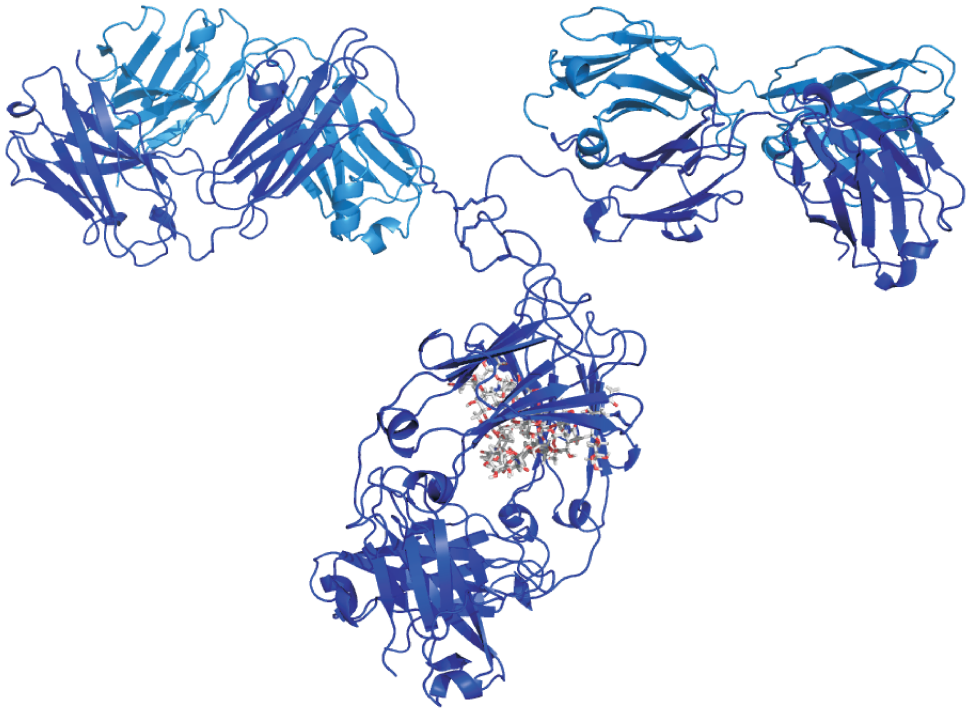
NMR spectroscopic investigations reveal insights into antibody secretion, amyloidogenic folding intermediates, and the process of focal adhesion

Sandra Groscurth



Dissertation 2009

TECHNISCHE UNIVERSITÄT MÜNCHEN

TECHNISCHE UNIVERSITÄT MÜNCHEN
Department Chemie
Lehrstuhl II
für Organische Chemie

NMR spectroscopic investigations reveal insights into antibody secretion, amyloidogenic folding intermediates, and the process of focal adhesion

Sandra Groscurth

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr. M. Sattler
Prüfer der Dissertation:	1. Univ.-Prof. Dr. H. Kessler
	2. Univ.-Prof. Dr. Th. Kiefhaber
	3. Univ.-Prof. Dr. J. Buchner

Die Dissertation wurde am 21.01.2009 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 03.03.2009 angenommen.

Those who are not shocked
when they first come across
quantum mechanics cannot
possibly have understood it

(Niels Henrik David Bohr)

Acknowledgement

The work presented in this thesis was prepared from May 2005 until October 2008 in the group of Prof. Dr. Horst Kessler at the Department of Chemistry of the Technical University of Munich, Germany.

I would like to thank my supervisor Prof. Dr. Kessler for giving me the opportunity to join his group, for the excellent research facilities, for the unrestricted support, and for giving me lots of personal freedom.

My further thank goes to:

- My protein factory Matthias Feige and Moritz Marcinowski. Thanks to them and Prof. Dr. Johannes Buchner for the great cooperation in the antibody projects. Furthermore, I would like to thank Matthias and Moritz for the excellent team work and a lot of fun besides the science. Thanks guys for the wonderful time!

- Dr. Jochen Klages for his kind to elate me for the NMR spectroscopy, for answering nearly every question, for the conversations about all the world and his brother during the morning coffee, and for discovering the advantage of a banana box.

- Prof. Dr. Thomas James and Dr. Mark Kelly, University of California, San Francisco, USA, for giving me the opportunity to join the research group for half a year. Additionally Prof. Dr. Diane Barber and Dr. Jyoti Srivastava for all the cell biological background and experiments on talin. And special thanks to Mark for involving me in the talin project and giving me a lot of freedom in every respect, turning my stay in San Francisco into a memorable time.

- Prof. Dr. Linda Hendershot and Dr. Yuichiro Shimizu, St. Jude Children's Research Hospital, Memphis, USA, for the *in vivo* experiments on the IgG antibody molecules.

- Dr. Emanuele Paci and Zu Thur Yew, University of Leeds, United Kingdom, for performing the MD simulations for the folding intermediate.

- Prof. Dr. Hermann Bauer, Georg Simon Ohm University of Applied Science, Nuremberg, for the enjoyable tea breaks at his office and for putting in a good word for me with Prof. Dr. Kessler.

- Dr. Vincent Truffault and Dr. Murray Coles for instructing me in setting up NMR experiments, for providing helpful scripts to facilitate the analysis of NMR data, and for their endless support regarding any NMR question.
- Sylvain Tourel for the perfect team work in the biochemistry lab and for his assistance in getting NMR amounts of pure and soluble protein.
- Dr. Axel Meyer, Dr. Jochen Klages, and Matthias Feige for critical reading of my thesis and providing constructive suggestions.
- Dr. Rainer Haessner for his efforts to keep the spectrometers running and thus enabling the excellent research facilities.

