucial roles in mediating biological and biochemical regulating processes in living cells and organisms. Despite the ample knowledge of ribosomal synthesis and analysis of interaction mechanisms fundamental elucidation of protein folding and misfolding is still challenging. The complexity of folding and misfolding reactions arises from the fact that one amino acid sequence codes for a single compact three-dimensional structure. Secondary and tertiary interactions are formed in less than a few milliseconds. Levinthal pointed out fundamentally that these time scales of folding imply that protein folding does not follow a random search of allowed conformations during synthesis (*1*).

In vivo correct folding is regulated by a multitude of cellular quality-control mechanisms (*2, 3*). Intrinsic and extrinsic factor influence protein folding, target proteins to different locations and play a crucial role in stabilizing possible unfolded or misfolded states. Dysregulation of these quality-control mechanisms is one of the key factors for protein folding diseases.

Unfolded states of proteins have been in the focus of biophysical investigations due to their importance for understanding protein folding diseases, in particular human neurodegenerative diseases (*3-6*). Most of these diseases are summarized under the global term "amyloidosis". The term was coined due to the observation that accumulated aggregates can be stained with Congo red in a similar manner as amylose in starch. Later, the terms "amyloid" and "amyloid fibrils" have been deduced. Formation of amyloid fibrils is a wide spread phenomenon being uncovered as a crucial pathological product during protein misfolding (*7, 8*). Amyloidosis like *Alzheimer's* disease, spongiform encephalopathies, systemic amyloidosis, lysozyme amyloidosis and type II diabetes for example are all associated with extracellular abnormal deposition of fibrillar aggregates in various organs like liver, kidney, heart and brain (*4, 5, 9-11*). Other protein folding diseases like Parkinson's and Huntington's disease (*10, 12*) are intracellular and therefore do not fulfill the strict definition of amyloidosis which is reduced to extracellular aggregates (*2*).

After protein synthesis at the ribosome there are several different states of structures that can be formed by the polypeptide chain depending on external and internal influences (Figure 1). There is only one path ending up with the correctly folded state but there are many other paths leading to aggregates. One form of aggregation is the formation of amyloid fibrils, highly organized structures with very similar properties and appearance in amyloid diseases mentioned above (*11*). The conversion of a soluble form of the protein to the aggregated state typically involves transient formation of unfolded or partially folded states.
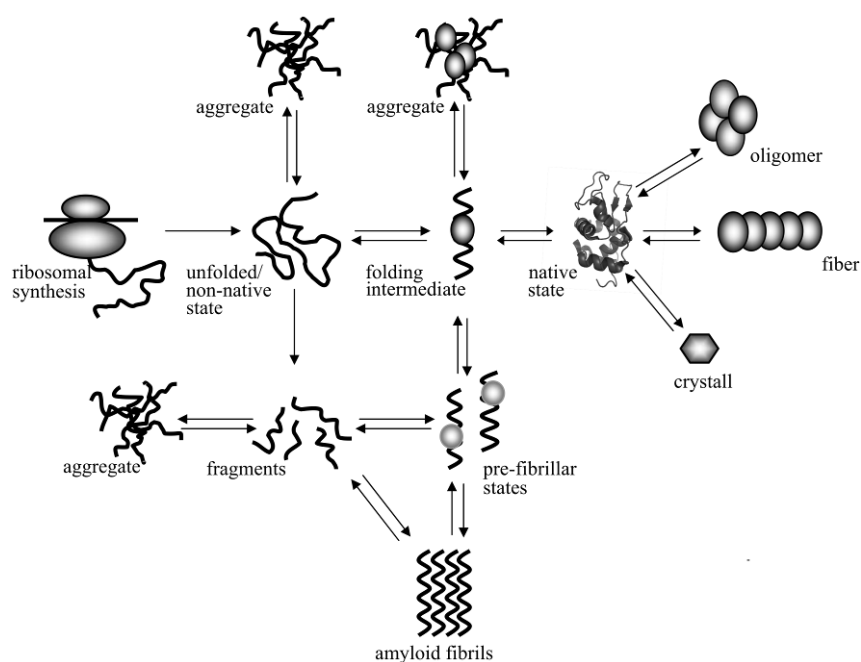
Figure 1: Protein states after synthesis at the ribosome. There is only one path ending up with a well-folded structure, but there are many other paths which lead to aggregates, fragments, pre-fibrillar states or amyloid fibrils. Modified after Dobson *et al.*, 2003.

The biological function of proteins has evolved during folding processes. Proteins in their native state have overall formed tertiary interactions whereas intrinsically unfolded proteins (IUPs) can often be described in conformational terms as premolten globule states. These states characteristically have by residual secondary structure including partially disordered regions but no truly random coil conformation (*13*). Some disease-associated proteins adopt

well-folded structures in their native states, while others belong to the class of intrinsically unstructured proteins, so the pathological origin is not only coded in unfolded regions (*14, 15*).

As described before in many protein folding diseases, single point mutations, often familial inherited, are linked to specific disease forms but the structural code for the onset of the disease is still unclear (*16, 17*). In the literature, single point mutations do not significantly change the structure of the native state (*18-21*) Their impact on the structural and dynamic characteristics of the unfolded or intrinsically unstructured state ensemble remains a key unsolved question in structural biology (*22*).

"In this communication I wish to draw a substance present in the tissues and secretions of the body, which is capable of rabidly dissolving certain bacteria. As this substance has properties akin to those of ferments, I have called it a Lysozyme…"; in 1922 Alexander Flemming invented the name and discovered the function of the lysozyme protein by accident on a patient suffering from acute coryza (*23*). From there on, lysozyme was and still is counted among one of the most studied proteins. It is therefore used as model system for structural and dynamical analysis of folding and unfolding processes. In this thesis, studies on lysozyme are extensively expanded including new insights in the analysis of single-point mutations of hen and human lysozyme in their unfolded states. In the focus are single point mutants which are chosen due to introducing significant differences in their hydrophobic behaviour. In hen egg white lysozyme hydrophobic tryptophans were replaced by unpolar glycines with small side chains showing an effect on residual secondary structure elements due to the deletion of hydrophobic side chain interactions. For human lysozyme, two mutants with high aggregation propensity and associated with diseases form of the protein, as reported in the literature (*18*), were investigated.

In the broad variety of different biophysical methods, heteronuclear multidimensional NMR spectroscopy in solution stands out as an important method to determine conformational and dynamic properties within protein ensembles representing the unfolded state of a protein (*24*) at atomic resolution.

To identify possible nucleation sites both for productive folding as well as unproductive misfolding, the random-coil behaviour of a polypeptide chain has to be defined to provide a

baseline for detection of non-random conformational and dynamical characteristics of regions in unstructured polypeptide chains. Using conformational models to describe the unfolded state, spectroscopic fingerprints can be predicted for random-coil conformation and dynamics and regions of polypeptide chains can be delineated that deviate from random-coil behaviour. Following this approach, random-coil expectation values for NMR parameters including homonuclear NOEs, relaxation rates, scalar coupling constants and residual dipolar couplings have been defined and applied with considerable success for the characterization of unfolded polypeptide chains. In developing such models for the prediction of random-coil behaviour, it is particularly interesting to note that differences in the time-scales of averaging are reflected in the prediction of NMR parameters and directly observable in the experimental data.

1.2   Hen and Human Lysozyme

1.2.1   Hen Egg White Lysozyme

In its native state, hen lysozyme consists of two domains, the α-domain (residue 1-35 and 85-129) and the β-domain (residue 36-84). It contains four disulfide bridges formed by eight cysteines (C6-C127, C30-C115, C64-C80 and C76-C94). An unusual high number of six tryptophans (W28, W62/63, W108, W111 and W123) can be found in hen lysozyme as shown in figure 2.
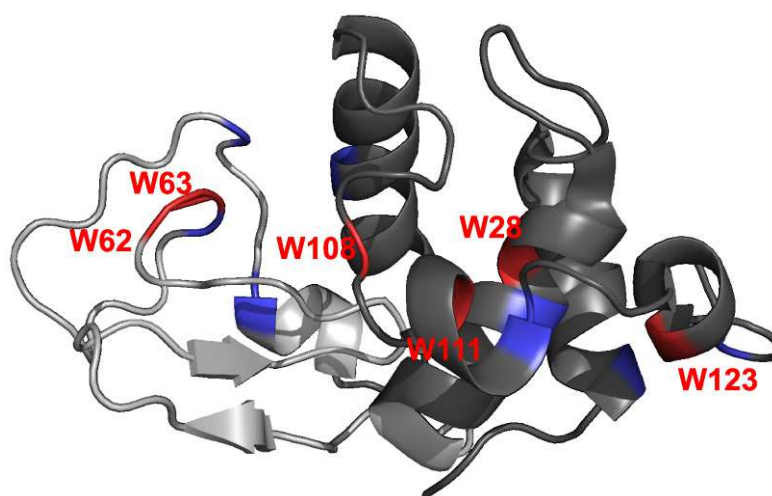
Figure 2: Hen egg white lysozyme in the native state (pdb 1E8L(*25*)); light grey: β-domain (residue 36-84); dark grey: α-domain (residue 1-35 and 85-129); red: tryptophans W28, W62/63, W108, W111 and W123; blue: positions of cysteine residues.

Folding of hen egg white lysozyme is rather complex following multiple, parallel folding pathways (*26*). In its disulfide-reduced state, the protein is unfolded even in the absence of denaturants (*27-29*). Residual secondary structure as well as the presence of long-range interactions persisting on the subnanosecond time scale were previously identified by an analysis of chemical shifts and R_2 relaxation rates in combination with single point mutations (*28, 29*). Native as well as non-native long-range interactions are present in the unfolded state of hen egg white lysozyme. Previous studies have been extended in this thesis where an all cysteine-to-alanine mutant of hen lysozyme (WT^Ala) was prepared. Chemical shift deviations

as well as coupling constants for both, backbone and side chain angles are presented. Also relaxation studies and spectral density functions have been analyzed and allow the identification of residual structure persisting on different timescales. This is supported by extensive diffusion measurements for the wildtype and all mutants and small angle x-ray scattering (SAXS). All data lead to conformational characteristics of unfolded states of proteins and will be discussed in detail in chapter 3.

1.2.2   Human Lysozyme

The structure of human lysozyme possesses also two domains: an α-domain (residues 1-38 and 86-130) with four α-helices and a β-domain (residues 39-85) with a triple-stranded β-sheet and a 3_{10} helix (figure 3). The protein contains four disulfide bridges at C6-C128, C30-C116, C65-C81 and C77-C95. Also a high number of hydrophobic amino acids can be found, in particular W28, W34, Y63, W64, W112 and Y124. The protein is highly homologous to hen lysozyme and is also used as model system for folding and unfolding processes.
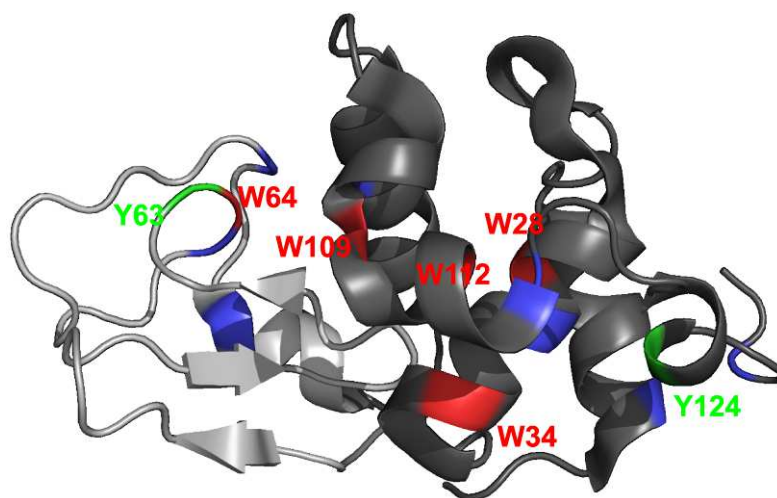
Figure 3: Human lysozyme in the native state (pdb 2ZIJ, Shoyama *et al.*, 2009). light grey: β-domain (residue 39-85); dark grey: α-domain (residue 1-38 and 86-130); red: tryptophans W28, W34, W64, W109, W112; green: Y63, Y124; blue: positions of cysteine residues.

One of the approximately 20 amyloid diseases known to date is associated with the deposition of amyloid plaques from single-point mutants (I56T, F57I, I59T, W64R or D67H) of human lysozyme (*30*) in various organs. Given the remarkable high stability of this protein family, the discovery of protein folding diseases associated with hLys came as a surprise. Initial studies revealed that the native state structures of wildtype and amyloidogenic mutants of human lysozyme are highly similar (*18*). Characterization of the thermal unfolding at pH 1.2 using CD, DSC and NMR spectroscopy showed that unfolding proceeds via loss of tertiary structure, resulting in a pre-molten globule state, which unfolds further, as the temperature increases.(*31*) The β-domain and the adjacent helix 3 (residues 90-100) unfold first. The same region is destabilized in the mutants as shown by hydrogen exchange (*32, 33*) leading to a lower thermal stability.(*18*) Also in the unfolded mutants a decreased stability is visible, as it is described in the results and discussion section (3.1). Such decrease in stability suggests that the relative stabilities of unfolded states are important in the process of amyloid formation. In fact, amyloids are formed fastest if the protein is incubated at very low pH (1.2) at temperatures near the midpoint of thermal unfolding (*31*), thus at a temperature where folded as well as molten globule and pre-molten globule states, summarized here as unfolded states, are populated. In contrast to the folded state of hLys, little is known about residual structure and dynamics in its non-native unfolded state. For a better understanding of the mechanism leading to the formation of amyloid diseases unfolded states are therefore extensively characterized here.

1.2.3   Human Lysozyme versus Hen Lysozyme

Human lysozyme and hen lysozyme share high sequence homology (~70%) and identity
(~60%). They differ in 51 amino acids and human lysozyme has one insertion at position 48
in the sequence. The native state structures of human lysozyme and hewl are very similar
with a backbone RMSD of 0.74 Å (*34*). An alignment of the two pdb structures of both
proteins illustrates the similarity in a concrete manner (figure 4).
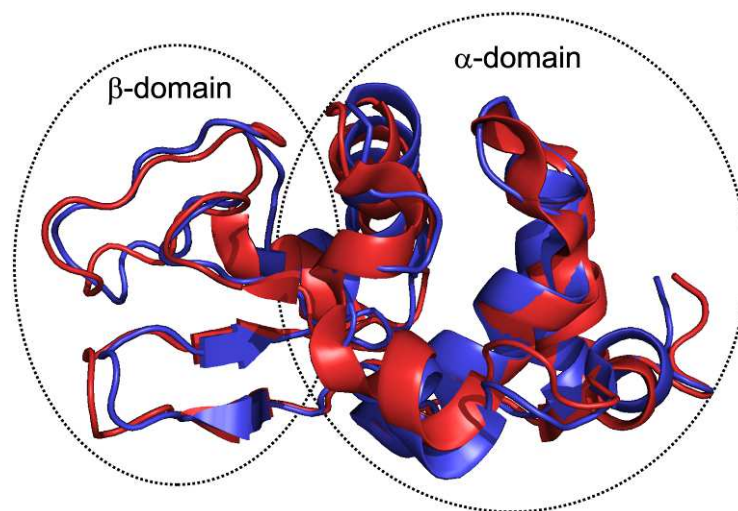
Figure 4: Alignment of the pdb structures of hen egg white
lysozyme 1E8L (*25*) (blue) and human lysozyme 2ZIJ (*35*) (red).
Both proteins share ~70% sequence homology and ~60% identity.

Despite the high homology, main differences in the amino acid sequence are predominantly
found between residues 70-80, 85-92 and 98-110, located at the interface between the α- and
the β-domain. As mentioned before both proteins have a remarkable high number of
hydrophobic amino acids, in particular six tryptophans in hen lysozyme, Trp28, Trp62/63,
Trp108, Trp111 and Trp123 and five in the human variant, Trp28, Trp34,, Trp64, Trp109,
Trp112. There are also two tyrosines in human lysozyme replacing two tryptophans in
significant regions in hewl, Tyr63 and Tyr124.

In their native state both proteins have different folding kinetics as it was shown by Radford *et al.*, 1992 for hen lysozyme and by Hooke *et al.*, 1994 for human lysozyme. The latter folds four times faster than hen lysozyme and the refolding is not fully cooperative, hence the α-domain is formed before the β-domain in both proteins (*36*). It could be shown partial previously and is shown here extensively that hen and human lysozyme have also main differences in their unfolded states concerning residual structure and dynamics (*37*).

In order to highlight the remarkable influence of the mutants on structure and dynamics in unfolded states, here nonnative variants of hen lysozyme and human lysozyme are compared. Surprisingly in the all-ala variants of the human protein an intrinsic higher aggregation propensity could be highlighted compared to hen lysozyme. This suggests that amyloidogenic mutants modulate structure and dynamics of the pre-molten globule state of human lysozyme much more than mutants in hen egg white lysozyme. A detailed structural and dynamical analysis of both proteins is described in the results and discussion section (chapter 3.1)

1.3   NMR spectroscopy

Several spectroscopical methods have been established over the years to analyse structural and dynamical features of nonnative protein states. To mention a few: FRET (Fluorescence Resonance Energy Transfer), CD spectroscopy (CD = Circular dichroism), SAXS (small angle x-ray scattering) and UV spectroscopy (UV = Ultraviolet). Furthermore Raman spectroscopy and IR sepectroscopy (IR = Infrarot) are generally used techniques for protein analysis. But with these methods it is still challenging to observe biological systems at atomic resolution. This is one of the key advantages of nuclear magnetic resonance and why it is commonly used and developed for the characterization of different protein states.

1.3.1   NMR spectroscopy under conformational averaging

Since a protein adopts a variety of different conformations during folding or misfolding, nonnative states consist of a broad ensemble of interconverting conformers. Those states interconvert on different rates and therefore NMR data represent the persistence of dynamics on different timescales. A general procedure to analyze unfolded states is the use of the so called random coil model. It assumes a polypeptide chain with no nonlocal interactions; different amino acids adopt different conformations and angle distributions in their local environment. Experimental NMR parameters of nonnative states compared to random coil values result in regions where residual α-helical or β-sheet structure is present and therefore a degree of structuredness derives. With this model it seems easy to uncover the structural and dynamical features of nonnative states but all data have to be analyzed under the large influence of conformational averaging of the unfolded ensemble.

NMR of proteins in their folded state is more or less unambiguous showing features from one individual structure. As mentioned before one challenging aspect of the characterization of nonnative states is still the existence of an ensemble of interconverting structures averaging NMR parameters over time and conformational space. But there are several NMR-based strategies for data interpretation to overcome this problem. It includes real-time NMR for kinetic information and experiments with multidimensional resolution. Suitable parameters

for the analysis are ^{1}H,^{15}N chemical shifts, heteronuclear relaxation rates (R_1, R_2, $R_{1\rho}$), relaxation dispersion, H/D exchange, Photo-CIDNP, diffusion measurements, J-couplings, residual dipolar couplings and paramagnetic relaxation enhancement (see table 1).

NMR parameter	conformational influence
chemical shit δ [ppm]	torsion angles ψ, φ, ω, χ_1
^{n}J coupling [Hz]	torsion angles from Karplus curve
heteronuclear NOE [a. U.]	distance, dependence on correlation time, motional properties
heteronuclear relaxation [Hz]	motional properties, dependance on τ_c, S^2, τ_e
residual dipolar coupling [Hz]	surface, dynamics, hydrogen bond persistence
H/D exchange	H^N exchange
diffusion	radius of hydration R_h
Photo-CIDNP	accessibilty (Trp, Tyr, His)
paramagnetic relaxation enhancement	long-range distance

Table 1: NMR parameters for the analysis of nonnative states of proteins (modified from (*24*)).

In chemical shifts (δ), relaxation rates, spectral density functions, angular distributions from scalar coupling constants (J), residual dipolar couplings (RDCs) and diffusion measurements it is still challenging to incorporate this averaging. To give a more detailed explanation, in folded states, the dynamics of a given N-H or C-H bond can be deduced from analysis of relaxation data using the Lipari-Szabo model (*38*), such approach is not feasible for unfolded proteins where overall and internal motions are on similar timescales and can therefore not be separated. However, it was previously shown that random-coil behaviour can be described using a simple segmental motion model. The model assumes that the relaxation properties of a given amide are governed by segmental motions of parts of the polypeptide chain, independent from the overall tumbling and, importantly in this context, independent from amino acid specific features (*27-29, 39*). With this model, the baseline of the relaxation rates is well described, however, the interpretation of relaxation rates deviating significantly from this baseline is still difficult and such interpretation is an important part of this thesis.

While ^{15}N relaxation rates are sensitive to motions faster than the overall tumbling (τ_c) time (subnanoseconds), structure persisting on supra-τ_c timescales can be monitored inter alia by chemical shifts, scalar coupling constants, and residual dipolar couplings. Thus, it has been shown in the past that in contrast to the predictions of the relaxation rates, predictions for chemical shifts, J couplings as well as RDCs have to take into account amino-acid specific sampling of the protein torsion space including both preferences in the backbone and side chain torsional distributions.

Typically, reference chemical shift values expected for a random-coil are derived from data for small unstructured peptides (*40, 41*). J-coupling constants expected for an unfolded protein can be calculated using the random-coil database which predicts angular preferences using a library of known torsions in coil regions of proteins (*42*), although random-coil values for unstructured peptides are also available (*42*). As mentioned before, deviations from these predicted chemical shifts or coupling constants are interpreted in terms of residual α-helical or β-sheet structure, for which reference data are also known. RDCs arise from dipolar interactions of two nuclei e.g. N and H. In isotropic solution, RDCs are averaged out, but can be measured in aligning media. The absolute values of the experimental RDCs thus depend on the structure of the protein as well as on the alignment of the protein in the anisotropic medium.

Different approaches exist for the prediction of RDCs in unfolded protein states. The random-flight chain algorithm (*43*) treats the polypeptide chain as a homopolymer predicting negative values which have a smaller absolute value at the termini of the polymer than in the middle, describing the shape of the overall base line well (*43*). The algorithm fails, however, to describe the modulation along the polypeptide chain observed experimentally (*44*). Similar to approaches for J coupling prediction (*42*), the flexible-meccano program constructs statistical ensembles of coil conformations and average sequence specific values are calculated over these ensembles resulting in a residue-specific modulation of the predicted RDCs (*45*), while the overall shape is similar to the one predicted by the random flight chain algorithm. The detailed interpretation of the deviations is still difficult, but deviating, mainly positive RDC values are interpreted in terms of residual structure and correlations with conformational ensembles by molecular dynamics simulations can be obtained (*46, 47*).

All these models explain the limited ability of interpreting NMR data on different timescales under the aspect of conformational averaging. Using the above mentioned models, residual structure and dynamics have been identified in a number of chemically or biochemically denatured proteins including SH3 of drkN (*48, 49*), ACBP (*50*), lysozyme (*29, 37*), α-lactalbumin (*51*) and also IUPs such as α-synuclein (*45, 52, 53*) and tau (*54, 55*). Comparison of the results for different proteins reveals that non-native states of proteins can vary widely in terms of the sampled conformations and dynamics. Only few systems exist where both relaxation rates and RDC data have been measured (ACBP (*50, 56*), Snase (*57, 58*), Apomyoglobin (*59, 60*), ubiquitin (*51, 61*)). An extensive study of dynamics on different timescales and mutational effects on structure and dynamics of unfolded hen and human lysozyme is presented in the results and discussion chapter (3.1).

One typical application on unfolded states is a two dimensional NMR experiment, the ^{1}H-^{15}N HSQC. As it is resolved in the ^{1}H and ^{15}N dimension a direct analysis of the backbone N-H resonances is possible. Also the degree of unfolding becomes visible through narrow chemical shift dispersion. Figure 5 shows hen egg white lysozyme in the native (a) and nonnative state (b). While the dispersion in the ^{1}H dimension ranges between 6 ppm and 11 ppm, in the unfolded state the ^{1}H dispersion is limited from 7.5 ppm to 8.5 ppm.
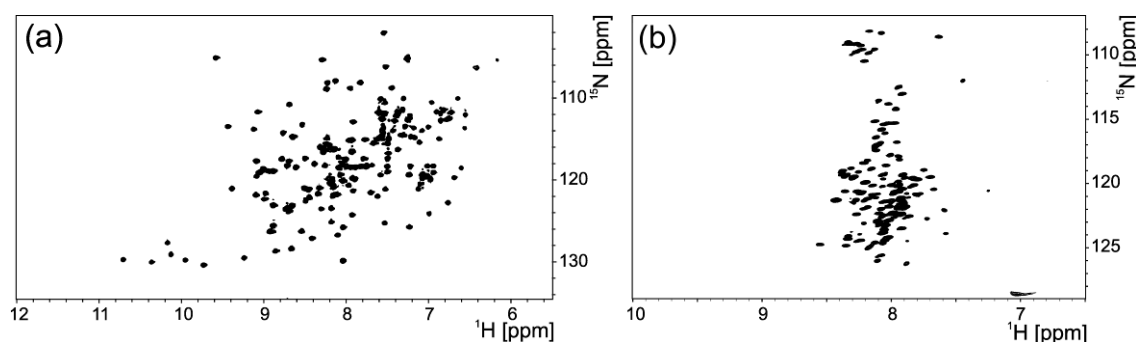
Figure 5: ^{1}H-^{15}N HSQC spectra of native (a) and nonnative hen egg white lysozyme (b); spectra were measured at a protein concentration of 300µM, 293K, pH 2, 600MHz (spectrum of native hewl kindly provided by R. Silvers, Goethe University, Frankfurt/Main, Germany)).

To circumvent the problem of small chemical shift dispersion in commonly used 2D-spectra, three and four dimensional NMR experiments have improved the resolution of backbone ^{1}H$_N$, ^{1}H$^\alpha$, ^{1}H$^\beta$, ^{15}N, ^{13}C$^\alpha$ and ^{13}C$^\beta$ nuclei. A differentiation between backbone and side chain proton and carbon resonances leads to a complete assignment of the protein even in unfolded states. In the past a large number of backbone and side chain experiments through scalar couplings have appeared. To mention a few: In ^{15}N labelled proteins HNN experiments (*62*) correlate ^{15}N$_{i-1}$ and ^{15}N$_{i+1}$ with ^{1}H$_{Ni}$ and ^{15}N$_i$ nuclei and therefore enables sequential assignment of the backbone amides. In the case of ^{13}C, ^{15}N labelled proteins HNCO (*63*), HNCACB (*63*), HN(CA)CO (*64*), HN(CO)HB (*65*) and CBCA(CO)NH (*66*) experiments are valuable tools for the assignment because they connect all ^{15}N backbone resonances with the ^{13}C resonances of their own and the preceding amino acid. Beside the amide resonances also C$^{\alpha i}$, C$^{\alpha(i-1)}$, C$^{\beta i}$, C$^{\beta(i-1)}$, CO$_i$ and CO$_{i-1}$ can be sequentially assigned. A new HN(COCA)HAHB experiment used in this thesis for stereospecific assignment of side chain H$^\beta$ protons was

recently developed by Hähnke *et al.*, 2009. It enables the determination of the $^3J(H^\alpha,H^\beta)$ coupling, one of the best parameterized coupling constants for small unstructured peptides defining the side chain torsion angle χ_1 (*67*).

1.3.2 Chemical shift

A series of parameters exists for unfolded proteins which can be investigated by NMR spectroscopy (see tab.1); amongst others there is the chemical shift parameter, typically analyzed for ^1H, ^{13}C and ^{15}N nuclei. Chemical shifts depend on the chemical environment of the observed nuclei containing information about secondary structure elements and tertiary contacts (*68, 69*), mainly influenced by noncovalent interactions like hydrogen bonding or aromatic stacking. In general nonnative states show smaller chemical shift dispersion than proteins in their native state. This is due to conformational averaging of the polypeptide chain with random coil like resonance behaviour. Secondary chemical shifts $\Delta\delta$ result from the difference of experimental values to random coil values from the literature (*68*) and give information about the content of residual structure (see equ. 1.0, δ_{exp}=experimental, δ_{rc}= random coil chemical shift).

$$\Delta\delta = \delta_{exp} - \delta_{rc} \tag{1.0}$$

It could be shown that a random coil polypeptide chain and a mixture of the corresponding amino acids have similar NMR spectra (*70*). This confirms the idea that there are no global but local interactions influencing random coil chemical shifts(*71*).

Chemical shift indexing (CSI) is a common method in NMR for characterizing secondary structure elements (*71-74*) and there have been a lot of research in determining random coil chemical shifts under various conditions (*75-78*). Differences to random coil are normalized by CSI to 1, 0 and -1 depending on the degree of the deviation (*71*). A sequence dependent plot of the CSI shows regions where secondary structure elements are present. Whereas α-helices have downfield shifts of $^{13}C_\alpha$ and $^{13}C'$ nuclei and upfield shifts of $^{13}C_\beta$ and ^1H nuclei, β-sheets show just contrary behaviour.

Another way of using chemical shifts is the SSP method (*53*). Secondary structure propensities (SSP) were obtained from $\Delta^{13}C_\alpha$ and $\Delta^{13}C_\beta$ chemical shifts and predicted. This

method weights and combines individual chemical shifts thereby calculating secondary structure propensies rather than absolute secondary structure. Values vary between -1 and 1 and display the mean percentage for α-helix or β-sheet.

Finally chemical shifts of nonnative proteins are one of a few fundamental parameters for determining the content of residual secondary structure.

1.3.3   Protein Dynamics and Relaxation

Proteins are involved in many biological processes like folding reactions or ligand binding for example. For these processes conformational changes in structure are often required which can only be performed by flexible regions in the protein. Solution NMR is a valuable tool for the analysis of dynamical motions in structures on different timescales, ranging from picoseconds up to seconds (figure 6).
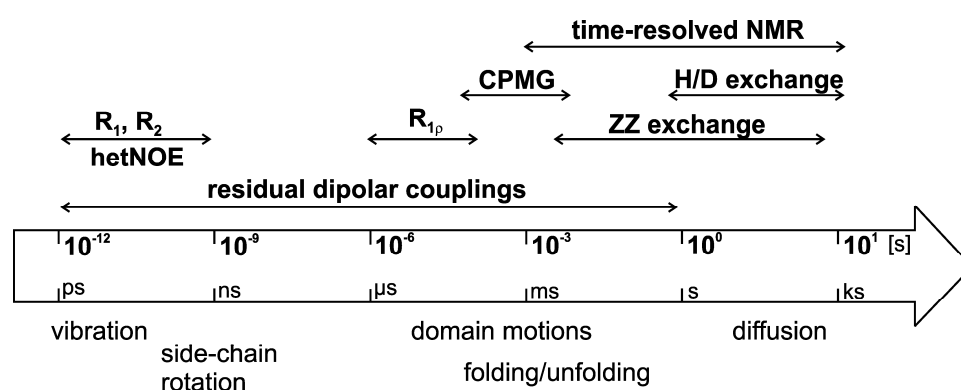
Figure 6: NMR timescales; Schematic representation of the NMR timescale of protein dynamics. Figure modified after [Boehr *et al.*, 2006].

Spherical molecules like proteins, DNA or RNA have faster conformational motions than the corresponding rotational correlation time τ_c enabling spin relaxation experiments. The rotational correlation time τ_c shows proportionality between the molecular weight of the protein and viscosity of the medium.

1.3.3.1 Heteronuclear relaxation

In particular longitudinal relaxation R_1, transversal relaxation R_2, spin-lattice relaxation $R_{1\rho}$ and the heteronuclear NOEs (hetNOE) are definable parameters depending mainly on dipolar interactions and chemical shift anisotropy (*79*). These relaxation parameters are described by the spectral density functions (equ. 1.1-1.3) (*80*).

$$R_I = \frac{d^2}{4}\left[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)\right] + c^2 J(\omega_N) \tag{1.1}$$

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

$$+ \frac{c^2}{6}\left[4J(0) + 3J(\omega_N)\right] + R_{ex} \tag{1.2}$$

$$hetNOE = 1 + \left(\frac{d^2}{R_I}\right)\left(\frac{\gamma_N}{\gamma_H}\right)\left[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)\right] \tag{1.3}$$

with ω_H and ω_N as the Lamor frequencies of ^{1}H, ^{15}N nuclei, γ_H and γ_N as the gyromagnetic ratio of the corresponding nucleus. $J(\omega)$ represents the spectral density function at the Lamor frequency, R_{ex} denotes conformational exchange. Dipolar interactions and the chemical shift anisotropy are described by *d* und *c* (equ. 1.4)

$$d^2 = 0.1\frac{\gamma_H^2 \gamma_N^2 h^2}{4\pi^2}\left(\frac{1}{r_{NH}^3}\right)^2 \quad \text{and} \quad c^2 = \frac{2}{15}\gamma_N^2 H_0^2 (\sigma_\mathrm{II} - \sigma_\perp)^2 \tag{1.4}$$

with the *Planck* constant *h*, the magnetic field strengh H_0 and the parallel and orthogonal component of the chemical shift tensor $\sigma_\mathrm{II} - \sigma_\perp$.

As nonnative states of proteins consist of ensembles of structures ^{15}N R_2 relaxation rates are one of the most promising data (figure 7, left). The R_2 rates are defined by the dipolar component $T_{2,DD}$ and the chemical shift anisotropy $T_{2,CSA}$ (equ. 1.5)

$$R_2 = \frac{1}{T_{2,DD}} + \frac{1}{T_{2,CSA}} + R_{ex} \tag{1.5}$$

Relaxation rates of folded proteins can be analyzed by the Lipari-Szabo *model-free* formalism (*38*). This method is used to fit the spectral density function $J(\omega)$ to experimentally obtained data and presumes that the rotation of the overall protein τ_c and the internal motions of the protein backbone τ_e range on different timescales. With known τ_c internal motions can be analyzed independently. Motional limitations of these fluctuations are described by the order parameter S^2 (*81*). S^2 results in values between 1 and 0, whereby the first value specifies a perfectly rigid state and the second value implicates a dynamic state. The dependency of S^2, τ_e and τ_c on relaxation rates is illustrated in figure 7 (right); the rigidity of side chains increases with increasing R_2 rates.
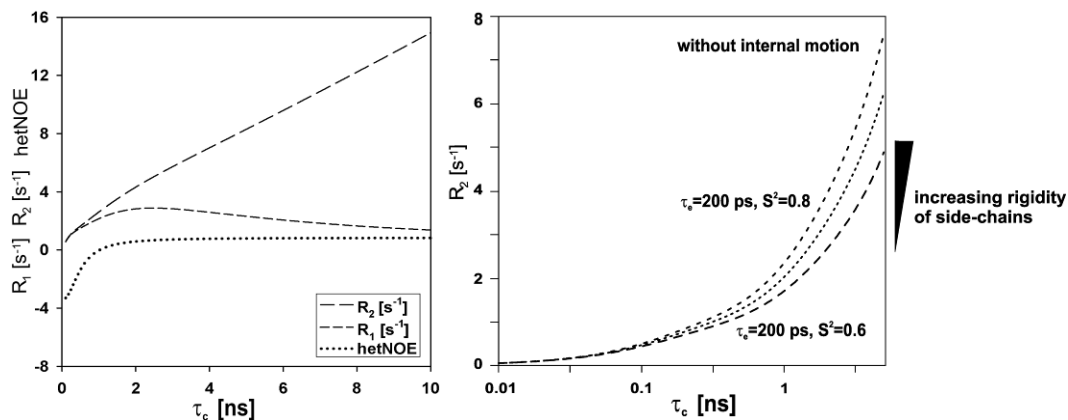
Figure 7: Relaxation; Correlation between R_1, R_2, hetNOE and the rotational correlation time τ_c (left) and the dependency of ^{15}N R_2 relaxation rates from S^2, τ_e and τ_c (right). Figure modified after [Wirmer *et al.*, 2005]

The interpretation of relaxation data in unfolded proteins remains still challenging. Here the Lipari-Szabo *model free* formalism is not appropriate since the approximation of independent internal and global motion fails. The overall and internal motions are on similar timescales and cannot be separated but it could be shown that randomcoil behaviour can be described using the segmental motion model (*27-29, 39*). There it is assumed that relaxation data of a given amide are governed by segmental motions only of parts of the polypeptide sequence,

independent from the overall tumbling and amino acid specific features. Experimentally obtained relaxation rates are dissected in individual relaxation terms (equ. 1.6).

$$R^{\exp}(i) = R^{RC}(i) + R^{Cluster}(i) + R^{DSB}(i) + R^{Ex}(i) \tag{1.6}$$

R^{RC} is the contribution of the random coil polypeptide chain, $R^{Cluster}$ of hydrophobic cluster, R^{DSB} of disulfide bridges, R^{Ex} of chemical exchange and $R^{\exp}$ the experimental relaxation rate. In the literature residual structure and dynamics have been identified in a number of chemically or biochemically denatured proteins (*37, 45, 49-51, 53-55*). Notably unfolded lysozyme is one example of a nonnative protein with regions of positive deviations from the baseline indicating residual structure (*27, 37, 51*). In general the baseline is described by a totally unstructured polymer with all over internal motion. Positive deviations suggest regions with less motion, as it is the case for six hydrophobic regions in hen and human lysozyme.

1.3.3.2 *Spectral density*

Besides the *model-free* analysis (*38, 82*) which is still challenging for unfolded protein ensembles, reduced spectral density mapping (*83-85*) is a convenient tool. The combination of ^{15}N R$_1$, R$_2$ and the heteronuclear [^{1}H]-^{15}N NOE restricts the number of independent dynamical parameters and minimize false descriptions of motions as it might be the case for the *model-free* formalism. Reduced spectral density functions J(0), J(ω_N) and J(0.87ω_H) are described in detail in the literature (*86, 87*). The graphical analysis of spectral densities assumes that dynamics of the N-H vector has *m* independent motional options at the correlation time τ_j. The result is a sum of Lorentzian lines (*86*) (see equ. 1.7, 1.8)

$$J(\omega) = \frac{2}{5} \sum_{j=0}^{m-1} a_j \frac{\tau_j}{1 + (\omega \tau_j)^2} \tag{1.7}$$

$$\text{and } \sum_{j=0}^{m-1} a_j = 1 \tag{1.8}$$

a_j represents the relative contribution of the individual motional options. The assumption that dynamics are limited to a single motion reduces equation (1.7) to

$$J(\omega) = \frac{J(0)}{(1 + 6.25(J(0)\omega))^2} \tag{1.9}$$

with $J(0)=4\tau$ and $J(0)$, $J(\omega)$ as reduced spectral density functions at frequency ω (*85*). Figure 8 shows a schematical representation of the Lorentzian line shape and the correlation of $J(0)$ and $J(\omega)$.
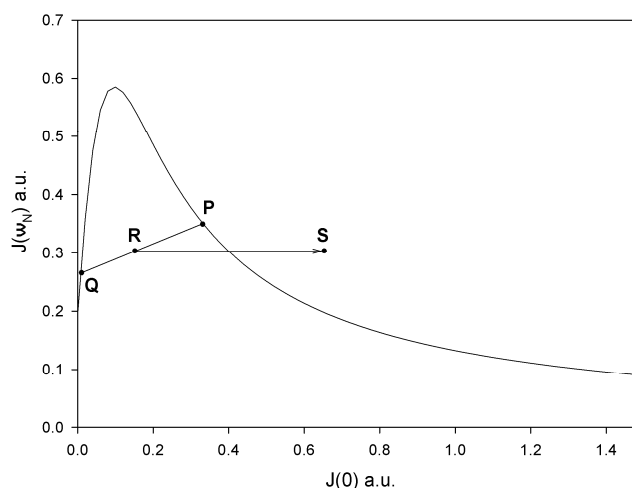
Figure 8: Spectral density; Lorentzian line shape, dynamics limited t a single motion, equ. (1.9) (*85*). P: dominated by overall tumbling; Q: dominated by fast internal motion; R: motions between P and Q; S: with conformational motions.

Motions dominated by the overall tumbling are indicated by P, Q represents motions dominated by a fast internal motion. Motions between these two points which result from measured spectral densities in the pico- to nanosecond time regime are assigned by R. Adding conformational motions, $J(0)$ can be increased by a parallel shift to S.

1.3.3.3   Relaxation dispersion

Information of structural data and slow exchange processes can be obtained by NMR even at low populated states. Relaxation dispersion, in particular Carr-Purcell-Meiboom-Gill experiments (CPMG) (*88, 89*) report on slow dynamics of backbone and side chain residues at the micro- to millisecond time regime. For low populated, so called "invisible" states with a population of around 0.5% and exchange in the slow time regime, relaxation dispersion is sensitive (*90*).

Residues undergoing conformational exchange due to nonnative hydrophobic interactions contribute to the ^{15}N line width and can be analyzed by CPMG-based experiments. Equation (1.8) shows the dependence of the effective transverse relaxation rate constant on the cross peak intensities with *I(νCPMG)* and without *I$_0$* the constant time CPMG interval (*91*).

$$R_{eff}(\nu CPMG) = -\frac{1}{T} ln \frac{I(\nu CPMG)}{I_0} \tag{1.8}$$

By deconvolution of the contribution to the apparent relaxation rate, ^{15}N relaxation dispersion data and ^{15}N longitudinal relaxation can be combined whereby slow exchange becomes visible (*92*). Therefore relaxation dispersion is a useful tool in analyzing unfolded protein states.

1.3.4   Diffusion

Besides J-couplings and torsion angles, considering particular amino acids, the overall extension of unfolded proteins is another interesting and helpful feature. The diffusion coefficient of a spherical molecule follows the *Stokes-Einstein* equation:

$$D = \frac{kT}{6\pi a\eta} \tag{1.9}$$

This equation is valid at a given temperature T in a continuous fluid; k is the Boltzmann constant, η the viscosity of the medium and the radius of the molecule. To show the expansion of nonnative states diffusion measurements exhibit the degree of compaction. Therefore the *Stokes-Einstein* equation is transformed into equ. (1.9) by replacing a with the

effective hydrodynamic radius R_h. The measured diffusion constant D is than inverse proportional to the hydrodynamic radius of the molecule (see equ. 1.10).

$$R_h^{protein} = \frac{D_{dioxane}}{D_{protein}} R_h^{dioxane} \qquad\qquad (1.10)$$

$R_h^{dioxane}$ is the radius of dioxane (2.12 Å) as internal reference, $D_{dioxane}$ and $D_{protein}$ are the decay rates, $R_h^{protein}$ is the hydrodynamic radius of the protein.

Diffusion measurements are usually performed by pulse field gradient NMR where the strength of the bipolar gradient for diffusion is varied percental (*93*). Thereby the signal intensity decreases with increasing gradient strength; the smaller the molecule the faster the decay of signals. Hydrodynamic radii of unfolded states are averaged over all radii in the conformational ensemble and therefore suggest the expansion of the overall ensemble. Due to loss of secondary and tertiary interactions more extended structures compared to the native states are expected.

1.3.5   Residual Dipolar Couplings (RDCs)
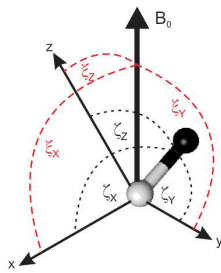
Residual dipolar couplings play a crucial role in structural analysis of biomacromolecules (*94*). As the literature reflects RDCs become more and more promising for structure prediction of unfolded proteins (*50, 58, 60, 95*). RDCs are parameters for dipolar interactions through space between spin ½ nuclei where the alignment of single bond vectors relative to the magnetic field is measurable. Through alignment molecular motions loose their isotropic behaviour whereby the averaging of all dipolar couplings is deleted. The aligment of the molecules can be generated by specific media like bicells (*96*), phages (*97*) or polyacrylamide gels (*98*) for example.

In unfolded proteins backbone and side chain dynamics result mainly from ^{1}H-^{15}N or ^{1}H-^{13}C bond vectors. The dipolar coupling D_{ij} of two involved nuclei *i* and *j* depend on the effective distance r_{ij} and the angle θ_{ij} between the internuclear vector and the external magnetic field (equ. 1.11) (*99, 100*).

$$D_{ij} = -\frac{\gamma_i \gamma_j \mu_0 h}{8\pi^3} \left\langle \frac{3\cos^2\theta_{ij}(t)-1}{2r_{ij}^3(t)} \right\rangle \tag{1.11}$$

μ_0 is the magnetic permeability, h the Planck constant, γ_i and γ_j the gyromagnetic ration of the two nuclei i and j.

The orientation of the bond vectors in the molecular frame in solution is averaged to zero and the magnetic field is aligned along the z-axis. The relative orientation of the molecule to the molecular frame can be converted into a principal axis system. Here the relative orientation between the principal axis system and the molecular frame is described by the angles ξ_x, ξ_y, ξ_z, whereas the angles ζ_x, ζ_y, ζ_z represent the orientation of the bond vector relative to the principal axis system (figure 9). The relation between the angle θ_{ij} and ξ_x, ξ_y, ξ_z and ζ_x, ζ_y, ζ_z is described by equation (1.12).

$$\cos\theta_{ij} = \begin{pmatrix} \cos\zeta_x \\ \cos\zeta_y \\ \cos\zeta_z \end{pmatrix} \begin{pmatrix} \cos\xi_x \\ \cos\xi_y \\ \cos\xi_z \end{pmatrix} = \cos\zeta_x \cos\xi_x + \cos\zeta_y \cos\xi_y + \cos\zeta_z \cos\xi_z \tag{1.12}$$

Figure 9: Molecular frame; Orientation of the bond vector relative to the magnetic field B_0 (ζ_x, ζ_y, ζ_z), and orientation of the principal axis system relative to the molecular frame (ξ_x ξ_y, ξ_z)(*101*).

Applying equ. (1.12) on equ. (1.11) and substituting $C_i = \cos\zeta_i$ and $C_i = \cos\zeta_i$ with the assumption of a constant orientation of the spins relative to the principal axis system results in equ. (1.13).

$$\left\langle \frac{3\cos^2\theta_{ij}(t)-1}{2r_{ij}^3(t)} \right\rangle = \frac{3}{2}\left[\langle c_x^2 \rangle C_x^2 + \langle c_y^2 \rangle C_y^2 + \langle c_z^2 \rangle C_z^2 + 2\langle c_x c_y \rangle C_x C_y + 2\langle c_x c_z \rangle C_x C_z + 2\langle c_y c_z \rangle C_y C_z \right] - \frac{1}{2} \tag{1.13}$$

Only terms aligned to the molecular frame are averaged considering the orientation of the spins as fixed concerning the principal axis system. The alignment tensor represents the

preferential orientation of the molecule in the *Saupé matrix* (1.14) containing elements A_{kl} and the *Kronecker delta* δ.

$$A_{kl} = \frac{3}{2}\langle\cos\xi_k\cos\xi_l\rangle - \frac{1}{2}\delta_{kl} \qquad (1.14)$$

Diagonalization of the Matrix results in the diagonal elements A_{xx}, A_{yy}, A_{zz}. Through a three-dimensional *Euler rotation* R(α, β, γ) the diagonalized tensor of the principal axis system is described by A_{xx}, A_{yy}, A_{zz} and the position of the vector by the polar angles θ and φ. From these values the axial and rhombic components are defined as $A_a = A_{xx}/2$ and $A_r = 1/3(A_{xx} - A_{yy})$. Applying all parameters, the residual dipolar couplings are expressed by equ. (1.15) (*99, 100*)

$$D_{ij}(\theta,\varphi) = -\frac{\gamma_i\gamma_j\mu_0 h}{16\pi^3 r_{ij,eq}^3}\left(A_a\left(3\cos^2\theta - 1\right) + \frac{3}{2}A_r\sin^2\theta\cos 2\varphi\right) \qquad (1.15)$$

The alignment tensor is described by five parameters A_a, A_r, α, β und γ which can be analyzed by at least five RDCs or by singular value decomposition (*102*). However the measured couplings are not clearly defined, every value is combined of endless possible combinations of the polar angles. Only values around the maxima and minima have smaller errors due to the degenerated orientation which can be constrained by consideration of multiple bond vectors with known orientation. For three vectors with minimal, maximal and medium range values and well known relative alignment only four orientations are possible. RDCs can also be measured by couplings in different alignment media. Five independent orientations lead to the experimental dipolar couplings.

Determination, prediction and interpretation of residual dipolar couplings in folded proteins are well-established but for unfolded states it is still challenging; every segment has an independent alignment tensor and has to be treated separately. Conformational dynamics in unfolded ensembles further complicate the analysis. One implemented alignment medium is

polyacryl amide gel which is described in detail in the materials and methods section (*43, 98, 103*).

1.3.6   J-couplings

Since the early fifties the measurement of scalar spin-spin couplings by NMR has been established as a powerful method for the determination of protein structures (*104, 105*). The information out of these experiments is extremely useful since the size of the coupling depends directly on the molecular conformation by the dihedral angles (*106, 107*).

Over the years nonnative proteins have come to fore of chemical and biochemical research (*2, 108, 109*). Partially close packing in unfolded states considerable create fluctuations of the angles but there are still conformational preferences lost in the residual structure (*42*). Defining backbone and side chain coupling constants in unfolded proteins is essential for determination of the torsion angles ϕ_i, ψ_i, ω_i for main chain and χ_1 for side chain conformations. Out of these data preferences for α-helical or β-sheet structural elements will derive. This information is crucial for the understanding of structure and stability of non-native states. As shown in figure 10, local conformations depend on three different backbone torsion angles ϕ_i, ψ_i and ω_i.
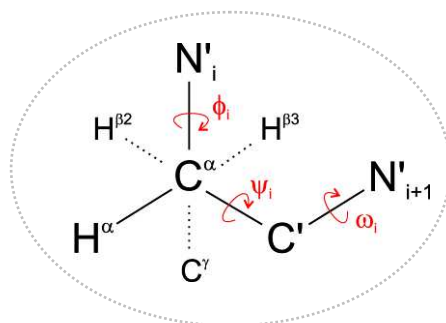
Figure 10: Main chain torsion angles; Schematical representation of a peptide backbone, torsion angles ϕ_i, ψ_i and ω_i are highlighted in red.

However the partially double bonded character of the amide limits rotations around ω_i. The *trans* rotamer is more stable than the *cis* rotamer and dominates the peptide backbone

conformation (*110, 111*). Hence ψ_i and ϕ_i are the essential parameters defining the conformations in the polypeptide backbone.

In particular vicinal ^3J coupling constants of proteins or peptide backbones provide useful data for structure determination (*28, 106*) where the angle dependence is reflected by the Karplus relationship (*112*). Karplus parameters have been determined for ^3J-couplings as well as ^1J- and ^2J-couplings (*113*). Six vicinal ^3J couplings $^3J_{HN,H\alpha}$, $^3J_{HN,C\beta}$, $^3J_{HN,CO}$, $^3J_{H\alpha,CO}$, $^3J_{CO,CO}$ and $^3J_{CO,C\beta}$ depend on the backbone torsion angle ϕ. In general these couplings can be described by equation (1.16) (*112, 114*).

$$^3J = A\,cos^2\phi + B\,cos\phi + C \qquad \text{with} \qquad \phi = \varphi + phase \qquad (1.16)$$

Applying the parameterization A = 6.4, B = -1.4, C = 1.9 from Vuister *et al.*(*115*) equation (1.16) is transformed into equation (1.17).

$$^3J = 6.4\,cos^2(\phi - 60°) - 1.4\,cos(\phi - 60°) + 1.9 \qquad (1.17)$$

It has to be pointed out that from the vicinal ^3J couplings the $^3J(H^N, H_\alpha)$ coupling is one of the precious coupling for structural investigation of backbone conformations because of the high gyromagnetic ratio of the nuclei. For α-helices, 3_{10}-helices and β-strands mean values for ^3J (H^N, H_α) couplings are 4.8 Hz, 5.6 Hz and 8.5 Hz, respectively.

secondary structure	Mean angle	[°]
	ϕ_i	ψ_i
α-helices	-64.7	-39.8
3_{10}-helices	-62.8	-16.5
β-strands	-112.6	123

Table 2: Torsion angles of secondary structures; Mean values for most populated torsion angles ϕ and ψ in regions of secondary structures; values for β-sheets include parallel and antiparallel sheets (*42*).

A data base prediction including 85 protein structures results in mean values for ψ_i and ϕ_i in regions of secondary structures (table 2) (*42*). Table 2 shows a significant range of dihedral angles for α-helix, 3_{10}-helix and β-sheet. Even though the global structure has variations through unfolding there are still conformational preferences in the polypeptide chain.

Indeed ^{3}J couplings are mainly used for structural characterization due to the dependence on torsion angles but ^{1}J(N,C$^\alpha$) and ^{2}J(N,$^{C\alpha-1}$) also show conformational dependence and therefore are a further indicator for secondary structure elements (*113*). ^{1}J(N,C$^\alpha$) coupling constants depend weakly on $\psi_{(i)}$ while ^{2}J(N,C$^\alpha$) coupling constants depend on the $\phi_{(i-1)}$ and the secondary structure determining angle $\psi_{(i-1)}$. This dependence of ^{1}J(N,C$^\alpha$) and ^{2}J(N,C$^\alpha$) allows differentiation between positive and negative ψ values. With a series of J-modulated HSQCs the dependence of the experimental peak volume on the mixing time can be analyzed (figure 11). The equation $I_{NC\alpha}^{\exp} = A\cos(\pi\,^1J\tau)\cos(\pi\,^2J\tau)\cdot e^{\frac{-\tau}{T_2}}$ can be dissolved by the ^{1}J(N,C$^\alpha$) and ^{2}J(N,$^{C\alpha-1}$) couplings (see materials and methods section).
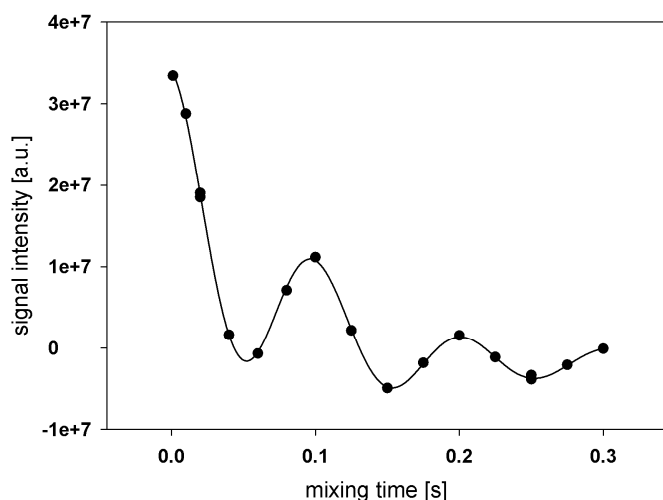
Figure 11: ^{1}J(N,C$^\alpha$) and ^{2}J(N,C$^{\alpha-1}$) couplings; the signal intensity of one exemplary residue of unfolded hen egg white lysozyme is plotted versus the mixing time. A intensity fit of the curve results in values for ^{1}J(N,C$^\alpha$) and ^{2}J(N,C$^{\alpha-1}$).

In peptides couplings of $^1J(N,C^\alpha) \sim 11Hz$ and $^2J(N,^{C\alpha-1}) \sim 7Hz$ have been reported in the literature (*116*) which can be confirmed with data from denatured ubiquitin (*113*) and the here presented data for unfolded hewl (see section 3.1). Beside main chain torsion angles also side chain couplings are useful parameters for structure determination of unfolded proteins. These couplings are challenging due to the limited dispersion of side chain proton resonances which were caused by conformational averaging of the structure ensembles. An improvement of this method is the distinction between $H^{\beta2}$ and $H^{\beta3}$ for stereospecific assignment (*117*) and the analysis of the χ_1 angles. There are in principle six J-couplings which define the χ_1 angle in the side chains of proteins: $^3J_{H\alpha,H\beta}$, $^3J_{N',H\beta}$, $^3J_{C',H\beta}$, $^3J_{H\alpha,C\gamma}$, $^3J_{N',C\gamma}$ and $^3J_{C',C\gamma}$. With a Pachler-type analysis the three probabilities for the populations $p_{-60°}$, $p_{60°}$, $p_{180°}$ (see equ. 1.18) can be verified by at least two J-couplings.

$$p_{-60} + p_{60} + p_{180} = 1 \tag{1.18}$$

In general a pachler-type analysis is based on J_t for the trans position of two hydrogens and J_g for the gauche position being constant and the probabilities a, b and c. Therefore in a classical staggered conformation the vicinal coupling for two systems is given by (*118*)

$$J_{12} = bJ_t + (a+c)J_g \tag{1.19}$$

$$J_{13} = aJ_t + (b+c)J_g \tag{1.20}$$

with $a + b + c = 1$ the equations above are transformed into

$$J_{12} = b(J_t - J_g) + J_g \tag{1.21}$$

$$J_{13} = a(J_t - J_g) + J_g \tag{1.22}$$

The distribution of the side chain angle χ_1 in unfolded proteins gives information about the compactness of the structure. The angle is defined by the relative orientation of N_i' and C^γ. If the two nuclei enclose a χ_1 angle of $\pm180°$ the conformation of the molecule is *gauche⁻*, if χ_1

is -60° the segment has *gauche*$^+$ conformation. In figure 12 the two possible staggered rotamers in *gauche* conformations g^+ (+60°) and g^- (-60°) are shown.
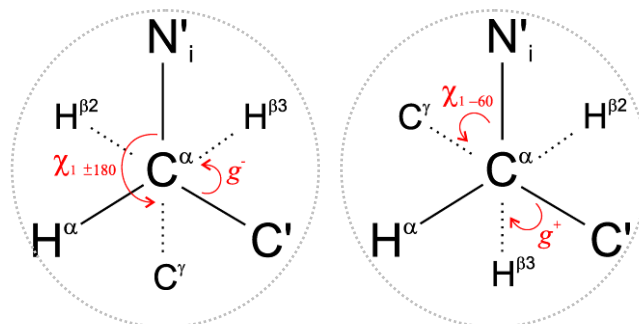
Figure 12: Side chain torsion angles; Fragments of two staggered rotamers with g+ (60°) and g- (-60°) conformation. The chi 1 angle varies with the angle of the gauche conformation: g+ = 60° with χ_1 = -60° and g- = -60° with χ_1 = ±180°.

With these findings a stereospecific assignment of the side chain protons of unfolded hen egg white lysozyme out of $^3J_{HN,H\alpha}$ and $^3J_{COH\beta}$ has been performed (see chapter 3.1).

2   Materials and Methods

2.1   Unfolded states of hen and human lysozyme and mutants

Since hen and human lysozyme share high sequence homology and identity, both proteins could be treated in a similar manner regarding biochemical synthesis and biophysical analysis. Therefore the methods used in this thesis are explained in general for both proteins in the following sections.

2.1.1   Cloning, expression, purification and NMR sample preparation

Site-directed mutagenesis

The genes encoding mutants W28G^{Ala}, W62G^{Ala}, W108G^{Ala}, W111G^{Ala} and W123G^{Ala} were synthesized by mutagenesis of the WT^{Ala} Plasmid-DNA (*119*) using the QuickChange Site-Directed Mutagenesis Kit$^®$ (Stratagene, CA; sequential analysis, Scientific Research and Development GmbH, Oberursel).

The plasmid-DNA encoding the all-alanine variant of human lysozyme (hLys) was purchased from Entelechon (Regensburg, Germany). The genes encoding the mutants hLys-I56T and hLys-D67H were synthesized by mutagenesis of the hLys template plasmid-DNA using the QuickChange Site-Directed Mutagenesis Kit (Stratagene, CA; sequential analysis, Scientific Research and Development GmbH, Oberursel). The appropriate DNA sequences for forward and backward primers for unfolded hen and human lysozyme can be found in the appendix (see 4.2). The RNA was sequenced by Scientific Research and Development GmbH, Oberursel.

Transformation and expression of isotope-enriched protein

^{15}N and ^{13}C/^{15}N labelled protein was expressed in *E. coli* strains BL21(DE3) with pET11a as expression vector. 1µL plasmid DNA was added to 50µL cell solution thawn on ice. The mix was incubated for 40 minutes on ice, a heat shock for 45 sec at 42°C was applied and than cooled for 2 minutes on ice. 250µL SOC medium (Invitrogen, Karlsruhe, Germany) was added and the cell mixture was allowed to incubate 30 minutes at 37°C on a shaker. Thereafter grown cells were plated on agar LB plates with an ampicillin resistance and

incubated over night. Cells were grown in M9 minimal medium with ^{15}N-labelled ammonium chloride ([15]NH$_4$Cl, Martek) and [13]C$_6$ Glucose (Cambridge Isotope Laboratories, Andover; USA) as sole nitrogen and carbon source, respectively. At OD$_{600}$= 0.8, induction of protein expression was initiated by addition of 1mM IPTG. The protein is thereby expressed in inclusion bodies. After 3 hours of induction, the cells were harvested by centrifugation at 6.500g for 20min (centrifuge: Beckmann Coulter Avanti J-20XP) and stored in buffer 1 (see appendix 4.1.2) at -80°C.

Protein purification and biochemical analysis

The cells were resuspended in 50 mM Tris/HCl, 25% sucrose and 1mM EDTA at pH 7.5 and cells were lysed by sonication. After centrifugation (10.000g, 30min), the supernatant was discarded and the pellet was resuspended in 20 mM Tris/HCl, 1% Triton X-100 and 1mM EDTA at pH 7.5. After second centrifugation (10.000g, 30min), the supernatant was again discarded and the pellet – which contains the inclusion bodies, was dissolved in 20 mM Tris/HCl, 50 mM NaCl, 5 mM EDTA and 8 M urea, pH 7.5. A last centrifugation step (10.000g, 30min) results in a pellet, which was discarded and the supernatant containing protein was loaded onto a CM-sepharose (Sigma-Aldrich) ion-exchange column, which was equilibrated with 50 mM Tris/HCl, 50 mM NaCl, 1 mM EDTA and 4 M urea, pH 7.5. The protein was eluted with a linear gradient between buffer A (50 mM Tris/HCl, 50 mM NaCl, 1 mM EDTA and 4 M urea, pH 7.5) and buffer B (50 mM Tris/HCl, 300 mM NaCl, 1 mM EDTA and 4 M urea, pH 7.5). Fractions with matching bands on a SDS-PAGE gel were pooled and dialyzed against water (pH 2) for several days (tubular membrane 3500 MWCO, ZelluTrans, Roth). After purification via HPLC with a reversed phase Vydac Grace column, TFA buffer and a linear water/acetonitrile gradient, the protein was freeze-dried and stored at -20°C until further usage.

NMR sample preparation

For NMR measurements, the protein was dissolved in sterile H$_2$O with 10% (v/v) D$_2$O at pH 2. RDC measurements were performed in 7% polyacryl amide in sterile H$_2$O with 10% (v/v) D$_2$O at pH 2. All sample concentrations for standard NMR experiments were at 300 µM, determined by UV light absorption at 280 nm using the extinction coefficient calculated from

the amino acid sequence; $\varepsilon = 37.470$ M^{-1}cm^{-1} for WT^{Ala} and $\varepsilon = 36,440$ M^{-1}cm^{-1} hLys (calculated via ExPASy, Swiss Institute of Bioinfomatics). Protein concentrations for diffusion measurements range between 100-300µM and for hLys and mutants between 300µM and 800µM.

2.2   NMR spectroscopy

2.2.1   Assignment of Hen Egg White Lysozyme mutants

The backbone assignment of all mutants was facilitated by the existing chemical shifts for wild-type WT^{Ala}. Therefore, measurements of ^{15}N-HSQCs (*120*) together with 3D-^{15}N-NOESY-HSQCs (*121, 122*) were sufficient. The 3D-NOESY-HSQC was recorded with 2048 points in the ^{15}N dimension 128 points in the ^{1}H dimension and eight scans. Side chain assignment was done with the performance of a 3D-HN(COCA)HAHB (*117*) and a 3D-^{15}N,^{13}C-HN(CO)HB (*123*) experiment. The former was recorded with 256 points the ^{1}H dimension and 300 points in the ^{13}C dimension, the latter was performed by the implemented Bruker pulse sequence. Diagonal peaks represent the ^{1}H$^{\alpha}$ or ^{13}C'-peak and the cross peak the coupled H$^{\beta}$. With these data an assignment of ^{13}CO and ^{1}H$^{\beta 2}$,^{1}H$^{\beta 3}$ resonances is possible. External referencing was done by 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS), ^{1}H direct and ^{13}C, ^{15}N indirect.

2.2.2   Assignment of Human Lysozyme and mutants

The backbone assignment of wt hLys was derived from analysis of 3D-^{15}N-NOESY-HSQC (*124*), HNHA- (*115*), HNCACB- (*125*), HNCO- (*63*) and HNN-experiments (*62*). The assignment of the mutants was done by comparing the existing chemical shifts for wild-type hLys from the ^{1}H-^{15}N HSQC with those of the mutant HSQCs. Therefore, measurements of ^{15}N-HSQCs together with 3D-^{15}N-NOESY-HSQCs were sufficient to obtain complete proton and nitrogen chemical shift assignment for hLys-I56T and hLysD67H. The 3D-NOESY-HSQC was recorded with 2048 points in the ^{1}H dimension, 128 points in both ^{15}N dimension and 8 scans (600 MHz), the HNHA-experiment with 2048

points in ^{15}N, 108 points in the ^{1}H dimension and 8 scans (600 MHz). The HNCACB-experiment was measured with 2048 points in the ^{15}N dimension, 88 points in ^{13}C and 16 scans (900 MHz), the HNCO with 2048 points in the ^{15}N dimension, 70 points in ^{13}C and four scans (700 MHz). For the HNN-experiment 2048 points in the ^{15}N dimension, 88 points in the ^{13}C dimension and 16 scans were recorded at 900 MHz. The backbone assignment was done by Sparky (SPARKY 3, University of California, San Francisco) and CARA (*126*) (www.nmr.ch). The direct ^{1}H chemical shifts were referenced external with 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS), from which ^{15}N and ^{13}C references derive indirectly.

2.2.3   Chemical shift deviations from random coil

Chemical shift deviations from random coil values (*78*) of hLys and hewl were calculated by $\Delta\delta=\delta^{exp}-\delta^{rc}$. The cut-offs for chemical shift indexing were adopted from Wishart and Sykes, 1994 (*71*). Secondary structure propensities were calculated from C^α and C^β chemical shifts following Marsh *et al.*, 2006. Due to a pH dependency of aspartate and glutamate, chemical shifts at pH 2 have to be omitted in the comparison with random coil values.

2.2.4   Relaxation experiments

All NMR relaxation experiments were measured under identical conditions: 600 MHz (Bruker, DRX 600, probe head: 5mm TXI HCN xyz-gradient), 293K, ^{15}N-labelled sample concentration of 300 µM, (H$_2$O/ 10% D$_2$O, pH 2) for WT^{Ala} and mutants. Concentrations of 300µM and 800µM were chosen for hLys and mutants. NMR data were processed using TopSpin 2.1 (Bruker Biospin) and analyzed by Sparky (SPARKY 3, University of California, San Francisco). For the determination of R$_2$ rates, the ^{15}N-CPMG-HSQC was recorded as a pseudo three-dimensional spectrum with nine increments ranging from 17ms to 306 ms including three repeats. The SPARKY routine was used to extract relaxation rates from peak heights. For R$_{1\rho}$ rates ten increments were measured ranging from 20 ms to 250 ms with two repeats. For the R$_1$ rates eight different relaxation times have been used ranging from 0.01s to 2s and two repeats. The SPARKY routine was used to extract relaxation rates from peak heights. CPMG relaxation dispersion experiments (*127*) were recorded as pseudo-three

dimensional spectra with the pulse sequence provided by Bruker and analyzed with the TOPSPIN 2.0 software. A dataset of 16 experiments was performed with different CPMG field strengths ranging from 25 to 750 Hz, including three repeats. All spectra were recorded at 293K with 128 real points and a sweep width of 22 ppm in the ^{15}N dimension at a protein concentration of 300µM. For data analysis the resulting effective relaxtion rates (R_2^{eff}) were plotted against the CPMG field strengths with

$$R_{eff}(\nu CPMG) = -\frac{1}{T} ln \frac{I(\nu CPMG)}{I_0} \tag{1.23}$$

Relaxation rates were fitted using the segmental motion model as described in Wirmer *et al.*, 2004 using the following formulae:

$$R_2^{int} = R_{int} \sum_{j=1}^{N} e^{\frac{|i-j|}{\lambda_0}} \tag{1.24}$$

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

$$R_2^{exp}(i) = R_{int} \sum_{j=1}^{N} e^{\frac{|i-j|}{\lambda_0}} + \sum_{cluster} R_{cluster} e^{-\left(\frac{i-x_{cluster}}{2\lambda_{cluster}}\right)^2} \tag{1.25}$$

The proposed model for lysozyme includes additionally a Gaussian term where the clusters are characterized by the position $x_{cluster}$, the width $\lambda_{cluster}$ and the rate $R_{cluster}$.

2.2.5 Spectral density mapping

Spectral densities of WT^{Ala} and mutants, hLys and D67H and I56T were determined from R_1, R_2 and hetNOE as described in Farrow *et al.*, 1997. The Lorentzian line is described as $J(\omega) = J(0)/(1+6.25(\omega J(0))^2$ at frequency ω (*85*).

2.2.6   RDC measurements

Anisotropic sample preparation

The proteins were aligned in stretched 7% polyacrylamide gels (*103, 128*) (ratio acryl amide to bisacryl-amide 37,5:1) at a concentration of 300 µM (in H_2O/ 10% D_2O, pH 2) to compare them to the isotropic samples in solution at equal concentration. The equipment for the preparation of the anisotropic samples was purchased from Newera Enterprises Inc., USA.

RDC experiments

The HDO quadrupolar splitting was measured from the quadrupolar splitting observable in the ^2H 1D spectrum. RDCs were measured on a 600 MHz spectrometer (Bruker, DRX 600, probe head: 5mm TXI HCN xyz-gradient) via the standard Bruker implementation of a 2D-^{15}N-IPAP-HSQC (*129, 130*) extracting the sum of $^1J(H_N,N)$ and $^1D(H_N, N)$. $^1J(H_N,N)$ were measured using the same experiments for unaligned samples.

2.2.7   Diffusion measurements

Translational diffusion measurements (*93*) were recorded in D_2O at pH 2 with protein concentrations ranging from 300µM to 100µM. A five fold molar excess in dioxane concentration was used as internal standard. The standard Bruker pulse sequence for DOSY measurements was used where the length of all pulses were held constant while the strength of the diffusion gradient was varied between 5 and 95%. The spectra were recorded with 256 scans for 300µM protein up to 512 scans for 100µM protein and 32 experiments. The 1D spectra were optimized to give a signal decay of 85-90% , thus a stimulated echo of 150ms, a diffusion gradient pulse of 6.0ms was used. Data analysis was done using topspin 2.0. Hydrodynamic radii were calculated by

$$R_h^{protein} = \frac{D_{dioxane}}{D_{protein}} R_h^{dioxane} \qquad\qquad (1.26)$$

with a known dioxane radius $R_h^{dioxane}$ of 2.12 Å. $D_{dioxane}$ and $D_{protein}$ are the decay rates, $R_h^{protein}$ is the hydrodynamic radius of the protein.

2.2.8   Measurements of J-coupling constants

All NMR experiments were measured on a 600 MHz spectrometer (Bruker, Avance II). The spectra were recorded at 293 K at a concentration of 300 µM (in H_2O/ 10% D_2O, pH 2) of ^{15}N/^{13}C labelled protein.

Backbone coupling constants

For the determination of $^1J(N_i,C_\alpha)$ and $^2J(N_i,C_{\alpha(i-1)})$ coupling constants a series of J-modulated $^1H,^{15}N$-HSQCs were measured with 15 increments (τ = 1ms - 300ms). To extract peak intensities the SPARKY routine was used. The intensities were fitted with

$$I_{NC\alpha}^{exp} = A\cos(\pi^1 J\tau)\cos(\pi^2 J\tau) \cdot e^{\frac{-\tau}{T_2}}$$ were I^{exp} are the experimental peak heights, A is a fitting factor, ^1J and ^2J are the couplings from N_i to $C_{\alpha i}$ and to $C_{\alpha(i-1)}$, respectively (*112, 114*). T_2 is the apparent transverse relaxation rate and τ the mixing time. For the determination of $^3J(H_N,H^\alpha)$ constants a 3D-^{15}N-HNHA was recorded with the implemented Bruker pulse sequence. The magnitude from relative peak intensities was calculated with $\frac{I_{cross}}{I_{diag}} = -\tan^2(2\pi J_{HH}\tau)$ (*115*) where I_{diag} and I_{cross} are the intensities of diagonal and cross peaks, respectively. Peak intensities were determined via the SPARKY routine for fitting the peak heights.

Sidechain coupling constants

Side chain coupling constants, in particular $^3J(H^\alpha,H^\beta)$ and $^3J(CO,H^\beta)$, were recorded with a 3D-HN(COCA)HAHB and a 3D-$^{15}N,^{13}C$-HN(CO)HB experiment (*123*) with the implemented Bruker pulse sequence. Diagonal peaks represent the $^1H^\alpha$ or ^{13}C'-peak and the cross peak the coupled H^β. The coupling constants result from intensities with

$\frac{I_{cross}}{I_{diag}} = -\tan^2(2\pi J_{HH}\tau)$ for $^3J(H^\alpha,H^\beta)$ and $\frac{I_{cross}}{I_{diag}} = -\tan^2(\pi J_{CH}\tau)$ for $^3J(CO,H^\beta)$. The measured coupling constants report on angular distributions of the three rotamers of the side chain angle χ_1 which can be calculated using a Pachler-type analysis (*118*) (for a detailed description see 1.3.6 J-couplings).

2.2.9   External methods: Asteroids and SAXS

Results from ASTEROIDS calculations were used in this thesis for prediction of secondary structure elements in unfolded WT^{Ala}. Calculations were performed by Malene Ringkjøbing-Jensen (group of Prof. Dr. Martin Blackledge, Institut de Biologie Structurale Jean-Pierre Ebel, Grenoble 38027, France) by the following: calculations were done as described previously (*131*) with additionally two iteration steps where φ/ψ distributions of individual residues are obtained followed by a selection of entire structures. On the basis of C$_\alpha$, C$_\beta$ and C' chemical shifts as well as the $^3J(H^NH_\alpha)$, $^1J(NC_{\alpha i})$ and $^2J(NC_{\alpha(i-1)})$ scalar couplings with Karplus parameterizations taken from Pardi *et al.* (*132*) and Wirmer *et al.* (*113*) 5×200 φ/ψ angles were selected for each residue in the iteration steps. Due to low pH in the unfolded ensembles and the influence on chemical shifts, residues Asp, Glu and His were left out in the ASTEROIDS calculations. Out of a pool of 15.000 structures from flexible meccano obtained by combining conformers from both iteration steps RDCs were calculated for each structure in the pool using PALES (*133*) employing the local alignment window (LAW) approach with a window size of 15 amino acids (*134, 135*). The resulting final selection of entire structures was done on the basis of chemical shifts, scalar couplings as well as ^1H-^{15}N RDCs. As mathematic feature RDC calculations for all residues were only performable by introducing seven dummy residues (Ala) at each end of the protein sequence. As described previously by Salmon *et al.* (*136*), a baseline was inserted in the RDCs. The outcome of these calculations was a selection of 200 structures, predicted on the basis of chemical shifts, scalar couplings and RDCs showing secondary structure propensities of the unfolded WT^{Ala} ensemble.

All practical aspects from SAXS measurements were performed by Frank Gabel and Jie-rong Huang (group of Prof. Dr. Martin Blackledge, Institut de Biologie Structurale Jean-Pierre Ebel, Grenoble 38027, France). Measurements were recorded on the ID14-3 BioSAXS beamline at the European Synchrotron Radiation Facility (ESRF Grenoble, France). Sample-detector distance was 2.6 m, the X-ray wavelength used 0.931 Å (13.32 keV). An amount of 50 µL of protein sample and buffer (using diluted protein concentrations from 300µM to 100µM in water, pH 2, T = 293K) were loaded in a flow-through quartz capillary cell. As total exposure time 100s per protein sample were used. The 2D diffraction patterns were

normalized to an absolute scale and azimuthally averaged to obtain the intensity profiles $I(q)$, within BSxCuBE (ESRF beamline data collection software). Solvent contributions (buffer backgrounds collected before and after every protein sample) were averaged and subtracted from the associated protein sample using the program PRIMUS (*137*). For data analysis two fitting routines were used, EOM (ensemble optimization method) (*138*) and the Porod-Debye relation (*139, 140*). EOM is based on averaging the individual scattering from every conformer in the conformational ensemble. Thereby a pool with random conformers is generated and only those were selected fitting the experimental data. The intensity of a set of conformations $I(s)$ with N conformers is given by

$$I(s) = \frac{1}{N} \sum_{n=1}^{N} I_n(s) \qquad\qquad (1.27)$$

For native proteins a Guinier approximation is the appropriate analysis but for unfolded ensembles the Porod-Debye law is used where a relation between the scattering angle q and the intensity $I(q)$ is described.

$$I(q) \approx \Delta\rho^2 . \frac{2\pi}{q^4 . S} \qquad\qquad (1.28)$$

Plotting $I(q) \cdot q^2$ versus q results in curves giving a degree of flexibility in the structures as seen in Rambo *et al.* (*141*).

3   Results and Discussion

3.1   Protein expression and NMR sample preparation

Isotopic labelling of the proteins enables fast investigation by NMR spectroscopy since natural abundance significantly reduces measurement time. The protein expression of hen and human lysozyme mutants, in particular WT^{Ala}, W28G^{Ala}, W62G^{Ala}, W108G^{Ala}, W111G^{Ala}, W123G^{Ala}, hLys,I56T and D67H, is performed in *E.coli* cells in good yields between 10-20 mg per liter M9 minimal medium after purification, ^{15}N or ^{13}C/^{15}N labelled.

3.1.1   Isotopic labelling

The results of unfolded human lysozyme are shown representative for all expressed proteins (unfolded hen lysozyme and mutants of both proteins). After transformation of plasmid DNA into competent BL21(DE3) *E.coli* cells protein expression was induced by IPTG. The largest amount of protein is expressed within the first hour of induction. The overall expression is finished after three hours. Figure 13A shows a SDS-PAGE analysis of human lysozyme before and after IPTG induction.
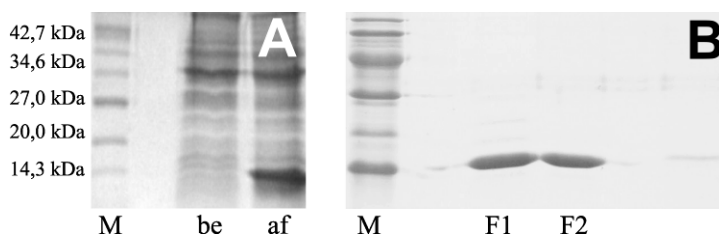
Figure 13: SDS-PAGE gel of all-ala-human lysozyme (representative also for WT^{Ala}); Both Gels have the same Marker (Broad Range, 2-212 kDa, New England BioLabs); A: SDS-PAGE before (be) and after (af) IPTG induction, the band at 14,3 kDa indicates positive expression; B: specific section of a SDS-PAGE gel of the FPLC purification of hLys, after elution of impurities (DNA and other cell components) protein fractions with identical molecular weight (eg. F1, F2) were pooled together.

The protein is expressed in inclusion bodies and purified after inclusion body solubilisation by fast protein liquid chromatography (FPLC) (figure 13B). At this step, the protein is only partially free of impurities, a subsequent HPLC purification yields in pure protein; thereby the loss through HPLC is about 30-40% of the total amount of raw protein from FPLC purification. The total purified protein yield is about 10-20mg per liter M9 medium; lyophilized protein is used for NMR sample preparation with a concentration of 300µM.

3.1.2   NMR samples

Samples were prepared with a final protein concentration of 300µM in water at pH 2, 10% D_2O, without any denaturant. Low concentrations are used to prevent unfolded states from aggregation (see concentration dependent studies of hLys and mutants in the following chapters). While the liquid NMR samples are stable over a few months, anisotropic polyacryl amide samples for RDC measurements last 2-3 days; air humidity causes irreversible damage and results in dehydration of the polyacryl amide gel material.

3.2   Analysis of unfolded states of Hen Lysozyme and mutants

Since unfolded states play a crucial role in neurodegenerative diseases (see 1.1) they are in the focus of biochemical and biophysical research. In the following chapters nonnative states of hen lysozyme and several mutants have been analyzed here by NMR with respect to structural and dynamical characteristics to elucidate the role of the unfolded ensemble in folding and misfolding processes.

3.2.1   Assignment and determination of coupling constants

The determination of chemical shift deviations, backbone and side chain coupling constants is the basis of further structural and dynamical characterization of the protein. These parameters are only accessible for a well assigned protein. As mentioned before unfolded proteins show narrow chemical shift dispersion especially in the ^{1}H-^{15}N HSQC which is one of the key common experiments for backbone ^{1}H$_N$ and ^{15}N assignment. Side chain coupling experiments for ^{1}H$^{\beta 2}$, ^{1}H$^{\beta 3}$, ^{13}C$^{\alpha}$, ^{13}C$^{\beta}$ and CO are even more complex because of the multitude of overlapping side chain proton and carbon resonances. Therefore assignment strategies by NMR are based on isotopic labelling and expansion from two dimensional to three or four dimensional space. In the following the basic assignment via HSQCs and the measurement and analysis of ^{3}J(H$_N$, H$^{\alpha}$) backbone and ^{3}J(H$^{\alpha}$,H$^{\beta}$), ^{3}J(CO,H$^{\beta}$) side chain coupling constants of unfolded states of hen lysozyme is described.

Backbone assignment from HSQC and NOESY-HSQC spectra

In figure 16 a compact overview of the ^{1}H-^{15}N HSQC spectra of the wildtype and all five mutants W28G^{Ala}, W62G^{Ala}, W108G^{Ala}, W111G^{Ala} and W123G^{Ala} is shown.

The backbone proton and nitrogen resonances ^{1}H$_N$ and ^{15}N for W28G^{Ala}, W62G^{Ala}, W108G^{Ala}, W111G^{Ala} and W123G^{Ala} have been assigned previously (*142, 143*) by comparing chemical shifts of their ^{1}H-^{15}N-HSQC spectra with the ^{1}H-^{15}N HSQC spectrum of WT^{Ala}. In figure 14 a compact overview of the ^{1}H-^{15}N HSQC spectra of the wildtype and all five mutants is shown to point out the narrow chemical shift dispersion and the similarity between the spectra.
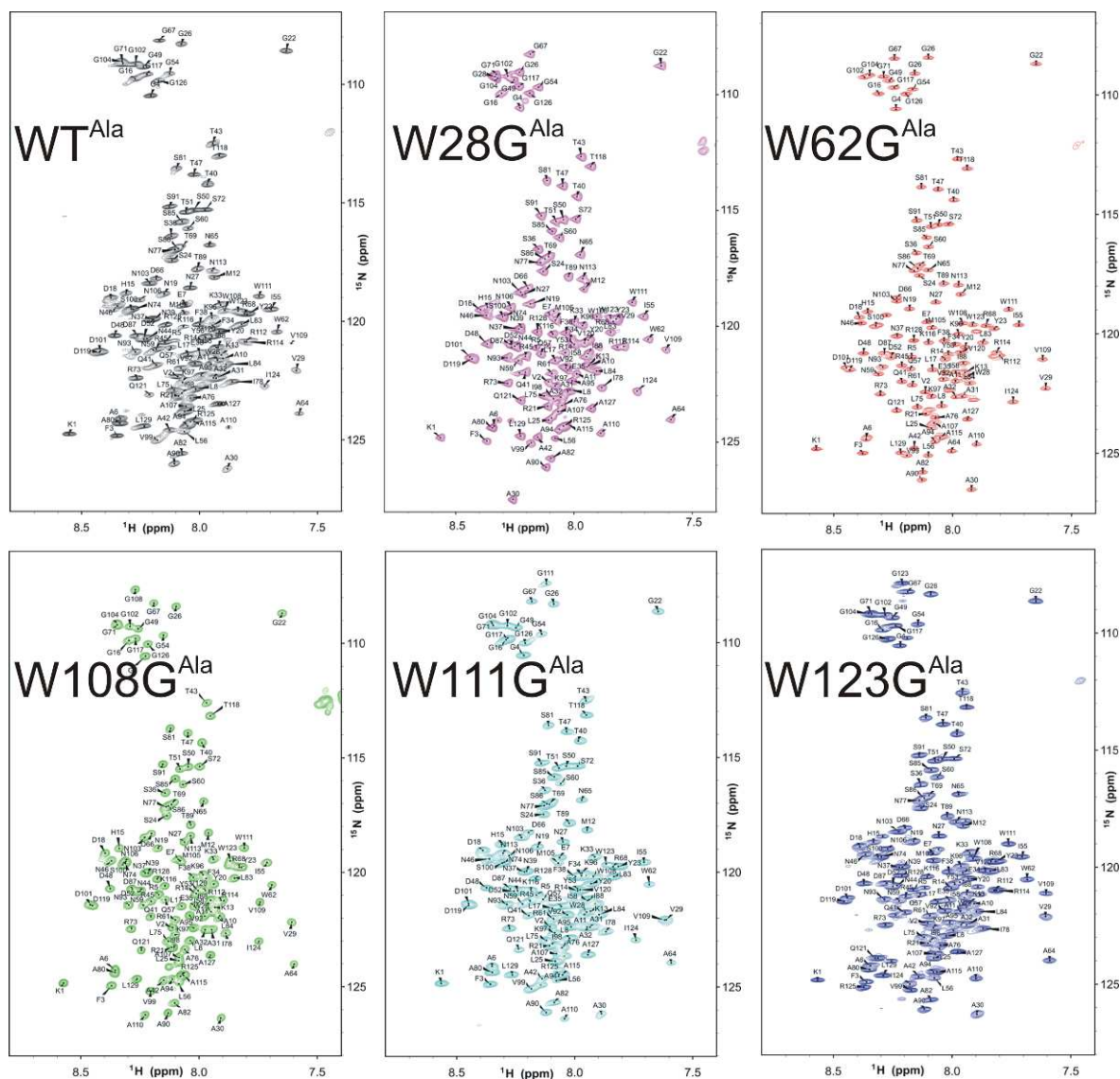
Figure 14: ^{1}H-^{15}N-HSQC spectra of WT^{Ala} and five mutants W28G^{Ala}, W62G^{Ala}, W108G^{Ala}, W111G^{Ala} and W123G^{Ala}. All spectra were recorded at 600MHz, 293K, 300µM protein concentration in water (pH2), 10% D$_2$O.

On the basis of the WT^{Ala} ^{1}H-^{15}N HSQC the assignment of the five W to G mutants was obtained by 1) comparing chemical shifts of their ^{1}H-^{15}N-HSQC spectra with the ^{1}H-^{15}N HSQC spectrum of WT^{Ala} and 2) by NOESY-HSQC spectra. Main changes in chemical shifts between wildtype and mutant were observed in regions around the mutation side (*142, 143*). Increasing influences of the mutants only on chemical shifts compared to

wildtype could be arranged in the following row: W123G^{Ala}, W28G^{Ala}, W111G^{Ala}, W108G^{Ala} and W62G^{Ala}.

Backbone coupling constants

Backbone coupling constants report on the conformation of the protein. As mentioned before, conformational averaging has to be taken into account in unfolded states. For nonnative hen egg white lysozyme here the HNHA experiment for determination of the $^3J(H_N, H^\alpha)$ coupling constants is shown in figure 15. Corresponding residues appear as cross peaks between the ^1H_N and $^1H^\alpha$ dimension; $^3J(H_N, H^\alpha)$ coupling constants are calculated by the ratio of the intensities of the cross and diagonal peaks (see 2.2.8)
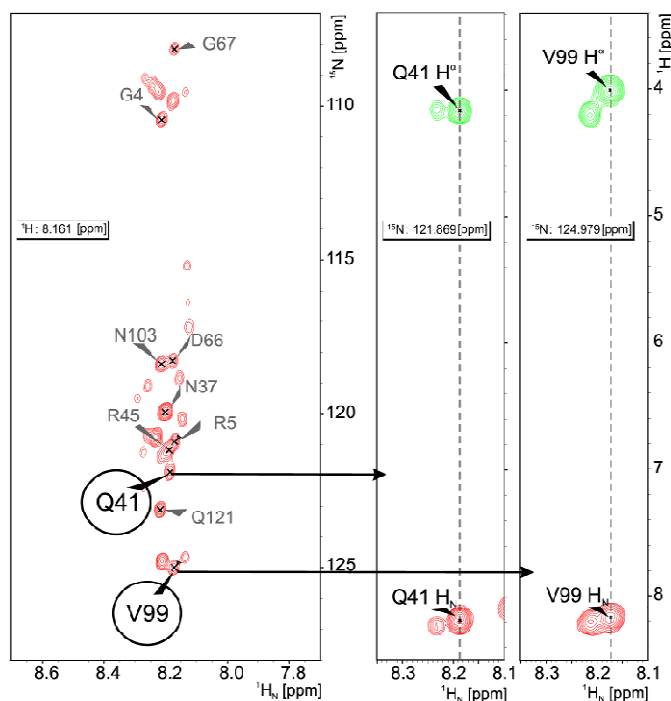
Figure 15: HNHA spectrum of WT^{Ala}; ^1H_N, ^{15}N and $^1H^\alpha$ dimension of WT^{Ala} (Q41, V99 as exemplary amino acids) left panel: ^1H_N diagonal peaks (red), shown in the ^1H_N, ^{15}N plane; right panels: $^1H^\alpha$ resonances (green); ^1H_N resonances (red); $^3J(H_N,H^\alpha)$ coupling constants can be calculated by peak intensities (see 2.2.8); spectrum was recorded at 600MHz, 293K, 300µM protein concentration in water (pH2), 10% D$_2$O.