Finally, I would like to thank my parents for their unlimited support at all times.

And last but not least I thank Axel for pushing me just a bit to finish writing, for always caring about exceptional alternations from science, and for all these wonderful memories of this time in Munich.

Parts of this thesis have been published:

M. Feige, S. Groscurth, M. Marcinowski, Z. Yew, V. Truffault, E. Paci, H. Kessler, J. Buchner, 'The structure of a folding intermediate provides insight into differences in immunoglobulin amyloidogenicity', *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 13373 – 13378.

J. Srivastava, G. Barreiro, S. Groscurth, A. Gingras, B. Goult, D. Critchley, M. Kelly, M. Jacobson, D. Barber, 'Structural model and functional significance of pH-dependent talin-actin binding for focal adhesion remodeling', *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 14436 – 14441.

M. Feige, S. Groscurth, M. Marcinowski, Y. Shimizu, L. Hendershot, H. Kessler, J. Buchner, 'A natively unfolded domain controls the secretion of murine IgG antibodies', *submitted*

Contents

1 Introduction **1**

2 Theoretical Aspects **5**
2.1 Protein Dynamics 7
2.1.1 Dynamics in the ps to ns Time Scale 8
2.1.2 Dynamics in the μs to ms Time Scale 13
2.2 Amide Proton Exchange 17
2.3 Residual Dipolar Couplings 18

3 The C_L Domain Explains Differences in Ig Amyloidogenicity **23**
3.1 Introduction 24
3.1.1 The Antibody Molecule Immunoglobulin G 24
3.1.2 The Immunoglobulin Fold 29
3.1.3 Immunoglobulin Amyloidogenicity 32
3.2 Folding Pathway of the C_L Domain 34
3.2.1 Backbone Assignment 35
3.2.2 Secondary Structure 35
3.2.3 Folding Kinetics 38
3.3 Structural Characterization of the Trapped Folding Intermediate 42
3.3.1 Backbone and Aliphatic Side Chain Assignment 42
3.3.2 Comparison of Secondary and Tertiary Structure of the Wild Type and the Mutant 44
3.3.3 NMR-restrained MD Simulations 47
3.4 Amyloidogenic Properties of the C_L Domain and β_2-microglobulin . . . 49
3.5 Discussion 50

4 The C_H1 Domain Controls the Secretion of IgG Molecules **53**
4.1 Introduction 54
4.1.1 Secretion of Antibody Molecules 54
4.1.2 Intrinsically Disordered Proteins 57
4.2 Investigation of the Structural Characteristics of the C_H1 Domain . . . 62
4.2.1 The Unassembled C_H1 Domain is Natively Unfolded 62
4.2.2 Association coupled Folding of the C_H1 Domain 65
4.3 Insights into the Secretion Control Mechanism 69
4.4 Discussion 73

5 pH Dependent Talin Binding to Actin Promotes FA Remodeling 75
5.1 Introduction . . . 76
5.1.1 Cell Migration . . . 76
5.1.2 The Focal Adhesion Protein Talin . . . 79
5.2 Investigation of the Dynamics of the Actin Binding Site of Talin . . . 82
5.2.1 Assignment of the pH Sensitive Actin Binding Domain of Talin 83
5.2.2 Titration Studies of the Monomeric and Dimeric Talin Construct 83
5.2.3 Conformational Coupling between the pH Sensor and the Actin Binding Site . . . 87
5.3 Insights into the pH Dependent Mechanism of Focal Adhesion Remodeling 93
5.4 Discussion . . . 94

6 Summary 97

A The C_L Domain of the Immunoglobulin G Antibody 101
A.1 Sample Preparation . . . 101
A.2 Experiments and Assignment of C_L^{wt} and C_L^{P35A} . . . 103

B The C_H1 Domain of the Immunoglobulin G Antibody 117
B.1 Sample Preparation . . . 117
B.2 Experiments and Assignment of the Folded and Unfolded C_H1 Domain 119

C The pH Sensitive USH-I/LWEQ Domain of Talin 127
C.1 Sample Preparation . . . 127
C.2 Experiments and Assignment of the Talin USH-I/LWEQ Domain . . . 129

Bibliography 135

List of Figures

2.1 Time scales for protein dynamics and NMR techniques 8
2.2 Schemes for the measurements of ^{15}N spin T_1, T_2, and ^{1}H–^{15}N NOE values 9
2.3 Schematic representation of irradiated transitions in the steady state NOE experiment 10
2.4 Schematic illustration of the internal correlation time and the overall correlation time 11
2.5 Schematic representation of the squared order parameter S^2 and the internal correlation time τ_i 13
2.6 Parameter dependence of relaxation dispersion curves 16
2.7 Summary of the motionally averaged spin interactions 18
2.8 Schematic representation of the dependence of RDC values on the orientation of the internuclear vector and the alignment tensor 19
2.9 Extracting dipolar coupling constants by comparing line splittings in isotropic and anisotropic solution 21