Figure 16 shows the results of the experimental $^3J(H_N, H^\alpha)$ coupling constants of WT^{Ala} as function of amino acid side chain. Also random coil values are included for a detailed comparison. The overall couplings are smaller than those of random coil which might be due to proportion of residual secondary structure.
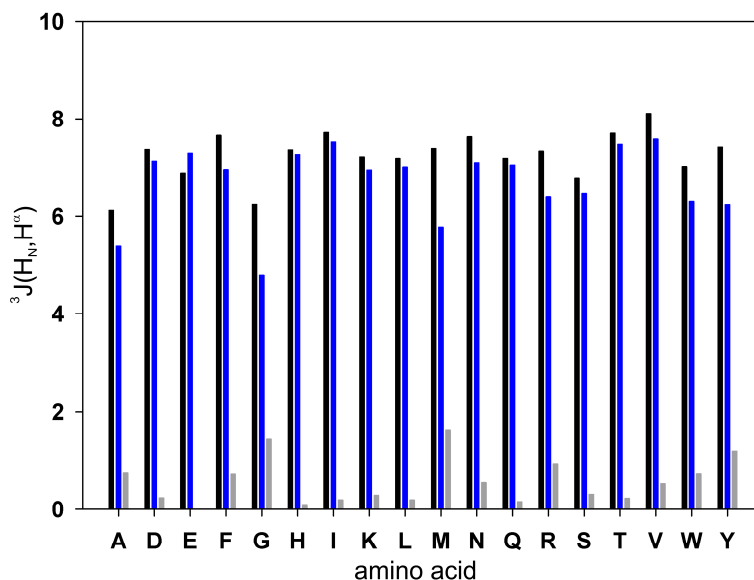
Figure 16: $^3J(H_N,H^\alpha)$ coupling constants of WT^{Ala} ; Comparison of the experimental $^3J(H_N, H^\alpha)$ coupling constants of WT^{Ala} and random coil; black: random coil data, blue: data of WT^{Ala}, grey: difference between random coil and WT^{Ala} coupling constants. $^3J(H_N,H^\alpha)$ coupling constants were extracted from peak intensities (also see chapter 2.2.8) of the HNHA spectrum shown in figure 17.

Large deviations from random coil are observed for hydrophobic amino acids like phenylalanine (F), tryptophan (W) and tyrosine (Y), also methionine (M) and arginine (R) deviate a lot. All these residues share sterically demanding side chains which might support the formation of residual structure elements and long range interactions. This will be further proofed by the analysis of other coupling constants of backbone and side chain in the following chapters. Medium changes are observable for amino acids with aliphatic side chains, isoleucine (I), leucines (L) and valine (V).

The $^1J(N_i,C_\alpha)$ and $^2J(N_i,C_{\alpha(i-1)})$ couplings also report on conformational preferences of the protein backbone $^1J(N,C^\alpha)$ coupling constants depend weakly on $\psi_{(i)}$ while $^2J(N,C^\alpha)$ coupling

constants depend on the $\phi_{(i-1)}$ and the secondary structure determining angle $\psi_{(i-1)}$ (see 1.2.6). Their determination results from a series of J-modulated ^{1}H,^{15}N-HSQCs with an exponential decay of the signal intensity. Figure 17 shows the results of the ^{1}J(N$_i$,C$_\alpha$) and ^{2}J(N$_i$,C$_{\alpha(i-1)}$) of WT$^{\text{Ala}}$.
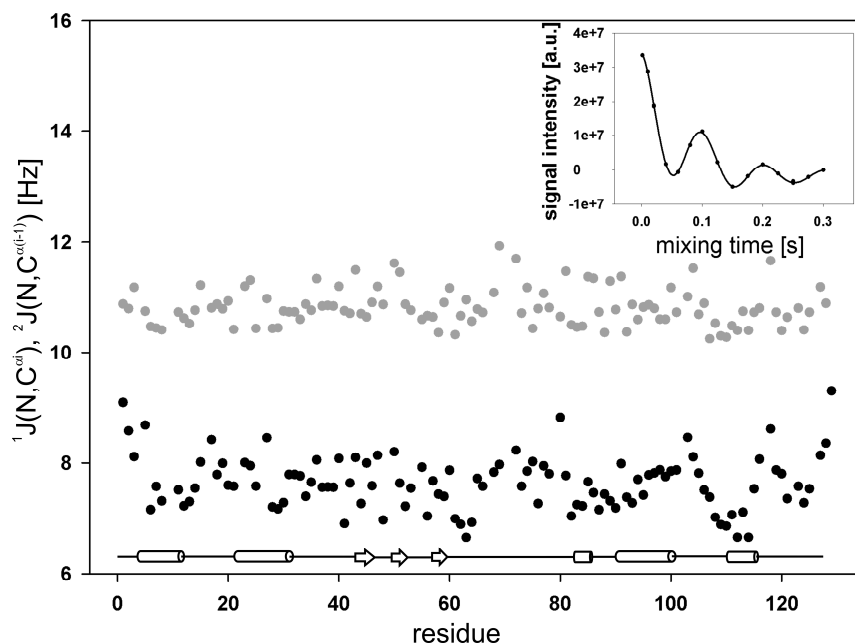
Figure 17: ^{1}**J(N,C$^{\alpha i}$),** ^{2}**J(N,C$^{\alpha(i-1)}$) coupling constants of WT$^{\text{Ala}}$** ; Grey: ^{1}J(N,C$^{\alpha i}$) couplings of WT$^{\text{Ala}}$, black: ^{2}J(N,C$^{\alpha(i-1)}$) couplings of WT$^{\text{Ala}}$, insert: exemplary function of the decay of signal intensity. Secondary structure elements of the folded state are schematically shown as cylinders (α-helices) and arrows (β-sheets). Coupling constants were extracted from of J-modulated HSQCs (see chapter 2.2.8), measured at 600MHz, 293K, pH 2 at a protein concentration of 300µM.

The ^{1}J(N,C$^{\alpha i}$) couplings for WT$^{\text{Ala}}$ range between 10-12Hz whereas the ^{2}J(N,C$^{\alpha(i-1)}$) couplings have values between 6Hz and 8Hz. In unstructured peptides values for ^{1}J(N,C$^{\alpha i}$) are around 11Hz and ^{2}J(N,C$^{\alpha(i-1)}$) around 7Hz (*116*). Higher values indicate β-sheet structure, lower values lead to α-helical structure. Unfolded hen egg white lysozyme shows on average coupling constants close to. This has been previously observed for unfolded ubiquitin (*113*). More details of side chian and backbone coupling constants and their influence on the unfolded ensemble are discussed in detail in chapter 3.2.2.

Side chain coupling constants

Side chain couplings in unfolded hen egg white lysozyme light up a detailed structural and dynamical characterization of the protein. Notably experiments reporting on side chain torsion angles are of great interest for structural predictions through angle populations of the conformational ensembles. An extended analysis of side chain couplings, in detail $^3J(H^\alpha,H^\beta)$ and $^3J(CO,H^\beta)$, has been performed in this thesis and is described in the following.

Figure 18 shows an exemplary strip plot of the HN(COCA)HAHB (*117*) experiment of WT^{Ala}. This NMR experiment reports on the $^3J(H^\alpha,H^\beta)$ coupling constant and therefore correlates the $^1H^\alpha$ with the $^1H^\beta$ proton of the particular amino acid. The proton resonances have different signs, H^α positive and H^β negative, which facilitates the analysis.
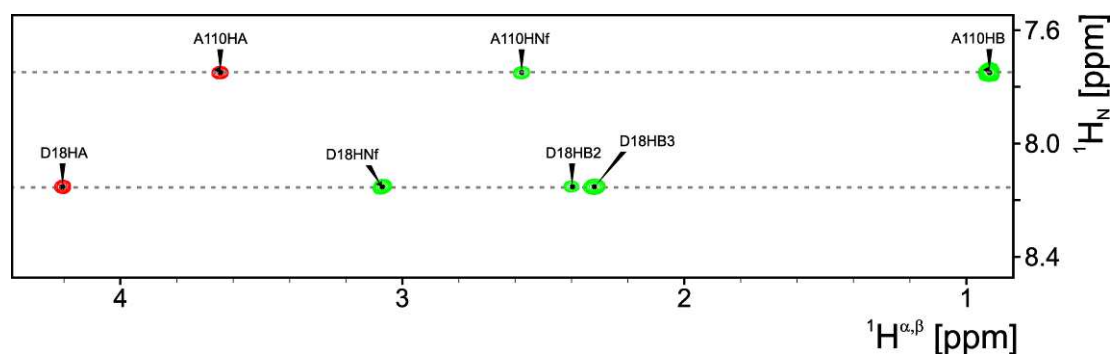
Figure 18: HN(COCA)HAHB spectrum of WT^{Ala} ; ^1H_N and $^1H^{\alpha,\beta}$ dimension of WT^{Ala} with A100 and D18 as exemplary amino acids; red peaks: $^1H^\alpha$ (indicated by HA) resonances, green peaks: $^1H^{\beta2}$, $^1H^{\beta3}$ (indicated by HB2, HB3) and folded ^1H_N (indicated by HNf) resonances; spectrum was recorded at 700MHz, 293K, 300μM protein concentration in water (pH2), 10% D_2O.

Despite small dispersion of some H^β protons, especially those of serine, glutamine and glutamate the overall dispersion was promising enough to analyze about 70% of all side chain coupling constants. Beta-branched amino acids, here valine and isoleucine show large dispersions in the H^β plane. The H^α resonances have been already assigned with the HNHA experiment described above. The coupling constants were determined by their intensity ratios (for equations see chapter 2.2.8).
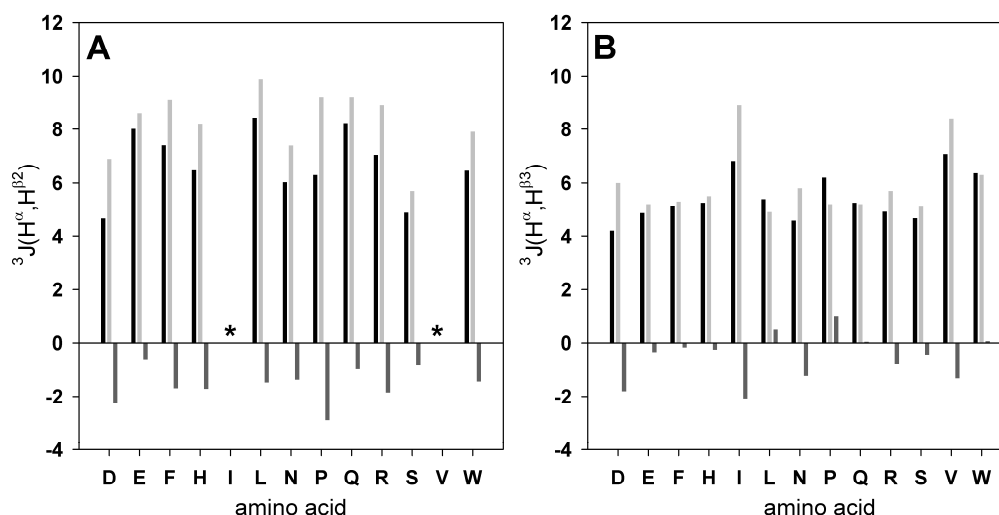
Figure 19: $^3J(H^\alpha,H^\beta)$ coupling constants of WT^{Ala}; amino acid specific plot of the available $^3J(H^\alpha,H^\beta)$ coupling constants from the HN(COCA)HAHB experiment; black: data of WT^{Ala}, grey: random coil (rc), dark grey: difference between WT^{Ala} and rc; A: couplings between $^1H^\alpha$ and $^1H^{\beta2}$; B: couplings between $^1H^\alpha$ and $^1H^{\beta2}$; asterisk: β-branched amino acids isoleucine and valine.

Analysis of the coupling constants (figure 19) shows lower values for hen egg white lysozyme compared to those of random coil. This indicates no clear side chain angle preference but an average of *gauche* and *trans* conformation in WT^{Ala}. The large deviation of charged amino acids can be explained by pH dependency. For stereospecific assignment of the H$^\beta$ protons at least one additional side chain coupling has to be measured. A detailed analysis of degree of deviation from random coil is performed in chapter 3.2.2 by a Pachler-type analysis where the angle populations are calculated. For this analysis two different couplings sharing one resonance are needed. Therefore the HN(CO)HB experiment for $^3J(CO,H^\beta)$ coupling constants has been measured. Figure 20 shows an exemplary strip plot of the spectrum in the ^1H_N and $1H^{\alpha,\beta}$ dimension. Here the amino acids I88 and R112 were depicted for a detailed view on the spectrum. Resolved peaks represent the ^{13}CO carbon resonances and the corresponding H$^\beta$ protons. The overall dispersion of the resonances $^1H^\beta$ and ^{13}CO was well resolved so that nearly 70% of all coupling constants could be analyzed.
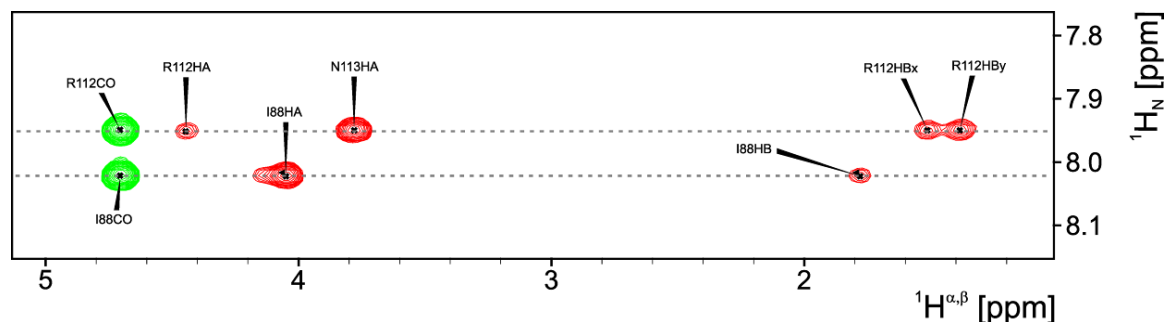
Figure 20: HN(CO)HB spectrum of WT^{Ala} ;^1H_N and $^1H^{\alpha,\beta}$of WT^{Ala} (exemplary R112 and I88); red peaks: $^1H^{\alpha}$ (indicated by HA) and $^1H^{\beta2}$, $^1H^{\beta3}$ (indicated by HBx, HBy) resonances; green peaks: ^{13}CO cross peaks with ^1H_N backbone resonances; spectrum was recorded at 900MHz, 293K, 300μM protein concentration in water (pH2), 10% D_2O.

Coupling constants were calculated as described in the materials and methods section (1.2.6). In figure 21 results of the analysis of WT^{Ala} compared to random data are shown. The $^3J(CO,H^{\beta})$ couplings are smaller than random coil values due to conformational averaging of the unfolded ensemble and to residual structure.
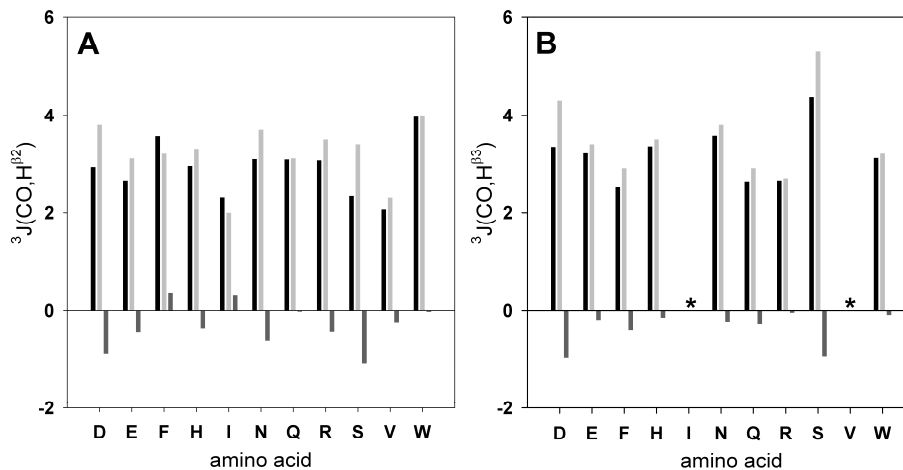
Figure 21: ^{3}J(CO,H$^{\beta}$) coupling constants of WT^{Ala}; amino acid specific plot of the available $^3J(CO,H^{\beta})$ coupling constants from the HN(CO)HB experiment; black: data of WT^{Ala}, grey: random coil (rc), dark grey: difference between WT^{Ala} and rc; A: couplings between $^1H^{\alpha}$ and $^1H^{\beta2}$; B: couplings between $^1H^{\alpha}$ and $^1H^{\beta2}$; asterisk: β-branched amino acids isoleucine and valine with only one H$^{\beta}$ proton.

3.2.2 Backbone conformational preferences

On the basis of the assignment from HSQCs and the data from coupling constant measurements, in the following section chemical shifts have been extracted and interpreted with respect to structural characteristics of WT^{Ala}. For the confirmation of residual structure in unfolded states of hen egg white lysozyme secondary structure propensities have been predicted with the free available software at http://pound.med.utoronto.ca/software.html (*53*). In figure 22, resulting secondary structure propensities (SSP) obtained from C^α and C^β chemical shifts are shown together with ^1H_N and $^1H^\alpha$ chemical shift differences from random coil (*76*) and $^3J(H_N,H^\alpha)$, $^2J(N,C^\alpha)$ and $^1J(N,C^\alpha)$ backbone coupling constants for WT^{Ala}.
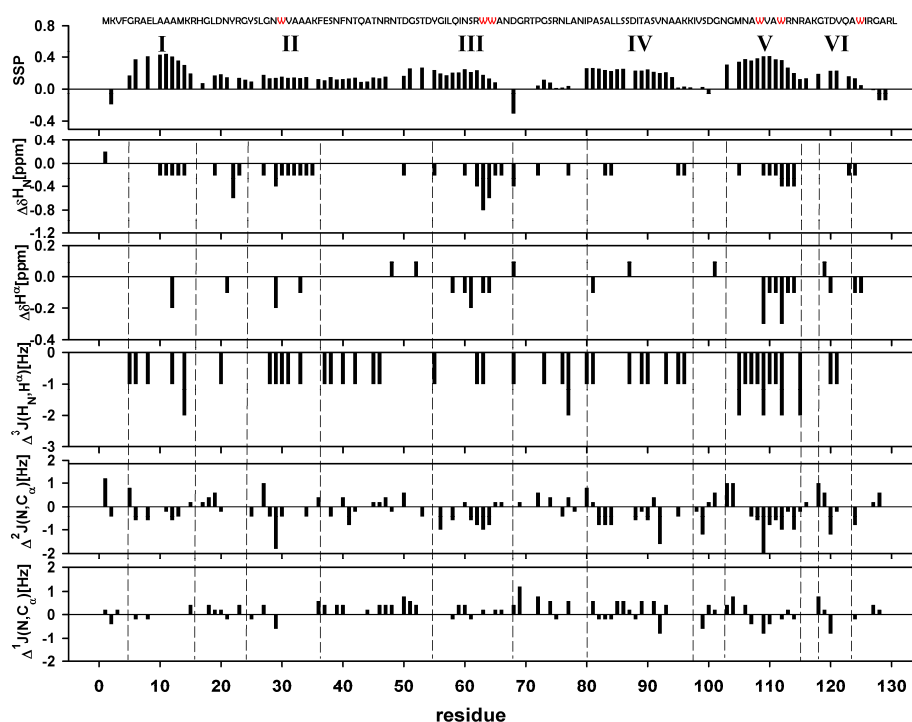
Figure 22: Residual secondary structure in WT^{Ala}; Secondary structure propensities (SSP), $\Delta\delta H_N$, $\Delta\delta H^\alpha$, $\Delta^3J(H_N,H^\alpha)$, $\Delta^2J(N,C^\alpha)$ and $\Delta^1J(N,C^\alpha)$ are indicated by black bars, black lines represent the hydrophobic clusters of WT^{Ala}. The values for $\Delta\delta H^\alpha$ and $\Delta\delta H^N$ were normalized by using value 0 ranging from 0.0 to 0.049, 0.1 from 0.05 to 0.149, value 0.2 from 0.15 to 0.249, etc., $\Delta\delta$ was calculated by $\Delta\delta=\delta_{exp}-\delta_{rc}$. All experiments have been carried out at 600MHz with 300µM protein at pH2, 293K.

WT^{Ala} mostly exhibits positive SSP values, indicating a slight propensity for α-helices in the unfolded protein, negative SSPs indicate β-sheet propensity. Positive values with structure propensities larger than 0.2, are observed for the regions ranging from residue 5-12 (region 1), 24-36 (region 2), 53-67 (region 3), 79-96 (region 4), 103-113 (region 5) and around 120 (region 6). In the ^1H_N and $^1H^\alpha$ chemical shifts compared to random coil significant perturbations are also observed in the regions mentioned above. $^3J(H_N,H^\alpha)$ coupling constants are slightly lower than the predicted random-coil values throughout the protein. These deviations from random-coil indicate small deviations of the ϕ angle values from the angles observed in the random-coil database; an observation which was already seen in other unfolded proteins(*144*). The dependence of $^1J(N,C^\alpha)$ on the angle $\psi_{(i)}$ and $^2J(N,C^\alpha)$ on the angles $\phi_{(i-1)}$ and $\psi_{(i-1)}$ allows differentiation between positive and negative ψ values. Here the propensity to form α-helical structure elements in the regions 1-6 is also observed by the $^2J(N,C^\alpha)$ coupling constants particularly in regions (2), (3), (5) and (6). All data shown in figure 22 consistently show that six regions adopt non-random conformation towards α-helical structure, four of these regions contain at least one of the 6 tryptophan residues.

3.2.3   Side chain conformational preferences

NMR experiments reporting on side chain torsion angles are of great interest for structural predictions through angle populations of the conformational ensembles. With the experiments described under 3.2.1, $^3J(H^\alpha,H^\beta)$ and $^3J(CO,H^\beta)$ couplings were analyzed. A Pachler-type analysis of the coupling constants results in a stereospecific assignment of the H$^\beta$ protons. Figure 23 shows a comparison of the $^3J(H^\alpha,H^\beta)$ and $^3J(CO,H^\beta)$ coupling constants with random coil predictions and the mean percentage populations of both.
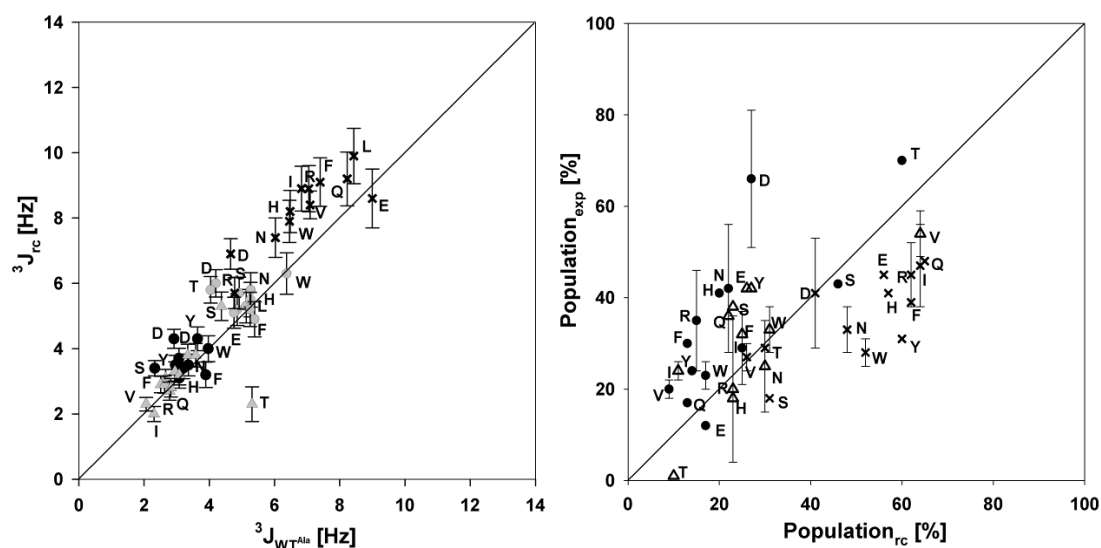
Figure 23: Side chain coupling constants of WT^{Ala}; Comparison of $^3J(CO,H^\beta)$ and $^3J(H^\alpha,H^\beta)$ coupling constants of WT^{Ala} with random-coil predictions; black circle: $^3J(H^\alpha,H^{\beta2})$, grey triangle: $^3J(H^\alpha,H^{\beta3})$, black cross: $^3J(CO,H^{\beta2})$, grey circle: $^3J(CO,H^{\beta3})$. Right: Mean percentage populations of WT^{Ala} vs. random-coil (cross: -60°, circle: 60°, triangle: 180°). All experiments have been carried out at field strength corresponding to a ^1H frequency of 600MHz, using 300µM protein at pH2, T=293K.

The comparison of the coupling constants of WT^{Ala} and random-coil shows small deviations from the predicted values. $^3J(H^\alpha,H^\beta)$ couplings and even more the $^3J(CO,H^\beta)$ couplings of WT^{Ala} are on average slightly lower compared to predicted random coil coupling constants (figure 23). For aspartate (D), deviations in both coupling constants are observable, presumably due to changes in the backbone side chain interactions caused by low pH. Phenylalanines (F), histidines (H) and arginines (R) show deviations mainly in the $^3J(H^\alpha,H^\beta)$ couplings, whereas serines (S) and threonines (T) deviate larger in $^3J(CO,H^\beta)$ couplings.

The angular distributions of the three rotamers of the side chain angle χ_1 were calculated using a Pachler-type analysis. Table 3 shows all populations of the rotamers for selcted amino acids. Here populations could be obtained for 26 residues were both $^3J(H^\alpha,H^\beta)$ and $^3J(CO,H^\beta)$ are available. The random-coil model predicts a preference for the χ=-60° rotamer for all amino acids but for valine, serine and threonine (*145*).

	WT^{Ala}			random-coil			
	-60°	**60°**	**180°**	**-60°**	**60°**	**180°**	
E (1)	45	12	42	56	17	26	al
I (3)	47 (±9)	29 (±8)	24 (±2)	64	25	11	al
Q (1)	48	17	36	65	13	22	al
R (4)	45 (±7)	35 (±11)	20 (±16)	62	15	23	al
D (2)	41 (±12)	66 (±15)	-8 (±5)	41	27	32	Asx
N (4)	33 (±5)	42 (±14)	25 (±10)	48	22	30	Asx
S (1)	18	43	38	31	46	23	OH
T (1)	29	70	1	30	60	10	OH
V (4)	27 (±3)	20 (±2)	54 (±5)	26	9	64	Val
H (1)	41	41	18	57	20	23	ar
F (1)	39	30	32	62	13	25	ar
W (2)	28 (±3)	38 (±3)	33 (±5)	52	17	31	ar
Y (1)	31	24	42	60	14	27	ar

Table 3: Population of side chain torsion angles; Population of different χ_1 angles in WT^{Ala} (left) and random coil (right). Counts in brackets indicate the numbers of amino acids for which the populations have been calculated; al: aliphatic residues, Asx: asparagine, aspartate, OH: threonine residue, Val: valine, ar: aromatic residues.

For long aliphatic amino acids (here E, R, Q, I), the predicted preference is not as pronounced as predicted: While the random coil model predicts populations of 56-65% for χ_1 =-60°, here the averaged χ_1 =-60° populations are 45-48%. Aspartate (D) and asparagine (N) residues show large deviations from random-coil: while the model predicts a very slight preference for the -60° conformer, here the +60 rotamer is the most populated one. As predicted in the model, the 180° rotamer is preferred in our experimental data for valines (V), allowing the positioning of the methyl groups furthest away from the backbone carbonyl group. This preference is not as pronounced as predicted, which differs from previous measurements on ubiquitin, protein G and oxidized lysozyme (*145, 146*) all unfolded in 7.4-8M urea at acidic pH. Interestingly, the difference could hint at an effect of urea on sidechain sampling in unfolded proteins.

For the OH containing amino acids serine (S) and threonine (T), the random-coil model predicts a preference for $\chi_1 = +60°$ of 46% for serine and 60% for threonine, most likely caused by polar interactions between the side chain and the backbone. In the data measured here, the preference is similar for the +60° angle with 43% for S and 70% for T, however, with the populations of the 180° and the -60° conformer varying significantly. On average significant deviations for serine and threonine residues were also observed in previous data (*145, 146*) and thus might be a general feature at acidic pH.

Aromatic residues deviate significantly: rather than exhibiting a clear preference for the -60° rotamer as predicted in the random-coil model, experimentally determined populations differ around 20% for almost each rotamer. For histidine (H), the -60° and the +60° rotamer are equally populated, in the case of phenylalanine (F) and tyrosin (Y), the populations of all three rotamers are similar. Most drastically, in tryptophan (W108, W63), the 180° rotamer is clearly preferenced over the other two rotameric states. While it was only possible to obtain both coupling constants for W108 and W63 allowing a Pachler-type analysis, coupling constants measured for the other tryptophans are in the same range, and it is thus very likely that the rotameric distribution for these residues also deviate from the random-coil. Similarly, deviations from the random-coil model were also observed for the tryptophan residues in urea denatured oxidized hewl (*145*), but not for urea denatured protein G and ubiquitin (*146*). The latter two have little residual structure in the denatured forms while the contrary is the case in lysozyme (*144, 147*). Thus, the here observed deviations emphasize the importance of aromatic residues and particular of tryptophan residues for the formation of non-random residual structure, also observed in the protein backbone.

Conformational preferences from ASTEROIDS

Conclusions from backbone and side chin conformational preferences are further maintained by the results obtained from ASTEROIDS (*131*) predictions. Calculations were performed incorporating H_α (used as cross-validation), C_α, CO and C_β chemical shifts, $^3J(H^N,H_\alpha)$, $^1J(N,C_{\alpha i})$, $^2J(N,C_{\alpha i-1})$ coupling constants and ^1H-^{15}N RDCs (see materials and methods, chapter 2.2.9). Figure 24 shows the correlation between experimental and predicted parameters.
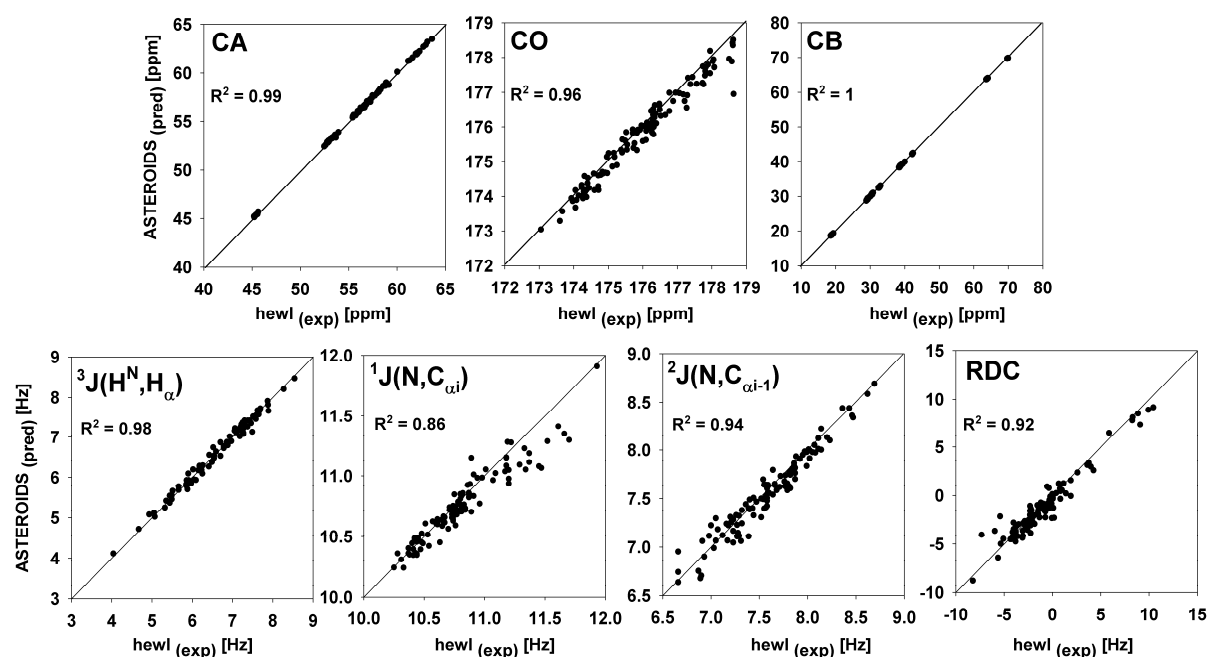
Figure 24: Chemical shift and coupling correlations from ASTEROIDS selection; Correlation of experimental and predicted (Asteroids) chemical shifts C_α, CO and C_β and coupling constants as $^3J(H^N,H_\alpha)$, $^1J(N,C_{\alpha i})$, $^2J(N,C_{\alpha i-1})$ and ^1H-^{15}N RDCs of WT^{Ala}. R^2 values indicate the accuracy of the data selected by ASTEROIDS calculations.

With a mean R^2 value between 0.83 and 1.0 for the backcalculation of the predicted chemical shifts and coupling constants the selected ensemble from ASTEROIDS matches very well with experimental data. Unfolded ensembles of WT^{Ala} selected by this algorithm show a more detailed view on secondary structure elements weighting populations of PPII-, α_R- and α_L-helix and β-sheet (figure 25).
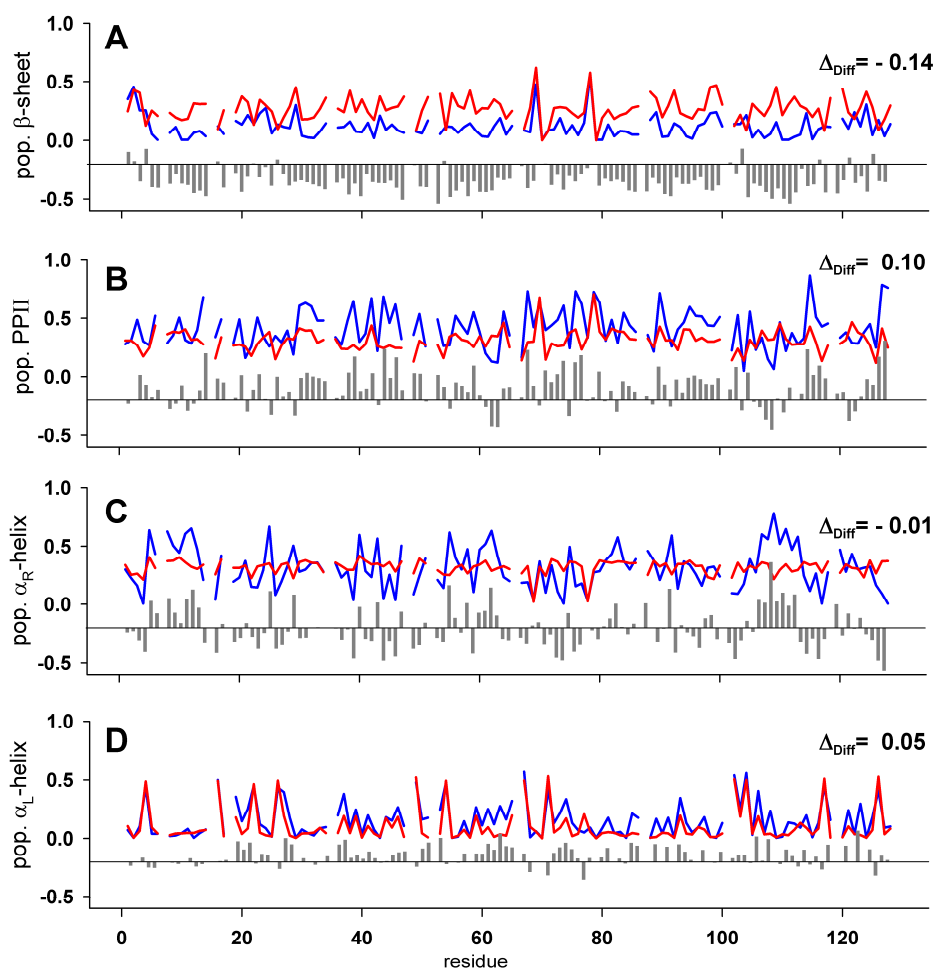
Figure 25: Selected populations of WT^{Ala} from ASTEROIDS calculations; Selected ASTEROIDS populations of secondary structure elements (A: β-sheet, B: polyproline II, C: $α_R$-helix, D: $α_L$-helix) of WT^{Ala} (excluding pH dependent amino acids D, E and H); red: flexible meccano ensemble (15.000 structures); blue: ASTEROIDS selection for WT^{Ala} (200 selected structures); grey bars: difference between selected structures and flexible meccano ensemble; $Δ_{Diff}$: mean difference values (calculated by $Δ_{Diff} = Diff_{selected}-Diff_{flecmec}$).

Deviations compared to the flexible meccano ensemble give information about the quality of selected structures. While polyproline II (figure 25B) is continuously over-represented in the ASTEROIDS selection with a positive mean difference value of $Δ_{Diff} = 0.1$, β-sheet structure (figure 25A) is in an analogous manner under-represented with a negative value of $Δ_{Diff} = -0.14$. Even though the mean difference values of a_R-helix ($Δ_{Diff} = -0.01$) (figure 25C)

and a_L-helix ($\Delta_{Diff} = 0.05$) (figure 25D) are in the same range, they deviate significantly in the population pattern. In the α_R-helix deviations from flexible meccano show clustered regions at residues 5-13, 55-63, 107-113 and 120-124 wich are in good agreement to hydrohobic regions in WT^{Ala} while these clusters are not visible in the α_L-helix. Positive deviations from the flexible meccano ensemble are also visible in the PPII population, these deviations are largest mainly around residues 14, 31, 44, 74, 96 and 115.

3.2.4   Protein dynamics by heteronuclear relaxation and residual dipolar couplings

Backbone as well as side chain conformational preferences indicate residual helical structure in unfolded hen egg white lysozyme. To confirm these findings, dynamics of WT^{Ala} have been analyzed in this thesis by comprehensive relaxation studies. NMR parameters used here are R_1, R_2 relaxation rates, hetNOE data (figure 26) and residual dipolar couplings (figure 27). Although all these parameters report on different timescales, data show the same sequence dependence: Low values can be found at the termini of the protein while in the central part a plateau value is reached from which deviations towards higher values are observed. As a general tendency in all relaxation data, strongly deviating residues are mainly located around hydrophobic tryptophan residues (figure 26A, B, C).
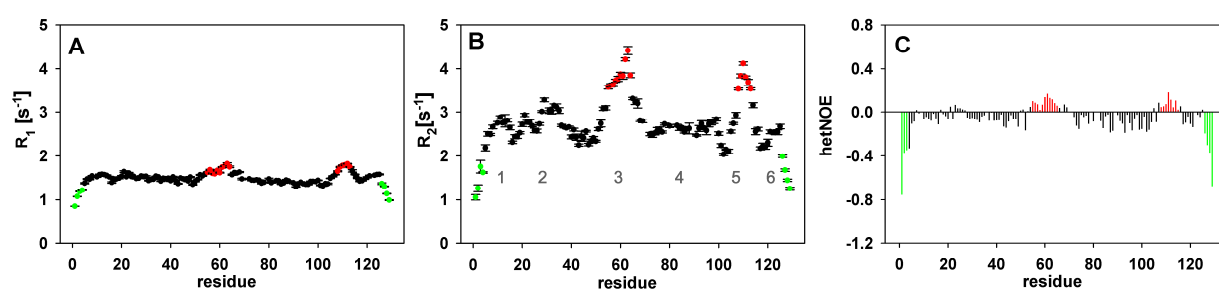
Figure 26: Relaxation and RDC data of WT^{Ala}; A: R_1 relaxation rates; B: R_2 relaxation rates (regions with high values are numbered with 1-6); C: hetNOE; green: amino acids with low rates at the N- and C-terminus (flexible regions) (600MHz, using 300µM protein at pH2, T=293K)

Most pronounced deviations are found in region 3 around W62 and W63 with a shoulder towards I56. Very high deviations are also observed in region 5 around W108 and W111. In

both of these regions also elevated R_1 relaxation rates and positive hetNOE values are observed, while hetNOE values are predominantly negative for the rest of the protein as expected for an unfolded polypeptide chain. High deviations of R_2 relaxation rates are also found in region 2 from 24 to 36 around W28. In this region, however, neither R_1 rates nor the hetNOE are significantly increased compared to the basal level. Regions with medium sized deviations are observed around A10, G22 and W123. For structural and dynamical details also residual dipolar couplings have been analyzed here. Thereby experimental residual dipolar couplings as a function of the protein sequence (figure 27A) are compared with predictions made from flexible meccano (figure 27B).
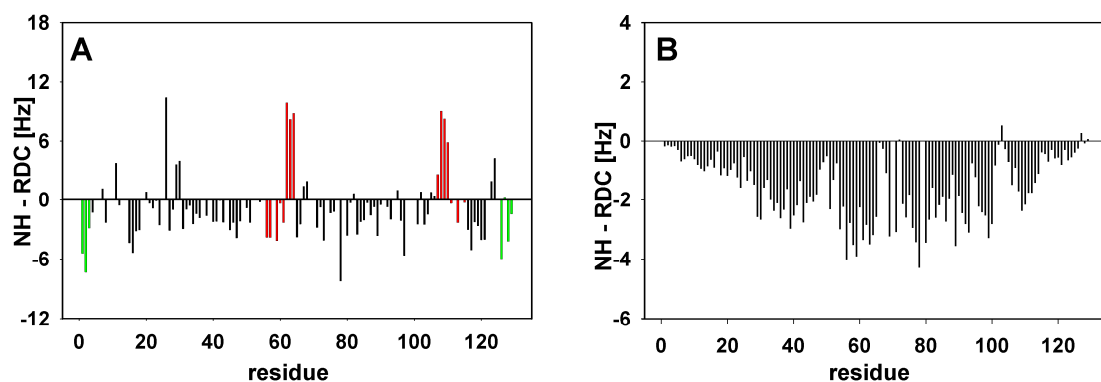
Figure 27: Relaxation and RDC data of WT^{Ala} A: experimental N-H RDCs and B: predicted NH-RDCs (by flexible meccano), red: amino acids with high rates (rigid regions), green: amino acids with low rates at the N- and C-terminus, flexible regions (600MHz, using 300µM protein at pH2, T=293K)

The overall shape of predicted data is bell-like in agreement with the smooth distributions expected for random flight chains. In contrast to the homopolymeric random flight chain, a rough landscape is predicted from flexible meccano due to residue-specific conformational preferences. Local or long-range interactions leading to residual structure are not included in the predictions. This is one explanation for significant deviation of experimental RDCs from predicted data. In general negative RDC values are expected for WT^{Ala}, but a number of positive values are found in the regions where hydrophobic interactions have been observed by relaxation analysis.

In particular these are the regions 1, 2, 3, 5 and 6. Interestingly, while RDC data are positive in the right part of region 3, RDCs in the left part are negative. These findings indicate that residual structure and dynamic restrictions at this part have different time scales. While in the right part around W62/W63 residual structure seems to persist on the millisecond time scale, dynamic restrictions persist on much smaller time scales in the left part around I56. A more detailed view into dynamics enables a reduced spectral density analysis (figure 28, 29).