3.1 Schematic illustration of an antibody molecule 25
3.2 Cartoon representation of immunoglobulin G 26
3.3 Cartoon representation of the glycosylated homodimer of the immunoglobulin domain C_H2 27
3.4 Cartoon representation of the heterodimer of the immunoglobulin variable domains 28
3.5 Structural tree of the Ig-fold 30
3.6 Topology diagrams of the Ig-fold 31
3.7 Sequence alignment of the distinct domains of IgG 32
3.8 Sequence alignment of C_L and β_2m 33
3.9 Cartoon representation of C_L compared to β_2m 34
3.10 Assigned ^{1}H-^{15}N HSQC spectrum of the C_L domain 36
3.11 Consensus chemical shift index of the C_L domain 37
3.12 Superimposed ^{1}H-^{15}N HSQC spectra of the native and the intermediate state of the C_L domain 39
3.13 Folding kinetics of the C_L domain 41
3.14 Superimposed ^{1}H-^{15}N HSQC spectra of the kinetic and the trapped intermediate of the C_L domain. 43
3.15 Prediction of the Φ and Ψ dihedral angles by the software TALOS . . . 45
3.16 Comparison of backcalculated and experimental NOESY-strip 46

3.17 Structural characterization of the folding intermediate of the C_L domain 48
3.18 Sequence of β_2m, C_L, and the $\beta_2 m^{toCL}$ exchange mutant 49
3.19 Amyloidogenic properties of β_2m and the $\beta_2 m^{toCL}$ exchange mutant . . 50

4.1 Organelles involved in the secretory pathway 55
4.2 Immunoglobulin heavy chain binding protein BiP 56
4.3 Schematic folding funnel 58
4.4 Schematic energy landscape for protein folding 59
4.5 Schematic illustration of the 'fly casting' mechanism 61
4.6 ^{1}H-^{15}N HSQC spectrum of the isolated C_H1 domain 63
4.7 ^{1}H-^{15}N HSQC spectrum of the C_H1 domain in complex with the C_L domain 66
4.8 Correlation between the backcalculated and the experimental RDC values of the C_H1 domain 67
4.9 Folding kinetics of the C_H1 domain 68
4.10 Verification of the formation of triple complexes between BiP, C_H1, and C_L 70
4.11 Secretion control mechanism of IgG antibodies 72

5.1 Schematic illustration of a migrating cell 77
5.2 Schematic illustration of the linkage between integrin receptors and the actin cytoskeleton 79
5.3 The focal adhesion protein talin 81
5.4 Assigned ^{1}H-^{15}N HSQC spectrum of the pH sensitive actin binding domain of talin 84
5.5 pH dependent conformational changes in talin USH-I/LWEQ 86
5.6 pH dependent residual dipolar couplings of talin USH-I/LWEQ 88
5.7 Speculation about a possible mechanism for the regulation of actin binding by talin 90
5.8 pH dependent CPMG relaxation dispersion experiments of talin USH-I/LWEQ 92
5.9 pH dependent focal adhesion stability 94

List of Tables

A.1 Experiments on the wild type C_L domain 103
A.2 Experiments on the C_L^{P35A} mutant 103
A.3 Backbone and side chain assignments of the wild type C_L domain . . . 104
A.4 Backbone and side chain assignments of the C_L^{P35A} mutant 108
A.5 TALOS dihedral angle restraints of the wild type C_L domain 112
A.6 TALOS dihedral angle restraints of the C_L^{P35A} mutant 114

B.1 Experiments on the assembled C_H1 domain 119
B.2 Experiments on the unassembled C_H1 domain 119
B.3 Backbone assignments of the assembled C_H1 domain 120
B.4 Backbone assignments of the unassembled C_H1 domain 122
B.5 RDC values for the assembled C_H1 domain 124

C.1 Experiments on the talin USH-I/LWEQ domain 129
C.2 Backbone assignments of the talin USH-I/LWEQ domain 130

Abbreviations and Symbols

Å	Ångstrøm
B_0	Static magnetic field strength
β_2m	β_2-microglobulin
$C(t)$	Rotational correlation function
C_H1	First constant domain of the heavy chain
$C_\mathrm{i}(t)$	Correlation function of the internal motion
C_L	Constant domain of the light chain
$C_\mathrm{o}(t)$	Correlation function of the overall motion
COSY	Correlation spectroscopy
CPMG	Carr-Purcell-Meiboom-Gill
CSI	Chemical shift index
D	Dipolar coupling constant
δ	Chemical shift difference
ECM	Extracellular matrix
ER	Endoplasmic reticulum
FA	Focal adhesion
γ	Gyromagnetic ratio
HetNOE	Heteronuclear Overhauser effect
HSQC	Heteronuclear single quantum correlation
Ig	Immunoglobulin
INEPT	Insensitive nuclei enhanced by polarization transfer
IPAP	Inphase antiphase
J	Scalar coupling constant
$J(\omega)$	Spectral density function
k_B	Boltzmann constant
MHC	Major histocompatibility complex
NHE	Na^+/H^+ exchanger

NOE	Nuclear Overhauser enhancement
NOESY	Nuclear Overhauser enhancement spectroscopy
ν	Frequency
ω	Larmor frequency
pH_i	intracellular pH
Φ, Ψ	Backbone dihedral angles
ppm	Parts per million (10^{-6})
r	Interatomic distance
RDC	Residual dipolar coupling
R_{ex}	Exchange contribution to the relaxation rate constant
rf	Radio frequency
R_1	Longitudinal or spin-lattice relaxation rate constant
R_2	Transverse or spin-spin relaxation rate constant
S^2	Generalized squared order parameter of the internal motion
S_f^2	Generalized squared order parameter of the fast internal motion
σ	Spin density matrix
S_s^2	Generalized squared order parameter of the slow internal motion
τ_c	Rotational correlation time
τ_f	Correlation time of the fast internal motion
τ_i	Correlation time of the internal motion
T_1	Longitudinal or spin-lattice relaxation time constant
TOCSY	Total correlation spectroscopy
T_2	Transverse or spin-spin relaxation time constant
τ_s	Correlation time of the slow internal motion