On the one hand there are very flexible residues at the termini of the protein where motions are only governed by internal motions (green), on the other hand residues with motions dominated by internal motions can be found in the central part of the protein and at least there are residues in which motions are determined by internal and overall motions (red), being thus more rigid than expected for a perfectly unfolded polypeptide chain.
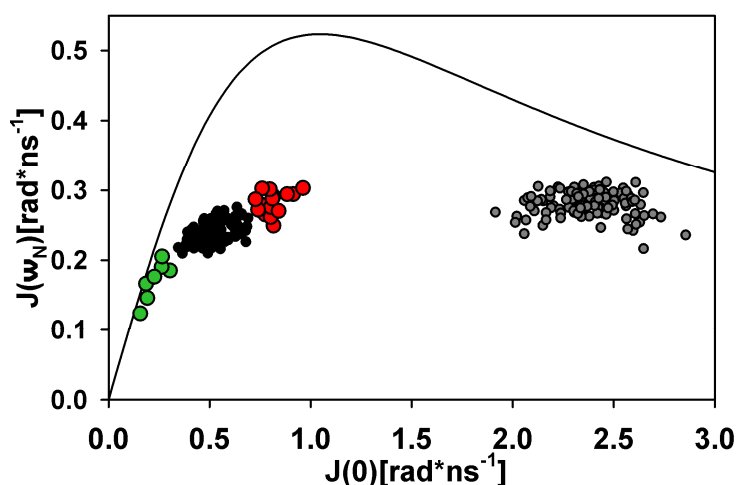
Figure 28: Relaxation and RDC data of WT^{Ala} spectral density mapping; Black line: spectral density function (dynamics limited to a single motion with $J(\omega) = J(0)/(1+6,25(\omega J(0))^2$ (*85*); grey circles: data of native hen egg white lysozyme (data taken from R. Silvers, 2011).green: amino acids with low rates at the N- and C-terminus (flexible regions) (600MHz, using 300µM protein at pH2, T=293K)

For comparison, in folded hen lysozyme (figure 29, right panels) motions are dominated by the overall tumbling for all residues but for G22 where chemical exchange is observed, shown by the significant shift towards a higher J(0) (figure 29, right panel, bottom). Results

of the detailed comparison of unfolded and folded hen lysozyme reflect the relaxation behaviour of the protein states.
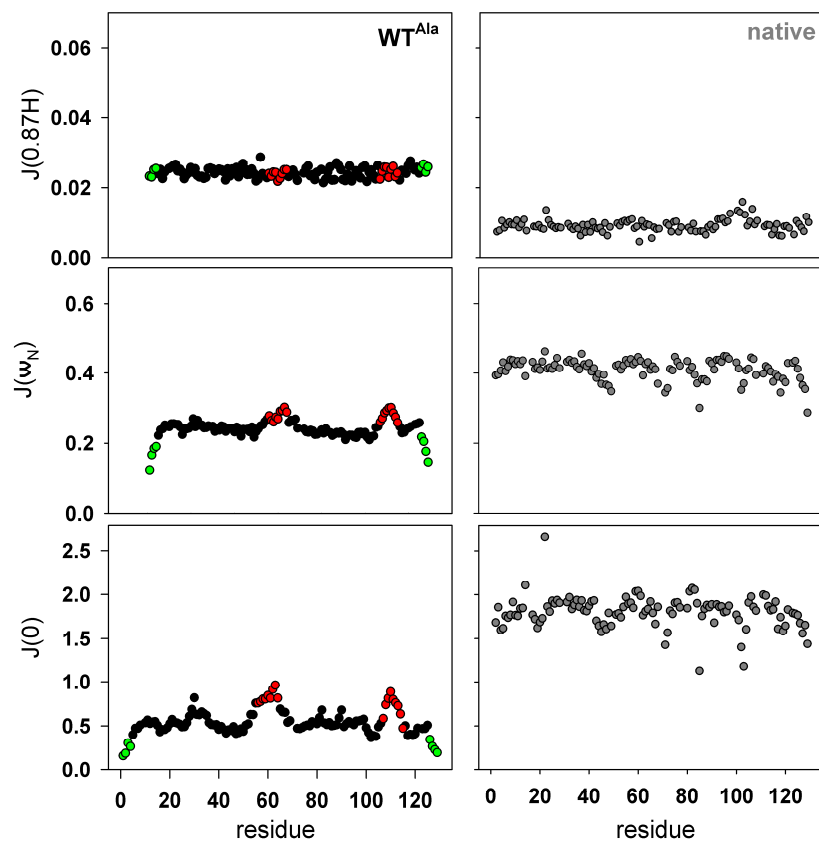
Figure 29: Spectral density data of WT^{Ala}; reduced spectral densities J(0.87H), J(ω_N) and J(0) of WT^{Ala} (left, black) and native lysozyme (right, grey, data taken from R. Silvers, 2011). Data derive from R_1, R_2 relaxation rates and hetNOE measurements, green: flexible regions, red: rigid regions (600MHz, 293K, pH 2 (WT^{Ala}), pH 3.8 (native hewl) 300µM protein concentration).

Regions with limited motional properties are confirmed. The presence of exchange for the unfolded polypeptide can be excluded by graphical analysis of the spectral density data for WT^{Ala}, where no parallel shift towards higher values of J(0) is observable. This is also shown by the analysis of relaxation dispersion measurements (figure 30) (*142*).
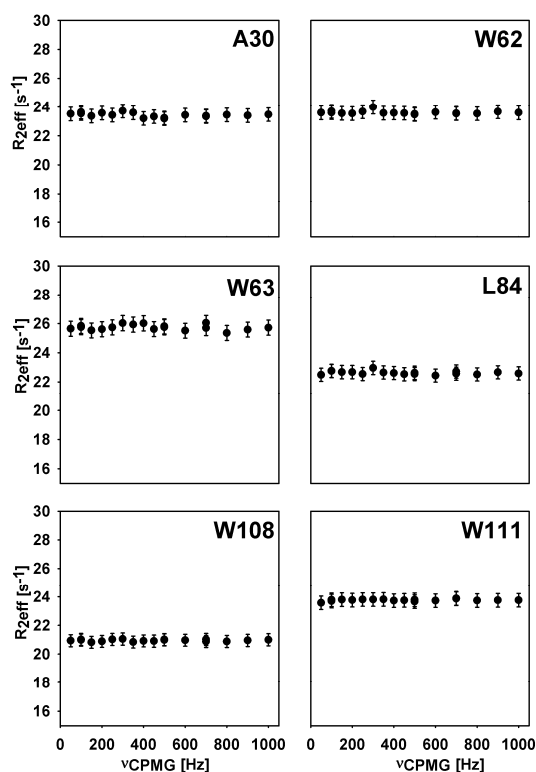
Figure 30: Relaxation dispersion measurements of WT^{Ala}; ^{15}N single-quantum relaxation rates R$_2$eff[s^{-1}] as a function of CPMG frequency for selected residues of WT^{Ala} (data taken and figure adapted from (*142*)).

A short summary of all relaxation data of WT^{Ala} clearly reveals residual structure to be present in at least 5 regions of the protein. By comparison of heteronuclear relaxation rates and residual dipolar couplings, it becomes clear that this residual structure persists from the subnanosecond time regime up to the millisecond time regime. These findings will be further supported by the analysis of single point mutants of WT^{Ala} and their structural and dynamical characteristics.

3.2.5 Modulation of the conformational ensemble by single point mutations

Further insight into long-range interactions within the ensemble of unfolded conformations was obtained here by characterization of non-conservative single point mutations as previously reported for other variants of lysozyme by Wirmer *et al.*, 2006. According to this, here five tryptophan-to-glycine mutants, W28G^{Ala}, W62G^{Ala}, W108G^{Ala}, W111G^{Ala} and W123G^{Ala}, within prominent regions 1-6 identified above in WT^{Ala} were analyzed in terms of chemical shifts, relaxation, residual dipolar couplings and diffusion measurements.

Chemical shift analysis

Since chemical shifts contain information about secondary structure elements and tertiary contacts (*68, 69*), mainly influenced by noncovalent interactions, it is used here for the delineation of mutational effects in WT^{Ala}. In figure 31 the correlation of the H^N chemical shifts from WT^{Ala} with the five mutants is shown. Chemical shifts were extracted from ^{1}H-^{15}N HSQCs by comparing them with the spectrum of the wildtype. All mutants have significant differences in the H^N chemical shifts around the mutation site. This is due to nearest neighbour effects. Larger effects are seen in the mutants W62G^{Ala} and W08G^{Ala}. In addition, the W108G^{Ala} mutant shows deviations in H^N chemical shifts in a region further away from the mutation site, at Gly71 and Ser72. This indicates long-range interactions of this region with the region around W108.
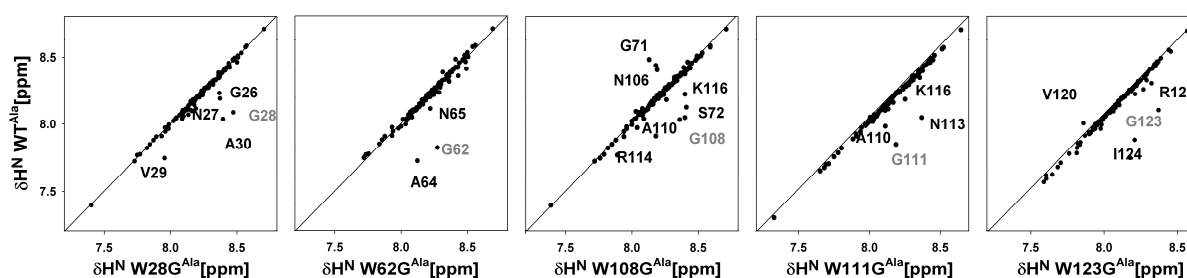
Figure 31: Chemical shift analysis of mutants; : H$_N$ chemical shift correlation of the mutants W28G^{Ala}, W62G^{Ala}, W108G^{Ala}, W111G^{Ala} and W123G^{Ala} versus wildtype WT^{Ala}. The mutation site is highlighted (grey) deviating residues are indicated. Experiments have been carried out at field strength corresponding to a ^{1}H frequency of 600MHz, 293K, using 300µM protein at pH2.

Because chemical shifts only report on structural changes of the mutants W28G^{Ala}, W62G^{Ala}, W108G^{Ala}, W111G^{Ala} and W123G^{Ala} compared to WT^{Ala} it is very interesting to analyze changes on the dynamical level. This is done in the following section where the dynamics of the mutants were analyzed with NMR experiments reporting on different timescales, in particular relaxation analysis and residual dipolar couplings.

Relaxation and residual dipolar couplings

Relaxation data and RDCs report on the subnanosecond to millisecond time regime. To compare both parameters, histograms of R$_2$ and RDCs of all mutants and the wildtype were generated. Before discussing the mutants, results of the R$_2$ and RDC tensor plots for the wildtype compared to random coil and native state lysozyme are shown (figure 32). The random-coil RDC distribution obtained from flexible-meccano contains only negative values and is rather slim with a minor tail towards lower values. This shape is even more pronounced in the distribution of the R$_2$ baseline with a significant tail towards lower values. In contrast, RDC and R$_2$ distributions are very different for the native protein. A broad asymmetric distribution is observed for the RDCs reflecting the angular distribution of NH vectors relative to the external alignment tensor within the folded protein, while a slim distribution at higher R$_2$ values is observed for the R$_2$ values reflecting the uniform dynamics and rigidity in folded proteins. Partially folded states or unfolded states with residual structure are characterized by broader distributions than observed for the random-coil states tailoring towards higher value. This is here the case for WT^{Ala} which ranges in between random coil and native state behaviour.
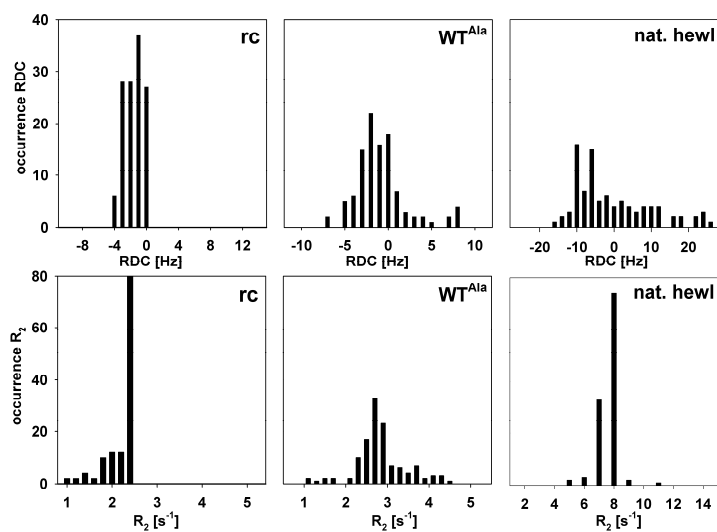
Figure 32: RDC and R_2 tensors of WT^{Ala}: Histograms of the RDC and R_2 tensors of WT^{Ala} (middle) compared to random coil predictions (left) and data of native hewl (right) (*144*).

For a better analysis of the degree of unfolding, histograms of mutants were also compared to random coil and native hen lysozyme data (figure 33). Since it was proposed by chemical shifts, also dynamical data show significant changes through single point mutation. The observed distributions for both, RDCs and R_2 values of the mutants are all asymmetric with a tail towards higher values reflecting residual structure. The center of the RDC distribution for all mutants falls within the range of the data obtained for the random coil from flexible meccano. In contrast, R_2 data are significantly higher than the values expected for a random coil. Here the center of the distribution ranges outside the expected for a random coil. This deviation might reflect a high degree of residual structure persisting on the subnanosecond time scale while the degree of residual structure is further averaged on the millisecond time scale and therefore scaled. One main finding is that in RDC tensors the mutants W62G^{Ala} and W108G^{Ala} show larger deviations than the other three mutants and a slightly modified profile including two centres. This behaviour is also observed in the R_2 tensors for W62G^{Ala}, but additionally for W28G^{Ala} and not for W108G^{Ala} as expected from RDC distribution.
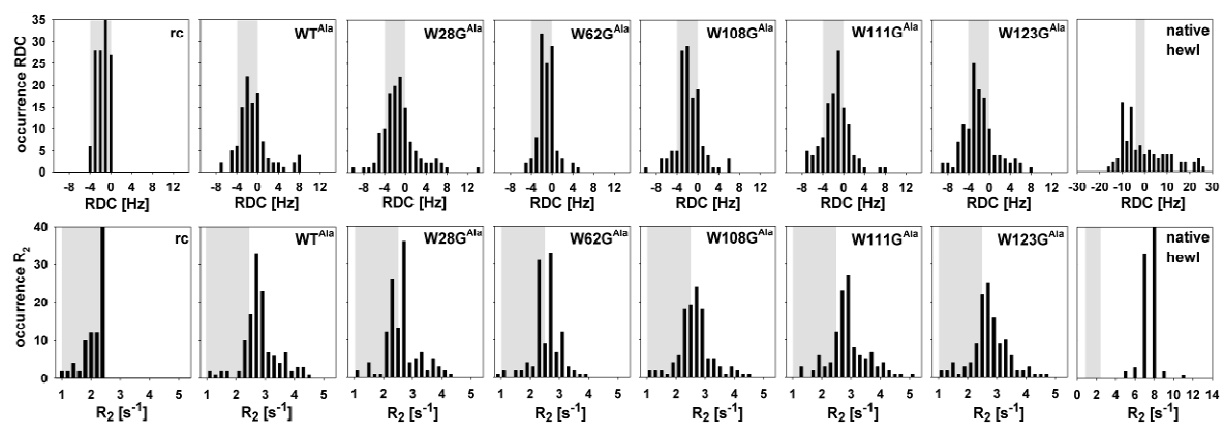
Figure 33: RDC and R_2 tensors of mutants: Tensor histograms of RDCs and R_2 rates of random-coil (rc) from flexible-meccano simulations and R_2 baseline predictions, WT[Ala] and five mutants W28G[Ala], W62G[Ala], W108G[Ala], W111G[Ala], W123G[Ala] and native hen lysozyme(*144*),; top: RDCs; bottom: ^{15}N R_2 relaxation rates. Grey bars indicate the shape of random coil predictions for RDCs (-4, 0) and for R_s rates (1-2, 4), respectively.

The distributions are rather narrow for W62G[Ala]. Broader distributions are observed for WT[Ala] and all other mutants for both, RDCs and R_2 data. One can conclude that the degree of residual structure is thus reduced in the W62G[Ala] mutants. For a detailed analysis of single point mutations and their effect on the individual hydrophobic clusters, RDCs and R_2 rates have been analyzed in a sequence-specific manner (figure 34).
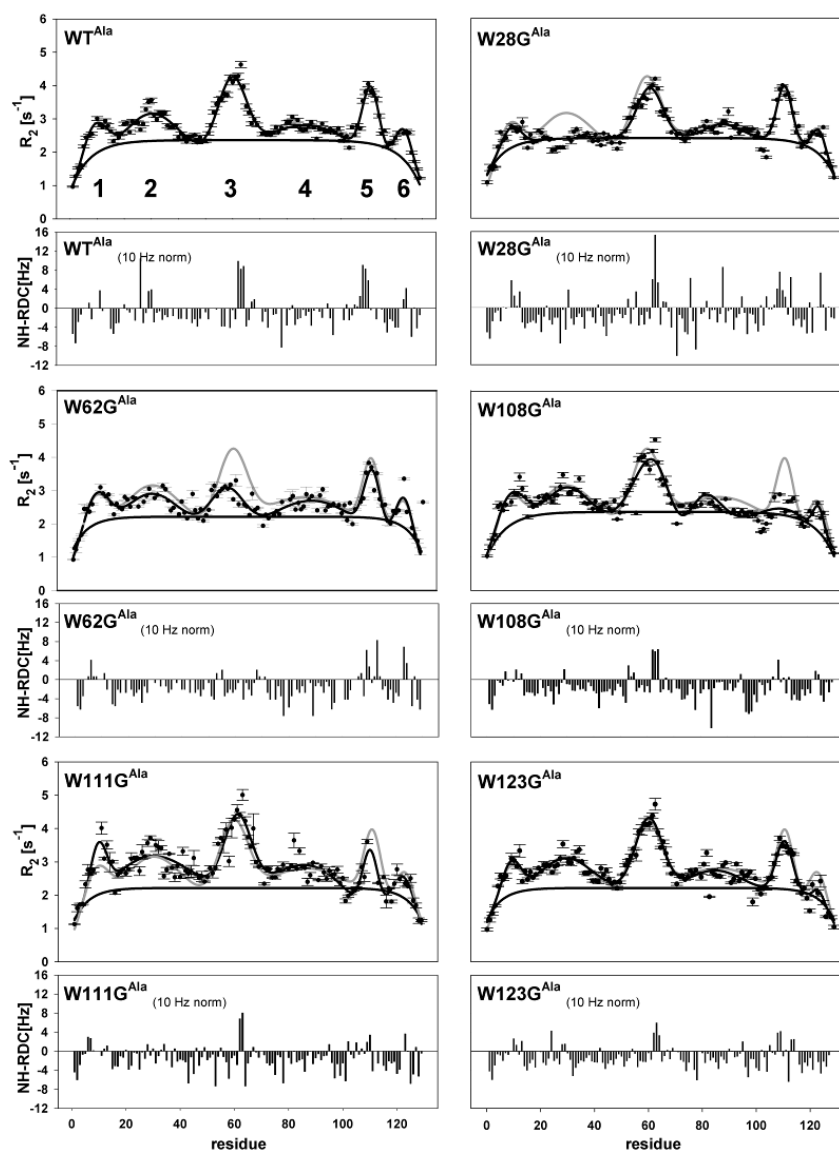
Figure 34: Sequence specific RDCs and R_2 rates of WT^{Ala} and mutants: ^{15}N R_2 rates and residual dipolar couplings (RDCs) ^{1}D(^{1}H,^{15}N) for wildtype WT^{Ala}, W28G^{Ala}, W62G^{Ala}, W108G^{Ala}, W111G^{Ala}, W123G^{Ala}. Fit of experimental R_2 rates are shown in black, fit of wildtype in grey; the baseline is coloured in black.^{15}N R_2 rates and RDCs were measured on 600MHz, 293K, using 300µM protein at pH2, via Bruker pulse sequences for relaxation experiments and IPAP-HSQCs, respectively.

In the wildtype sequence, 5 clusters can be identified around hydrophobic tryptophans. Single point mutations lead to changes within the local, perturbed cluster and remote clusters distant from the mutation site, indicating long-range interactions between these regions. In

the W28G^{Ala} mutant, the cluster around the mutation site has vanished; cluster 3 is attenuated while the other 3 regions remain unchanged by the mutation. Only minor changes are observed in the W123G^{Ala} mutant: while the own cluster becomes less pronounced and minor changes are observed in the neighboring region around W108 and W111, the remainder of the protein remains unchanged. Interesting changes are observed in the W108G^{Ala} and the W111G^{Ala} mutant. In these mutants, the flexibility around the mutation site increases as do residues located in cluster 3, the main cluster, in the protein as monitored by decreasing R$_2$ relaxation rates and decreasing RDCs. Furthermore, RDCs change from positive values indicating residual structure into negative values indicative for random coil around W28 while the R$_2$ relaxation rates remain the same at this position. The mutation in the 5th cluster therefore retains the rigidity in the 2nd region on the subnanosecond time scale, while residual structure on the millisecond time scale is destabilized. The largest changes are observed in the W62G^{Ala} mutant. The mutation disrupts its own cluster as well as it diminishes the other clusters in the protein. As observed for W108G^{Ala} and W111G^{Ala}, the subnanosecond structure, as monitored by the R$_2$ relaxation rates, is more persistent than the structural preferences on the millisecond time regime.

Figure 35 show sequence specific experimental RDCs in comparison with predictions made from flexible meccano. As mentioned for WT^{Ala} (see chapter 3.2.4) the overall shape of predicted data is bell-like in agreement with the smooth distributions expected for random flight chains. In contrast to the homopolymeric random flight chain, a rough landscape is predicted from flexible meccano due to residue-specific conformational preferences. Local or long-range interactions leading to residual structure are not included in the predictions. Also the predictions for the mutants largely deviate from experimental data. Since negative RDC values are expected for unfolded ensembles, a number of positive values are found in the mutants, in particular the regions where hydrophobic interactions have been observed by relaxation analysis. As previously seen in the R$_2$ rates, single point mutations have diminishing effects on hydrophobic clusters, especially at the mutation site but also at other regions, confirming long range interactions. These effects can be mainly found at W62G^{Ala} and W108G^{Ala}.
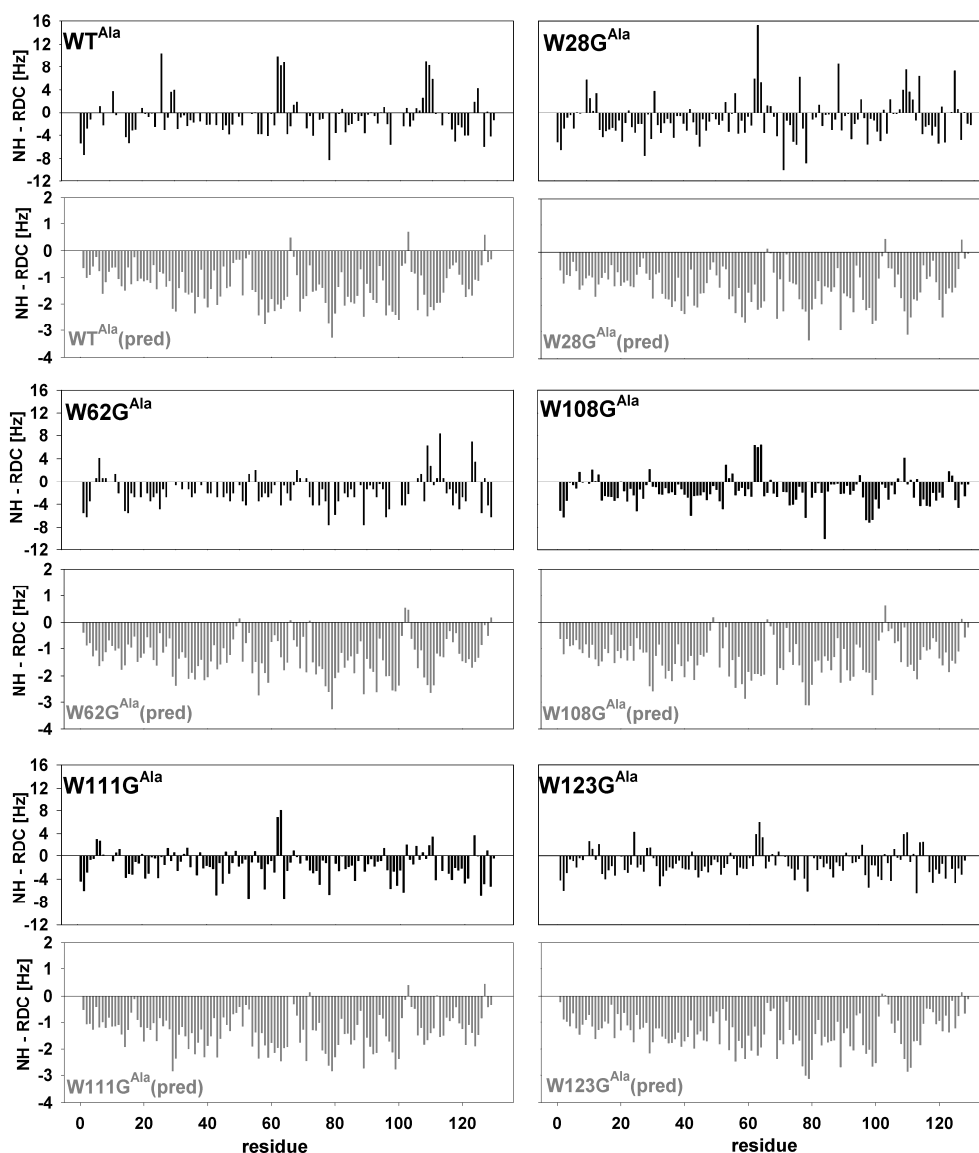
Figure 35: Flexible meccano predictions for RDCs of WT^{Ala} and mutants: Comparison of experimental (black) and predicted (grey) residual dipolar couplings of WT^{Ala} and the mutants W28G^{Ala}, W62G^{Ala}, W108G^{Ala}, W111G^{Ala}, W123G^{Ala}.

The degree of residual structure in the W123G^{Ala} mutant is only slightly reduced compared to WT^{Ala}. Residual secondary structure remains the same as in WT^{Ala}, the dynamics on both the subnanosecond and millisecond timescales around the mutation site and the neighbouring cluster 5 increases slightly. In the W28G^{Ala} mutant, deviations from the random coil disappear completely around the mutation site and only minor changes are seen around

cluster 3 of the protein. The residual structure around W28 is thus not important for the overall stabilization of the protein but is stabilized by the clusters around W111G^{Ala} and W108G^{Ala} as seen in the additional mutants investigated. In both mutants, dynamics increase around the mutation site and the neighbouring cluster. Furthermore, the dynamics on the millisecond time scale around W28 are significantly altered. In the W108G^{Ala} mutant furthermore residual structure and dynamics in the region of region 4 are affected as seen by R$_2$ rates and analysis of chemical shift differences. The most flexible mutant is the W62G^{Ala} mutant - both RDC and R$_2$ data are closest to random coil.

Degree of structure by diffusion measurements and small angle x-ray scattering

In the previous chapters a detailed analysis of chemical shifts and scalar coupling constants, R$_2$ and RDC data indicate areas of restricted motions coinciding with 5 regions displaying residual secondary structure. These areas interact through long-range interactions as identified by the analysis of the mutants. So a significant extent of residual structure is thus present in WT^{Ala} stabilized by local as well as by non local interactions. To further proof these findings the degree of compaction of WT^{Ala} and mutants was measured by translational diffusion and small angle x-ray scattering (SAXS). Table 4 illustrate the hydrodynamic radii determined by DOSY measurements of WT^{Ala} and five mutants.

protein	WT^{Ala}	W28G^{Ala}	W62G^{Ala}	W108G^{Ala}	W111G^{Ala}	W123G^{Ala}
R$_h$ [Å]	31.8 ±0.5	30.9 ±0.6	35.7 ±0.9	35.4 ±0.6	32.5 ±1.0	31.0 ±0.5

Table 4: Hydrodynamic radii of unfolded hen lysozyme; Experimental hydrodynamic radii R$_h$ of WT^{Ala}, W28G^{Ala}, W62G^{Ala}, W108G^{Ala}, W111G^{Ala} and W123G^{Ala} determined by diffusion measurements (native = folded hen egg white lysozyme)

A hydrodynamic radius (R$_h$) of 31.8 ±0.5 Å was obtained for WT^{Ala} from concentration dependent measurements. The observed value is considerably (14 %) smaller than observed for reduced and methylated hen lysozyme in 8M urea (R$_h$=34.6 Å) which is in line with other strongly denatured proteins(*93*). In analogy to data from chemical shifts, relaxation and residual dipolar couplings, the mutants show variable deviations in diffusing characteristics

compared to wildtype. While W28G^{Ala} (R_h = 30.9 ±0.6 Å) and W123G^{Ala} (R_h = 31.0 ±0.5 Å) have similar hydrodynamic radii than WT^{Ala}, the other three mutants show larger deviations. While medium range deviations are observable in W111G^{Ala} with an R_h of 32.5 ±1.0 Å, W108G^{Ala} (R_h = 35.4 ±0.6 Å) and W62G^{Ala} (R_h = 35.7 ±0.9 Å) show more drastic effects. In figure 36 the proposed degree of structure for WT^{Ala} and mutants is shown. The most extended structure, based on DOSY data, is proposed for W62G^{Ala} which is consistent with relaxation data and RDCs.
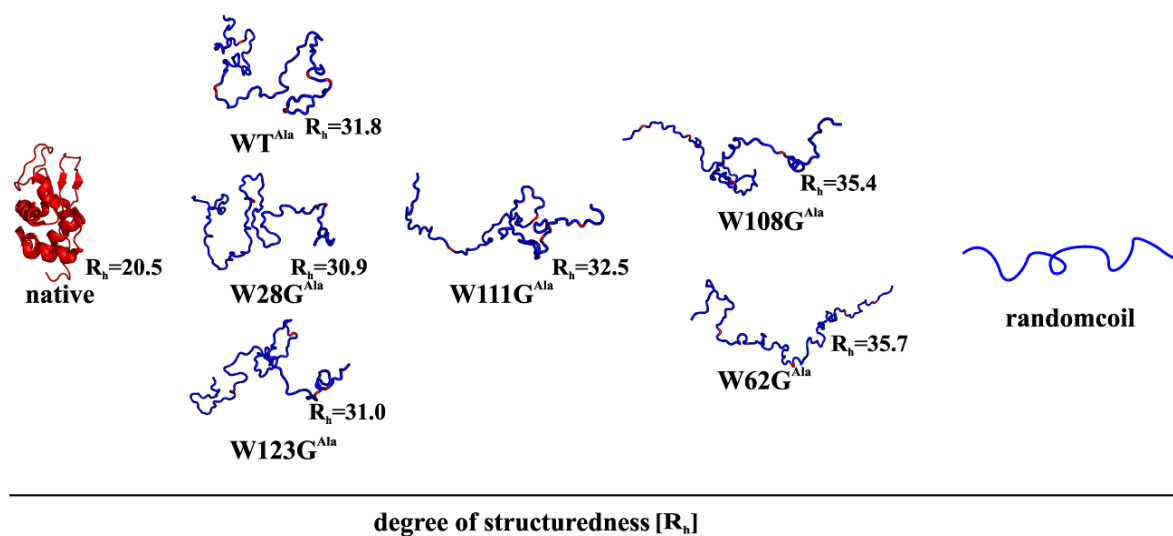
Figure 36: Degree of structure in unfolded hen lysozyme: Schematical representation of the degree of structure in hen egg white lysozyme from the folded state to unfolded states with residual structure. Structures were calculated by MD-simulations. Experimental R_h values from diffusion experiments were used to choose structures with smiliar R_h calculated by HydroproNMR data. (tryptophans in unfolded states are highlited in red).

The decrease of residual structure in mutants of WT^{Ala} has also been seen in SAXS measurements. As mentioned in materials and methods raw data can be analyzed by two different fitting routines, EOM (ensemble optimization method) (*138*) and the Porod-Debye relation (*141*) at which here EOM was chosen for discussion of the data. The results of SAXS are shown in table 5 and figure 37.

protein	WT^{Ala}	W28G^{Ala}	W62G^{Ala}	W108G^{Ala}	W111G^{Ala}	W123G^{Ala}
R_h [Å] (DOSY)	31.8	30.9	35.7	35.4	32.5	31.0
R_g [Å] (SAXS)	35.7	41.4	40.8	47.9	47.4	38.8

Table 5: Comparison of R_h (DOSY) and R_g (SAXS); Experimental hydrodynamic radii R_h and gyromagnetic radii R_g of WT^{Ala}, W28G^{Ala}, W62G^{Ala}, W108G^{Ala}, W111G^{Ala} and W123G^{Ala} determined by diffusion measurements (DOSY) and small angle x-ray scattering (SAXS).

To get an idea of the range of R_h and R_g in unfolded ensembles, ten thousand structures, predicted by flexible meccano, were additionally used to compare the results with general nonnative characteristics. All measured proteins have a higher R_g than R_h, but with remarkable differences. With an R_h of WT^{Ala} of 31.8Å the R_g is with a value of 35.67Å in an expected range. This is not the case for the W111G^{Ala} mutant which has an R_g of 47.43Å with an extremely smaller radius of hydration of 32.5Å. There is also a large deviation in the fitting results. Similar behavior is observable for W108G^{Ala} with an R_h of 35.4 and an R_g of 47.97Å (EOM). In general those effects are only visible during aggregation, so that there is no clear conclusion for these two mutants. As SAXS was performed at a maximum concentration of 300µM aggregation should be excluded and the W62G^{Ala} mutant should have the largest R_g. However the same effect is observable in the DOSY data, where both mutants, W111G^{Ala} and W108G^{Ala}, also show higher R_h values at higher concentration than W62G^{Ala}. At low concentrations this effect is deleted. In the SAXS measurements protein concentrations from 300-100µM have been incorporated in the analysis of R_g.
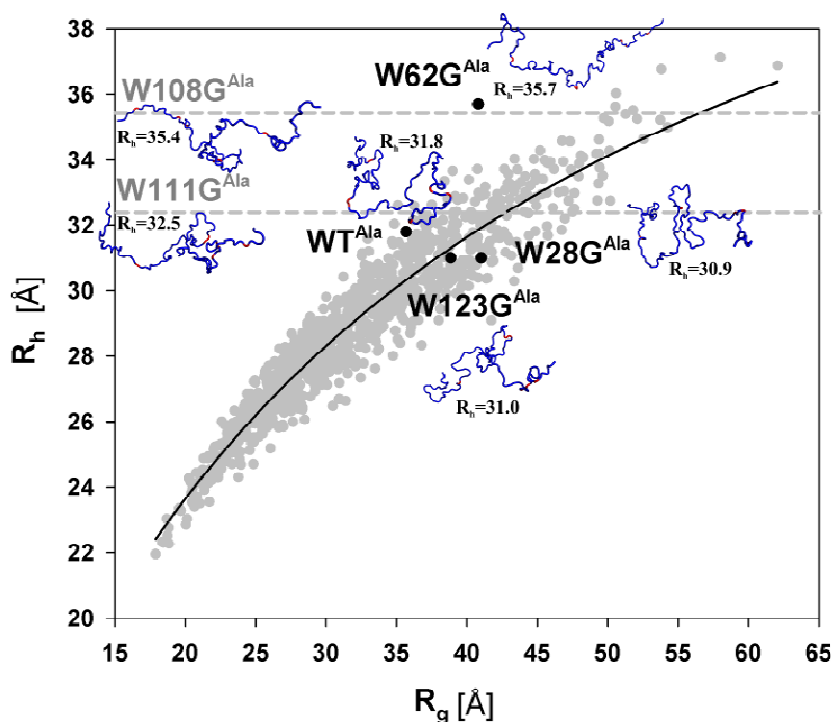
Figure 37: SAXS data of WT^{Ala} and mutants: The hydrodynamic radius R_h of WT^{Ala} and five mutants (black circles) is plotted against their radius of gyration R_g. Grey data points: ensemble of 10.000 structures of WT^{Ala} from flexible meccano simulations; black line: regression of these simulation data. R_h values result from DOSY measurements, R_g values derive from small angle x-ray scattering (SAXS); data fit was done by EOM.(*138*). For W108G^{Ala} and W111G^{Ala} only R_h data are shown. Structures were calculated by MD-simulations. Experimental R_h values from diffusion experiments were used to choose structures with smiliar R_h from the simulations.

To shortly summarize the analysis of unfolded hen lysozyme in chapter 3.2, a number of novelties are presented here. In detail $^1J(N,C^\alpha)$ and $^2J(N,C^\alpha)$ coupling constants with dependence on the angle (ψ) have been used to further identify residual secondary structure. Scalar side chain couplings $^3J(H^\alpha,H^\beta)$ and $^3J(CO,H^\beta)$ reveal information about the side chain angle χ_1 which is influenced by the overall compaction of the protein. Further insight into the overall behaviour of unfolded proteins was obtained by the analysis of R_2 and RDC distributions. This analysis allows identification of structured parts in the protein variants while it does not rely on the assignment of the protein. Using a $J(\omega_N)/J(0)$ correlation allows

differentiation between residues whose motions are governed by internal or overall motions or both. It could also be shown that both the analysis of R_2 and RDC data in a conjoint manner and the use of single point mutants can be used to identify regions within the proteins with differential stabilities.

3.3   Unfolded state dynamics of Human vs. Hen Lysozyme

Hen and human lysozyme share high structural and sequential identity. Therefore human lysozyme is as hen lysozyme a very useful model protein to study protein folding processes. Here in this thesis unfolded states of both proteins have been analyzed. These states have been stabilized by introducing alanines instead of all cysteines for the inhibition of disulfide formation. Both so called all-ala variants were stable at pH2 in water without any denaturant. In the following chapters nonnative hen and human lysozyme and disease related mutants of human lysozyme are compared with respect to structural and dynamical characteristics; in particular aggregation propensities of both proteins and the onset of pathological behaviour.

3.3.1   Protein resonance assignment

Since the all-ala variant of human lysozyme is a construct which is synthesized and analyzed for the first time in this thesis, complete assignment of the protein was necessary. The assignment of the backbone is the basis of further structural and dynamical characterization of the protein and two unfolded mutants with amyloidogenic potential in the native state, I56T and D67H. Since the narrow chemical shift dispersion in the ^{1}H-^{15}N HSQC for backbone ^{1}H$_N$ and ^{15}N assignment, strategies by NMR are based on isotopic labelling and 2D- and 3D-experiments. In this thesis backbone assignment was performed by generally used backbone NMR experiments; HNN, HNCACB and ^{1}H-^{15}N HSQC (via the implemented Bruker pulse sequences).

Backbone assignment

In the following the complete backbone assignment of hLys is described. Figure 38 shows the HNN experiment of hLys where the ^{15}N$_i$ amide resonances with the ^{1}H$_{Ni-1}$ and ^{1}H$_{Ni+1}$ proton resonances of the predecessor and successor of sequenced amino acids are correlated. In addition a HNCACB experiment was measured (figure 39) which correlates the ^{13}C$^\alpha$ and ^{13}C$^\beta$ resonances with the backbone ^{1}H$_N$ and ^{15}N frequencies. By assigning ^{13}C$^\alpha$ resonances, the experiment also reports on the corresponding ^{13}C$^\beta$ resonances. Figure 41

shows the complete assignment of the ^{1}H-^{15}N HSQC of hLys which is the basis for the assignment of the two mutants I56T and D67H.
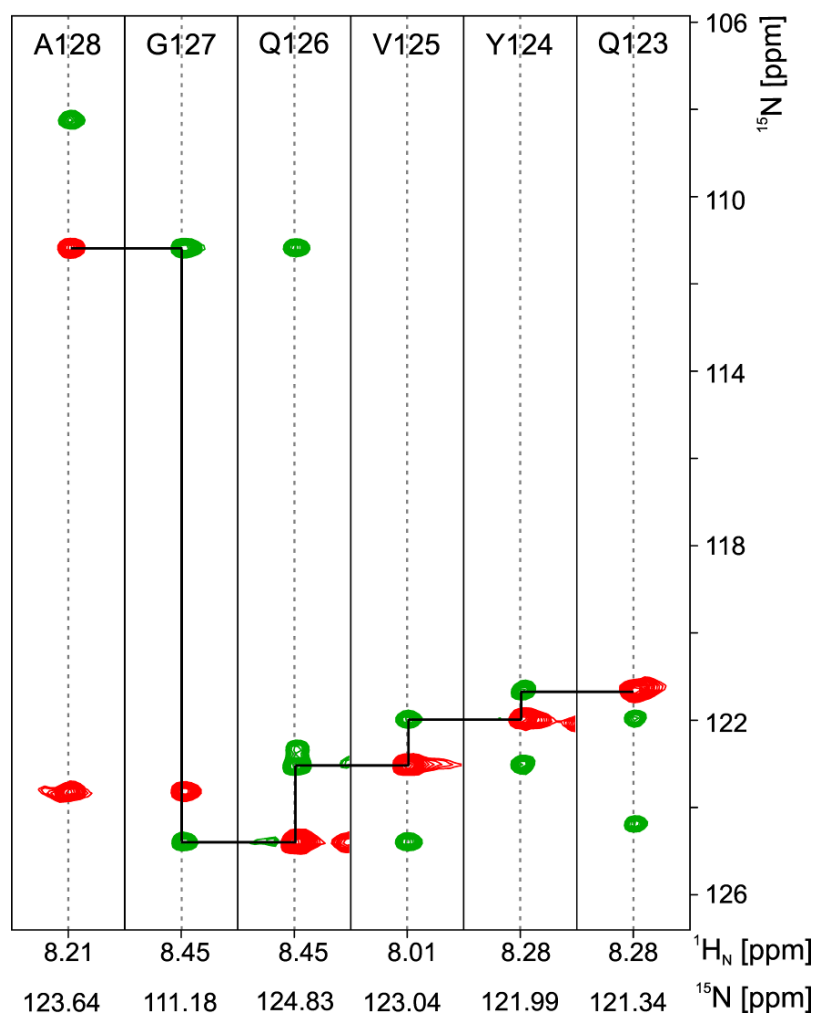
Figure 38: HNN spectrum of hLys; striplot of the HNN experiment of the amnio acids G127, Q126, V125, Y124 and Q123 of hLys in the ^{1}H$_N$, ^{15}N and ^{1}H$_N$ dimension; red: ^{15}N diagonal peaks; green: ^{1}H$_N$/^{15}N cross peaks between predecessor and successor in the amino acid sequence; black lines: sequential walk through the sequence. Spectrum was recorded at 900MHz, 293K, pH 2 at a protein concentration of 300µM.

As mentioned before the HNN experiment provides ^{15}N$_i$ amide resonances correlated with the ^{1}H$_{Ni-1}$ and ^{1}H$_{Ni+1}$ proton resonances. The black line in figure 38 indicates one sequential

walk through the sequence of all-ala human lysozyme. By means of this spectrum about 85% of all NH backbone peaks could be assigned. To complement these results the HNCACB was analyzed. The results of this experiment is shown in figure 39 where the black line guides through the $^{13}C^{\alpha}$ resonances of an exemplary section of the amino acid sequence of hLys.
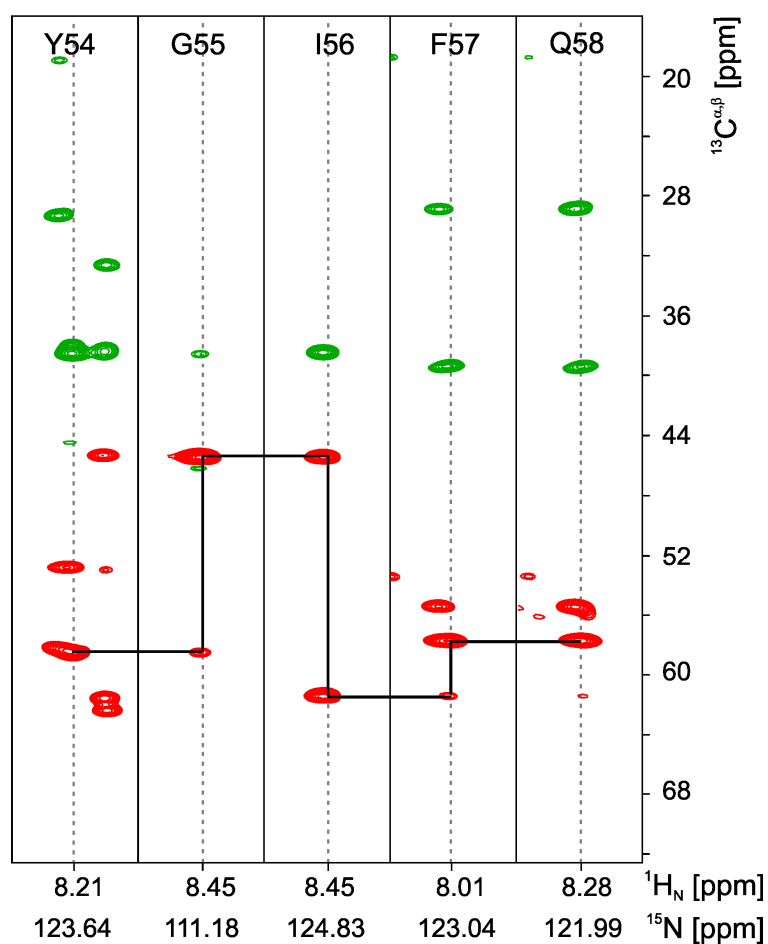
Figure 39: HNCACB spectrum of hLys; striplot of the HNCACB experiment of the amnio acids Y54, G55, I56, F57 and Q58 of hLys; red: $^{13}C^{\alpha}$ resonances at defined $^{1}H_N$ ^{15}N frequencies from the HNN experiment described above; green: $^{13}C^{\beta}$ resonances at corresponding $^{13}C^{\alpha}$, $^{1}H_N$ and ^{15}N frequencies. Spectrum was recorded at 900MHz, 293K, pH 2 at a protein concentration of 300μM.