Chapter 1

Introduction

'The world of the nuclear spins is a true paradise for theoretical and experimental physicists. But not only for physicists is nuclear magnetic resonance of great fascination. More and more chemists, biologists, and medical doctors discover NMR, not so much for its conceptual beauty but for its extraordinary usefulness.' Richard Ernst started his Nobel lecture with these words, when he was awarded with the Nobel prize in chemistry in 1991 for 'his contribution to the development of the methodology of high resolution nuclear magnetic resonance (NMR) spectroscopy'. Within a few decades, NMR spectroscopy has developed from the first theoretical approaches into a well established and routinely applied technique covering a wide range of distinct fields. Every organic chemistry lab requires the NMR spectroscopy to verify the synthesized compounds. Moreover, NMR imaging represents an essential application in the medical sector complementing the classical X-ray examination as it also allows investigation of the different soft tissues. Complementing the X-ray crystallography holds also for the study of biomacromolecules. Even though it is generally true that a picture is worth a thousand words, this statement does not completely account for proteins and nucleic acids. Whereas X-ray crystallography elucidates reliable three dimensional structures in a reasonable amount of time, these static structures often lack explanations for the biologically relevant, functional activity of the protein. Although an averaged NMR structure in the same way represents a static picture of the protein, the technique itself possesses the capacity to additionally study structural changes over time. Hence, NMR spectroscopy is the only experimental technique with abilities to determine atomic resolution structures as well as investigate dynamics and intermolecular interactions of biological macromolecules at atomic detail.

Many proteins frequently undergo structural rearrangements to complete their functions. Transitions from low energy, ground state conformations to higher energy, excited state conformations play a vital role in protein function. Thus, to fully comprehend

the biological mechanism of a protein, it is essential to obtain information on the time dependent conformational fluctuations of the protein which generate the excited state although with low population. This conformational exchange can be exquisitely investigated by a powerful suite of relatively new heteronuclear NMR relaxation techniques. Recent methodologies have focused on μs–ms time scales, since many biologically relevant processes occur in this time regime, such as enzyme catalysis, protein folding, and allosteric transitions. NMR spectroscopy is unique in this respect, allowing the investigation of important biological events that are relatively inaccessible to other techniques.

Since the discovery that numerous human diseases are caused by protein aggregation, the biophysical characterization of misfolded states and their aggregation mechanisms has attracted enormous interest. In general, proteins possess the intrinsic property to fold rapidly and spontaneously to their native structures that are required for their specific functions. How the disordered polypeptide chains manage to navigate through the rough folding landscape and in most cases reach the functional native state within a reasonable amount of time is still a fascinating and challenging subject to current research. Since the native conformation is not the only fold available to a polypeptide chain, but an alternative structure may represent another ground state of protein folding, research into the specific features that favor productive folding into the correct, functional native state over off pathway species that run the risk of aggregation has received increased attention.

Structural characterization of the folding pathway and in addition of the intermediate state of the antibody domain C_L by NMR spectroscopy in Chapter 3 provides insights into these specific features that might guide an unfolded polypeptide chain successfully to its biologically active structure. Furthermore, the detailed picture of the major folding intermediate of the C_L domain allows identification of structural properties that might have evolved over the course of evolution to avoid pathogenic misfolding reactions while preserving an identical protein topology.

Although many proteins fold to their native structures rapidly and spontaneously, recent studies indicate that a surprisingly high number of gene sequences in eukaryotic genomes encode intrinsically disordered proteins. These proteins that lack a well structured three dimensional fold are biologically active, in contrast to the classic notion that proteins require a well defined globular structure to be functional. Due to the persistence of these natively unfolded proteins throughout evolution despite the permanent risk of degradation or causing diseases, these non-globular structures seem to play a pivotal role in cellular biology. In general, intrinsically disordered proteins fold into ordered structures only upon binding to their cellular targets. The mechanism by which folding is coupled to binding remains poorly understood, but it has been hypothesized on theoretical grounds that the disordered protein binds weakly and non-specifically to its target and folds as it approaches the cognate binding site.

NMR spectroscopic investigation on the structural characteristics of the constant domain of the heavy chain, C_H1, of the antibody molecule in Chapter 4 extends the growing class of natively unfolded proteins by a prominent member of the immunoglob-

ulin superfamily. The unstructured nature of the C_H1 domain results in retention of the whole antibody heavy chain in the endoplasmic reticulum (ER). Only upon association with the antibody light chain, the interaction with the C_L domain induces structure formation in the C_H1 domain. A detailed description of this induced folding reveals insights into the mechanism of folding upon binding by which intrinsically disordered proteins perform their diverse biological functions like in the case of antibody molecules the secretion control from the ER.

Intrinsically structured proteins must exhibit a distinct mechanism to fulfill their biological functions and transfer cellular signals. Instead of inducing folding, binding of a ligand induces conformational changes. Allosteric processes involve coupling of conformational changes between two widely separated binding sites. Since the binding of a ligand to one site can affect the other through a propagated change in the protein shape, nature uses this strategy to regulate protein activity.

Chapter 5 studies the role of allosteric regulation in the process of cell migration. Actin filament binding by the focal adhesion associated protein talin stabilizes cell-substrate adhesion. Because increased intracellular pH promotes cell migration, a hallmark of metastatic carcinomas, and in addition actin binding by talin is pH sensitive, focal adhesion remodeling might be increased through lower affinity talin actin binding. NMR spectroscopic investigation of the pH sensitive USH-I/LWEQ module of talin reveals a structural mechanism for pH dependent actin binding, suggesting that focal adhesion turnover is in part allosterically mediated by pH dependent affinity of talin for actin binding.

Chapter 2

Theoretical Aspects

In NMR spectroscopic investigations of proteins and nucleic acids, nuclear spin relaxation is a critical factor for optimizing the set-up of the NMR experiments, provides key data for de novo structure determination, and can also provide a wealth of information on global and intramolecular motions that may be crucial for macromolecules to adapt their structures to particular functions.

In a static magnetic field B_0, the thermal equilibrium state of a statistical ensemble of nuclear spins gives rise to a macroscopic magnetization along B_0 (longitudinal magnetization). The application of an external perturbation, typically one or a sequence of several radio frequency pulses irradiated perpendicular to B_0, generates transverse magnetization in the xy-plane. Subsequently, the longitudinal magnetization aims to recover to the equilibrium state - a process named longitudinal relaxation and the rate at which this process occurs is determined by the longitudinal relaxation time T_1. The disappearance of transverse magnetization by transverse relaxation, with a rate given by the transverse relaxation time T_2, may be governed entirely by reestablishment of the equilibrium state of the longitudinal magnetization with $T_2 = T_1$. However, for proteins and nucleic acids generally additional relaxation processes contribute to the transverse relaxation, resulting in $T_2 < T_1$.

In general, reestablishment of the equilibrium state of the longitudinal magnetization could be achieved by spontaneous and stimulated emission of a photon. Though, the probability W for transition from the higher to the lower energy state of an isolated magnetic dipole by spontaneous emission of a photon of energy $\Delta E = \hbar\omega_0$ is given by[1]

$$W = \frac{\mu_0 \gamma^2 \hbar \omega_0^3}{6\pi c^3} \tag{2.1}$$

in which c is the speed of light. For instance, the probability of an ^{1}H spin with a Larmor frequency of 500 MHz for this transition would be $W \approx 10^{-21}$ s^{-1}. Therefore,

contributions to the overall relaxation rate from transitions between the discrete energy levels of the spin systems by spontaneous emission of photons are negligible. Nuclear spin relaxation can thus entirely be accounted for by suitable coupling of the spin systems with the surroundings which allows the exchange of energy between the spins and the lattice. The lattice is represented by random molecular motions and hence nuclear spin relaxation in solution can be treated as arising from spin interactions modulated by random rotational and translational motions of the molecules in which the spins are located. In longitudinal relaxation, only direct energy exchange (non-adiabatic relaxation processes) between the spin systems and the lattice contributes to the observed magnetization recovery rate. In this case, transition of the spin system from a higher (lower) energy state to a lower (higher) energy state is accompanied by an energy conserving transition of the lattice from a lower (higher) to a higher (lower) energy state. Since the lattice presumably resides in a permanent thermal equilibrium with a larger population in the lower energy state, a transition of the spin system from higher energy to lower energy is more probable. Thus, exchange of energy between the spin system and the lattice brings the spin system into a state of thermal equilibrium in which the populations of the stationary states have the Boltzmann distribution. On the contrary, in transverse relaxation besides direct energy exchange loss of phase coherence by adiabatic relaxation processes also accounts for the decay rate of transverse relaxation. These adiabatic processes do not alter the populations of the states and no energy is exchanged between the spin system and the lattice, as transitions between stationary states do not occur. Rather, fluctuating magnetic fields at the sites of the nuclear spins determine the adiabatic relaxation. Typically, a nuclear spin experiences many different sources of local magnetic fields which fluctuate as the molecule rotates in solution. For instance, neighboring nuclear spins create magnetic fields, that constitute the basis for the dipole-dipole relaxation mechanism. Furthermore, local magnetic fields are also generated by induced electron currents, leading to spin relaxation through chemical shift anisotropy. All of these local fields are modulated by the rotation of the molecule.

From the point of view of physical chemistry, proteins are long chained polymers whose structures are mainly based on weak non-covalent interactions. Due to thermal motions, the resulting structures can thus undergo large fluctuations and sample a range of conformations. These fluctuations include among others collision with solvent molecules, libration of closely packed interior groups, hinge bending motion between elements of well defined secondary structures, rolling and sliding of helices and β-sheets, and concerted fluctuations and jumps of neighboring dihedral angles. Such motions may cover a wide range of amplitudes, energies, and time scales.[2]

A typical sample for NMR experiments with a protein concentration of 1 mM in 500 μL of solvent contains 10^{17} protein molecules performing thermally activated random walks colliding with 10^{22} water molecules.[3] An analytic description of this scenario exceeds by far the scope of current theoretical and computational methods. However, as already evident in the early and formative years of gaining an understanding of NMR phenomena, the nuclear spins couple only weakly with the molecular surroundings.[4] Consequently, basic thermodynamic arguments were sufficient to predict the

behavior of an ensemble of such 'non-interacting' spins. In the meantime, a variety of nuclear spin relaxation theories have been developed to characterize the macroscopic response of spin systems to time dependent external magnetic fields.