Bringing together the results from spectra analyzed above the assignment of the ^{1}H-^{15}N HSQC was completed (figure 40) to about 90%.
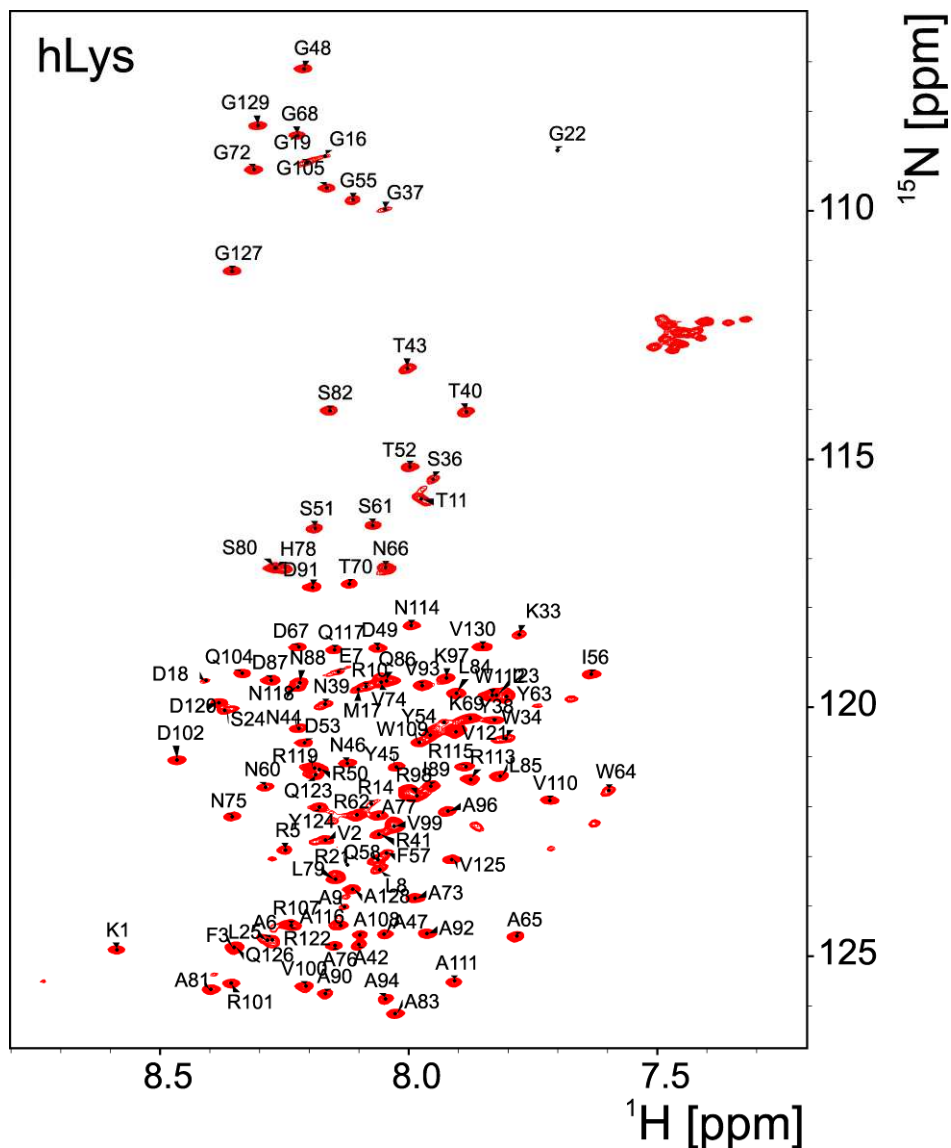
Figure 40: ^{1}H-^{15}N-HSQC spectra of hLys; spectrum was recorded at 600MHz, 293K, 300µM protein concentration in water (pH2), 10% D_2O.

In addition to hLys, two mutants have been analyzed here by NMR spectroscopy. Since the pattern of their ^{1}H-^{15}N HSQCs deviate only in a few NH resonances, it was possible to assign the all-ala mutants I56T and D67H by comparing their spectra with the spectrum of the

wildtype (figure 41). Both mutants were measured at low protein concentration (300µM) since they have aggregation potential in their native states.
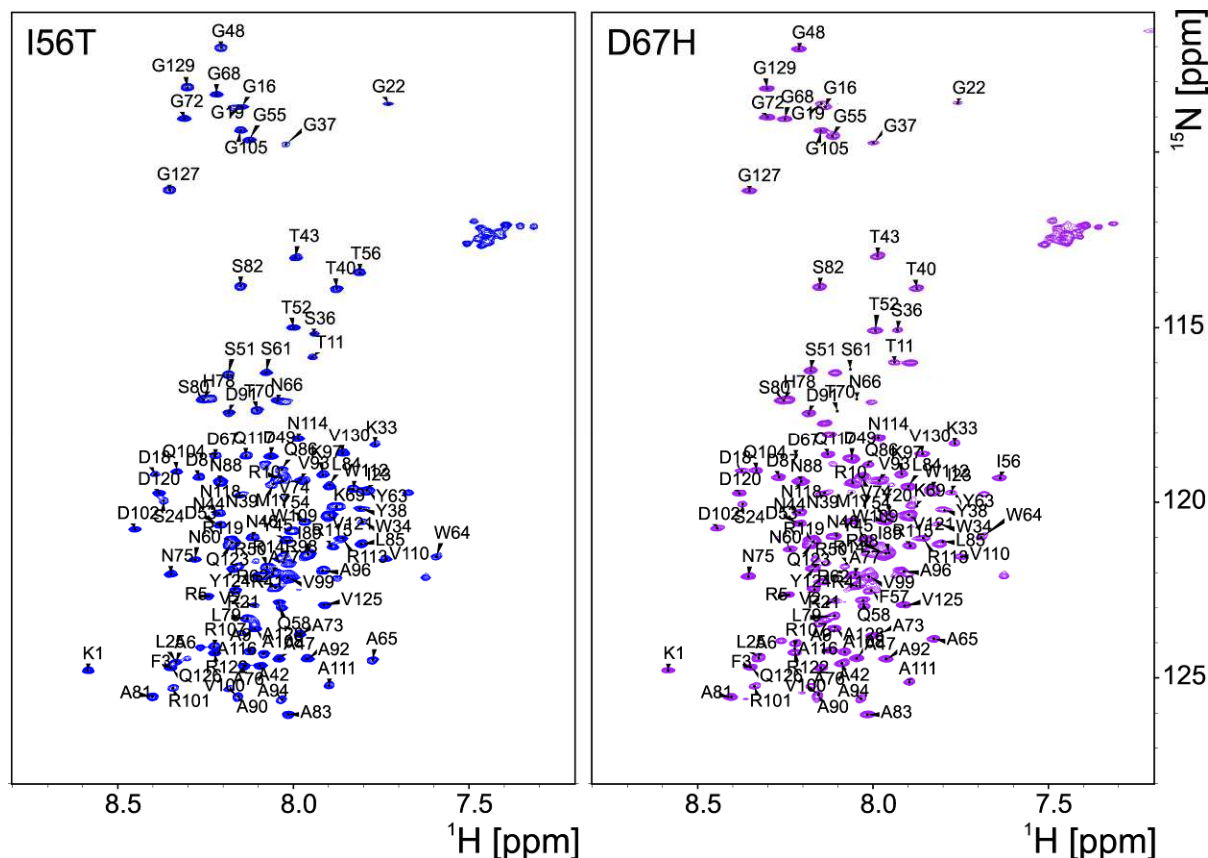
Figure 41: ¹H-¹⁵N-HSQC spectra of I56T and D67H; Both spectra were recorded at 600MHz, 293K, 300µM protein concentration in water (pH2), 10% D₂O.

Both mutants mainly show deviations from the wildtype in the region around the mutation site. While in I56T large shifts of the amino acids G37, Y54, G55, and W64 are observable, D67H shows comparable chemical shift deviations of S61, W64, A65, N66, G68 and T70. From these data one can assume that there are also structural and dynamical differences present between mutants and hLys. This will be discussed in detail by means of relaxation behavior in chapter 3.3.4.

3.3.2   Protein sequence comparison between hen and human lysozyme

The amino acid sequences of hen and human lysozyme are compared to show main analogies and differences in these proteins and to further evaluate structural and dynamical characteristics. Since unfolded states are in the focus, only the all-cysteine to all-alanine variants of both proteins are shown here (figure 42).
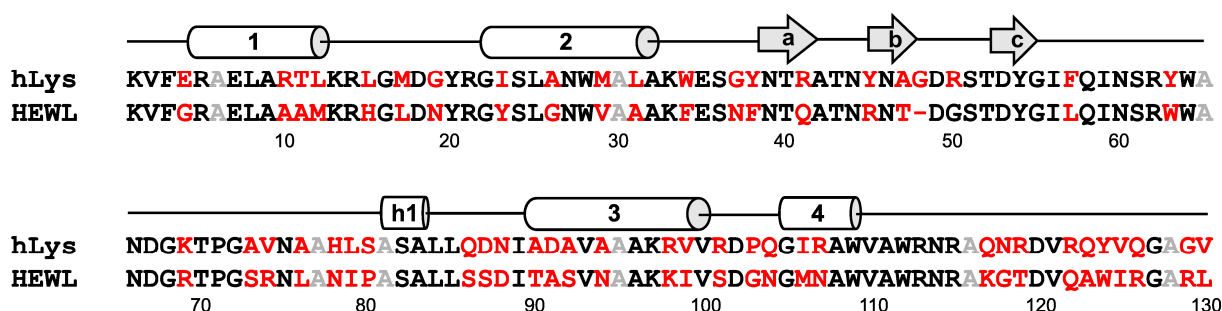
Figure 42: Sequence alignment of human and hen lysozyme(*148*) with residue numbers of hLys. Red: differing amino acids; grey: cysteine to alanine mutations;. α-helices in the native states are indicated by cylinder 1, 2, 3 and 4; the 3_{10}-helix by h1; β-sheets are indicated by arrows a, b and c.

As mentioned in the introduction human and hen lysozyme share high sequence homology (~70%) and identity (~60%). The native state structures of hLys and hewl are very similar with a backbone RMSD of only 0.74 Å (*149*). The structure of human lysozyme possesses two domains: an α-domain (residues 1-38 and 86-130) with four α-helices and a β-domain (residues 39-85) with a triple-stranded β-sheet and a 3_{10} helix. Differences in the amino acid sequence are predominantly found between residues 70-80, 85-92 and 98-110, located at the interface between the α- and the β-domain. Since human lysozyme shows higher aggregation potential than hen lysozyme in their native state, it accounts for the elucidation of folding processes to analyze the unfolded state behaviour of both proteins. By reason of apparent structural resemblance of the native states, analogue structural and dynamical behaviour is expected. In the following section the properties of the all-ala variant of human lysozyme are analyzed on the basis of the insights of unfolded hen lysozyme. Similarities and differences will be discussed here in detail.

3.3.3   Structural and dynamical protein characteristics

Since unfolded hen lysozyme has regions with residual structure it seems obvious that unfolded human lysozyme shows an analogous behaviour. To confirm this hypothesis, secondary structure propensities of both (SSP, http://pound.med.utoronto.ca/software.html (*53*)) have been analyzed using carbon C^{α} and C^{β} chemical shifts compared to random coil (figure 43).
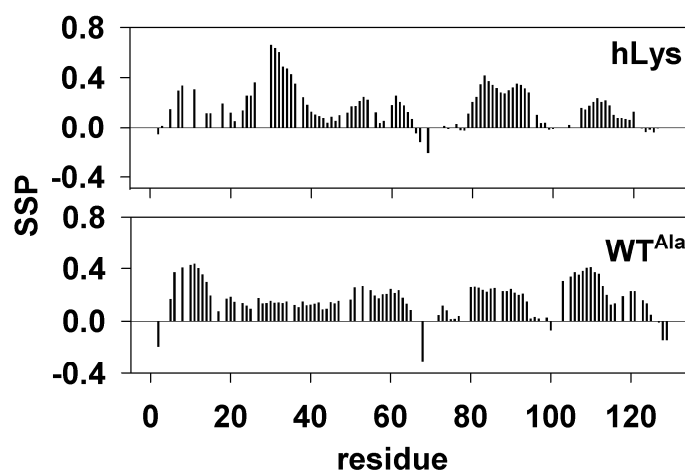
Figure 43: Secondary structure propensities of hLys; secondary structure propensities (SSP) of hLys and WT^{Ala}; SSPs were obtained from C^{α} and C^{β} chemical shifts and calculated by http://pound.med.utoronto.ca/software.html (*53*).

As the figure above shows, human lysozyme exhibit regions with significant deviations from random coil. Especially the regions where secondary structure is present in the native state show high SSP values in the unfolded state. Here positive deviations indicate transient formation of α-helical structure and cluster around residues 5-15, 23-36, 53-65, 79-94 and 107-118. To further proof these findings, $^{1}H_N$ chemical shifts of hen and human lysozyme were compared to random coil data (figure 44). Regions with negative deviations, indicating a-helical structure, appear at similar areas in the amino acid sequence as already presumed by SSPs. Interestingly, even regions which have β-sheet structure in the folded protein seem here to have more α-helical character, mainly the region around residues 40-60.
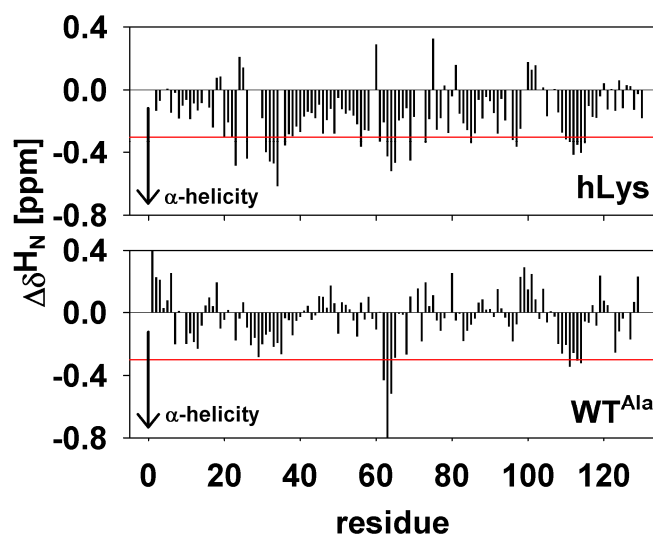
Figure 44: Chemical shift difference $\Delta\delta H_N$ of hLys and WT^{Ala}; Chemical shift deviations from random coil values ($\Delta\delta=\delta_{exp}-\delta_{rc}$) of hLys and WT^{Ala} are compared; red lines: cut-offs for chemical shift indexing for α-helices at -0.3 (*71*), δH_N were obtained from the ^{1}H-^{15}N HSQC, measured at 600MHz, 293K, 10% D_2O, pH2 at a protein concentration of 300μM

The cut-offs for chemical shift indexing for a-helices are highlighted at -0.3 ppm (*71*). Interestingly hLys has a higher cut-off than unfolded hen lysozyme suggesting a larger a-helical propensity. Indeed both proteins cluster around regions within hydrophobic amino acids, the overall average is still higher for hLys. For a more precise interpretation, figure 45 shows a detailed overview. In panel A to C, ^{1}H-^{15}N hetNOEs and ^{15}N longitudinal R_1 and transverse R_2 relaxation rates of the two proteins are shown. The general tendency of hLys to a higher degree of a-helicity is continued in relaxation data. For hLys, mainly positive hetNOE values are observed particularly in the regions from 11-24, 33-69, 81-91 and 108-116, negative values are found at the termini and between residues 72-80 and 97-107. The hetNOE is mainly negative for hewl with few exceptions. R_1 and R_2 relaxation rates of hLys are compared to WT^{Ala}, with higher relaxation rates observed for hLys.
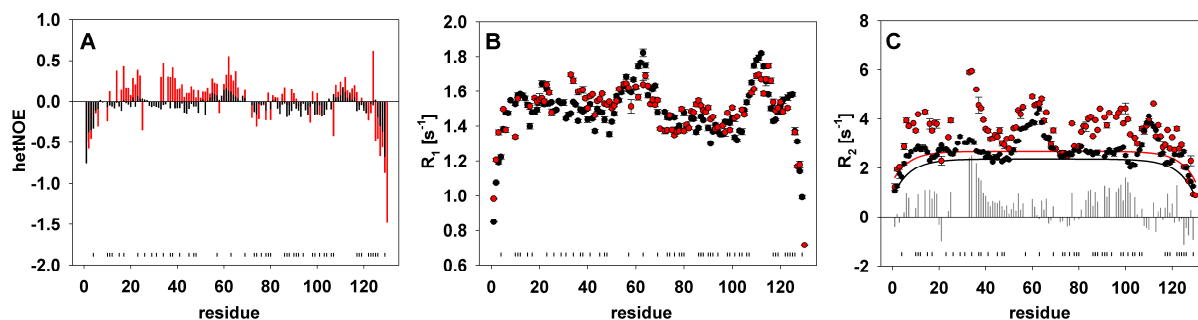
Figure 45: Comparison of hLys and WT^{Ala}; A-C: Comparison of heteronuclear ^{1}H-^{15}N NOE, R$_1$ and R$_2$ relaxation rates of hLys (red) and hewl (black); grey bars: difference in R$_2$ rates of hLys vs. hewl. Small vertical marks indicate differing amino acids in the sequences of hlys and hewl. All experiments have been carried out at 600MHz with 300μM protein at pH2 (water, 10%D$_2$O), 293K

This is in good agreement with the higher α-helical propensity compared to unfolded hen lysozyme. More stabilized residual structure seems to be present in hLys. Relaxation rates show smooth profiles, variations along the sequence have similar shape for R$_1$ and R$_2$. R$_1$ data range up to ~1.8 s^{-1} while R$_2$ rates range up to 6s^{-1} for hLys and 5 s^{-1}for hewl. It is remarkable that variations along the sequence are thus more pronounced for R$_2$ data. Such behaviour could be either due to increased rigidity or the presence of conformational exchange processes. This will be discussed in the chapter at a later point of time (see figures 47 and 48).
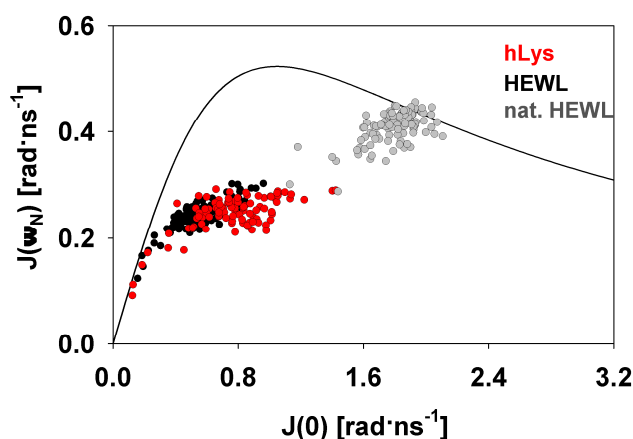
Figure 46: Comparison of hLys and WT^{Ala};; c.) spectral densities of hewl (black), hLys (red) and native hewl (grey) (*150*). All experiments have been carried out at 600MHz with 300μM protein at pH2 (water, 10%D$_2$O), 293K

In figure 46 the graphical analysis of spectral density values is shown. For a more precise differentiation also spectral density values for the folded state are shown. In general, residues with $J(\omega_N)/J(0)$ correlation near the ascending part of the Lorentzian line are dominated by fast internal motions, while motions in residues near the descending part (with $J(0)=0.4\tau_c$) are rigid and dominated by the overall tumbling. For residues exhibiting chemical exchange, $J(0)$ would be increased by the exchange contribution resulting in $J(0)$ values shifted to the right (for a detailed description see materials and methods). One striking attribute found out here is that the variation for individual residues is larger for hLys than for hewl, ranging from very flexible residues to residues which have contributions from both, fast and slow motions. Particular high values for $J(0)$ in hLys, indicating higher rigidity, are found for residues 33, 34, 56, 60-64, 89, 98, 111 and 112. This stands in good agreement with the findings from R_2 relaxation data, where at least these regions seem rigid due to residual structure.

The presence of exchange for the unfolded hen wgg white lysozyme could be excluded by graphical analysis of the spectral density data for WT^{Ala}, where no parallel shift towards higher values of $J(0)$ is observable (see 3.2.4). This is also shown by the analysis of relaxation dispersion measurements (see 3.2.4). The existence of chemical exchange processes in hLys has also been analyzed in this thesis. The $J(\omega_N)/J(0)$ analysis, the shape of the relaxation rates and the absence of significant scatter let assume that no chemical exchange is present in unfolded human lysozyme at a concentration of 300µM. This hypothesis was further proofed by relaxation dispersion and $R_{1\rho}$ measurements to exclude that the elevated level of relaxation rates stems from chemical exchange processes. Figure 47 shows the results of relaxation dispersion measurements of unfolded human lysozyme. ^{15}N single-quantum relaxation rates as a function of CPMG frequency have been analyzed as described in materials and methods, chapter 2.2.4. Indeed selected residues A6, G16, G22, N60 and A65 show small variations in R_{2eff}, the overall variation of R_{2eff} is almost zero.
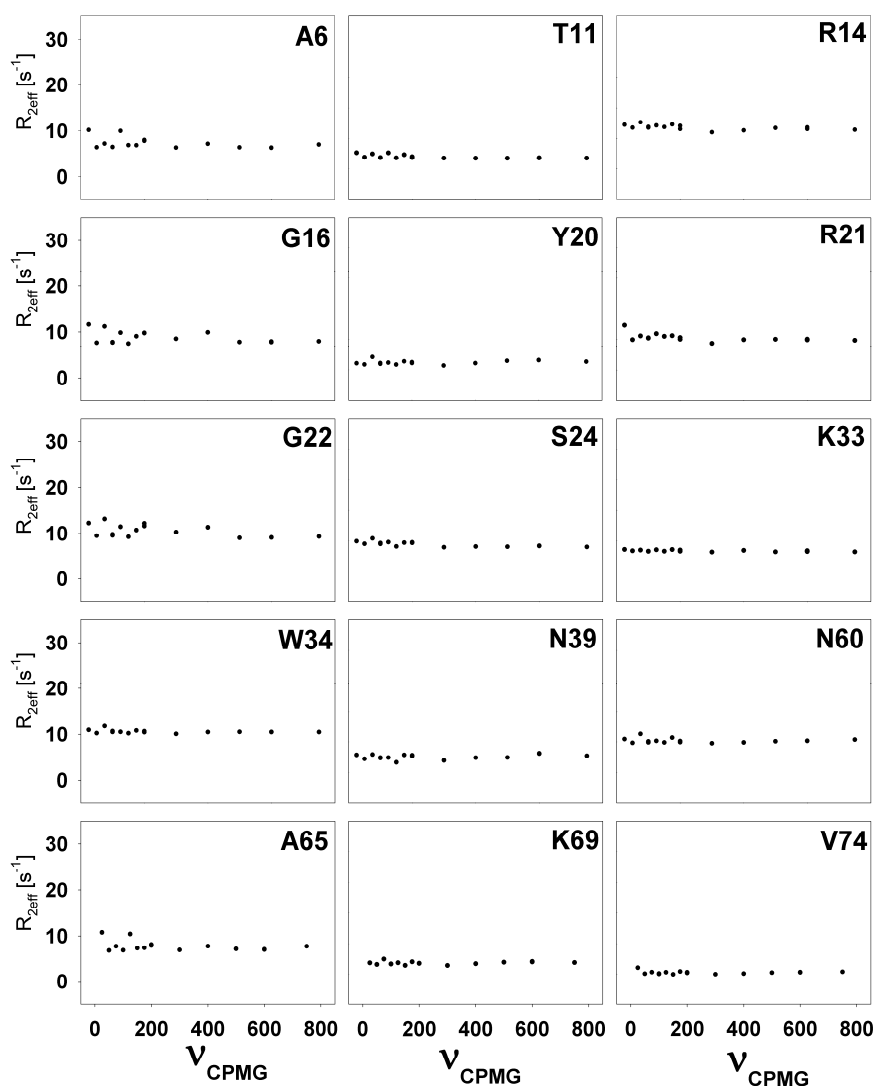
Figure 47: Relaxation dispersion measurements of hLys; ^{15}N single-quantum relaxation rates R_2eff [s[-1]] of hLys as a function of CPMG frequency for selected residues A6, T11, R14, G16, Y20, R21, G22, S24, K33, W34, N39, N60, A65, K69 and V74; experiments were performed at 300μM protein concentration in water pH2 (10% D_2O), 293K, 600MHz.

Further confirmation results from $R_{1\rho}$ measurements shown in figure 48. Differences between R_2 and $R_{1\rho}$ relaxation rates are negligible small and in the range of the error bars at a protein concentration of 300μM. Together with relaxation dispersion data this leads to the assumption of an absence of chemical exchange in hLys.
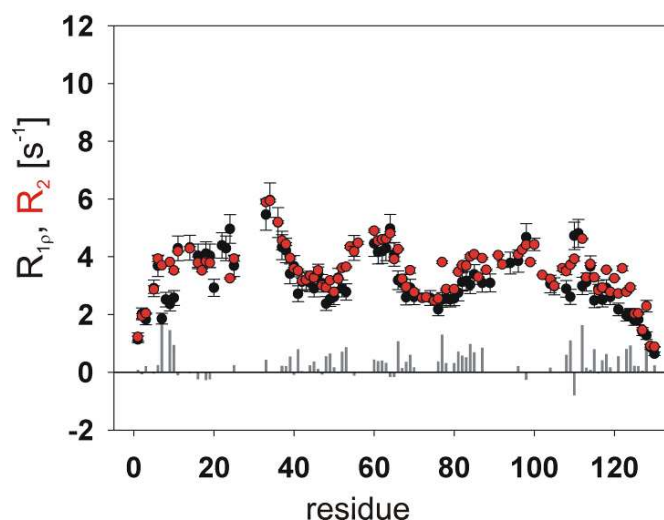
Figure 48: $R_{1\rho}$ relaxation data of hLys; $R_{1\rho}$ (black) and R_2 (red) relaxation rates of hLys; grey bars: difference of R_2 and $R_{1\rho}$. All data were measured at 600MHz, 293K, pH2 (water, 10% D_2O) and 300µM protein concentration.

To shortly summarize the chapter of chemical shifts and relaxation behaviour of hLys compared to WT^{Ala} at protein concentrations of 300µM, similarly, six clusters are found in hLys around (1) L12, (2) W28/W34, (3) Y63/W64, (4) V93, (5) W109/W112, and (6) Y124. As expected by the very similar sequence, the overall cluster architecture remains the same. But the relative sizes of the clusters differ enormous for the two proteins. The cluster with the largest relaxation rates in hLys is cluster (2) with W28 and W34, while cluster 2 is not dominant in WT^{Ala} containing a phenylalanine on position 34 rather than a tryptophan. Clusters 1, 3, 4 and 5 have approximately the same amplitudes with R_2-rates ranging from $1.8s^{-1}$ to $3.2s^{-1}$, while cluster 6 is very small in hLys. The most striking difference is found in cluster 4 around residues 78-105. This cluster is only marginally present in WT^{Ala}, while it is very prominent in hLys, despite the fact that it does not contain any aromatic residue. Residues in this region of the polypeptide chain are at the interface between the two domains forming a 3_{10} helix (residues 81-85) and an α-helix (helix 3, residue 90-100) in the native proteins. This is an important finding discovered in the thesis and will be drawn through the following sections.

3.3.4   Concentration dependency of chemical shift and relaxation

Since human lysozyme in its native state has a higher aggregation propensity than hen lysozyme (*18*), the dependency of chemical shift and relaxation on low and high (aggregation competent) concentrations of unfolded hLys is analyzed. It was found out here that unfolded human lysozyme is stable until 850µM, at higher concentration values the protein precipitates. At first, chemical shifts derived from ^{1}H-^{15}N HSQCs have been compared.
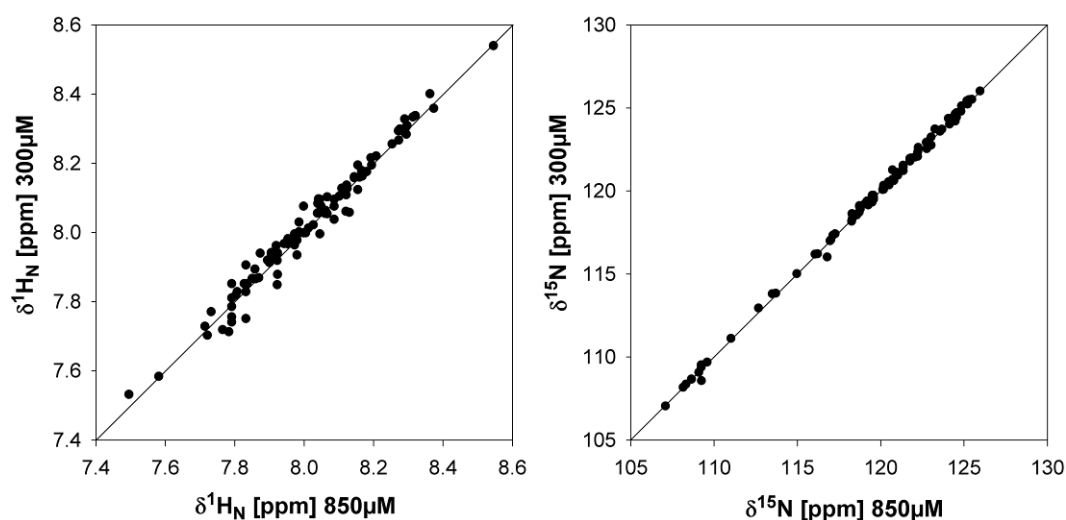
Figure 49: δ^1H$_N$ and δ^{15}N of hLys at different concentrations; Comparison of ^{1}H$_N$ and ^{15}N chemical shifts at 300µM and 850µM protein (hLys) from ^{1}H-^{15}N HSQCs, both measured at 600MHz, 293K, water pH2, 10% D$_2$O.

Figure 49 shows the correlation of ^{1}H$_N$ and ^{15}N chemical shifts of hLys at 300µM (low) and 850µM (high) protein concentration. In principle there are only small deviations visible at the two concentrations mentioned before, in the ^{1}H$_N$ as well as in the ^{15}N dimension. With respect to the structure this could be interpreted as only small overall structural changes in the enfolded ensemble or one major conformation superimposing other minor populated, aggregated conformations, only visible in the ^{1}H-^{15}N HSQC by missing peaks due to exchange (at 300µM there are 113 peaks visible, whereas at 850µM only 95 peaks appear). For a more detailed view on the dynamics at different concentrations, relaxation behaviour was analyzed. Figure 50 (A-D) shows R$_1$, R$_2$ and R$_{1\rho}$ relaxation rates and hetNOE data of hLys at 300µM (red) and 850µM (black) protein concentration.
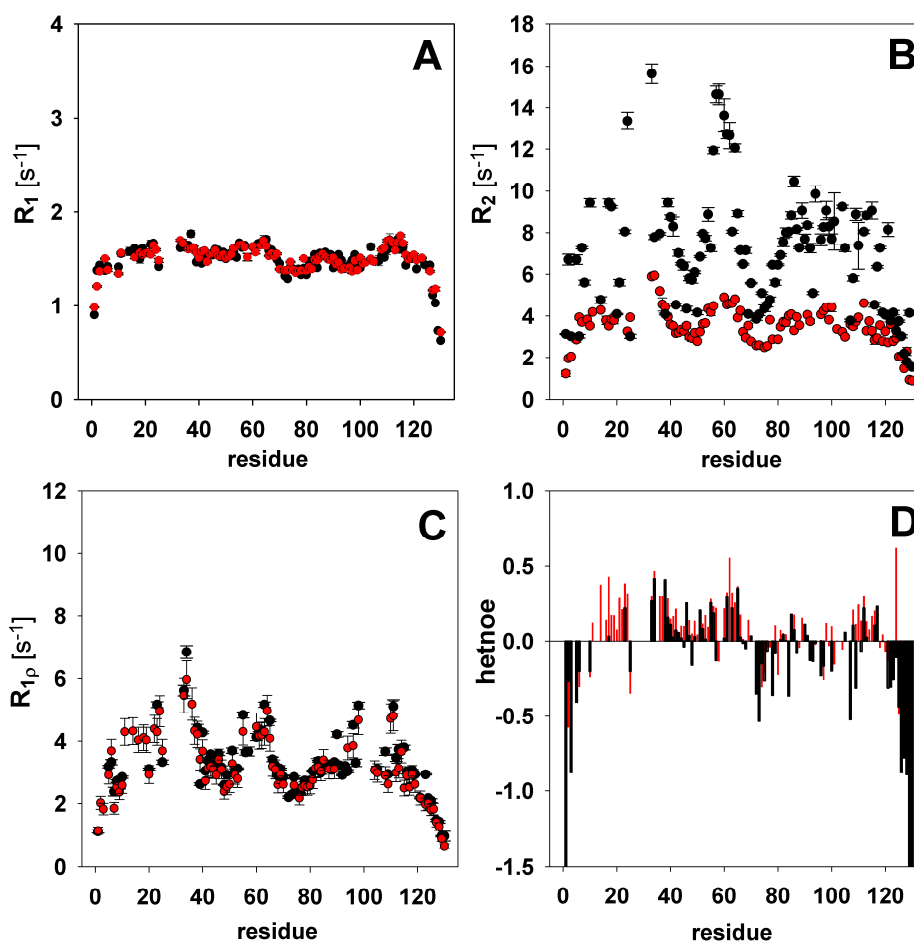
Figure 50: Concentration dependent relaxation data of hLys; R_1, R_2, $R_{1\rho}$ relaxation rates and hetNOE data of hLys at 300uM (red) and 850µM (black) measured at 600MHz, 293K, pH2 (water, 10% D_2O)

While R_1 and $R_{1\rho}$ indicate no changes due to higher concentration (figure 50 A and C), the R_2 rates at 850µM deviate significantly from those at lower concentration. Chemical exchange appears at high concentrations, visible through large differences between R_2 and $R_{1\rho}$. Figure 50B shows that the higher the protein concentration the larger the R_2 rates due to exchange. This effect was also observable for unfolded hen lysozyme but not at that extend. The maxima of R_2 (850µM) ranges between 14.6[s^{-1}] for residue K33 and 15.6[s^{-1}] for residue F57, for example. ^{1}H-^{15}N hetNOE experiments (figure 50D) further confirm findings from relaxation rate analysis. Since the native state of human lysozyme shows aggregation it seems likely that the unfolded state aggregates at high concentrations.

3.3.5   Influence of single point mutations on the aggregation behaviour of hLys

In order to highlight the remarkable influence of the mutants on structure and dynamics in unfolded human lysozyme, two all-ala mutants I56T and D67H have been analyzed in this thesis. Figure 51 shows ^1H_N and ^{15}N chemical shifts of these mutants.
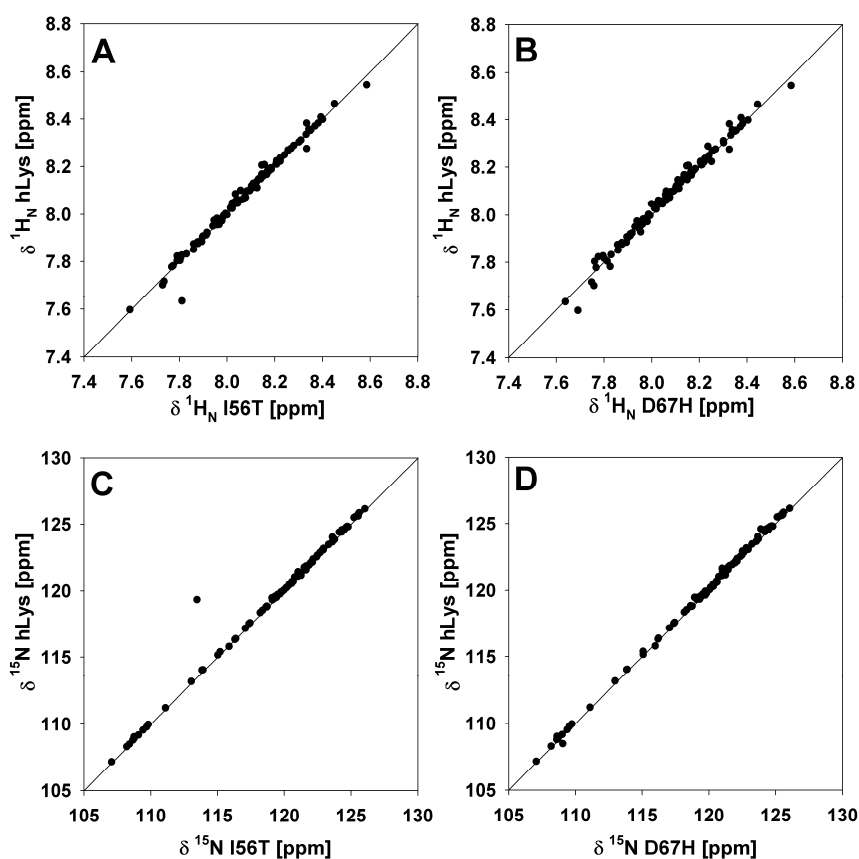
Figure 51: Chemical shifts δ ^1H_N and δ ^{15}N of I56T and D67H; Comparison of the ^1H_N (A, B) and ^{15}N (C, D) chemical shifts of the all-ala mutants I56T and D67H with those of hLys (600MHz, 293K, water pH 2, 10%D_2O, 300µM protein).

As previously seen in the ^1H-^{15}N HSQCs (see 3.3.1) of the mutants both show small deviations in chemical shifts compared to wildtype hLys. Only the isoleucine to threonine mutation at residue 56 deviates large in the ^{15}N chemical shift (figure 51C). One major reason is the substitution of an aliphatic by a polar amino acid. This seems not the case for the aspartate to histidine substitution where only very small variations are observable.

The influence on the dynamics of unfolded hLys mutants was analyzed by R_2 relaxation studies of I56T and D67H. The results are shown in figure 52. While wildtype hLys clearly indicates six clustered regions with high R_2 rates, both mutants disrupt this cluster formation; even though they were all measured at low protein concentration (300µM). Also the baseline is elevated which results in higher intrinsic rates with $R_2^{int} = 3.2$ [s^{-1}] for I56T and $R_2^{int} = 3.0$ [s^{-1}] for D67H; the value of the intrinsic R_2 rate for hLys is 2.7 [s^{-1}]. Highest relaxation rates are observed for the mutant D67H with 7.7 s^{-1} and 7.9 s^{-1} for K33 and W34 in cluster 2. In contrast to the wild type, significant scatter is observed. The mutation at residue 56 cause a disruption of region 3 compared to hLys and also the other regions are affected by this substitution. Similar behaviour is observed for D67H where also the main cluster formation is deleted by decreasing R_2 rates, except of region 3. Here the mutational effect seems little compared to wildtype hLys. Cluster 4, 5 and 6 are clearly separated in the wt sequence, the clustered structure in this area vanishes and rather scattering with values up to 5.9 s^{-1} is observed.
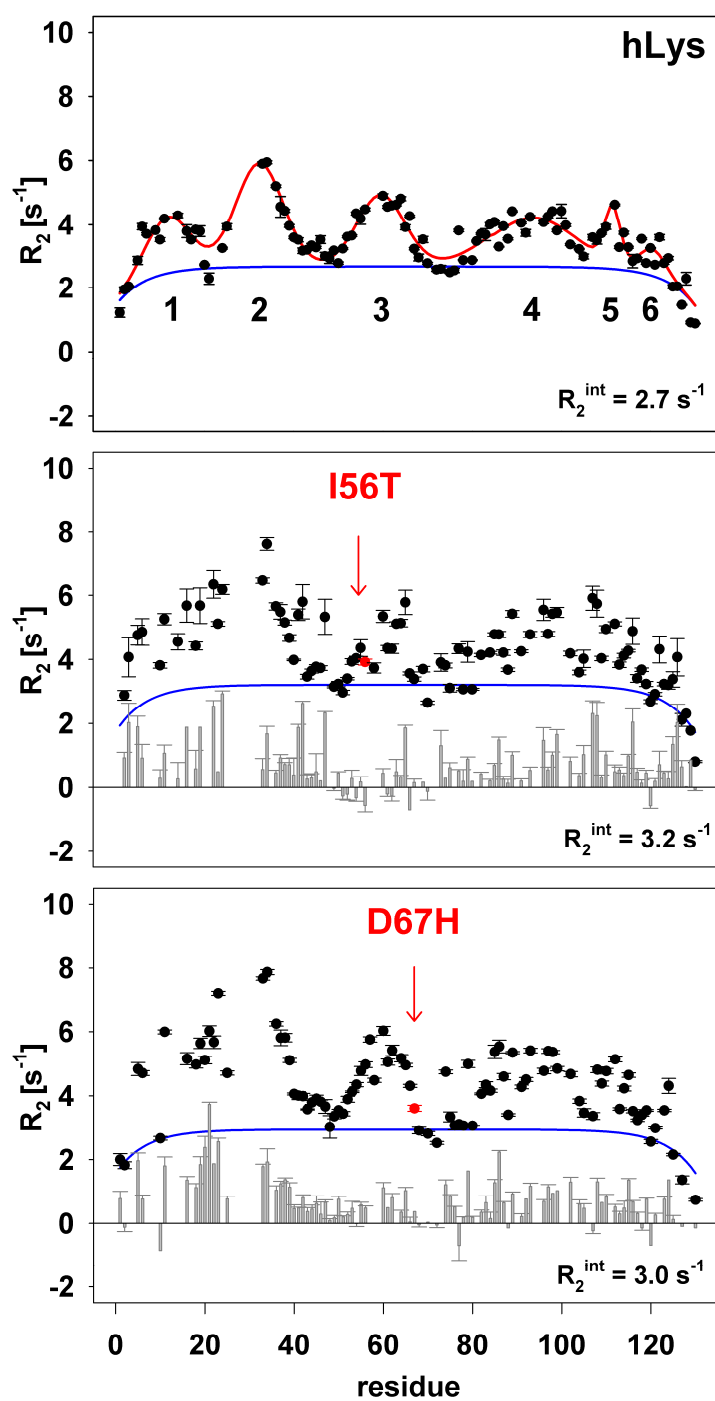
Figure 52: R$_2$ relaxation rates of hLys and mutants I56T and D67H; R$_2$ relaxation rates of hLys(black: 300µM, grey: 850µM), hLys-I56T (middle) and hLysD67H (bottom) (600MHz using 300µM protein at pH2, 293K). Blue line: R$_2^{rc}$ baseline; red line: fit for the wildtype; grey bars: R$_2$ difference of the mutants to the wildtype, respectively.

To get a more detailed view on the mutants among themselves, differences in R_2 relaxation rates of I56T and D67H are compared (figure 53). As one might expect from previous results, these differences are restricted to the regions at the mutation site. But additional effects can be observed around residue 100, where the scattering part is located. Thereby significant scattering is indicative for chemical exchange processes in the mutants. Two tryptophans, W109 and W112, are localized in this region and might interplay with cluster 3. The scattering part is slightly more pronounced in I56T. Both mutants thus show reduced flexibility and exchange even at low concentration..
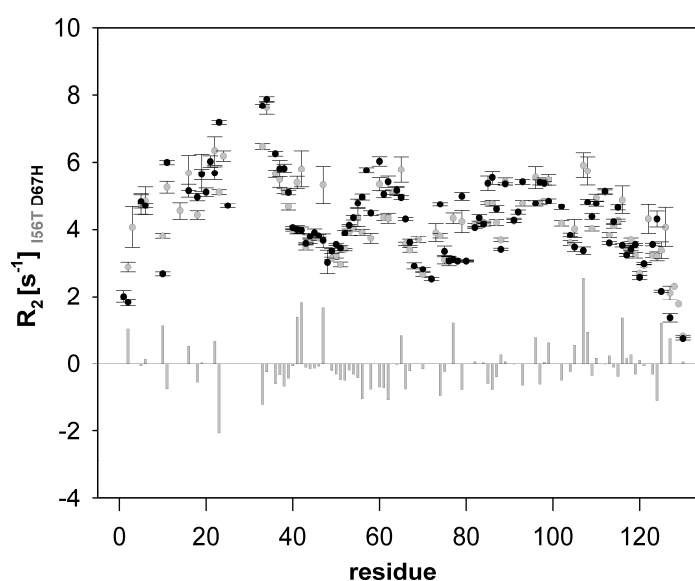
Figure 53: Comparison of R_2 relaxation rates of I56T and D67H; R_2 relaxation rates of I56T (grey) and D67H (black); grey bars: difference of R_2 of I56T and D67H (all data have been collected from experiments at 600MHz, 293K, pH 2, water 10% D_2O, 300µM protein concentration).

The hetNOE of I56T and D67H also provides information about the degree of structure. Figure 54 shows hetNOE data of both mutants compared to wildtype hLys. The regions with positive hetNOE values for hLys are observed particularly from 11-24, 33-69, 81-91 and 108-116, whereas negative values are found at the termini and between residues 72-80 and 97-107. Overall this tendency is reflected in I56T and D67H but with a few exceptions. A clear decrease in intensity is visible in the regions around the mutation site, in detail in region

3. Interestingly in hLys residue 124 has a large positive hetNOE while this residue is negative in both mutants. This might be due to a long range interaction of region 3, where the mutations are located, and the hydrophobic cluster around tyrosine 124.
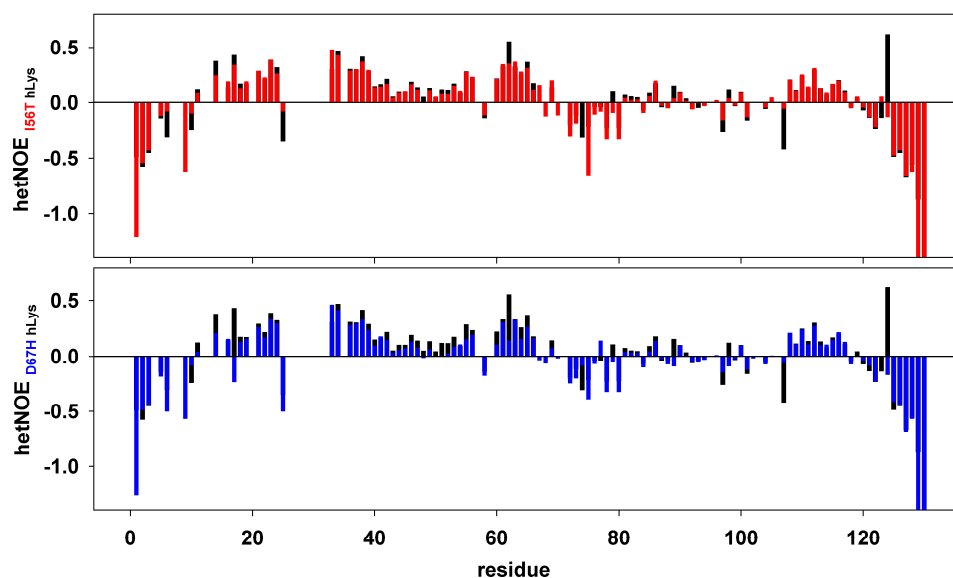
Figure 54: HetNOE of I56T and D67H; Comparison of the heteronuclear NOE of hLys (black bars in the background), I56T (top; red bars) and D67H (bottom; blue bars); all data have been collected from experiments at 600MHz, 293K, pH 2, water 10% D_2O, 300µM protein concentration.