In 1946 Bloch introduced the concept of relaxation time into NMR with his phenomenological theory of spin relaxation.[5] This theory assumes that the magnetization along the external magnetic field decays exponentially to the equilibrium state with a time constant T_1 or the spin-lattice or longitudinal relaxation time. The magnetization perpendicular to the external field similarly decays exponentially to an equilibrium value of zero characterized by the spin-spin or transverse relaxation time T_2. Shortly afterwards in 1948, the first microscopic theory of spin relaxation proposed by Bloembergen, Purcell, and Pound (BPP) followed the Bloch equations.[6] The BPP theory relates the spin relaxation times to transition probabilities between nuclear spin energy levels and the lattice corresponds to a random field that interacts with the spins to give rise to the relaxation of these nuclear spins. Most subsequent theories have built on the ideas in this fundamental paper. In 1953 Wangsness and Bloch replaced the existing models with a rigorous quantum mechanical treatment[7] and finally Redfield formulated the foundation of current nuclear relaxation theory for molecules in solution.[8] In this approach the lattice acts as a reservoir of infinite heat capacity persistent in thermal equilibrium. The BWR formulation restrictively holds for fast motion and weak collision limits. In the fast motion limit, the correlation time of the motion τ_c (the correlation time τ_c represents the time it takes for a molecule to reorient by one degree[9]) has to be much shorter than T_1 and T_2; the weak collision limit incorporates that many collisions are required to change the orientation of the interaction tensor. Mostly, the description of macromolecular dynamics in solution satisfies these conditions.

By themselves, nuclear spin relaxation rates are of little interest. Their importance lies in the information they convey about the molecular system. These nuclear spin relaxation rates provide information about the amplitudes and rates of molecular dynamics. Measurements of relaxation rates are based on perturbing the spin systems away from their equilibrium states and then monitoring their return to equilibrium.

2.1 Protein Dynamics

Function in biological systems exquisitely depends on spatial and temporal changes in biomacromolecules. Various biological processes rely on transduction of information through conformational changes in proteins and nucleic acids associated with folding and assembly, ligand binding and molecular recognition, and catalysis. NMR spectroscopy offers the unique ability to investigate dynamic properties of molecules over a range of different time scales with atomic resolution (Fig. 2.1).

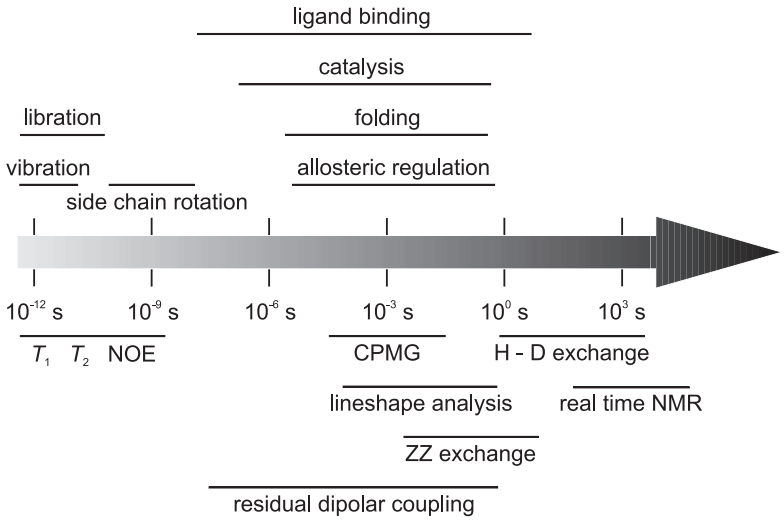


Figure 2.1: Time scales for protein dynamics and NMR techniques. Protein motions and the established NMR techniques for studying these motions span more than 15 orders of magnitude in time scale. Many biologically relevant dynamics in proteins occur on microsecond to millisecond time scales and manifest as chemical exchange line broadening in NMR spectra.

2.1.1 Dynamics in the ps to ns Time Scale

Conformational dynamics on time scales comparable to or faster than the overall rotational correlation times for biomacromolecules influence spin relaxation rate constants by modulating dipole-dipole interactions, chemical shift anisotropy, and quadrupolar Hamiltonians. The measurement of ^{15}N relaxation rates is particularly beneficial to probe backbone motions because the relaxation of these nuclei is governed predominantly by the dipolar interaction with directly bound protons and to a much smaller extent by the chemical shift anisotropy mechanism,[10] whereas the quadrupolar relaxation mechanism does not play any role since this kind of relaxation is restricted to spins with $I > \frac{1}{2}$ such as ^{2}H.[11]

Analysis of rapid backbone motions in the ps to ns time scale requires T_1, T_2, and heteronuclear steady state NOE experiments (Fig. 2.2).[12] T_1 or T_2 values are extracted in a straightforward way by measuring the intensities of cross-peaks as a function of a relaxation delay, T. Fitting these exponential decays reveals the relaxation rates R_1 and R_2, respectively for each analyzed spin. In contrast, the phenomenon of the steady state NOE is based on a rather exceptional effect. Application of a weak radio frequency field at the Larmor frequency of one spin for a sufficiently long time strongly effects the longitudinal magnetization of the non-irradiated spin (Fig. 2.3). This radio frequency field equalizes the populations across the irradiated transitions and after a sufficiently long interval of resonant irradiation the populations settle into steady state values. Saturation of the spin leads to perturbed energy level populations and thus the spin aims to reach the equilibrium Boltzmann distribution again. The necessary spin state transitions are caused by relaxation. Whereas relaxation of the saturated