For a more extended analysis of exchange processes in the mutants, reduced spectral densities of I56T and D67H are compared to hLys (figure 55). As explained in materials and methods (chapter 1.2.3.2) spectral densities in the pico- to nanosecond time regime are located at the ascending part of the Lorentzian line while rigid residues can be found at the descending part. The latter are dominated by the overall tumbling. For residues exhibiting chemical exchange, J(0) would be increased by a parallel shift. While the motions in hLys range from very flexible residues to residues which have contributions from both, fast and slow motions, both mutants show clear conformational dynamics indicated by parallel shifts on the J(0) axis. In particular residues I23, K33 and W34 (figure 55, left) shift most. Compared to hLys, both mutants show deviations in the overall profile (figure 55, right) with largest deviating values around region 2, which is also the region with highest R_2 relaxation

rates. Somewhat smaller in intensity are the deviations in region 4, where scattering was observed in relaxation data.
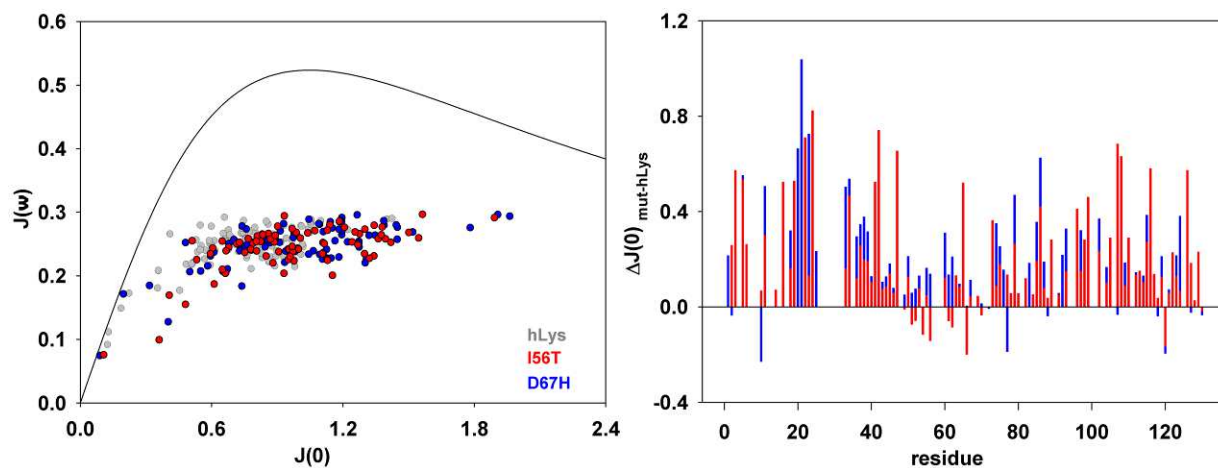
Figure 55: Spectral densities of I56T and D67H; Left: reduced spectral density functions of hLys (grey), I56T (red) and D67H (blue) are compared; right: sequence specific difference plot of J(0) between hLys and the particular mutant (I56T: red, D67H: blue).

The results from spectral density mapping stand in good agreement with the findings from R_2 relaxation data, where at least these regions seem to form oligomeric aggregates.

3.3.6   Degree of compaction by translational diffusion

To test the hypothesis of aggregation of unfolded human lysozyme and mutants, translational diffusion measurements were performed. Thereby the hydrodynamic radii have been determined. Since the aggregation tendency should increase with higher concentration, proteins were analyzed at 300µM and 850µM (table 6).

	R_h [Å] 300 µM	R_h [Å] 850 µM	R_h increase
hLys	**26.7** ±0.4	**30.3** ±0.2	+13.5%
I56T	**31.9** ±0.4	**33.5** ±0.3	+10.5%
D67H	**30.3** ±0.5	**35.2** ±0.5	+16.2%

Table 6: Hydrodynamic radii R_h of hLys, I56T and D67H; Comparison of the hydrodynamic radii R_h of hLys, I56T and D67H at two different concentrations, 300µM and 850µM. Data were obtained from translational diffusion measurements (600MHz, 293K, pH2, water 10% D_2O).

The R_h of hLys at low concentration is with a value of 26.7±0.4 Å significantly smaller than those of I56T, R_h=31.9±0.4Å and D67H, R_h=30.3±0.5Å. Increasing radii by the introduction of single point mutations at crucial positions in the protein sequence were expected. Both mutations increase the aggregation behaviour and the measured R_h seems to trace back to a mixture of aggregated and not aggregated protein in the sample, while this is not the case for hLys. It is known that both mutants show amyloid formation in the native state which implicates an increasing R_h at higher protein concentrations. Also hLys aggregates at higher concentration as shown by a 13.5% increase of R_h at 850µM. The mutants already show higher hydrodynamic radii even at 300µM but increase further as the concentration is increased up to R_h=33.5±0.3Å for I56T and R_h=35.2±0.5Å for D67H. The data suggest that the mutations move the onset for aggregation towards lower concentration while the concentration dependent increase is in the same range (around 10%) for all three proteins.

3.3.7   Conformational dynamics

Chemical shifts, relaxation data and diffusion measurements indicate a higher compaction of hLys compared to unfolded hen lysozyme. The mutants I56T and D67H also show higher aggregation propensities than mutants of WT^{Ala}. This leads to the assumption that human lysozyme has a higher aggregation propensity through insertion of single point mutations than hen lysozyme. In the following changes in conformational dynamics on the basis of the native structure of hLys will be explained. Therefore R_2 relaxation rates are mapped by a running average onto the native state structure of human lysozyme to provide insights about structural features located in the core or at the outer sphere of the protein (figure 56).
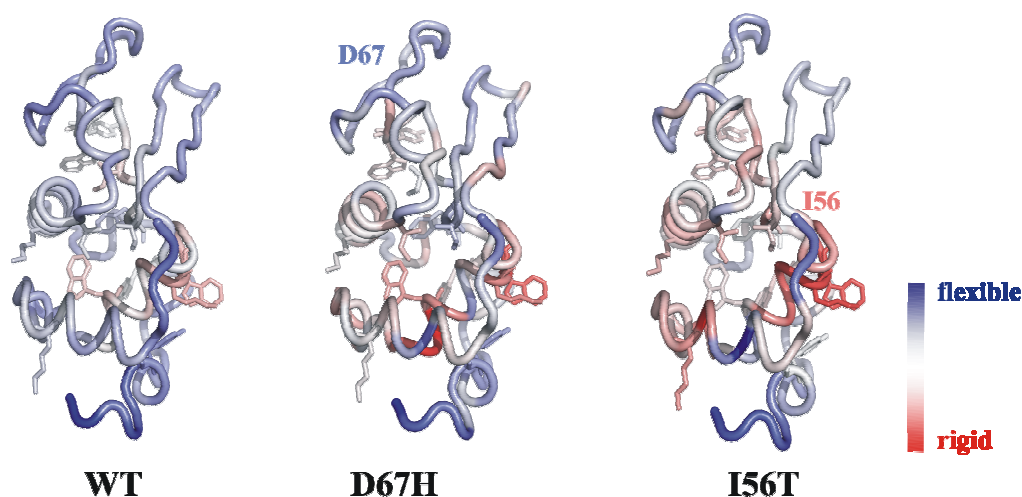
Figure 56: Dynamical characteristics of hLys, I56T and D67H; Illustration of dynamical characteristics of hLys, I56T and D67H; R_2 relaxation rates are mapped onto the native state structure of human lysozyme (2zij.pdb(*34*)); blue: low R_2 rates, flexible regions; white: medium R_2 rates; red: high R_2 relaxation rates, rigid regions.

With this form of mapping, flexible and rigid regions in the protein can be highlighted. In the wildtype residues that would be in the core of the native protein have lower flexibility (coloured in white-red) than those on the outside of the protein (coloured in blue). One exception is the region around W62/Y63 with only rigid properties. Similar behaviour but in a more pronounced intensity show both amyloidogenic mutants. There changes in the

relaxation rates are induced towards higher rigidity or increased exchange. Interestingly D67H shows a smaller increase in R_2 rates in the region around the mutation site than I56T. This might be due to a larger unfolding through substitution of an aliphatic isoleucine by a polar threonine, whereas in D67H the change of polarity is negligible. From this observation one can conclude that I56T has a higher aggregation propensity than D67H. The overall changes of both mutants towards higher relaxation rates cannot be explained by interactions present in the native structure, since relaxation rates are influenced in regions of the protein that have no interactions with the mutation site but point towards non-native interactions in the pre-molten globule state of the protein. Neither can they be explained by using a random coil model without any interactions between parts of the protein distant in sequence. This is the reason why formation of transient oligomers may be a suitable explanation, taking in account the observed chemical exchange. The incorporation of these findings is shown in a free energy landscape linking native state, unfolded non-native states and aggregates (figure 57).
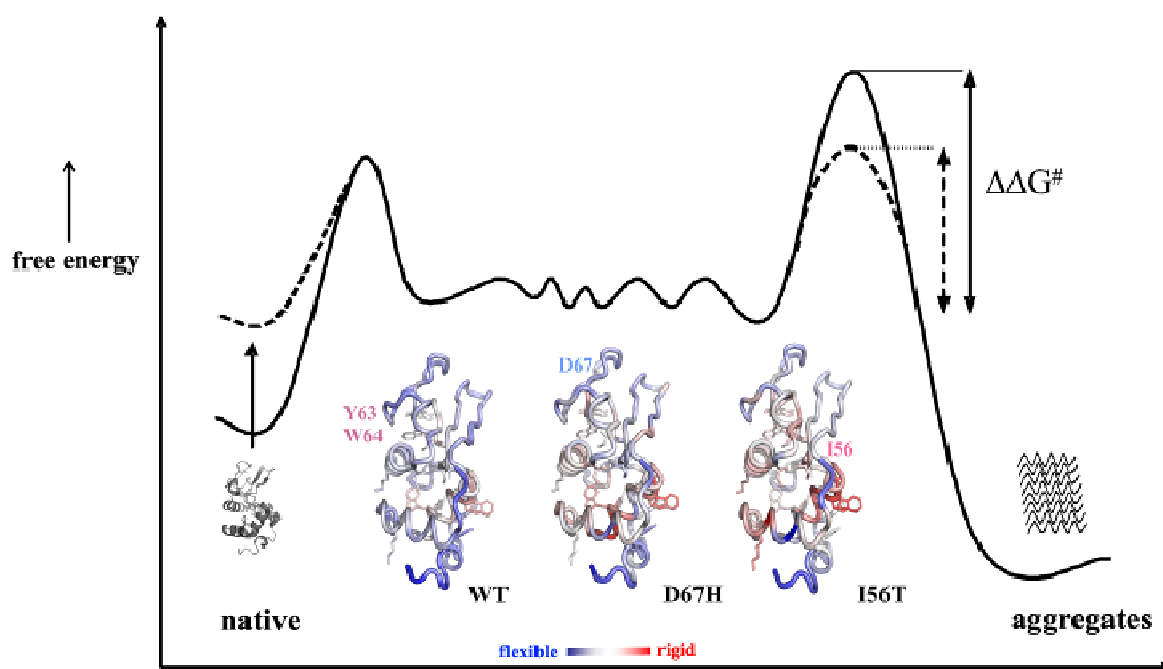
Figure 57: Energy landscape of unfolded human lysozyme; Proposed free energy landscape for the unfolding and aggregation of hLys, I56T and D67H; dashed lines: landscape of human lysozyme; solid line: landscape of hen lysozyme; structures: see chapter 3.3.7, figure 51 (2zij.pdb (*34*)).

As previously reported (*18*), single point mutations do not change the structure of the native state but lead to a destabilization resulting in significantly reduced melting temperatures of the mutants. This results in an elevated free energy level of native human lysozyme. In the unfolded state ensemble single point mutants have a pronounced effect on the free energy landscape. The relative populations are changed within the unfolded non-native state ensemble towards transient population of states that are aggregation-competent, here the mutants. These states than consequently have a lower activation energy $\Delta\Delta G^{\#}$ for the transition from the unfolded pre-molten globule state towards the aggregated state. From this one can conclude that the ruggedness of the folding landscape connecting different members of the ensemble of pre-molten states is affected by single point mutations and a change in energetics of the folding landscape leads to aggregation.

To shortly summarize the findings in this chapter about unfolded human lysozyme and two amyloidogenic mutants, the overall compactness and formation of long-range hydrophobic clusters characteristic within the family of lysozymes is maintained, the dynamics of the conformational ensemble is markedly different in the disease-related single-point mutants; even if initial studies revealed that the native state structures of wildtype and amyloidogenic mutants are highly similar (*18*). The model, analyzed in this thesis on the basis of hen and human lysozyme, therefore implies a very subtle and fragile balance between maintaining the functional native state of the protein and avoidance of mutations that, by changing the energetics of the folding landscape, would lead to aggregation.

3.4 Conclusion of the thesis

Recapitulating the introduction, lysozyme is one of the most studied model protein for folding and unfolding processes over several decades. However, the question of misfolding and aggregation linked with neurodegenerative diseases is still not completely resolved and leaves large space for further research. In the focus of the present thesis was an extensive expansion of the studies on two proteins from the lysozyme family, hen and human lysozyme in their nonnative states by nuclear magnetic resonance. New insights into the unfolded ensembles of both proteins have been delivered with very detailed and astonishing new results.

At first, a nonnative all cysteine-to-alanine variant of hen lysozyme (WT^{Ala}) was prepared and analyzed here at low pH in the absent of any denaturant. Chemical shift deviations revealing residual structure of the backbone as well as coupling constants revealing torsion angle preferences for both, backbone and side chain angles have been presented. The backbone angle ϕ was tested by $^3J(H_N,H^\alpha)$ coupling constants, the backbone angle ψ by $^1J(N,C^\alpha)$ and $^2J(N,C^\alpha)$ coupling constants. The side chain angle χ_1 was evaluated using $^3J(H^\alpha,H^\beta)$ and $^3J(CO,H^\beta)$ coupling constants and compared to database results. Dynamical characteristics of this unfolded system have been evaluated by relaxation measurements, in particular R_1, R_2, $R_{1\rho}$ relaxation rates and the heteronuclear NOE. Furthermore relaxation dispersion data have been analyzed which include information about chemical exchange.

Besides the standard treatment of presenting relaxation data as function of the protein sequence, the analysis of reduced spectral densities with a $J(\omega_N)/J(0)$ correlation was used to identify residues whose motions are governed by internal motions only, residues that relax predominantly due to global tumbling and residues that are influenced by both, global and internal motions. Besides this, ^{15}N relaxation rates and residual dipolar coupling constants, both anisotropic parameters, have been analyzed here in a conjoint matter by analysis of their distribution. The combined per residue analysis of R_2 and RDC data, which has been done for the first time in this work, allows the identification of residual structure persisting on different timescales. In addition, the conformational landscape of hen lysozyme by introducing single point mutations in the six most prominent clusters of the protein and

delineating their influence on overall structure and dynamics was probed. This was supported by extensive diffusion measurements and small angle x-ray scattering (SAXS) whereby the hydrodynamic and gyration radii have been determined. Single point mutations W28G^{Ala}, W62G^{Ala}, W108G^{Ala}, W111G^{Ala} and W123G^{Ala} are shown to have substantial effects on local and long-range structural dynamics in the unfolded states, further supporting the notion of their impact on the rate of productive folding and unproductive misfolding and aggregation by modulating the conformational characteristics of unfolded states of proteins.

The second model protein analyzed in this thesis was nonnative human lysozyme. As previously described for hen lysozyme, the unfolded state of human lysozyme was similarly stabilized by an all cysteine-to-alanine mutation (hLys) in the absence of any chemical denaturant. This nonnative system was investigated for the first time in the present thesis evaluating structural and dynamical properties. Especially aggregation behaviour compared to unfolded hen lysozyme was analyzed. Previous studies on human lysozyme revealed that the native state structures of wildtype and amyloidogenic mutants of hLys are highly similar (*18*). Characterization of the thermal unfolding at pH 1.2 using CD, DSC and NMR spectroscopy showed that unfolding proceeds via loss of tertiary structure, resulting in a pre-molten globule state, which unfolds further, as the temperature increases (*31*). Here the pre-molten globule unfolded states of the reduced states of hLys and two mutants have been analyzed with respect to residual structure elements and aggregation behaviour. Therefore ^{15}N relaxation rates and chemical shift deviations have been determined and the concentration dependency was tested. Even though the overall structural characteristics of human and hen lysozyme are remarkably similar, with six hydrophobic regions and residual secondary structure, the wildtype sequence of hLys reveals greater chemical shift deviations, higher heteronuclear relaxation rates, a larger spread in these rates and through a reduction of flexibility as evidenced by the analysis of $J(\omega_N)/J(0)$. All results of the hLys wildtype analysis indicate higher α-helical residual structure with high stability in clustered regions compared to WT^{Ala}. The effects of the disease-related mutations in hLys are different compared to the W-to-G mutants in hen lysozyme. Partial disruption of the cluster structure, larger relaxation rates and larger scatter of the rates, indicative for chemical exchange, are observed in these mutants. These effects are also observed in the wt sequence but only at

high concentration. Concentration dependent data in hLys show increased R_2 relaxation rates and in particular scatter in the regions of cluster 1, 2, 4, 5 and 6 are observed at higher protein concentration while R_1, hetNOE and R_{1rho} remain qualitatively the same as at low concentration. Scatter, the height of the rates and the big difference between R_2 and R_{1rho} data are indicative for exchange. Both, the increase of rigidity by the overall higher relaxation rates and the presence of exchange indicate most likely the transient formation of oligomers in the sample at higher concentration. Formation of amyloid fibrils neither occurs from the folded conformation nor from a completely unstructured state but from some partially structured, aggregation-competent state. From the present thesis one can conclude that the degree of residual structure and restriction of dynamics in hLys compared to the homologous hen lysozyme variant correlates with the increased tendency of nonnative human lysozyme to form amyloids. One reason for this might be the regions where the two proteins differ most in their primary sequence of the polypeptide chain. While positively charged and neutral residues are found in cluster 4 in hen lysozyme at acidic pH, the region is predominantly neutral in hLys, explaining the observed differences by the absence of charge repulsion and a potential increase in hydrophobic interactions. As it is described in the literature starting with the native state protein, it is indeed the region of the β-domain and the adjacent helix 3 (*33*) which is found to be the first region involved in amyloid formation.

4 Appendix

4.1 Media and buffers

M9 minimal medium

Before autoclaving	42 mM Na_2HPO_4 22 mM KH_2PO_4 8,5 mM NaCl 1 g/L $^{15}NH_4Cl$ or ^{13}C Glucose
After autoclaving	1 mM $CaCl_2$ 2 mM $MgSO_4$ 10 g/L Glucose (sterile) 1 mL/L trace elements (1 mM Cu^{2+}, Zn^{2+}, Mn^{2+}, Co^{2+}, Ni^{2+}, MoO^+) 30 μM $FeCl_3$ 5 ng/L Thiamin 1 mL/L Ampicillin Stammlösung

Solubilisation and purification buffer

Buffer 1 (sonification)
50 mM Tris/HCl pH 7,5
25 % Saccharose (w/v)
1 mM EDTA

Buffer 2 (washing)
20 mM Tris/HCl pH 7,5
1 % Triton X-100
1 mM EDTA

Buffer 3a (denaturing)
20 mM Tris /HCl pH 7,5
50 mM NaCl
5 mM EDTA
8 M Urea

Buffer 3b (native conditions)
20 mM Tris /HCl pH 7,5
50 mM NaCl
5 mM EDTA
15,5 g/L DTT

Buffer A (loading)
20 mM Tris /HCl pH 7,5
8 M Urea
5 mM EDTA

Buffer B (elution)
20 mM Tris /HCl pH 7,5
8 M Urea
5 mM EDTA
300 mM NaCl

4.2 Primer sequences, DNA and protein sequences

I56T forward

 5'-CTG AGC ACT GAC TAC GGC ACC TTC CAG ATC AAC TCC-3'

I56T reverse

5'-GGA GTT GAT CTG GAA GGT GCC GTA GTC AGT GCT ACG-3'

D67H forward

5'-CGC TAT TGG GCC AAC CAC GGT AAA ACC CCG GGT GCG-3'

D67H reverse

5'-CGC ACC CGG GGT TTT ACC GTG GTT GGC CCA ATA GCG-3'

WT^{Ala}

```
        NdeI
DNA     CATATGAAAGTTTTTGGCCGTGCTGAGTTAGCGGCGGCAATGAAACGTCAT   (1-51)
Protein    M   K   V   F   G   R   A   E   L   A   A   A   M   K   R   H   (1-16)


DNA     GGTTTGGACAACTATCGCGGCTATAGTCTGGGGAATTGGGTTGCTGCGGCC   (52-102)
Protein G   L   D   N   Y   R   G   Y   S   L   G   N   W   V   A   A   A   (17-33)


DNA     AAATTTGAAAGCAATTTCAATACCCAAGCGACCAACCGTAATACGGATGGC   (103-153)
Protein K   F   E   S   N   F   N   T   Q   A   T   N   R   N   T   D   G   (34-50)


DNA     TCGACGGACTACGGTATTCTGCAGATCAACSGCCGCTGGTGGGCGAACGAC   (154-204)
Protein S   T   D   Y   G   I   L   Q   I   N   S   R   W   W   A   N   D   (51-67)


DNA     GGTCGTACGCCGGGCAGCCGTAATCTGGCAAATATCCCAGCCTCGGCGCTT   (205-255)
Protein G   R   T   P   G   S   R   N   L   A   N   I   P   A   S   A   L   (68-84)


DNA     CTCAGTTCGGATATTACCGCAAGTGTTAACGCGGCAAAAAAGATCGTGAGC   (256-306)
Protein L   S   S   D   I   T   A   S   V   N   A   A   K   K   I   V   S   (85-101)


DNA     GATGGCAACGCGATGAACGCGTGGGTGGCCTGGCGTAACCGCGCCAAAGGT   (307-357)
Protein D   G   N   G   M   N   A   W   V   A   W   R   N   R   A   K   G   (102-118)
                                                    BamHI
DNA     ACCGATGTGCAGGCCTGGATTCGCGGGGCTCGTCTGTAAGGATCC   (358-401)
Protein T   D   V   Q   A   W   I   R   G   C   R   L   Stop   (119-130)
```

W28G^{Ala}

```
        10          20          30          40          50          60
MKVFGRAELA  AAMKRHGLDN  YRGYSLGNGV  AAAKFESNFN  TQATNRNTDG  STDYGILQIN
        70          80          90         100         110         120
SRWWANDGRT  PGSRNLANIP  ASALLSSDIT  ASVNAAKKIV  SDGNGMNAWV  AWRNRAKGTD
       130
VQAWIRGARL
```

W62G^{Ala}

```
        10          20          30          40          50          60
MKVFGRAELA  AAMKRHGLDN  YRGYSLGNWV  AAAKFESNFN  TQATNRNTDG  STDYGILQIN
        70          80          90         100         110         120
SRGWANDGRT  PGSRNLANIP  ASALLSSDIT  ASVNAAKKIV  SDGNGMNAWV  AWRNRAKGTD
       130
VQAWIRGARL
```

W108G^{Ala}

```
        10          20          30          40          50          60
MKVFGRAELAA  AMKRHGLDNY  RGYSLGNWVA  AKFESNFNT  QATNRNTDGS  TDYGILQINS
        70          80          90         100         110         120
RWWANDGRTP  GSRNLANIPA  SALLSSDITA  SVNAAKKIVS  DGNGMNAGVA  WRNRAKGTDV
       130
QAWIRGARL
```

W123G^{Ala}

```
        10          20          30          40          50          60
MKVFGRAELA  AAMKRHGLDN  YRGYSLGNWV  AAAKFESNFN  TQATNRNTDG  STDYGILQIN
        70          80          90         100         110         120
SRWWANDGRT  PGSRNLANIP  ASALLSSDIT  ASVNAAKKIV  SDGNGMNAWV  AWRNRAKGTD
       130
VQAGIRGARL
```

W111G^{Ala}

```
        10        20        30        40        50        60
MKVFGRAELA AAMKRHGLDN YRGYSLGNWV AAAKFESNFN TQATNRNTDG STDYGILQIN
        70        80        90        100       110       120
SRWWANDGRT PGSRNLANIP ASALLSSDIT ASVNAAKKIV SDGNGMNAWV AGRNRAKGTD
        130
VQAWIRGARL
```

hLys-all-ala

```
        NdeI
DNA      CATATGAAAGTCTTTGAACGTGCTGAACTGGCGCGTACCCTGAAACGTCTG   (1-51)
Protein    M   K   V   F   E   R   A   E   L   A   R   T   L   K   R   L   (1-16)


DNA      GGTATGGACGGTTACCGTGGCATCTCTCTGGCCAACTGGATGGCACTGGCG   (52-102)
Protein  G   M   D   G   Y   R   G   I   S   L   A   N   W   M   A   L   A   (17-33)


DNA      AAATGGGAGTCTGGTTACAACACGCGTGCCACCAACTATAACGCGGGTGAC   (103-153)
Protein  K   W   E   S   G   Y   N   T   R   A   T   N   Y   N   A   G   D   (34-50)


DNA      CGTAGCACTGACTACGGCATCTTCCAGATCAACTCCCGCTATTGGGCCAAC   (154-204)
Protein  R   S   T   D   Y   G   I   F   Q   I   N   S   R   Y   W   A   N   (51-67)


DNA      GACGGTAAAACCCCGGGTGCGGTTAACGCAGCTCACCTGTCTGCGTCTGCT   (205-255)
Protein  D   G   K   T   P   G   A   V   N   A   A   H   L   S   A   S   A   (68-84)


DNA      CTGCTCCAGGACAACATTGCGGATGCAGTAGCAGCTGCCAAACGCGTTGTT   (256-306)
Protein  L   L   Q   D   N   I   A   D   A   V   A   A   A   K   R   V   V   (85-101)


DNA      CGTGATCCGCAGGGTATCCGTGCATGGGTTGCTTGGCGTAACCGTGCGCAG   (307-357)
Protein  R   D   P   Q   G   I   R   A   W   V   A   W   R   N   R   A   Q   (102-118)

                                                             BamHI
DNA      AACCGTGATGTGCGTCAGTACGTCCAGGGTGCTGGTGTGTAAGGATCC   (358-408)
Protein  N   R   D   V   R   Q   Y   V   Q   G   A   G   V   Stopp   (119-131)
```

hLysI56T

```
          10         20         30         40         50         60
MKVFERAELA RTLKRLGMDG YRGISLANWM ALAKWESGYN TRATNYNAGD RSTDYGTFQI
          70         80         90        100        110        120
NSRYWANDGK TPGAVNAAHL SASALLQDNI ADAVAAAKRV VRDPQGIRAW VAWRNRAQNR
         130
DVRQYVQGAG V
```

hLysD67H

```
          10         20         30         40         50         60
MKVFERAELA RTLKRLGMDG YRGISLANWM ALAKWESGYN TRATNYNAGD RSTDYGIFQI
          70         80         90        100        110        120
NSRYWANHGK TPGAVNAAHL SASALLQDNI ADAVAAAKRV VRDPQGIRAW VAWRNRAQNR
         130
DVRQYVQGAG V
```

4.3 Chemical shifts, J-couplings, relaxation data and RDCs of WT^{Ala}

	^1H_N	^{15}N	$^1H^\alpha$	$^1H^{\beta 2}$	$^1H^{\beta 3}$	$^{13}C^\alpha$	$^{13}C^\beta$	$^{13}C'$	$H^\alpha H^{\beta 2}$	$H^\alpha H^{\beta 3}$	$C'H^{\beta 2}$	$C'H^{\beta 3}$	R_1	$R_{1\rho}$	R_2	hetN.	RDC
K1	8.70	126.11	4.41	1.24		56.23	32.82	173.33	7.86				0.85	0.94	1.05	-0.75	-5.41
V2	8.25	123.71	4.13		1.52	61.83	32.79	173.41		7.31		2.01	1.08	1.35	1.27	-0.37	-7.33
F3	8.50	126.24	4.70	2.54	2.68	57.5	39.8	173.95	7.40	5.13	3.59	2.79	1.19	1.48	1.75	-0.35	-2.86
G4	8.35	111.82	3.99				45.02	171.91					1.22	1.67	1.62	-0.33	-1.29
R5	8.30	122.25	4.25	1.41	1.85	57.02	30.31	174.81	7.66	5.36	3.72		1.38	2.15	2.18	-0.11	-0.06
A6	8.49	125.48	4.29	0.96		53.22	18.52	176.44	8.68		7.03		1.48	2.47	2.50	-0.09	
E7	8.21	120.72	4.29	1.56		56.62	28.01	174.69	7.78		3.68		1.47	2.42	2.48	0.02	1.15
L8	8.17	124.22				55.85	41.92	175.66					1.55	2.60	2.66	-0.01	-2.31
A9				0.98					8.68								
A10	8.03	122.87	4.23	0.98		53.47	18.47	176.48	8.86				1.53	2.72	2.75	-0.07	
A11	8.10	123.13	4.25	0.99		53.39	18.41	176.47	8.67		6.93		1.58	2.86	2.88	-0.05	3.76
M12	8.09	119.51	4.21	1.63		56.02	32.46	176.49	7.74				1.58	2.66	2.75	-0.06	-0.55
K13	8.06	122.81	4.29			56.69	32.51	174.74					1.58	2.75	2.80	-0.08	
R14	8.14	122.21	4.29	1.31		56.38	30.38	174.15	7.05		3.67		1.56	2.78	2.81	-0.02	
H15	8.46	120.17	4.73	2.85	2.74	55.38	28.7	172.65	6.48	5.26	3.36	2.95	1.52	2.51	2.64	-0.06	-4.34
G16	8.42	111.09	4.00				45.13	171.83					1.40	2.37	2.31	-0.11	-5.35
L17	8.20	122.7	4.39	1.18		55.23	42.2	175.13	7.93		3.24		1.48	2.43	2.43	0.02	-3.16

D18	8.53	120.36	4.68	2.31	2.39	52.82	37.64	172.56	5.96	4.78		3.55	1.48	2.39	2.44	-0.02	-3.05
N19	8.29	120.16	4.70	2.23	2.28	52.91	38.3	172.57	6.40	5.67		3.82	1.49	2.35	2.52	-0.04	-2.09
Y20	8.07	121.73	4.54	2.53	2.62	58.1	38.3	173.62	8.48	6.94	4.59		1.54	2.63	2.71	-0.07	0.82
R21	8.24	124.54	4.23	1.18	1.31	56.09	30.11	174.14	7.39	5.11	3.04	2.68	1.64		2.93	0.05	-0.36
G22	7.77	109.88	3.88				45.03	171.46					1.53	2.54	2.76	-0.06	-0.87
Y23	7.94	121.02	4.61	2.46	2.54	57.77	38.78	173.69	7.16	5.59	3.87	3.03	1.61	2.82	2.77	0.07	-0.11
S24	8.27	118.77	4.47	3.66	3.35	57.92	63.9	172.17	5.84			4.44	1.53	2.66	2.68	0.03	-2.57
L25	8.22	125.17	4.34	1.18		55.51	42.01	175.62	7.88		3.49		1.54	2.67	2.58	0.03	
G26	8.23	109.59	3.86				45.21	171.79					1.49	2.54	2.56	0.02	10.44
N27	8.19	119.93	4.73			53.08	38.41	172.87					1.49	2.63	2.72	0.02	-3.09
W28	8.08	122.98	4.61	2.80		57.91	29.07	174.1	7.38		5.06		1.54	2.84	3.01	-0.05	-0.99
V29	7.74	123.47	3.87		1.45	62.52	32.58	173.57		7.33		2.12	1.53	3.25	3.28	-0.07	3.61
A30	8.03	127.55	4.22	0.91		52.62	18.66	175.69		6.90			1.54	3.28		-0.05	3.98
A31	8.09	123.75	4.22	0.91		52.64	18.7	175.93	8.91	6.81			1.44	2.94	3.03	-0.06	-2.93
A32	8.11	123.79	4.23	0.88		52.64	18.69	175.93	8.95	6.97			1.47	3.03	3.03	-0.07	-0.97
K33	8.07	120.7	4.20	1.23		56.62	32.54	174.35	7.43				1.47	2.98	3.15	-0.10	-0.59
F34	8.10	121.59	4.61	2.88	2.97	57.89	39.25	173.59			4.17	2.69	1.54	3.16	3.11	-0.04	-2.43
E35	8.15	122.68	4.36	1.48	1.63	55.55	28.53	173.6	9.00	4.86	3.20	2.66	1.52	2.97	3.03	-0.03	-1.42
S36	8.27	117.77	4.38	3.6	3.37	58.59	63.67	172.09		4.99			1.43	2.71	2.69	0.01	-1.84
N37	8.35	121.28	4.70	2.23		53.32	38.54	172.82	6.52			4.07	1.50	2.60	2.65	-0.08	
F38	8.15	121.48	4.60	2.84	2.98	58.01	39.09	173.4	7.75		3.92	2.05	1.48	2.71	2.63	-0.01	-1.62
N39	8.34	121.25	4.75	2.26	2.38	53.24	38.49	173.38	5.10	4.90	3.31	3.92	1.50	2.60	2.65	-0.08	
T40	8.12	115.6	4.32	4.13		62.13	69.5	172.64				5.53	1.44	2.49	2.43	-0.08	-2.22
Q41	8.33	123.18	4.35	1.55	1.68	55.9	28.96	173.78	8.23	5.25	3.08	2.63	1.53	2.63	2.59	-0.08	-2.19
A42	8.28	125.97	4.37	0.96		52.64	18.82	175.89	8.76	6.92			1.44	2.33	2.42	-0.02	-0.09
T43	8.10	113.86	4.32	4.19		61.89	69.7	172.26					1.37	2.28	2.25	-0.13	-2.26
N44	8.38	122.09	4.75	2.35	2.38	53.13	38.46	173.04	5.12	5.00		4.11	1.50	2.42	2.44	-0.15	
R45	8.33	122.54	4.36	1.29	1.41	56.14	30.34	173.96	7.38	4.65	2.97	2.83	1.48	2.39	2.41	-0.09	-3.07
N46	8.50	120.7	4.80	2.33	2.41	53.34	38.49	173.57	4.82	4.30			1.48	2.39	2.56	-0.03	-2.31
T47	8.18	115.17	4.36			61.9	69.47	172.57				5.60	1.44	2.24	2.26	-0.06	-3.84
D48	8.51	121.95	4.75	2.45	2.79	53.08	37.67	173.49	5.71			3.96	1.52	2.37	2.34	-0.06	-2.17
G49	8.38	110.62	4.02				45.31	172.15					1.35	2.13	2.41	-0.14	-0.01
S50	8.17	116.63	4.54	3.45	3.42	58.5	63.79	172.77	5.36	5.32		4.24	1.42	2.31	2.33	0.01	-0.83
T51	8.21	116.76	4.33			61.89	69.59	172					1.47	2.50	2.60	0.02	-2.30
D52	8.38	122.12	4.75	2.29	2.41	53.01	38.33	172.6	5.33	4.98	2.46	2.68	1.57	2.71	2.73	-0.17	
Y53	8.13	121.98	4.50	2.78	2.93	58.52	38.11	174.25	7.8		3.41	2.62	1.55	3.08	3.08	0.01	
G54	8.27	110.88	3.92				45.37	172.09					1.59	3.06	3.10	0.05	-0.23
I55	7.84	120.85	4.17	1.42		61.41	38.3	174.2	6.48		2.38		1.63	3.35	3.58	0.10	0.05
L56	8.22	126.06	4.36	1.20	1.10	55.33	42.01	175.15	8.42	5.30	3.32		1.68	3.47	3.62	0.09	-3.78
Q57	8.27	122.46	4.33			55.66	28.83	174.04					1.61	3.57	3.63	0.07	-3.81
I58	8.09	122.63	4.05	1.37		61.74	38.21	173.95	6.94		2.45		1.59	3.54	3.73	0.02	
N59	8.35	122.71	4.75	2.32	2.65	53.29	38.41	173.27	5.72			4.07	1.67	3.86	3.77	0.07	-4.12
S60	8.20	117.45	4.29	3.41	3.33	58.95	63.54	172.83	4.79	4.41	2.41	4.24	1.62	3.72	3.87	0.14	-0.37
R61	8.23	123.34	4.06	1.08		56.44	29.5	174.34	7.03		3.17		1.74	3.95	3.83	0.17	-2.32
W62	7.82	121.8	4.56	2.75		57.71	29.02	174.16	7.21		4.85		1.76	4.02	4.21	0.13	9.93
W63	7.39	121.87	4.47	2.49	1.95	57.26	28.96	173.82	5.62	5.1	3.97	3.11	1.82	4.18	4.41	0.12	8.26

A64	7.72	125.22	4.17	0.7		52.57	18.66	175.3	8.31		6.88		1.75	3.67	3.85	0.08	8.86
N65	8.11	118.12	4.65	2.23	2.39	53.16	38.41	173.11	6.43	5.66	3.24	3.69	1.58	3.25	3.31	0.06	-3.78
D66	8.33	119.53	4.66	2.66	2.76	53.09	37.56	173.35	5.60		2.95	3.53	1.62	3.07	3.22	0.04	-2.46
G67	8.31	109.46	3.88			45.44		171.94					1.61	2.82	3.20		1.39
R68	7.96	120.99	4.45	1.30	1.38	55.75	30.63	174.13	6.55	3.99	2.84	2.59	1.63	2.82	2.81	0.07	1.90
T69	8.25	118.25				59.82	69.51	170.91					1.48	2.62	2.79	0.04	
P70																	
G71	8.48	110.54	4.00			45.13		172.15									-2.81
S72	8.12	116.63	4.47	3.43		58.53	63.83	172.58	5.24			4.34	1.48	2.49	2.49	-0.04	-0.74
R73	8.42	123.75	4.36	1.58	1.69	56.08	30.42	173.72			2.49	2.02	1.45	2.46	2.47	-0.13	-4.09
N74	8.44	120.85	4.72	2.26	2.40	53.09	38.39	173.02	6.28	5.87	2.98	3.64	1.45		2.51	-0.16	
L75	8.26	124.28	4.34	1.19		55.21	41.98	174.92	8.67		3.01		1.46	2.49	2.57	-0.03	-1.34
A76	8.19	124.58	4.31	0.93		52.4	18.87	175.17	8.66				1.45	2.44	2.61	-0.05	-1.20
N77	8.28	118.42	4.70	2.24	2.35	52.73	38.54	172.44			3.19	3.36	1.38	2.40	2.54	-0.03	
I78	8.02	123.96				58.66	38.43	172.23					1.40	2.66	2.72	-0.11	-8.23
P79																	
A80	8.49	125.66	4.26	0.98		53.16	18.53	176.36	8.71		7.02		1.48	2.72	2.68	-0.09	-3.61
S81	8.25	114.97	4.36	3.46	3.40	58.68	63.41	172.69	4.60	4.24		4.63	1.40	2.89			-0.32
A82	8.23	126.97	4.37	0.9		52.62	18.78	175.75	8.62		6.79						0.65
L83	7.97	121.5	4.33	1.21		55.38	41.97	175.42	8.71		3.55		1.44	2.79	2.69	-0.08	-3.50
L84	8.04	123.18	4.42	1.23		55.3	42.04	175.59	8.77		3.26		1.44	2.68	2.63	-0.15	-2.24
S85	8.23	117.14	4.44	3.48	3.42	58.69	63.57	172.7	5.21	4.97			1.37	2.51	2.64	-0.04	-2.07
S86	8.27	118.36	4.48	3.47	3.43	58.68	63.59	172.3	4.75	4.69		4.31	1.39	2.57	2.54	-0.02	-0.31
D87	8.40	121.99	4.79	2.42	2.49	53.01	37.58	173.12	3.72	3.68			1.45	2.50	2.60	-0.19	-1.56
I88	8.08	122.49	4.23	1.47		61.77	38.16	174.51	7.07		2.27		1.43	2.58	2.62	-0.17	-0.72
T89	8.16	119.2	4.29	4.13		62.07	69.54	172.27				5.20					-3.65
A90	8.26	127.34	4.37	0.98		52.7	18.84	175.91	8.7		7.00		1.39	2.61		-0.03	-0.50
S91	8.28	116.55	4.47	3.93	3.74	58.65	63.57	172.98	6.21		2.25	4.50	1.30	2.40	2.46	-0.09	0.09
V92	8.17	123.21	4.11		1.66	62.93	32.24	174.18		6.54		2.18	1.37	2.59	2.62	-0.11	-0.73
N93	8.42	122.59	4.71	2.60	2.68	53.44	38.48	173.23	7.21		2.72	3.26	1.40	2.66	2.73	-0.19	-1.98
A94	8.20	125.74	4.28	0.97		52.74	18.74	175.6	8.51		6.94		1.37	2.53	2.63	-0.03	0.05
A95	8.14	123.6	4.29	0.96		52.6	18.7	175.81	8.4		6.97		1.35	2.55	2.57	-0.10	0.98
K96	8.10	121.23	4.29	1.31		56.25	32.73	174.38	7.77				1.35	2.71	2.72	-0.17	-2.10
K97	8.21	123.85	4.34	1.30		56.18	32.69	174.16	7.33		3.48		1.42	2.64	2.76	-0.06	-5.62
I98	8.22	124.42	4.20	1.39		60.94	38.3	174.17	7.86		2.11		1.42	2.68	2.78	-0.03	
V99	8.32	126.35	4.20			62.09	32.53	173.9					1.42	2.62	2.83	-0.16	
S100	8.45	120.92	4.53	3.70	3.39	58.03	63.75	172.19	6.07			4.40	1.45		2.50	-0.16	
D101	8.58	122.59	4.83	2.77		52.68	37.7	173.47				4.49	1.34	2.17	2.24	-0.04	-2.47
G102	8.41	110.46	4.01			45.44		171.91					1.32	1.96	2.03	-0.18	0.83
N103	8.35	119.71	4.76		2.38	53.19	38.62	173.86		5.56			1.37	2.07	2.14	-0.16	-2.49
G104	8.47	110.41	3.99			45.36		172.46					1.37	2.25	2.10	-0.10	-1.47
M105	8.21	120.99	4.46	1.56		55.73	32.19	174.23	7.41		3.70		1.49	2.52	2.55	0.04	0.78
N106	8.40	120.43	4.64	2.27		53.73	38.29	173.35	7.67		3.33		1.54		2.73	-0.10	0.39
A107	8.21	124.94	4.24	0.88		53.46	18.51	175.81	8.92		6.87		1.57	2.85	2.93	0.09	2.59
W108	8.04	120.93	4.58	2.86	2.82	58.35	28.9	174.63	6.46	6.37	3.97	3.08	1.65	3.38	3.53	0.05	9.08
V109	7.76	122.35	3.78		1.44	63.42	32.34	174.21		7.61		1.95	1.75	3.78	3.84	0.05	8.31

A110	8.03	125.82	4.13	0.91		53.44	18.33	176.46	8.54		6.92		1.78	3.96	4.12	0.07	5.86
W111	7.90	120.32	4.48	2.85		58.37	28.75	175.03	7.79				1.79	3.62	3.81	0.18	-0.36
R112	7.97	122.18	3.95	1.09	1.21	57.05	29.92	174.34	7.10	5.52	3.46	3.02	1.82	3.68	3.69	0.11	
N113	8.09	119.17	4.62	2.23	2.39	53.47	38.36	173.26	8.05	5.52			1.74	3.44	3.54	0.04	-2.31
R114	7.90	122.14	4.23	1.44		56.44	30.17	173.99	6.91				1.66	3.06	3.16	0.11	
A115	8.18	125.59	4.27	0.88		52.56	18.72	175.67	8.38		6.81		1.58	2.47	2.55	0.02	-0.28
K116	8.22	121.58	4.28	1.31	1.39	56.56	32.65	175.07	7.26	4.69	3.50		1.50	2.75	2.60	0.05	-3.03
G117	8.37	110.91	4.04				45.15	172.13					1.43	2.24	2.17	-0.12	-5.07
T118	8.06	114.32	4.38	3.77		61.82	69.75	172.23	4.04			4.92	1.44	2.25	2.22	-0.10	-2.26
D119	8.57	122.63	4.79	2.38	2.50	52.68	37.71	172.94	4.29	4.13	3.35	3.81	1.48	2.21	2.20	-0.04	-2.64
V120	8.10	122.04	4.01		1.58	62.77	32.19	174.12		6.64		2.07	1.52	2.24	2.30	-0.11	-4.03
Q121	8.36	124.49				55.97	28.67	173.69					1.55		2.54	-0.14	-4.01
A122			0.89						8.65		6.76						
W123	7.99	120.96	4.67	3.08	2.78	57.2	29.08	174.09		6.44	4.74		1.56	2.38	2.53	-0.03	1.90
I124	7.87	124.16	4.04	1.28		61.11	38.43	173.73	8.21				1.58	2.49	2.54	-0.05	4.25
R125	8.19	125.59	4.14	1.27	1.35	56.5	30.07	174.63	6.85	5.31			1.58	2.57	2.66	0.02	
G126	8.32	111.2	3.96				45	171.53					1.37	2.04	1.99	-0.20	-5.96
A127	8.06	124.83	4.32	0.89		52.26	19.13	175.24	8.83		6.90		1.31	1.66	1.68	-0.30	0.18
R128	8.29	121.51	4.32			55.72	30.36	173.98					1.14	1.54	1.45	-0.37	-4.19
L129	8.38	125.83				53.98	41.69	176.97					0.99	1.15	1.26	-0.67	-1.44
													0.85	0.94	1.05	-0.75	-5.41

Tab.S11: table of chemical shifts ^1H_N, ^{15}N, $^1H^\alpha$, $^1H^{\beta2\beta3}$, $^{13}C^\alpha$, $^{13}C^\beta$, $^{13}C'$ [ppm], $^3J(H^\alpha,H^\beta)$, $^3J(CO,H^\beta)$ [Hz], R_1, $R_{1\rho}$, R_2 [s^{-1}], hetNOE [rad*s^{-1}] and RDCs [Hz] of WT^{Ala}.