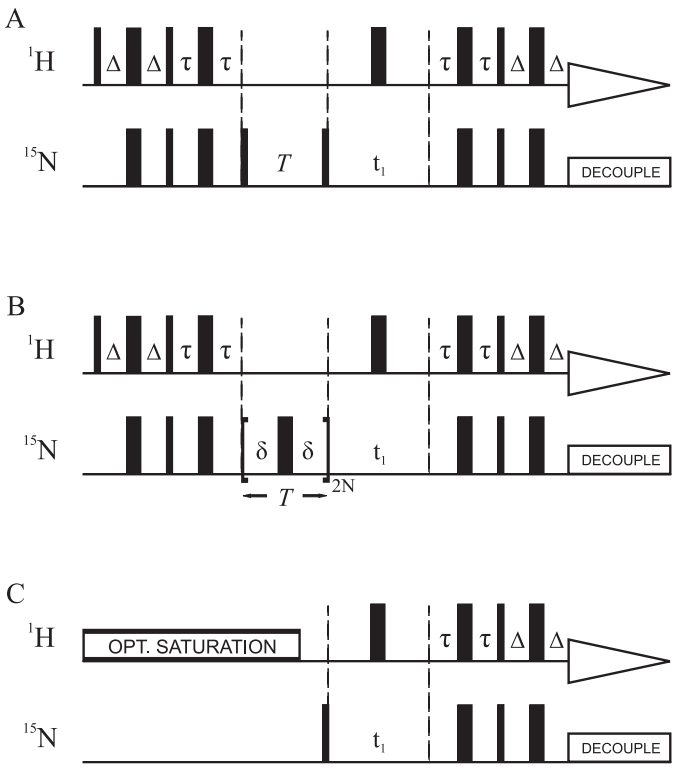


Figure 2.2: Schemes for the measurements of ^{15}N spin (A) T_1, (B) T_2, and (C) ^{1}H–^{15}N NOE values.[12] The delay Δ is adjusted to $1/(4J_{NH})$ as well as the delay τ due to the relatively long transverse relaxation times for backbone ^{15}N spins. In order to measure R_1 or R_2 relaxation rates of the heterospin, two refocused INEPT-type sequences are employed to transfer magnetization from the directly bound proton to the low γ heteronucleus and back to proton for detection. T_1 or T_2 values are extracted in a straightforward way by measuring the intensities of cross-peaks as a function of the relaxation delay, T. Spectra for measuring ^{1}H–^{15}N NOEs are recorded in the presence and absence of ^{1}H saturation. Narrow and wide bars represent 90° and 180° pulses, respectively.

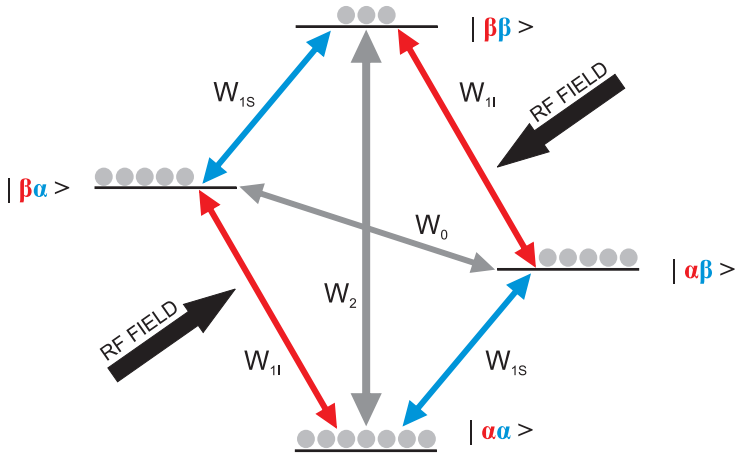


Figure 2.3: Schematic representation of irradiated transitions in the steady state NOE experiment. After a sufficient amount of time without any disturbance, a coupled spin ensemble adopts a state of thermal equilibrium with respect to the molecular environment. In the case of a continuous radio frequency field at the I spin Larmor frequency, this irradiation induces transitions across two pairs of energy levels. The populations across these irradiated transitions equalize after a sufficiently long interval of resonant irradiation. These varied populations also effect the S spin magnetization compared to its thermal equilibrium value by a factor ε_{NOE}. This NOE enhancement factor can be calculated according to $\varepsilon_{NOE} = 1 + \frac{\gamma_I}{\gamma_S}\frac{(W_2 - W_0)}{W_0 + 2W_{1S} + W_2}$, where W_i denotes the transition probability between the energy levels of the associated states.[9]

spin via the W_1 mechanism has no effect on nearby spins, the mechanisms W_0 and W_2 however cause the so far unperturbed spins to deviate from their Boltzmann equilibrium towards different populations of the α and β states (Fig. 2.3). This dipolar cross-relaxation alters the longitudinal magnetization compared to its thermal equilibrium state resulting in an increase or decrease of the magnetization. The change in signal intensity depends on the correlation time and the gyromagnetic ratio of the interacting spins. In the case of very rapid molecular motions the ^{15}N magnetization changes sign when the ^{1}H spins are saturated. Whereas in the case of slow molecular motions the signal intensity of the ^{15}N spins remains more or less equal to the reference experiment without the saturation.