4.4 ^{15}N-^1H RDCs and R_2 rates of all-ala mutants

	^{15}N-^1H RDC [Hz]					R_2 [s^{-1}]				
	W28G^{Ala}	W62G^{Ala}	W108G^{Ala}	W111G^{Ala}	W123G^{Ala}	W28G^{Ala}	W62G^{Ala}	W108G^{Ala}	W111G^{Ala}	W123G^{Ala}
K1	-4.92	-5.54	-5.10	-4.38	-5.88	1.07	1.23	1.04	1.13	0.97
V2	-6.18	-6.24	-6.24	-6.00	-8.40	1.52	1.52	1.30	1.61	1.26
F3	-2.70	-3.47	-3.36	-2.82	-4.14	1.47	1.67	1.54	1.72	1.45
G4	-0.96	0.00	-0.18	-0.66	-0.84	1.54	2.45	1.67	1.73	1.68
R5	-0.54	0.69	-0.60	-0.48	-1.20	2.14	2.44	2.52	2.33	2.27
A6	-2.70	4.13	-1.20	3.00	-2.76	2.58	2.37	2.59	2.76	2.58
E7	0.06	0.70	1.68	2.70	-0.54	2.33	2.59	2.40	2.64	2.47
L8	-0.12	0.70	-0.18	0.12	-1.02	2.41			2.57	2.75	2.58
A9										
A10	5.46		-0.36		3.72	2.75	3.10	2.66		3.10

A11	2.40	1.39	2.10	-0.90	1.86	2.72	2.91	2.93	4.01	3.04
M12	0.30	-2.08	0.12	0.48	-0.96	2.60	2.83	2.88	3.10	2.86
K13	3.24		1.26	1.14	3.00	2.68	2.88	3.41	3.51	3.33
R14	-2.94	-5.15	-3.30	-17.22	-4.32	2.88	2.60	3.06	2.99	2.41
H15	-4.08	-5.53	-2.52	-3.72	-5.64	2.40	2.28	2.81	3.01	2.71
G16	-3.12	-2.08	-2.64	-3.06	-3.48	2.11	2.61	2.21	2.07	2.47
L17	-2.88	-2.78	-2.70	-3.18	-2.46	2.51	2.49	2.66	2.61	2.70
D18	-2.58		-3.36	-1.02	-4.74	2.27	2.44	2.51	2.68	2.53
N19	-3.06	-2.77	-2.88	-1.26	36.48	2.31	2.81	2.56	2.78	2.62
Y20	-1.38	0.00	0.06	0.24	-0.18	2.63	2.73	2.80	2.65	2.81
R21	-4.86	-2.09	-1.68	-3.84	-3.36	2.57	2.51	2.65	2.78	2.72
G22	-1.80	-3.47	-3.48	-3.00	-4.08	2.43	2.79	2.77	3.09	2.91
Y23	0.36	-2.77	-1.26	-0.24	-2.16	2.42	2.76	2.87	3.12	2.92
S24	-2.46	-2.08	-2.40	-0.42	6.00	2.35	2.52	2.78	3.13	2.89
L25	-3.36	-4.85	-5.16	-3.78	-2.82	2.03		2.67	2.72	2.68
G26	-1.98	-1.38	-1.38	-0.18	-2.16	2.05	2.66	2.85	3.30	2.86
N27	-1.98	-2.77	-3.00	-1.50	-3.78	2.12	2.84	2.85	3.16	2.90
G28	-7.14		-0.60	1.44	2.10	2.12	3.05	3.11	3.56	3.10
V29	-0.42	0.00	2.16	-0.84	2.16	2.13		3.47	3.72	3.53
A30	-4.38	-0.69	-0.90	0.54	-0.60		2.77			
A31	3.60		-0.96	-2.52	-2.46	2.40	2.87	2.91	3.50	2.99
A32	-2.16	-1.38	-2.22	-0.96	-7.26	2.32		2.91	3.44	2.98
K33	-3.42		-2.28	0.24	-4.86	2.59	3.14	3.04	3.42	3.11
F34	-1.86	-1.39	-1.14	1.50	-3.54	2.67	3.04	3.02	2.57	3.30
E35	-1.14	-2.77	-2.04	-1.92	-2.70	2.46	2.70	3.35	2.76	3.36
S36	-1.86	-2.08	-1.86	0.00	-3.00	2.45	2.40	2.83	3.24	3.01
N37	-4.20		-2.34	-3.30	-1.92	2.49	2.48	2.63	2.82	2.66
F38	-0.66	-0.69	-0.48	0.54	-1.14	2.55		2.60	2.54	2.65
N39	-0.66		-1.32	-2.10	-2.94	2.42	2.29	2.63	2.87	2.70
T40	-1.92	-2.08	-2.28	-1.68	-3.24	2.31	2.67	2.46	2.54	2.43
Q41	-3.06	-2.09	-2.82	-1.86	-3.30	2.50	2.33	2.68	3.32	2.44
A42	0.60	-0.01	-5.94	-2.22	0.96	2.33	2.17	2.39	2.73	2.41
T43	-1.68	-2.77	-2.52	-6.72	-3.36	2.19	2.90	2.50	2.65	2.80
N44	-3.72		-2.46	-1.20	-5.16	2.54		2.60	2.72	2.62
R45	-5.64	-2.77	-2.40	-4.74	-3.60	2.38	2.41	2.63	3.11	2.76
N46	-1.20	-2.08	-1.86	0.66	-2.58	2.44	2.16	2.47	2.63	2.48
T47	-3.06	-3.47	-3.24	-3.00	-3.96	2.26	2.44	2.38	2.52	2.35
D48	-1.56	-2.08	-2.22	-1.20	-2.22	2.44	2.10	2.69	2.56	2.39
G49	-0.36	-0.01	-0.78	0.78	-0.90	2.08	2.26	2.14	2.41	2.21
S50	-1.20	-0.69	-1.44	-1.80	-1.50	2.26	2.42	2.34	2.55	2.39
T51	-2.10	-3.48	-3.42	-1.26	-4.50	2.42	3.02	2.57	2.81	2.62
D52	-1.44	-4.17	-4.80	-0.60	-2.88	2.61	3.14	2.90	2.66	2.90
Y53	1.74	1.39	2.94	-7.32	-1.98	3.00	2.83	3.02	2.82	3.03
G54	-3.24	0.00	0.66	-1.08	0.66	3.02	2.97	3.16	3.54	3.17
I55	-0.30	2.08	1.44	0.72	-0.72	3.32	3.06	3.33	3.71	3.39
L56	3.24	-3.46	-2.40	-1.14	-4.62	3.39	3.12	2.33	3.64	3.20

Q57	-3.54	-2.77	-1.62	-2.22	-2.82	3.52	3.06	3.94	3.97	3.91
I58	-1.44	-2.09	-1.08	-5.70	-3.06	3.35	3.18	4.00	3.02	3.96
N59	-3.36	-2.77	-2.52	-1.38	-3.12	3.64	2.74	4.03	4.02	4.15
S60	-0.78	-2.08	-1.32	-0.84	-0.54	3.57	3.30	3.80	4.27	4.11
R61	-2.22	-0.69	-2.64	-2.82	-2.46	3.91	2.59	3.63	4.56	4.15
W62	5.64	0.00	6.30	6.84	5.46	3.98		4.19	4.41	4.39
W63	14.52	-4.17	5.94	8.04	8.40	4.18	2.58	4.52	5.01	4.73
A64	5.04	-0.70	6.36	-7.32	4.68		2.53	3.85	4.23	3.88
N65	-3.42	-2.08	-2.58	-2.52	-3.00	3.19	2.26	3.34	3.74	3.43
D66	1.20	-3.32	-1.98	-1.14	-1.38	3.08	2.90	3.18	3.46	3.60
G67	1.08	-0.69	0.36	0.84	0.36	3.26	2.41	3.29	4.00	3.25
R68	-0.78	2.08	-2.10	-0.18	-2.28	2.97	2.48	2.87	2.99	2.92
T69	-3.96	0.69	-2.70	-1.26	0.96	2.58		2.62	2.85	2.65
P70										
G71	-9.60	0.68	-1.80	-0.78	-0.24	2.50	1.94	2.00	2.34	2.34
S72	-1.44	-2.77	-1.86	-2.46	-2.52	2.54	2.27	2.55	2.82	2.56
R73	-2.16	-4.16	-4.14	-2.94	-3.06	2.39	2.17	2.54	2.78	2.59
N74	-4.86		-4.02	-2.52	-5.82	2.31	2.27	2.32	2.53	2.47
L75	-5.34	-4.16	-3.18	-4.92	-3.30	2.52	2.38	2.34	2.53	2.53
A76	5.94	-1.38	-0.90	-0.78	-0.06	2.61	2.30	2.55	2.78	2.73
N77	-2.70	-3.47	-1.92	-1.20	-5.46	2.54	2.66	2.50	2.75	2.65
I78	-8.34	-7.62	-6.30	-6.66	-8.52	2.67		2.56	2.91	2.76
P79										
A80	-1.26	-5.83	-2.82	-1.32	-0.54	2.58	2.73	2.59	2.79	2.55
S81	-0.90	-3.47	-1.98	-2.58	-3.42	2.76	2.83	2.88	2.89	2.93
A82	1.32	-0.01	-0.30	-0.06	-0.72	2.91	2.43		3.65	3.27
L83	-2.28	-2.08	-1.98	-2.22	-2.70	2.71	2.63	2.62	2.79	1.95
L84	-0.48	-2.77	-10.14	-1.86	-1.86	2.67	2.44	2.66	3.33	2.69
S85	-0.42	-1.38	-1.74	-1.62	-3.18	2.56	2.52	2.52	2.85	2.62
S86	-2.94	-2.77	-0.48	-4.26	-5.16	2.56	2.89	2.52	2.82	2.57
D87	-1.26		-0.48	-0.84	-1.56	2.80	2.72	2.42	2.40	2.56
I88	8.10	-0.69	-0.36	0.06	-2.46	2.91	2.83		2.60	2.88
T89	-3.00	-7.63	-2.16	-2.64	-3.54	2.89		2.74	2.97	2.97
A90	-0.66	-1.38	-2.10	-1.38	0.72	3.20	2.40			
S91	-0.30	-0.70	-0.72	-0.54	-0.12	2.41	2.85	2.33	2.45	2.37
V92	-4.44	-1.39	-2.28	-1.80	-1.74	2.74	3.03	2.42	2.87	2.48
N93	-1.98	-2.78	-1.62	-0.96	-1.50	2.82	2.40		2.68	2.93
A94	-1.26	-0.48	-0.48	-0.84	-0.72	2.57	2.52	2.27	2.49	2.43
A95	2.22	-1.39	1.14	1.44	2.88	2.64	2.55	2.40	2.73	2.69
K96	-1.02	-6.23	-2.76	-2.40	-4.62	2.70	2.60	2.32	2.60	2.57
K97	-5.28	-4.84	-6.72	-5.64	-7.62	2.67		2.29	2.63	2.60
I98	-1.14		-7.14	-2.64	-2.22	2.64	2.57		2.77	2.70
V99	-1.38		-6.66	-5.10	-2.28	2.67	2.32	2.29	2.36	1.80
S100	-3.18		-3.12	-2.46	-4.92	2.47	2.56	2.27	2.51	2.50
D101	-4.74	-4.16	-4.68	-6.24	-5.76	2.55	2.11	2.09	1.83	2.17
G102	0.48	-4.16	-0.48	2.04	0.36	2.07	2.63	1.75	1.98	2.03

N103	-3.54	-2.20	-1.08	-0.66	-1.74	2.02	1.99	1.84	2.05	2.20
G104	2.22	0.00	-3.18	-1.44	-6.00	1.83	2.44	2.00	2.15	2.22
M105	-0.30	0.00	-0.66	1.80	1.80	2.56		2.25	2.21	2.48
N106	-0.30	0.69	-2.22	-0.66	-0.48	2.77	2.66	2.20	2.36	2.54
A107	0.54	1.39	0.60	0.54	-0.96	2.98	2.77	2.80	2.84	2.95
W108	3.78	-3.47	0.06	-0.48	5.46	3.55	3.53		2.54	3.22
V109	7.14	6.24	4.14	1.92	5.88	3.86	3.83	2.88	3.61	3.71
A110	3.48	2.77	-0.48	3.42	-1.44	3.98	3.70			3.50
W111	2.16	-0.68	0.36	-4.14	0.42	3.69	3.01	2.36		3.46
R112	-1.38	0.69	-2.82		-8.94	3.74	3.51	2.67		3.28
N113	6.06	8.31	0.48	-2.52	3.48	3.56	2.74	2.68	2.37	3.35
R114	-3.60	0.70	-4.26		3.60	3.15	2.55	2.77		3.23
A115	-2.40	-2.08	-3.12	-3.00	-0.06	2.58	2.57	2.33	2.54	2.43
K116	-2.16	-1.38	-4.20	-4.08	-3.90	2.76	2.09	2.20	1.81	2.33
G117	-3.84	-4.17	-4.32	-2.04	-6.42	2.13	2.01	2.00	2.21	2.05
T118	-2.40	-2.09	-1.98	-2.46	-3.30	2.29	2.38	1.91	1.81	2.21
D119	-5.16	-4.84	-3.30	-2.16	-4.26	2.29	2.33	2.09	2.34	1.91
V120	1.02	-2.78	-1.92	-4.68	-1.92	2.45	2.40	2.38	2.77	1.53
Q121	-4.98	-3.47		-3.84		2.58		2.29	2.49	2.32
A122			-2.82		-5.46		3.35			
W123	0.12	6.93	1.80	3.66	-3.12	2.52	2.36	2.63	2.36	2.06
I124	6.96	3.47	1.08	-0.06	-6.48	2.48		2.18	2.26	2.42
R125	0.60		-3.24	-6.78	-3.00	2.73	1.88	2.33	2.51	1.78
G126	-4.56	-5.53	-4.56	-4.80	-4.50	1.96	1.68	1.77	1.84	1.35
A127	0.12	0.69	-0.48	0.84	-1.14	1.70	1.48	1.53	1.69	1.47
R128	-1.92	-4.17	-2.58	-5.22	0.00	1.55	1.17	1.30	1.23	1.34
L129	-2.16	-6.24	-0.48	-0.42		1.22	2.65	1.12	1.24	1.04

Tab.S12: table of ^{15}N-^{1}H RDCs [Hz] and R_2 [s^{-1}] relaxation rates of W28G^{Ala}, W62G^{Ala}, W108G^{Ala}, W111G^{Ala}, W123G^{Ala}.

4.5   Chemical shifts and relaxation data of unfolded hLys

	^1H_N	^{15}N	$^{13}C^\alpha$	$^{13}C^\beta$	$^1H^\alpha$	R_1	R_2	hetN.	$R_{1\rho}$
K1	8.55	124.80	56.36	32.99		0.98	1.22	-0.49	1.14
V2	8.17	122.67	62.23	32.95		1.21	1.95	-0.58	2.02
F3	8.35	124.83	57.73	39.65	4.25	1.36	2.04	-0.45	1.83
E4					4.49				
R5	8.25	122.87	56.84	30.47		1.50	2.87	-0.14	2.91
A6	8.28	124.67	53.04		4.19	1.38	3.94	-0.31	3.69
E7	8.14	119.29	56.19	28.63			3.71		1.85
L8	8.06	123.27	55.65	42.14	4.15				2.52
A9	8.13	124.01	53.22	18.73			3.82		2.36
R10	8.09	119.59	55.77			1.34	3.52	-0.24	2.57
T11	7.98	115.80	62.64	69.64		1.56	4.19	0.12	4.30
L12	8.15		56.19		4.14				
K13	8.20		56.60						
R14	8.08	121.95	56.07	30.47	4.14		4.29	0.38	4.32
L15			55.47	42.32					
G16	8.17	108.90	45.34			1.51	3.79	0.14	4.04
M17	8.10	119.64	56.72	38.17			3.52	0.43	
D18	8.41	119.47	53.52			1.56	3.84	0.17	4.11
G19	8.21	109.03	45.34	38.58		1.54	3.79	0.17	4.04
Y20	7.93	120.31	58.20	30.53		1.55	2.73	0.08	2.92
R21	8.12	123.17	56.13			1.56	2.29	0.29	
G22	7.70	108.79	45.28	38.94		1.54		0.21	4.39
I23	7.82	119.76	61.16	63.98		1.64		0.39	4.30
S24	8.37	120.07	57.95	41.90		1.59	3.26	0.32	4.96
L25	8.38		56.24	18.44		1.48	3.94	-0.35	3.69
A26	7.98		53.46						
N27									
W28									
M29									
A30	8.24		55.53	18.67					
L31	7.84		56.07	42.14					
A32	7.96		53.64	18.26					
K33	7.78	118.54	57.49	32.24		1.70	5.89	0.30	5.46
W34	7.80	120.63	58.26	29.10		1.66	5.94	0.47	5.96
E35			56.30	28.63					
S36	7.95	115.40	59.16	63.67		1.60	5.20	0.30	5.19
G37	8.05	109.97	45.31			1.60	4.56	0.30	4.33
Y38	7.83	120.26	58.62	38.64		1.61	4.43	0.42	4.22
N39	8.17	119.94	53.25	38.64	4.29	1.56	3.96	0.28	3.41
T40	7.88	114.04	62.29	69.58	4.55	1.50	3.59	0.15	3.68
R41	8.06	122.56	56.24	30.47	4.12	1.57	3.52	0.17	2.72
A42	8.10	124.77	52.63	19.15	4.18	1.58	3.17	0.21	3.13
T43	8.00	113.16	62.05	69.82	4.18	1.46	3.20	0.05	3.18
N44	8.22	120.43	52.93	38.58	4.18	1.54	3.34	0.10	3.10

Y45	8.03	121.20	58.50	38.64	4.58	1.56	3.28	0.10	2.90
N46	8.12	121.13	52.81	38.88	4.36	1.59	3.52	0.19	3.40
A47	8.05	124.56	52.99	18.97	4.52	1.53	3.00	0.14	3.08
G48	8.21	107.14	45.34	38.05	4.07	1.52	2.93	0.05	2.38
D49	8.06	118.82	52.99	30.53	3.78	1.54	3.18	0.13	2.53
R50	8.18	121.25	56.24	63.71	4.58	1.50	2.78	0.04	2.61
S51	8.19	116.39	58.73	69.64	4.19	1.52	3.24	0.12	3.27
T52	8.00	115.15	62.05	37.87	4.33	1.54	3.61	0.12	2.90
D53	8.21	120.72	52.87	38.41	4.15	1.61	3.65	0.17	2.78
Y54	7.96	120.56	58.50	38.41	4.58		4.35	0.09	
G55	8.11	109.78	45.42	39.30		1.64	4.19	0.28	4.30
I56	7.63	119.34	61.40	28.87	3.73	1.65	4.48	0.23	
F57	8.05	122.93	57.73		3.97	1.62		0.22	
Q58	8.06	123.06	55.47	38.64	4.50	1.51		-0.14	
I59				63.59	4.14				
N60	8.29	121.61	53.50	30.11		1.63	4.90	0.22	4.46
S61	8.07	116.33	58.97	38.53	4.58	1.57	4.56	0.33	4.16
R62	8.10	122.17	56.60	29.50			4.60	0.55	4.19
Y63	7.80	119.79	58.38	18.95		1.64	4.62	0.32	4.30
W64	7.60	121.67	57.13	38.53		1.69	4.81	0.26	4.98
A65	7.78	124.60	52.62	37.99		1.63	3.92	0.37	4.09
N66	8.05	117.17	53.22			1.53	4.26	0.18	3.19
D67	8.22	118.79	53.16	33.01		1.55	3.24		3.08
G68	8.22	108.48	45.40	69.64		1.52	2.96		2.60
K69	7.88	120.22	56.07		3.78	1.55	3.53	0.14	2.92
T70	8.12	117.53	59.86		4.27	1.38	2.78		2.61
P71				19.27					
G72	8.31	109.17	45.04	32.60		1.37	2.57	-0.20	
A73	7.99	123.84	52.51	38.76	3.80	1.38	2.60	-0.12	
V74	8.06	119.50	62.59	19.09	4.22	1.46		-0.31	2.57
N75	8.36	122.20	53.12	18.89	3.92	1.38	2.49	-0.22	
A76	8.15	124.79	52.57	28.51	4.59	1.35	2.55	-0.04	2.18
A77	8.06	122.18	52.74	42.61		1.37	3.81	-0.04	2.51
H78	8.27	117.18	55.06	64.01		1.37	2.88	-0.23	2.56
L79	8.15	123.46	55.12	18.70	4.57	1.50		0.11	2.52
S80	8.27	117.18	58.14	63.42	4.29	1.37	2.88	-0.23	2.56
A81	8.40	125.67	53.63	18.89	4.28	1.41	3.48	0.07	2.75
S82	8.16	114.02	59.33	42.01	4.17	1.39	3.71	0.05	3.13
A83	8.03	126.16	53.50	42.02	4.17	1.49	3.69	0.05	3.16
L84	7.90	119.72	56.05	28.93	4.17	1.48	4.01	-0.07	3.03
L85	7.82	121.38	56.07	37.76	4.14	1.53	4.09	0.09	3.39
Q86	8.05	119.47	56.42	38.53	4.14	1.55	3.30	0.18	
D87	8.28	119.46	53.70	38.41	4.14	1.52	3.95	-0.04	3.09
N88	8.22	119.51	53.46	18.55	4.54	1.49	3.55	-0.05	
I89	7.96	121.59	62.29	37.81	3.94	1.48	4.42	0.15	3.09
A90	8.17	125.75	53.52	18.83	4.14	1.51		0.10	
D91	8.19	117.60	53.58	32.36	4.54	1.44	4.06	0.03	
A92	7.96	124.55	53.63	18.67	4.18	1.45	3.74	-0.06	
V93	7.97	119.56	63.53		3.86	1.38	4.24	-0.05	

A94	8.05	125.86	53.16	18.73	4.11	1.39		-0.03	3.78
A95				32.78					
A96	7.92	122.09	52.87	30.64	4.11	1.40	4.07		3.86
K97	7.92	119.42	56.42	32.66	4.11	1.36	4.25	-0.26	
R98	7.98	121.78	56.24	32.66	4.15	1.48	4.42	0.12	4.68
V99	8.03	122.39	62.35	30.82		1.38	3.81	-0.03	
V100	8.21	125.60	62.23	38.07		1.51	4.43	0.10	
R101	8.36	125.54	55.83		4.20	1.44	3.98	-0.16	
D102	8.47	121.06	51.28	29.18		1.46	3.37		
P103									
Q104	8.34	119.32	56.24		4.22	1.49	3.23	-0.06	3.06
G105	8.16	109.54	45.40	30.41		1.47	2.99		2.98
I106			62.35	18.97	3.98				
R107	8.24	124.39	56.30	29.34			3.60	-0.42	
A108	8.10	124.57	52.81	32.67	4.14	1.59	3.50	0.21	2.89
W109	7.98	120.71	58.20	18.67	4.44	1.61	3.72	0.11	2.61
V110	7.72	121.87	62.91	29.16		1.69	3.93	0.24	4.73
A111	7.91	125.49	53.16	30.29		1.70		0.14	4.81
W112	7.83	119.76	58.20	38.64		1.67	4.62	0.30	2.99
R113	7.88	121.45	56.72	30.41		1.59	3.28	0.13	3.13
N114	8.00	118.36	53.40	18.85		1.67	3.74	0.08	3.66
R115	7.89	121.19	56.48	29.22	4.44	1.74	3.28	0.17	2.49
A116	8.14	124.38	52.75	38.58		1.66	2.84	0.20	2.89
Q117	8.15	118.84	56.07	30.47		1.54	2.93	0.11	2.51
N118	8.22	119.59	53.58	38.01		1.49	3.55	-0.05	2.92
R119	8.19	121.22	56.30			1.50	2.78	0.04	2.61
D120	8.38	119.91	53.00	30.41		1.54	3.26	-0.07	
V121	7.91	120.50	61.52	29.46		1.48	2.73	-0.14	2.17
R122	8.24	124.39	56.30	38.77			3.60	-0.23	
Q123	8.19	121.35	55.89	32.95		1.50	2.78	-0.14	1.97
Y124	8.18	122.01	57.82	29.22			2.94	0.62	2.01
V125	7.91	123.06	62.05				2.03	-0.49	1.80
Q126	8.35	124.83	56.24			1.36	2.04	-0.45	1.83
G127	8.36	111.21	45.16			1.17	1.46	-0.67	1.41
A128	8.11	123.66	52.51			1.18	2.29	-0.56	1.27
G129	8.30	108.29	45.16				0.92	-0.87	0.88
V130	7.85	118.79	61.28			0.72	0.88	-1.49	0.64

Tab.S13: table of chemical shifts ^1H_N, ^{15}N, $^1H^\alpha$ [ppm], R_1, $R_{1\rho}$, R_2 [s^{-1}], hetN. (hetNOE) [rad*s^{-1}] of hLys.

4.6   Chemical shifts and relaxation data of all-ala mutants I56T and D67H

	I56T						D67H					
	^1H_N	^{15}N	R_1	R_2	$R_{1\rho}$	hetN.	^1H_N	^{15}N	R_1	R_2	$R_{1\rho}$	hetN.
K1	8.59	124.79	0.94		1.28	-1.27	8.59	124.79	0.95	2.00	1.44	-1.22
V2	8.17	122.49	1.23	2.88	1.73	-0.49	8.17	122.49	1.23	1.83	2.17	-0.55
F3	8.35	124.72	1.33	4.07	1.84	-0.44	8.35	124.70	1.34		1.78	-0.43
E4												
R5	8.24	122.68	1.47	4.76	3.17	-0.18	8.24	122.63	1.50	4.83	3.69	-0.12
A6	8.33	124.56	1.32	4.85	2.55	-0.50	8.33	124.43	1.40	4.71	3.79	-0.08
E7												
L8												
A9	8.11	123.61	1.03		1.35	-0.57	8.11	123.66	1.08		1.28	-0.62
R10	8.04	119.32	1.43	3.81	2.98	-0.09	8.06	119.45	1.29	2.67	2.51	-0.10
T11	7.95	115.87	1.54	5.26	4.52	0.04	7.94	115.98	1.55	6.00	5.68	0.09
L12												
K13												
R14	8.08	121.86	1.59	4.55	3.83	0.21	8.08	121.75	1.61		7.27	0.25
L15												
G16	8.15	108.75	1.54	5.68	3.87	0.15	8.14	108.72	1.51	5.15	4.68	0.19
M17	8.06	119.49	1.52		4.06	-0.24	8.06	119.73	1.66		7.04	0.34
D18	8.39	119.22	1.56	4.42	4.24	0.13	8.38	119.10	1.53	4.97	4.84	0.13
G19	8.15	108.75	1.54	5.68	3.87	0.15	8.15	108.63	1.57	5.64	5.18	0.19
Y20							7.95	120.33	1.59	5.11	4.11	0.11
R21	8.11	122.94	1.61		4.70	0.26	8.11	122.81	1.62	6.02	5.97	0.29
G22	7.73	108.67	1.55	6.34	3.93	0.17	7.76	108.60	1.53	5.67	4.85	0.23
I23	7.79	119.68	1.67	5.11	4.69	0.33	7.78	119.72	1.60	7.18	6.39	0.39
S24	8.37	119.98	1.57	6.18	5.48	0.29	8.37	120.05	1.55		6.68	0.26
L25	8.33	124.56	1.32		2.55	-0.50	8.33	124.43	1.40	4.71	3.79	-0.08
A26												
N27												
W28												
M29												
A30												
L31												
A32												
K33	7.77	118.37	1.70	6.46	5.88	0.46	7.77	118.33	1.69	7.67	6.82	0.48
W34	7.80	120.57	1.68	7.62	6.27	0.41	7.81	120.61	1.69	7.87	6.71	0.43

E35											
S36	7.94	115.20	1.64	5.66	5.12	0.28	7.93	115.07	1.58	6.25	6.06	0.29
G37	8.02	109.82	1.60	5.48	4.47	0.29	8.00	109.74	1.64	5.81	5.18	0.30
Y38	7.81	120.19	1.61	5.15	4.63	0.33	7.80	120.24	1.66	5.81	5.20	0.37
N39	8.15	119.83	1.60	4.67	4.07	0.24	8.14	119.73	1.59	5.10	4.66	0.30
T40	7.88	113.92	1.53	3.98	3.64	0.09	7.88	113.88	1.51	4.06	4.01	0.14
R41	8.05	122.43	1.59	5.39	3.76	0.18	8.05	122.38	1.54	4.00	4.08	0.15
A42	8.10	124.65	1.55	5.80	3.27	0.14	8.09	124.59	1.53	3.99	3.54	0.17
T43	7.99	113.04	1.44	3.46	3.09	0.03	7.99	112.97	1.46	3.58	3.43	0.04
N44	8.21	120.34	1.52	3.64	3.23	0.06	8.21	120.30	1.53	3.79	3.41	0.09
Y45	8.02	121.11	1.54	3.77	3.13	0.07	8.02	121.07	1.52	3.91	3.43	0.11
N46	8.12	121.03	1.58	3.73	3.11	0.13	8.10	120.95	1.59	3.81	3.54	0.17
A47	8.04	124.47	1.50	5.33	2.92	0.08	8.05	124.44	1.52	3.66	3.24	0.12
G48	8.21	107.07	1.49		2.69	-0.02	8.21	107.07	1.51	3.03	2.94	
D49	8.07	118.72	1.54	3.15	2.98	0.06	8.06	118.75	1.52	3.36	3.36	0.11
R50	8.18	121.15	1.50	3.23	2.84	0.00	8.17	121.12	1.51	3.54	3.02	0.05
S51	8.19	116.37	1.49	2.97	2.92	0.00	8.18	116.22	1.50	3.44	3.43	0.08
T52	8.00	115.03	1.52	3.40	3.19	0.02	7.99	115.08	1.55	3.89	3.91	0.08
D53	8.21	120.68	1.61	3.93	3.24	0.09	8.21	120.61	1.62	4.13	3.93	0.15
Y54	7.97	120.56	1.59	4.03	3.77	0.11	7.96	120.49	1.61	4.35	4.11	0.11
G55	8.13	109.70	1.61	4.35	3.28	0.15	8.12	109.55	1.65	4.77	4.38	0.28
I/T56	7.81	113.47	1.53	3.92	3.57	0.19	7.64	119.31	1.63	4.97	4.89	0.23
F57							8.01	122.51	1.63	5.76	5.29	0.26
Q58	8.04	122.87	1.40	3.73	2.58	-0.18	8.03	122.80	1.48	4.48	4.19	-0.11
I59												
N60	8.28	121.64	1.62	5.34	4.25	0.10	8.24	121.34	1.66	6.03	5.34	0.22
S61	8.08	116.31	1.57	4.35	4.11	0.30	8.07	116.19	1.59	5.05	4.99	0.35
R62	8.10	122.11	1.64	4.33	4.63	0.14	8.10	121.97	1.72	5.41	4.93	0.35
Y63	7.79	119.68	1.67	5.11	4.69	0.33	7.76	119.83				0.37
W64	7.59	121.56	1.71	5.13	4.27	0.15	7.69	121.00	1.71	5.17	5.28	0.28
A65	7.77	124.49	1.63	5.78	3.70	0.26	7.83	123.89	1.67	4.95	4.54	0.31
N66	8.05	117.09	1.54	3.56	3.28	0.16	8.05	117.04	1.56	4.31	4.01	0.12
D/H67	8.22	118.70	1.54	3.39	3.19	-0.03	8.22	118.70	1.51	3.61	3.60	0.16
G68	8.22	108.40	1.51		2.68	-0.07	8.25	109.06	1.44	2.91	2.83	-0.13
K69	7.88	120.14	1.54	3.70	3.06	0.07	7.88	120.11	1.63		5.85	0.20
T70	8.10	117.36	1.38	2.65	2.59	-0.02	8.10	117.37	1.34	2.81	2.76	-0.12
P71												
G72	8.31	109.09	1.36		2.45	-0.25	8.30	109.01	1.32	2.52	2.45	-0.30
A73	7.98	123.78	1.39	3.90	2.45	-0.20	7.98	123.71	1.32		2.47	-0.19
V74	8.04	119.32	1.43	3.81	2.98	-0.09	8.03	119.28	1.45	4.75	4.47	-0.01

N75	8.35	122.06	1.32	3.11	2.47	-0.39	8.36	122.11	1.24	3.34	2.27	-0.66
A76	8.14	124.66	1.32		2.54	-0.07	8.15	124.72	1.27	3.07	2.66	-0.11
A77	8.06	122.06	1.45	4.33	2.89	0.14	8.05	122.06	1.30	3.11	2.71	-0.08
H78	8.26	117.08	1.35	3.07	2.93	-0.33	8.25	117.07	1.33	3.06	3.02	-0.32
L79	8.13	123.32	1.38	4.23	2.82	-0.06	8.11	123.23	1.42	4.99	5.21	-0.09
S80	8.26	117.08	1.35	3.07	2.93	-0.33	8.25	117.07	1.33	3.06	3.02	-0.32
A81	8.40	125.53	1.41		3.35	0.04	8.40	125.55	1.39		3.47	0.04
S82	8.15	113.84	1.37	4.13	3.80	0.03	8.15	113.84	1.35	4.06	3.95	0.02
A83	8.02	126.04	1.48		3.86	0.03	8.02	126.05	1.48	4.34	3.95	0.03
L84	7.90	119.52	1.50	4.21	3.94	-0.10	7.90	119.55	1.48	4.16	4.01	-0.09
L85	7.81	121.21	1.53	4.78	4.09	0.04	7.80	121.12	1.54	5.37	4.21	0.06
Q86	8.03	119.08	1.50	4.78	4.48	0.14	8.01	118.93	1.57	5.54	5.10	0.20
D87	8.27	119.31	1.48	4.21	3.96	-0.02	8.27	119.29	1.47	4.60	4.12	-0.03
N88	8.21	119.42	1.48	3.68	3.26	-0.07	8.21	119.41	1.48	3.40	3.53	-0.05
I89	7.96	121.50	1.45	5.42	4.51	-0.09	7.96	121.44	1.49	5.35	4.95	0.03
A90	8.16	125.54	1.52		5.09	0.09	8.16	125.50	1.50		5.73	0.09
D91	8.18	117.44	1.46	4.25	4.11	0.00	8.18	117.44	1.46	4.27	4.06	0.02
A92	7.96	124.47	1.42		3.84	-0.05	7.96	124.46	1.44	4.51	3.87	-0.06
V93	7.98	119.38	1.39	4.78	4.49	-0.06	7.98	119.39	1.39	5.41	4.67	-0.01
A94	8.03	125.62	1.42		5.38	-0.03	8.03	125.59	1.42		5.91	-0.03
A95												
A96	7.92	121.96	1.41	5.55	4.21	0.01	7.91	121.96	1.39	4.77	4.38	0.02
K97	7.92	119.23	1.40	4.80	4.67	-0.15	7.92	119.19	1.38	5.40	5.31	-0.16
R98	7.96	121.50	1.45	5.42	4.51	-0.09	7.96	121.46	1.53	5.36	4.95	0.03
V99	8.02	122.16	1.40	5.47	4.37	-0.02	8.01	122.12	1.40	4.84	4.49	-0.03
V100	8.16	125.54	1.52		5.09	0.09	8.15	125.44	1.50		5.73	0.09
R101	8.34	125.30	1.44		4.25	-0.12	8.34	125.24	1.49		5.19	-0.13
D102	8.45	120.78	1.41	4.18	3.48	-0.02	8.44	120.74	1.43	4.67	3.76	0.00
P103												
Q104	8.33	119.11	1.49	3.59	3.24	-0.07	8.33	119.09	1.51	3.83	3.56	-0.05
G105	8.15	109.42	1.45	4.02	3.07	0.01	8.15	109.39	1.47	3.47	3.12	0.04
I106												
R107	8.23	124.12	1.56	5.91	4.43	-0.06	8.22	124.17	1.58	3.37	2.37	-0.06
A108	8.08	124.32	1.57	5.74	3.96	0.21	8.08	124.26	1.60	4.81	4.27	0.21
W109	7.97	120.56	1.59	4.03	3.77	0.11	7.97	120.55	1.61	4.38	4.11	0.11
V110	7.74	121.63	1.66	4.95	4.36	0.24	7.75	121.56	1.66	4.77	4.62	0.25
A111	7.90	125.22	1.68		5.00	0.11	7.90	125.12	1.69		5.68	0.15
W112	7.83	119.64	1.68	5.11	3.86	0.27	7.83	119.63	1.67	5.14	4.03	0.31
R113	7.86	121.03	1.59	3.83	3.43	0.10	7.86	121.02	1.64	3.59	3.57	0.13
N114	7.99	118.21	1.67	4.12	3.83	0.10	7.99	118.16	1.70	4.23	4.08	0.09

R115	7.89	121.26	1.76	4.27	3.92	0.15	7.89	121.24	1.72	4.65	4.21	0.17
A116	8.12	124.23	1.61	4.88	3.06	0.21	8.12	124.21	1.65	3.52	3.32	0.20
Q117	8.14	118.70	1.52	3.41	3.07	0.13	8.13	118.63	1.51	3.24	3.09	0.10
N118	8.21	119.42	1.48	3.68	3.26	-0.07	8.21	119.41	1.48	3.40	3.53	-0.05
R119	8.18	121.15	1.50	3.23	2.84	0.00	8.17	121.12	1.51	3.54	3.02	0.05
D120	8.38	119.76	1.57	2.68	2.50	-0.05	8.38	119.74	1.55	2.56	2.51	-0.05
V121	7.90	120.38	1.45	2.92	2.76	-0.09	7.90	120.36	1.48	2.99	2.75	-0.13
R122	8.22	124.31	1.56	4.31	2.40	-0.23	8.22	124.17	1.60		2.37	-0.22
Q123	8.18	121.15	1.50	3.23	2.84	0.00	8.17	121.12	1.51	3.54	3.02	0.05
Y124	8.17	121.90	1.50	3.20	2.30	-0.17	8.17	121.82	1.48	4.31	2.33	-0.13
V125	7.91	122.93	1.43	3.38	2.07	-0.42	7.91	122.95	1.41	2.16	2.11	-0.48
Q126	8.35	124.72	1.33	4.07	1.84	-0.44	8.35	124.70	1.34		1.78	-0.43
G127	8.35	111.11	1.15	2.11	1.45	-0.69	8.35	111.10	1.16	1.37	1.54	-0.66
A128	8.11	123.61	1.03	2.30	1.35	-0.57	8.11	123.56	1.08		1.28	-0.62
G129	8.30	108.21	0.79	1.77	0.90	-1.59	8.30	108.19	0.79		0.88	-1.54
V130	7.86	118.63	0.67	0.81	0.77	-2.10	7.86	118.62	0.66	0.74	0.79	-2.12

Tab.S14: table of chemical shifts ^1H_N, ^{15}N [ppm] and R_2 [s^{-1}] of the all-ala mutants I56T and D67H (300µM protein concentration, pH 2, water 10% D_2O, 293K, 600MHz).

.

5   References

1.   Levinthal, C. (1959) Coding Aspects of Protein Synthesis, *Rev Mod Phys 31*, 249-255.
2.   Dobson, C. M. (2001) The structural basis of protein folding and its links with human disease, *Philosophical transactions of the Royal Society of London 356*, 133-145.
3.   Dobson, C. M. (2003) Protein folding and misfolding, *Nature 426*, 884-890.
4.   Pepys, M. B., Rademacher, T. W., Amatayakul-Chantler, S., Williams, P., Noble, G. E., Hutchinson, W. L., Hawkins, P. N., Nelson, S. R., Gallimore, J. R., Herbert, J., and et al. (1994) Human serum amyloid P component is an invariant constituent of amyloid deposits and has a uniquely homogeneous glycostructure, *Proceedings of the National Academy of Sciences of the United States of America 91*, 5602-5606.
5.   Prusiner, S. B. (1997) Prion diseases and the BSE crisis, *Science (New York, N.Y 278*, 245-251.
6.   Serpell, L. C., Sunde, M., and Blake, C. C. (1997) The molecular basis of amyloidosis, *Cell Mol Life Sci 53*, 871-887.
7.   Thomas, P. J., Qu, B. H., and Pedersen, P. L. (1995) Defective protein folding as a basis of human disease, *Trends in biochemical sciences 20*, 456-459.
8.   Dobson, C. M. (1999) Protein misfolding, evolution and disease, *Trends in biochemical sciences 24*, 329-332.
9.   Kelly, J. W. (1998) The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways, *Current opinion in structural biology 8*, 101-106.
10.   Lansbury, P. T., Jr. (1999) Evolution of amyloid: what normal protein folding may tell us about fibrillogenesis and disease, *Proceedings of the National Academy of Sciences of the United States of America 96*, 3342-3344.
11.   Sunde, M., and Blake, C. (1997) The structure of amyloid fibrils by electron microscopy and X-ray diffraction, *Advances in protein chemistry 50*, 123-159.
12.   Perutz, M. F. (1999) Glutamine repeats and neurodegenerative diseases: molecular aspects, *Trends in biochemical sciences 24*, 58-63.
13.   Uversky, V. N. (2002) What does it mean to be natively unfolded?, *European journal of biochemistry / FEBS 269*, 2-12.
14.   Turoverov, K. K., Kuznetsova, I. M., and Uversky, V. N. (2010) The protein kingdom extended: ordered and intrinsically disordered proteins, their folding, supramolecular complex formation, and aggregation, *Progress in biophysics and molecular biology 102*, 73-84.
15.   Sugase, K., Dyson, H. J., and Wright, P. E. (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein, *Nature 447*, 1021-1025.
16.   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., and et al. (1993) Human lysozyme gene mutations cause hereditary systemic amyloidosis, *Nature 362*, 553-557.

17.   Morozova-Roche, L. A., Zurdo, J., Spencer, A., Noppe, W., Receveur, V., Archer, D. B., Joniau, M., and Dobson, C. M. (2000) Amyloid fibril formation and seeding by wild-type human lysozyme and its disease-related mutational variants, *Journal of structural biology 130*, 339-351.

18.   Booth, D. R., Sunde, M., Bellotti, V., Robinson, C. V., Hutchinson, W. L., Fraser, P. E., Hawkins, P. N., Dobson, C. M., Radford, S. E., Blake, C. C., and Pepys, M. B. (1997) Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis, *Nature 385*, 787-793.

19.   Barghorn, S., Zheng-Fischhofer, Q., Ackmann, M., Biernat, J., von Bergen, M., Mandelkow, E. M., and Mandelkow, E. (2000) Structure, microtubule interactions, and paired helical filament aggregation by tau mutants of frontotemporal dementias, *Biochemistry 39*, 11714-11721.

20.   Jenkins, D. C., Sylvester, I. D., and Pinheiro, T. J. (2008) The elusive intermediate on the folding pathway of the prion protein, *The FEBS journal 275*, 1323-1335.

21.   Trivella, D. B., Bleicher, L., Palmieri Lde, C., Wiggers, H. J., Montanari, C. A., Kelly, J. W., Lima, L. M., Foguel, D., and Polikarpov, I. (2010) Conformational differences between the wild type and V30M mutant transthyretin modulate its binding to genistein: implications to tetramer stability and ligand-binding, *Journal of structural biology 170*, 522-531.

22.   Koo, E. H., Lansbury, P. T., Jr., and Kelly, J. W. (1999) Amyloid diseases: abnormal protein aggregation in neurodegeneration, *Proceedings of the National Academy of Sciences of the United States of America 96*, 9989-9990.

23.   Fleming, A. (1922) On a remarkable bacteriolytic element found in tissues and secretions., *Proceedings of the Royal Society of London Series B-Containing Papers of a Biological Character 93*, 306-317.

24.   Wirmer, J., Schlörb, C., and Schwalbe, H. (2005) *Handbook of Protein Folding (eds. J. buchner and T. Kiefhaber),* chap. "Conformation and Dynamics of Nonnative States of Proteins studied by NMR Spectroscopy" *Wiley-VCH, Weinheim, Germany*, 737-808.

25.   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.

26.   Matagne, A., Radford, S. E., and Dobson, C. M. (1997) Fast and slow tracks in lysozyme folding: insight into the role of domains in the folding process, *Journal of molecular biology 267*, 1068-1074.

27.   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.

28.   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 (New York, N.Y 295*, 1719-1722.

29.   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, *Angewandte Chemie (International ed 43*, 5780-5785.

30. Rocken, C., and Wilhelm, S. (2005) Influence of tissue fixation on the microextraction and identification of amyloid proteins, *The Journal of laboratory and clinical medicine 146*, 244-250.

31. Dhulesia, A., Cremades, N., Kumita, J. R., Hsu, S. T., Mossuto, M. F., Dumoulin, M., Nietlispach, D., Akke, M., Salvatella, X., and Dobson, C. M. (2010) Local cooperativity in an amyloidogenic state of human lysozyme observed at atomic resolution, *Journal of the American Chemical Society 132*, 15580-15588.

32. Dumoulin, M., Canet, D., Last, A. M., Pardon, E., Archer, D. B., Muyldermans, S., Wyns, L., Matagne, A., Robinson, C. V., Redfield, C., and Dobson, C. M. (2005) Reduced global cooperativity is a common feature underlying the amyloidogenicity of pathogenic lysozyme mutations, *Journal of molecular biology 346*, 773-788.

33. Canet, D., Last, A. M., Tito, P., Sunde, M., Spencer, A., Archer, D. B., Redfield, C., Robinson, C. V., and Dobson, C. M. (2002) Local cooperativity in the unfolding of an amyloidogenic variant of human lysozyme, *Nature structural biology 9*, 308-315.

34. Artymiuk, P. J., and Blake, C. C. (1981) Refinement of human lysozyme at 1.5 A resolution analysis of non-bonded and hydrogen-bond interactions, *Journal of molecular biology 152*, 737-762.

35. Shoyama, Y., Tamada, T., Nitta, K., Kumagai, I., Kuroki, R., and Koshiba, T. (2009) Preparation and characterization of methyonyl-lysine attached human lysozyme expressed in Escherichia coli and its effective conversion to the authentic-like protein, *to be published*.

36. Hooke, S. D., Radford, S. E., and Dobson, C. M. (1994) The refolding of human lysozyme: a comparison with the structurally homologous hen lysozyme, *Biochemistry 33*, 5867-5876.

37. Sziegat, F., Wirmer-Bartoschek, J., and Schwalbe, H. (2011) Characteristics of Human Lysozyme and Its Disease-Related Mutants in their Unfolded States, *Angewandte Chemie (International ed*, 5628-5632.

38. Lipari, G., and Szabo, A. (1982) Analysis of Nmr Relaxation Data on Macromolecules Using the Model-Free Approach, *Biophysical Journal 37*, A380-A380.

39. Peti, W., Meiler, J., Bruschweiler, R., and Griesinger, C. (2002) Model-free analysis of protein backbone motion from residual dipolar couplings, *Journal of the American Chemical Society 124*, 5822-5833.

40. 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, *Journal of the American Chemical Society 123*, 2970-2978.

41. Tamiola, K., Acar, B., and Mulder, F. A. (2010) Sequence-specific random coil chemical shifts of intrinsically disordered proteins, *Journal of the American Chemical Society 132*, 18000-18003.

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

43. Obolensky, O. I., Schlepckow, K., Schwalbe, H., and Solov'yov, A. V. (2007) Theoretical framework for NMR residual dipolar couplings in unfolded proteins, *Journal of biomolecular NMR 39*, 1-16.

44.   Louhivuori, M., Fredriksson, K., Paakkonen, K., Permi, P., and Annila, A. (2004) Alignment of chain-like molecules, *Journal of biomolecular NMR 29*, 517-524.

45.   Bernado, P., Bertoncini, C. W., Griesinger, C., Zweckstetter, M., and Blackledge, M. (2005) Defining long-range order and local disorder in native alpha-synuclein using residual dipolar couplings, *Journal of the American Chemical Society 127*, 17968-17969.

46.   Bouvignies, G., Markwick, P., Bruschweiler, R., and Blackledge, M. (2006) Simultaneous determination of protein backbone structure and dynamics from residual dipolar couplings, *Journal of the American Chemical Society 128*, 15100-15101.

47.   Bouvignies, G., Markwick, P. R., and Blackledge, M. (2007) Simultaneous definition of high resolution protein structure and backbone conformational dynamics using NMR residual dipolar couplings, *Chemphyschem 8*, 1901-1909.

48.   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, *Journal of molecular biology 284*, 1153-1164.

49.   Marsh, J. A., Neale, C., Jack, F. E., Choy, W. Y., Lee, A. Y., Crowhurst, K. A., and Forman-Kay, J. D. (2007) Improved structural characterizations of the drkN SH3 domain unfolded state suggest a compact ensemble with native-like and non-native structure, *Journal of molecular biology 367*, 1494-1510.

50.   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, *Journal of molecular biology 339*, 1191-1199.

51.   Wirmer, J., Berk, H., Ugolini, R., Redfield, C., and Schwalbe, H. (2006) Characterization of the unfolded state of bovine alpha-lactalbumin and comparison with unfolded states of homologous proteins, *Protein Sci 15*, 1397-1407.

52.   Bussell, R., Jr., and Eliezer, D. (2001) Residual structure and dynamics in Parkinson's disease-associated mutants of alpha-synuclein, *The Journal of biological chemistry 276*, 45996-46003.

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

54.   Mukrasch, M. D., Markwick, P., Biernat, J., Bergen, M., Bernado, P., Griesinger, C., Mandelkow, E., Zweckstetter, M., and Blackledge, M. (2007) Highly populated turn conformations in natively unfolded tau protein identified from residual dipolar couplings and molecular simulation, *Journal of the American Chemical Society 129*, 5235-5243.

55.   Jeganathan, S., von Bergen, M., Mandelkow, E. M., and Mandelkow, E. (2008) The natively unfolded character of tau and its aggregation to Alzheimer-like paired helical filaments, *Biochemistry 47*, 10526-10539.

56.   Modig, K., and Poulsen, F. M. (2008) Model-independent interpretation of NMR relaxation data for unfolded proteins: the acid-denatured state of ACBP, *Journal of biomolecular NMR 42*, 163-177.

57.   Alexandrescu, A. T., Jahnke, W., Wiltscheck, R., and Blommers, M. J. (1996) Accretion of structure in staphylococcal nuclease: an 15N NMR relaxation study, *Journal of molecular biology 260*, 570-587.

58.   Sallum, C. O., Martel, D. M., Fournier, R. S., Matousek, W. M., and Alexandrescu, A. T. (2005) Sensitivity of NMR residual dipolar couplings to perturbations in folded and denatured staphylococcal nuclease, *Biochemistry 44*, 6392-6403.

59.   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.

60.   Mohana-Borges, R., Goto, N. K., Kroon, G. J., Dyson, H. J., and Wright, P. E. (2004) Structural characterization of unfolded states of apomyoglobin using residual dipolar couplings, *Journal of molecular biology 340*, 1131-1142.

61.   Meier, S., Grzesiek, S., and Blackledge, M. (2007) Mapping the conformational landscape of urea-denatured ubiquitin using residual dipolar couplings, *Journal of the American Chemical Society 129*, 9799-9807.

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

63.   Grzesiek, S., and Bax, A. (1992) Improved 3d Triple-Resonance Nmr Techniques Applied to a 31-Kda Protein, *J Magn Reson 96*, 432-440.

64.   Clubb, R. T., Thanabal, V., and Wagner, G. (1992) A Constant-Time 3-Dimensional Triple-Resonance Pulse Scheme to Correlate Intraresidue H-1(N), N-15, and C-13(') Chemical-Shifts in N-15-C-13-Labeled Proteins, *J Magn Reson 97*, 213-217.

65.   Schmidt, J. M., Lohr, F., and Ruterjans, H. (1996) Heteronuclear relayed E.COSY applied to the determination of accurate (3)J(H-N,C') and (3)J(H-beta,C') coupling constants in Desulfovibrio vulgaris flavodoxin, *Journal of biomolecular NMR 7*, 142-152.

66.   Grzesiek, S., and Bax, A. (1992) Correlating Backbone Amide and Side-Chain Resonances in Larger Proteins by Multiple Relayed Triple Resonance Nmr, *Journal of the American Chemical Society 114*, 6291-6293.

67.   West, N. J., and Smith, L. J. (1998) Side-chains in native and random coil protein conformations. Analysis of NMR coupling constants and chi1 torsion angle preferences, *Journal of molecular biology 280*, 867-877.

68.   Wishart, D. S., Sykes, B. D., and Richards, F. M. (1991) Relationship between nuclear magnetic resonance chemical shift and protein secondary structure, *Journal of molecular biology 222*, 311-333.

69.   Spera, S., and Bax, A. (1991) Empirical Correlation between Protein Backbone Conformation and C-Alpha and C-Beta C-13 Nuclear-Magnetic-Resonance Chemical-Shifts, *Journal of the American Chemical Society 113*, 5490-5492.

70.   McDonald, C. C., and Phillips, W. D. (1969) Proton magnetic resonance spectra of proteins in random-coil configurations, *Journal of the American Chemical Society 91*, 1513-1521.

71.   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, *Journal of biomolecular NMR 4*, 171-180.

72.   Williamson, M. P. (1990) Secondary-structure dependent chemical shifts in proteins, *Biopolymers 29*, 1423-1431.

73. 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.

74. Szilagyi, L. (1995) Chemical-Shifts in Proteins Come of Age, *Prog Nuc Magn Reson Spect 27*, 325-443.

75. Bundi, A., and Wuthrich, K. (1979) ^{1}H 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.

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

77. Plaxco, K. W., Morton, C. J., Grimshaw, S. B., Jones, J. A., 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, *Journal of biomolecular NMR 10*, 221-230.

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

79. Wagner, G. (1993) NMR Relaxation and Protein Mobility, *Current opinion in structural biology 3*, 748-754.

80. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation, *Biochemistry 33*, 5984-6003.

81. Lipari, G., and Szabo, A. (1981) A Model-Free Approach to the Interpretation of NMR Relaxation in Macromolecules, *Biophysical Journal 33*, A307-A307.

82. Clore, G. M., Szabo, A., Bax, A., Kay, L. E., Driscoll, P. C., and Gronenborn, A. M. (1990) Deviations from the Simple 2-Parameter Model-Free Approach to the Interpretation of N-15 Nuclear Magnetic-Relaxation of Proteins, *Journal of the American Chemical Society 112*, 4989-4991.

83. Farrow, N. A., Zhang, O., Forman-Kay, J. D., and Kay, L. E. (1995) Comparison of the backbone dynamics of a folded and an unfolded SH3 domain existing in equilibrium in aqueous buffer, *Biochemistry 34*, 868-878.

84. Lefevre, J. F., Dayie, K. T., Peng, J. W., and Wagner, G. (1996) Internal mobility in the partially folded DNA binding and dimerization domains of GAL4: NMR analysis of the N-H spectral density functions, *Biochemistry 35*, 2674-2686.

85. Krizova, H., Zidek, L., Stone, M. J., Novotny, M. V., and Sklenar, V. (2004) Temperature-dependent spectral density analysis applied to monitoring backbone dynamics of major urinary protein-I complexed with the pheromone 2- sec-butyl-4,5-dihydrothiazole, *Journal of biomolecular NMR 28*, 369-384.

86. Ishima, R., and Nagayama, K. (1995) Quasi-Spectral-Density Function-Analysis for N-15 Nuclei in Proteins, *J Magn Reson B 108*, 73-76.

87. Mandel, A. M., Akke, M., and Palmer, A. G., 3rd. (1996) Dynamics of ribonuclease H: temperature dependence of motions on multiple time scales, *Biochemistry 35*, 16009-16023.

88. Carr, H. Y., and Purcell, E. M. (1954) Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments, *Physical Review 94*, 630-638.

89. Meiboom, S., and Gill, D. (1958) Modified Spin-Echo Method for Measuring Nuclear Relaxation Times, *Review of Scientific Instruments 29*, 688-691.

90. Palmer, A. G., 3rd, Kroenke, C. D., and Loria, J. P. (2001) Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules, *Methods in enzymology 339*, 204-238.

91. Mulder, F. A., Skrynnikov, N. R., Hon, B., Dahlquist, F. W., and Kay, L. E. (2001) Measurement of slow (micros-ms) time scale dynamics in protein side chains by (15)N relaxation dispersion NMR spectroscopy: application to Asn and Gln residues in a cavity mutant of T4 lysozyme, *Journal of the American Chemical Society 123*, 967-975.

92. Farrow, N. A., Zhang, O., Forman-Kay, J. D., and Kay, L. E. (1994) A heteronuclear correlation experiment for simultaneous determination of 15N longitudinal decay and chemical exchange rates of systems in slow equilibrium, *Journal of biomolecular NMR 4*, 727-734.

93. 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.

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

95. Marsh, J. A., and Forman-Kay, J. D. (2009) Structure and disorder in an unfolded state under nondenaturing conditions from ensemble models consistent with a large number of experimental restraints, *Journal of molecular biology 391*, 359-374.

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

97. Prestegard, J. H., Bougault, C. M., and Kishore, A. I. (2004) Residual dipolar couplings in structure determination of biomolecules, *Chemical reviews 104*, 3519-3540.

98. Tycko, R., Blanco, F. J., and Ishii, Y. (2000) Alignment of biopolymers in strained gels: A new way to create detectable dipole-dipole couplings in high-resolution biomolecular NMR, *Journal of the American Chemical Society 122*, 9340-9341.

99. Blackledge, M. (2005) Recent progress in the study of biomolecular structure and dynamics in solution from residual dipolar couplings, *Prog Nucl Magn Reson Spectrosc 46*, 23-61.

100. Bouvignies, G., Bernado, P., and Blackledge, M. (2005) Protein backbone dynamics from N-HN dipolar couplings in partially aligned systems: a comparison of motional models in the presence of structural noise, *J Magn Reson 173*, 328-338.

101. Mathieu, D. (2005) Entwicklung neuer Methoden zur Messung von Residualen Dipolaren Kopplungen an Naturstoffen", , *Diploma Thesis, Johann Wolfgang Goethe-Universität, Frankfurt am Main*

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

103. 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, *Journal of biomolecular NMR 18*, 303-309.

104. Gutowsky, H. S., and Mccall, D. W. (1951) Nuclear Magnetic Resonance Fine Structure in Liquids, *Phys Rev 82*, 748-749.

105. Hahn, E. L., and Maxwell, D. E. (1952) Spin Echo Measurements of Nuclear Spin Coupling in Molecules, *Phys Rev 88*, 1070-1084.

106. Lendel, C., and Damberg, P. (2009) 3D J-resolved NMR spectroscopy for unstructured polypeptides: fast measurement of 3J HNH alpha coupling constants with outstanding spectral resolution, *Journal of biomolecular NMR 44*, 35-42.

107. Eberstadt, M., Gemmecker, G., Mierke, D. F., and Kessler, H. (1995) Scalar Coupling-Constants - Their Analysis and Their Application for the Elucidation of Structures, *Angewandte Chemie (International ed 34*, 1671-1695.

108. Dyson, H. J., and Wright, P. E. (2002) Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance, *Advances in protein chemistry 62*, 311-340.

109. Uversky, V. N. Targeting intrinsically disordered proteins in neurodegenerative and protein dysfunction diseases: another illustration of the D(2) concept, *Expert review of proteomics 7*, 543-564.

110. Andrews, P. R. (1971) Cis-trans isomerism of the peptide bond, *Biopolymers 10*, 2253-2267.

111. Zimmerman, S. S., and Scheraga, H. A. (1976) Stability of cis, trans, and nonplanar peptide groups, *Macromolecules 9*, 408-416.

112. Karplus, M. (1963) Vicinal Proton Coupling in Nuclear Magnetic Resonance, *Journal of the American Chemical Society 85*, 2870-&.

113. 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, *Journal of biomolecular NMR 23*, 47-55.

114. Hoch, J. C., Dobson, C. M., and Karplus, M. (1985) Vicinal coupling constants and protein dynamics, *Biochemistry 24*, 3831-3841.

115. Vuister, G. W., and Bax, A. (1993) Quantitative J Correlation - a New Approach for Measuring Homonuclear 3-Bond J(H(N)H(Alpha) Coupling-Constants in N-15-Enriched Proteins, *Journal of the American Chemical Society 115*, 7772-7777.

116. Bystrov, V. P. (1979) Spin-Spin Couplings and the Conformational State of Peptide Systems, *Proc NMR Spectrosc 10*, 41-81.

117. Hähnke, M. J., Richter, C., Heinicke, F., and Schwalbe, H. (2010) The HN(COCA)HAHB NMR experiment for the stereospecific assignment of Hbeta-protons in non-native states of proteins, *Journal of the American Chemical Society 132*, 918-919.

118. Pachler, K. G. R. (1963) Nuclear Magnetic Resonance Study of Some Alpha-Amino Acids .1. Coupling Constants in Alkaline and Acidic Medium, *Spectrochim Acta 19*, 2085-2092.

119. Schlörb, C., Ackermann, K., Richter, C., Wirmer, J., and Schwalbe, H. (2005) Heterologous expression of hen egg white lysozyme and resonance assignment of

tryptophan side chains in its non-native states, *Journal of biomolecular NMR 33*, 95-104.

120. Bodenhausen, G., and Ruben, D. J. (1980) Natural Abundance N-15 Nmr by Enhanced Heteronuclear Spectroscopy, *Chem Phy Lett 69*, 185-189.

121. Palmer, A. G., Cavanagh, J., Wright, P. E., and Rance, M. (1991) Sensitivity Improvement in Proton-Detected 2-Dimensional Heteronuclear Correlation Nmr-Spectroscopy, *J Magn Reson 93*, 151-170.

122. 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, *Journal of biomolecular NMR 4*, 301-306.

123. Löhr, F., and Ruterjans, H. (1996) Novel Pulse Sequences for the Resonance Assignment of Aromatic Side Chains in 13C-Labeled Proteins, *J Magn Reson B 112*, 259-268.

124. Davis, A. L., Keeler, J., Laue, E. D., and Moskau, D. (1992) Experiments for Recording Pure-Absorption Heteronuclear Correlation Spectra Using Pulsed Field Gradients, *J Mag Reson 98*, 207-216.

125. 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.

126. Wüthrich, K. (2005) Optimizing the process of nuclear magnetic resonance spectrum analysis and computer aided resonance assignment *Diss., Naturwissenschaften, Eidgenössische Technische Hochschule ETH Zürich*, Nr. 15947.

127. Tollinger, M., Skrynnikov, N. R., Mulder, F. A., Forman-Kay, J. D., and Kay, L. E. (2001) Slow dynamics in folded and unfolded states of an SH3 domain, *Journal of the American Chemical Society 123*, 11341-11352.

128. Tycko, R. (2000) Solid-state NMR as a probe of amyloid fibril structure, *Current opinion in chemical biology 4*, 500-506.

129. Ishii, Y., Markus, M. A., and Tycko, R. (2001) Controlling residual dipolar couplings in high-resolution NMR of proteins by strain induced alignment in a gel, *Journal of biomolecular NMR 21*, 141-151.

130. Ottiger, M., Delaglio, F., and Bax, A. (1998) Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra, *J Magn Reson 131*, 373-378.

131. Jensen, M. R., Salmon, L., Nodet, G., and Blackledge, M. (2010) Defining conformational ensembles of intrinsically disordered and partially folded proteins directly from chemical shifts, *Journal of the American Chemical Society 132*, 1270-1272.

132. Pardi, A., Billeter, M., and Wuthrich, 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, *Journal of molecular biology 180*, 741-751.

133. Zweckstetter, M. (2008) NMR: prediction of molecular alignment from structure using the PALES software, *Nature protocols 3*, 679-690.

134. Marsh, J. A., Baker, J. M., Tollinger, M., and Forman-Kay, J. D. (2008) Calculation of residual dipolar couplings from disordered state ensembles using local alignment, *Journal of the American Chemical Society 130*, 7804-7805.

135. Nodet, G., Salmon, L., Ozenne, V., Meier, S., Jensen, M. R., and Blackledge, M. (2009) Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from NMR residual dipolar couplings, *Journal of the American Chemical Society 131*, 17908-17918.

136. Salmon, L., Nodet, G., Ozenne, V., Yin, G., Jensen, M. R., Zweckstetter, M., and Blackledge, M. (2010) NMR characterization of long-range order in intrinsically disordered proteins, *Journal of the American Chemical Society 132*, 8407-8418.

137. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., and Svergun, D. I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis, *Journal of Applied Crystallography 36*, 1277-1282.

138. Bernado, P., Mylonas, E., Petoukhov, M. V., Blackledge, M., and Svergun, D. I. (2007) Structural characterization of flexible proteins using small-angle X-ray scattering, *Journal of the American Chemical Society 129*, 5656-5664.

139. Debye, P., Anderson, H. R., and Brumberger, H. (1957) Scattering by an Inhomogeneous Solid .2. The Correlation Function and Its Application, *J App Phys 28*, 679-683.

140. Porod, G. (1951) Die Röntgenkleinwinkelstreuung von dichtgepackten kolloiden Systemen, I. Teil, *Kolloid Zeitschrift 124*, 31.

141. Rambo, R. P., and Tainer, J. A. (2011) Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law, *Biopolymers 95*, 559-571.

142. Schlörb, C. (2007) NMR-based Structural and Dynamical Studies on Non-Native Variants of Hen Egg White Lysozyme, *Dissertation, Johann Wolfgang Goethe-Universität, Frankfurt am Main*

143. Heinicke, F. (2007) NMR-spektroskopische Untersuchungen nicht-nativer Zustände von Punktmutanten des Hühnereiweiß-Lysozyms *Diploma Thesis, Johann Wolfgang Goethe-Universität, Frankfurt am Main*

144. 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, *Journal of biomolecular NMR 19*, 153-165.

145. 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, *Journal of molecular biology 288*, 705-723.

146. Vajpai, N., Gentner, M., Huang, J. R., Blackledge, M., and Grzesiek, S. (2010) Side-chain chi(1) conformations in urea-denatured ubiquitin and protein G from (3)J coupling constants and residual dipolar couplings, *Journal of the American Chemical Society 132*, 3196-3203.

147. Mishima, T., Ohkuri, T., Monji, A., Imoto, T., and Ueda, T. (2007) A particular hydrophobic cluster in the residual structure of reduced lysozyme drastically affects the amyloid fibrils formation, *Biochemical and biophysical research communications 356*, 769-772.

148. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0, *Bioinformatics (Oxford, England) 23*, 2947-2948.

149. Redfield, C., and Dobson, C. M. (1990) 1H NMR studies of human lysozyme: spectral assignment and comparison with hen lysozyme, *Biochemistry 29*, 7201-7214.

150. Buck, M., Boyd, J., Redfield, C., MacKenzie, D. A., Jeenes, D. J., Archer, D. B., and Dobson, C. M. (1995) Structural determinants of protein dynamics: analysis of 15N NMR relaxation measurements for main-chain and side-chain nuclei of hen egg white lysozyme, *Biochemistry 34*, 4041-4055.

Zusammenfassung

Seit einigen Jahrzehnten ist Lysozym eines der am meisten erforschten Proteine in der Literatur und wird hauptsächlich als Modell Protein zur Aufklärung der Faltungs- und Entfaltungsprozesse genutzt. Da die Frage nach Fehlfaltung und deren Verknüpfung mit neurodegenerativen Krankheiten bis zum heutigen Tag nicht vollständig geklärt ist, besteht hier ein großer Spielraum für weitere Forschungsansätze. In der vorliegenden Arbeit wurden daher zwei Modellsysteme verwendet, Hühereiweiß-Lysozym und menschliches Lysozym, jeweils in ihrem nicht-nativen ungefalteten Zustand. Diese ungefalteten Ensembles wurden mit Hilfe NMR spektroskopischer Methoden untersucht und ergaben sehr detaillierte, zum Teil auch überraschende neue Einblicke in Struktur und Dynamik der beiden Proteine und liefern somit wichtige Erkenntnisse zu Faltungs- und Aggregationsprozessen.

Als erstes Modellprotein wurde eine Variante des Hühnereiweiß-Lysozyms analysiert, bei welcher alle Cysteine durch Alanine ersetzt wurden (WT^{Ala}). Dies führt zu einer Eliminierung der Disulfidbrücken und somit zu einer Stabilisierung des nicht-nativen Zustandes.
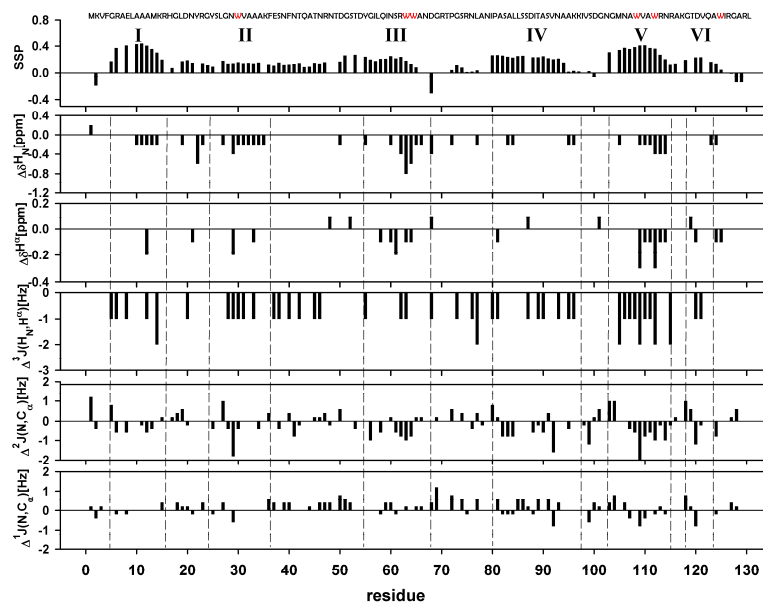
Abb. S1: Sekundäre Reststruktur in WT^{Ala}; Schwarze Balken (von oben nach unten): Sekundäre Reststruktur (SSP), $\Delta\delta H_N$, $\Delta\delta H^\alpha$, $\Delta^3 J(H_N, H^\alpha)$, $\Delta^2 J(N, C^\alpha)$ und $\Delta^1 J(N, C^\alpha)$, schwarze Linien: hydrophobe Regionen in WT^{Ala}. Werte für $\Delta\delta H^\alpha$ und $\Delta\delta H^N$ wurden normalisiert mit dem Wert 0 für 0.0 bis 0.049, dem Wert 0.1 von 0.05 bis 0.149, 0.2 von 0.15 bis 0.249, usw., $\Delta\delta$ wurde mit Hilfe von $\Delta\delta = \delta_{exp} - \delta_{rc}$ berechnet. Alle Experimente wurden aufgenommen bei 600MHz mit einer Proteinkonzentration von 300µM bei pH2, 293K.

Chemische Verschiebungsdifferenzen, welche auf Reststruktur des Proteinrückgrats hinweisen, wurden ebenso untersucht, wie Kopplungskonstanten des Rückgrats und der Seitenketten, welche Erkenntnisse über die Torsionswinkel liefern (Abb. S1). Der Rückgratwinkel ϕ wurde über die $^3J(H_N,H^\alpha)$ Kopplung bestimmt, ψ über $^1J(N,C^\alpha)$ und $^2J(N,C^\alpha)$ Kopplungskonstanten und der Seitenkettenwinkel χ_1 wurde mittels $^3J(H^\alpha,H^\beta)$ und $^3J(CO,H^\beta)$ bestimmt. Eine stereospezifische Analyse nach *Pachler* lieferte die Populationen von χ_1 für ausgewählte Aminosäuren. Zum Vergleich wurden zusätzlich Randomcoil Werte für Polypeptide herangezogen. Neben der allgemein verwendeten Auswertung von Relaxationsdaten als Funktion der Aminosäuresequenz, wurden auch die reduzierten spektralen Dichtefunktionen anhand der $J(\omega_N)/J(0)$ Korrelation analysiert. Diese Analyse ermöglichte hier eine Aussage über interne und globale Bewegungen des Ensembles und solche, die sich aus beiden Bewegungen zusammensetzen. Relaxationsraten und residuale dipolare Kopplungen, beides anistrope Parameter, wurden ebenfalls untersucht.
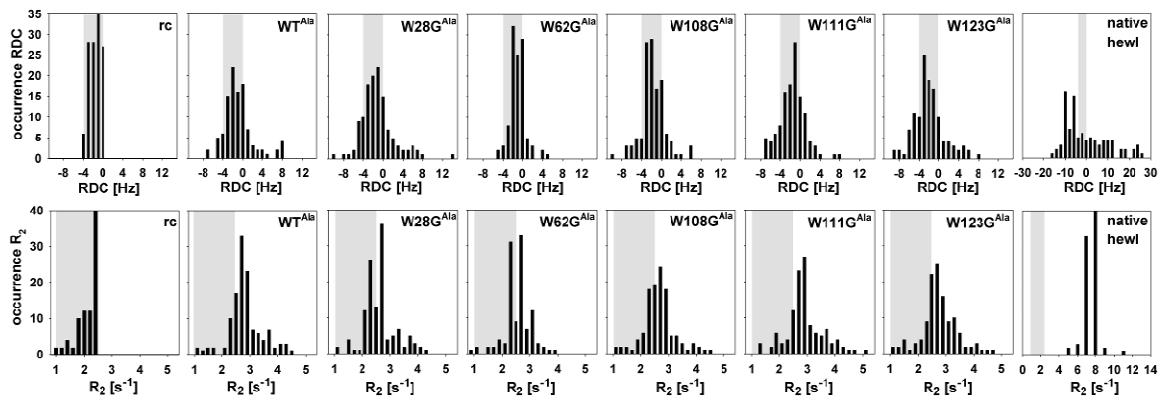
Abb. S2: RDC und R_2 Tensoren der Mutanten: Tensorhistogramm der RDCs und R_2 Relaxationsraten von random-coil (rc), WT^{Ala}, W28G^{Ala}, W62G^{Ala}, W108G^{Ala}, W111G^{Ala}, W123G^{Ala} und nativem Lysozym (*144*),; oben: RDC Tensoren; unten: Tensoren der ^{15}N R_2 Relaxationsraten; graue Balken: Profil der random coil Vorhersagen für RDCs (-4, 0) und R_2 Raten (1-2, 4).

Erstmals wurde dies hier sowohl sequenzspezifisch als auch hinsichtlich ihrer Häufigkeit für ^{15}N Relaxationsraten durchgeführt. Unter Miteinbeziehung der RDCs lieferten diese Ergebnisse schließlich Hinweise auf Regionen mit sekundärer Reststruktur, welche auf verschiedenen Zeitskalen bestehen bleibt. In einem weiteren Ansatz wurden Struktur und

Dynamik von entfaltetem Hühnereiweiß-Lysozym durch Einführung von Punktmutationen beeinflusst. Dies geschah durch einen Austausch hydrophober Tryptophane gegen Glycine in den sechs Regionen mit sekundärer Reststruktur; bei den Punktmutanten handelt es sich um W28G^{Ala}, W62G^{Ala}, W108G^{Ala}, W111G^{Ala} und W123G^{Ala}. Unterstützt wurde dieser Ansatz durch die Durchführung von SAXS- (small angle x-ray scattering) und Diffusionsmessungen, welche die Bestimmung der hydrodynamischen Radien und der Gyrationsradien ermöglichte. Weitreichende Effekte auf lokaler und globaler Ebene waren beobachtbar und verdeutlichen wiederum die Auswirkungen auf die Faltungs- und Fehlfaltungsraten bzw. auf das Aggregationsverhalten von Hühnereiweiß-Lysozym durch konformationelle Veränderungen und Beeinflussung der Faltungsenergielandschaft.

Als zweites Modellprotein diente in dieser Arbeit das menschliche Lysozym. Ebenso wie Hühnereiweiß-Lysozym wurde auch dieses in seinem nicht-nativen Zustand untersucht, welches durch die Substitution aller Cysteine durch Alanine erfolgte (hLys). Dieses System wurde hier zum ersten Mal synthetisiert und mit Hilfe NMR spektroskopischer Methoden analysiert, ebenso die zwei Punktmutanten I56T und D67H in ihrem entfalteten Zustand. Vorhergehende Studien belegen, dass unter thermischer Entfaltung bei pH 1,2 die tertiäre Struktur verschwindet und ein Premolten-Globule Zustand entsteht. Hier wurden nun die Premolten-Globule Zustände von hLys und den beiden Punktmutanten auf sekundäre Reststruktur und Aggregationsverhalten geprüft. Dies geschah mittels ^{15}N Relaxationsraten und chemischen Verschiebungsdifferenzen. Ebenso wurde die Konzentrationsabhängigkeit der Aggregation getestet. Obwohl ungefaltetes Hühnereiweiß- und humanes Lysozym insgesamt sehr ähnliche strukturelle und dynamische Eigenschaften zeigten, welche sechs hydrophobe Regionen mit Reststruktur und weitreichenden Wechselwirkungen beinhalten, gibt es dennoch wichtige Unterschiede. So weist hLys signifikant größere chemische Verschiebungsdifferenzen, höhere heteronukleare Relaxationsraten mit breiterer Verteilung und höhere Rigidität auf, was auch durch die Analyse von $J(\omega_N)/J(0)$ gezeigt werden konnte (Abb. S3).
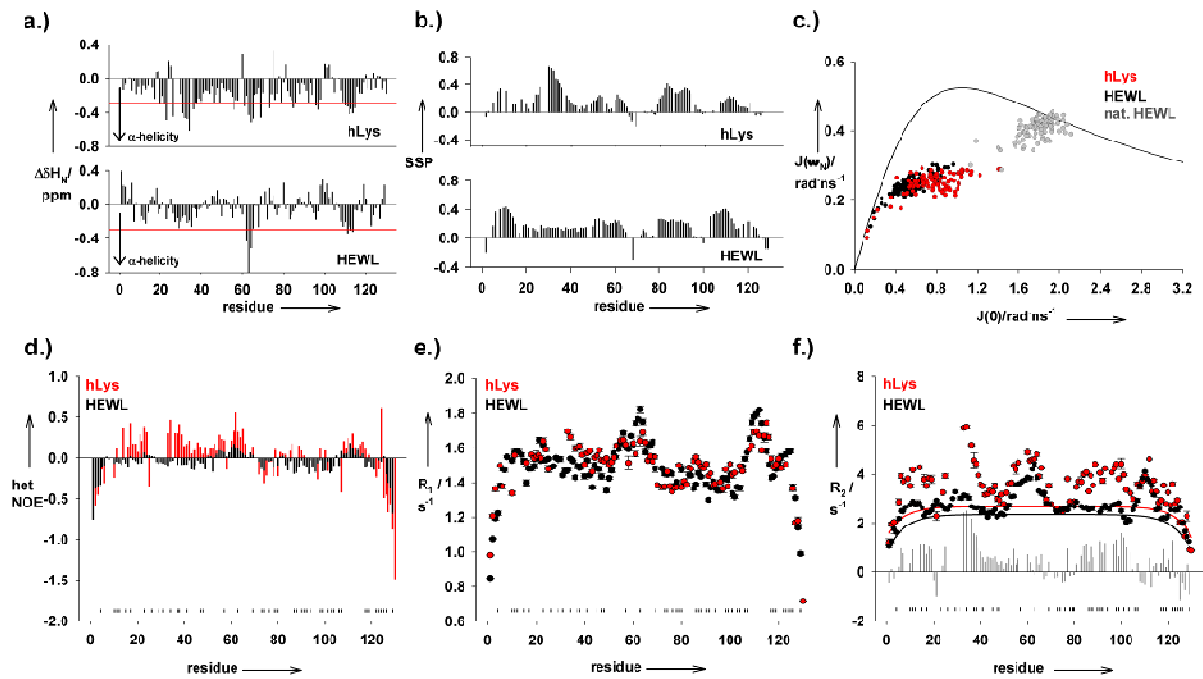
Abb. S3: Relaxation und sekundäre Reststruktur von hLys und WT^{Ala}: a.) Chemische Verschiebungsdiffrenzen zu random coil Werten ($\Delta\delta=\delta^{exp}-\delta^{rc}$) von hLys und HEWL; rote Linien: chem. Verschiebungsindex für α-helices bei -0.3; b.) Sekundäre Reststruktur von hLys and HEWL; c.) Spektrale Dichtefunktionen für HEWL (schwarz), hLys (rot) und natives HEWL (grau) d.-f.) Vergleich von ^{1}H-^{15}N NOE, R_1 und R_2 Relaxationsraten von hLys (rot) und HEWL (schwarz); graue Balken: Differenz in R_2 Raten; schwarze vertikale Markierungen: unterschiedliche Aminosäuren in der Sequenz von hLys und HEWL. Alle Experimente wurden aufgenommen bei 600MHz mit einer Proteinkonzentration von 300µM bei pH2, 293K

Diese Ergebnisse deuten bei hLys auf einen wesentlich höheren Anteil an α-Helizität hin als bei WT^{Ala}. Im Vergleich zu den Mutanten von WT^{Ala} zeigten die Mutanten I56T und D67H deutliche Unterschiede vor allem in den Relaxationsraten (Abb. S4). Die teilweise sehr hohen Raten deuten auf eine große Aggregationstendenz hin, welches auch schon bei niedriger Konzentration auftritt. Dies konnte im Vorfeld auch schon für hLys gezeigt werden, jedoch nur bei hoher Konzentration. Des Weiteren weisen die Mutanten in den R_2 Raten einen Bereich starker Streuung auf, welcher auf Austauchprozesse hindeutet. Alle diese Ergebnisse lassen darauf schlissen, dass es sich hierbei um die transiente Ausbildung von Amyloidfibrillen handelt.
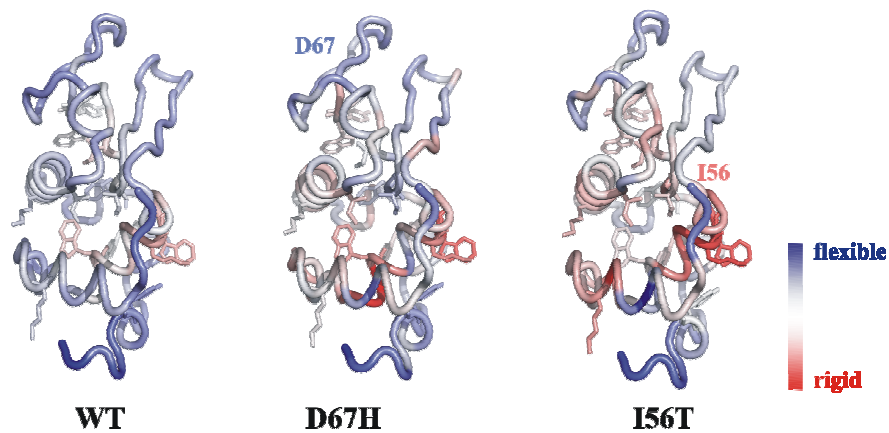
Abb. S4: Dynamische Eigenschaften von hLys, I56T und D67H; R_2 Relaxationsraten von hLys, I56T und D67H wurden auf die native Struktur des humanen Lysozyms abgebildet (2zij.pdb(*34*)); blau: niedrige R_2 Raten, flexible Regionen; weiß: mittlere R_2 Raten; rot: hohe R_2 Raten, strukturierte Regionen.

Amyloide Fibrillenbildung entsteht aus teilweise strukturierten, aggregationskompetenten Proteinen. Dies ist sowohl in Hühnereiweiß-Lysozym als auch in humanem Lysozym der Fall. Dennoch ist der Grad der sekundären Reststruktur in hLys signifikant höher als in WT[Ala], was auch durch Diffusionsstudien und die resultierenden hydrodynamischen Radien belegt werden konnte. Eine Erklärung hierfür beinhaltet möglicherweise die Region, in der sich die beiden Proteine hinsichtlich ihrer Aminosäuresequenz am meisten unterscheiden. So befinden sich in Region 4 in Hühnereiweiß-Lysozym hauptsächlich positive und neutrale Aminosäuren, während diese Region in hLys neutral geladen ist und daher keine Ladungsabstoßung stattfinden kann; die hydrophobe Wechselwirkungen sind hier verstärkt. Wie schon vorher in der Literatur beschrieben, ist es tatsächlich die Region um die β-Domäne und die anschließende Helix 3, welche zuerst in die Ausbildung der Amyloidfibrillen involviert ist.

Danksagung

Ich danke meinem Doktorvater Prof. Dr. Harald Schwalbe für die interessante, abwechslungsreiche Themenstellung und umfassende Betreuung meiner Doktorarbeit, außerdem für die hervoragenden Arbeitsbedingungen in seinem Arbeitskreis durch eine exzellente Laborausstattung.

Julia Wirmer-Bartoschek danke ich für das unermüdliche Engagement bei Fragen und Korrekturen aller Art rund um meine Doktorarbeit und das Ausarbeiten von Manusskripten, für interessante Denkanstöße, tatkräftige Unterstützung am Spektrometer und das umfassende Einbringen Ihres Fachwissens.

Bei Robert Silvers bedanke ich mich für ein allzeit offenes Ohr und ständige Diskussionsbereitschaft, für die unermüdliche Hilfsbereitschaft im Labor und am Spektrometer und bei der Auswertung und Aufarbeitung schwieriger Daten.

Ich danke dem „boysroom" für die angenehme, motivierende, produktive Arbeitsatmosphäre, die ständige Hilfsbereitschaft bei allen chemischen, biologischen und computertechnischen Fragen und die stetige „Gute-Laune-Luft".

Prof. Dr. Martin Blackledge und seiner Arbeitsgruppe danke ich für die Mitarbeit im Hühnereiweiß Lysozym Projekt, insbesondere für die RDC-Vorhersagen, ASTEROIDS Rechnungen und die SAXS-Messungen.

Bei Julia Wernsdorfer bedanke ich mich für die „physikalische" Unterstützung, für ein offenes Ohr in allen Situationen und die produktiven, diskussionsreichen Kaffeepausen

Ich danke meiner Familie, besonders meinen Eltern, die mir das Studium, die Doktorarbeit und ein Leben in Frankfurt erst ermöglicht haben, alle Veränderungen währenddessen aktiv mit durchlebten und immer an mich geglaubt haben.

Übersicht meiner akademischen Lehrer an der

Goethe-Universität Frankfurt am Main (alphabetisch)

Prof. Dr. Norbert Auner, Prof. Dr. Wolf Aßmus, Prof. Dr. Bernd Brutschy, Prof. Dr. Ernst Egert, Prof. Dr. Joachim Engels, Prof. Dr. Michael Göbel, Prof. Dr. Max Holthausen, Prof. Dr. Bernd Kolbesen, Prof. Dr. Bernd Ludwig, Prof. Dr. Magnus Rüping, Prof. Dr. Thomas Prisner, Prof. Dr. Martin U. Schmidt, Prof. Dr. Gisbert Schneider, Prof. Dr. Harald Schwalbe, Prof. Dr. Robert Tampé, Prof. Dr. Andreas Terfort, Prof. Dr. Matthias Wagner, Prof. Dr. Josef Wachtveitl

List of Publications

Hähnke M.J., Richter C., Heinicke F., Schwalbe H. (**2010**), *The HN(COCA)HAHB NMR experiment for the stereospecific assignment of Hbeta-protons in non-native states of proteins*, JACS, 132; 918-9
http://dx.doi.org/10.1021/ja909239w

Sziegat F., Wirmer-Bartoschek J., Schwalbe H. (**2011**), *Characteristics of Human Lysozyme and Its Disease-Related Mutants in their Unfolded States*, Angew Chem Int Ed Engl, 50; 5514-8
http://onlinelibrary.wiley.com/doi/10.1002/anie.201008040/pdf

Sziegat F., Silvers R., Hähnke M., Ringkjobing-Jensen M., Gabel F., Huang J., Wirmer-Bartoschek J., Blackledge M., Schwalbe H. (**2011**), *Biophysical studies of protein unfolded states: precise mapping local and long-range structure and dynamics*, submitted, JACS.

Silvers R., Sziegat F., Tachibana H., Segawa S., Whittaker S., Günther U., Gabel F., Huang J., Blackledge M., Wirmer-Bartoschek J., Schwalbe H. (**2011**), *Modulation of Structure and Dynamics by DisulfideBond Formation in Unfolded States*, submitted, PNAS