These relaxation data (T_1, T_2, and heteronuclear steady state NOE) contain the information concerning the nature of internal motions that occur in the system under investigation. The simplest possible description of these internal dynamics involves specifying the rate (time scale) and the spatial restriction of the motion of the ^{15}N–H vectors.[13] Because of intrachain and interchain interactions, the vector cannot adopt all possible orientations. Consequently, the backbone atoms experience restricted motions rather than isotropic motions. Lipari and Szabo developed a model free approach that allows complete characterization of the fast internal motions contained in NMR relaxation experiments by two independent quantities, a generalized order parameter, S, which represents the degree of spatial restriction of the motion, and an effective correlation time, τ_i, which describes the rate of reorientation.[13,14]

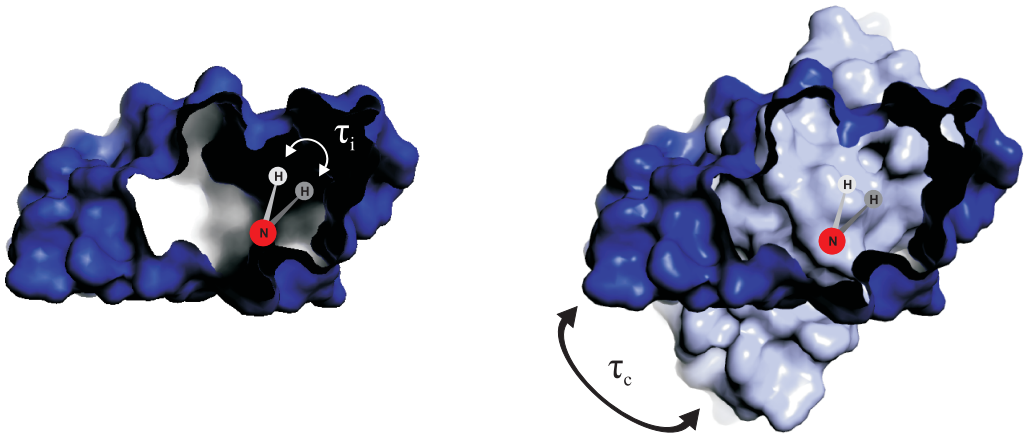


Figure 2.4: Schematic illustration of the internal correlation time, τ_i, and the overall correlation time, τ_c. In both cases the correlation time τ presents the reorientation of the N–H vector as a function of time. However, τ_i indicates the internal dynamics of each single N–H vector within the molecule, whereas it is assumed that the overall tumbling can be characterized by a single τ_c value for all N–H vectors located within the molecule.

Fluctuations of the ^{15}N–H vectors with respect to the external magnetic field determine the relaxation of the nitrogen nucleus. This bond vector samples a variety of distinct orientations over the time, a motional process that is mathematically stated by a time dependent correlation function. The correlation function holds for spherical molecules and depends on the rotational correlation time (τ_c), a parameter that characterizes the rate of reorientation of the bond vector. Typically, large molecules exhibit sluggish motions and hence their correlation time is rather long as compared to small molecules that rotate quite fast. Assumed that the overall motion of a macromolecule can be described by a single correlation time and that the overall and internal motions are independent (Fig. 2.4), the total correlation function can be quantified as

$$C(t) = C_o(t)C_i(t) \tag{2.2}$$

where the correlation function for the overall motion is

$$C_o(t) = \frac{1}{5}\exp^{-t/\tau_c}. \tag{2.3}$$

Assuming a spherical molecule, the simplest theoretical approach calculates the isotropic rotational correlation time from Stokes' law

$$\tau_c = \frac{4\pi\eta_w r_{hydro}^3}{3k_B T} \tag{2.4}$$

with the viscosity of the solvent η_w, the hydrodynamic radius of the molecule r_{hydro}, the Boltzmann constant k_B, and the temperature T. The internal correlation function

can be approximated by

$$C_i(t) = S^2 + \left(1 - S^2\right) \exp^{-t/\tau_i} \tag{2.5}$$

where τ_i is the effective correlation time and S^2 the generalized order parameter. A combination of equation 2.3 and 2.5 gives the total correlation function as

$$C(t) = \frac{1}{5} S^2 \exp^{-t/\tau_c} + \frac{1}{5} \left(1 - S^2\right) \exp^{-t/\tau} \tag{2.6}$$

with

$$\tau^{-1} = \tau_c^{-1} + \tau_i^{-1} \tag{2.7}$$

and subsequently Fourier transformation results in the corresponding spectral density function

$$J(\omega) = \frac{2}{5} \left[\frac{S^2 \tau_c}{1 + (\omega\tau_c)^2} + \frac{(1 - S^2)\tau}{1 + (\omega\tau)^2} \right]. \tag{2.8}$$

Although this simple two parameter model free approach emerged as remarkably successful in the interpretation of the relaxation data (T_1, T_2, and HetNOE), occasional discrepancy in the attempt to determine a single parameter set (S^2, τ_i) for certain spins indicates that the evaluation of these relaxation experiments must additionally account for slow components.[15] Hence, the correlation function for the internal motions has to be expanded to accommodate at least two parameters, the slow correlation time τ_s in the order of 1-3 ns that is still faster than the overall rotational correlation time (around 10 ns for medium size proteins in aqueous solution at room temperature) and the fast correlation time τ_f that reflects the fast random motions. Assuming that the fast internal motions are independent of the slow motions and differ by at least one order of magnitude, the total generalized order parameter can be decomposed as

$$S^2 = S_f^2 S_s^2. \tag{2.9}$$

Furthermore, since in the majority of cases τ_f is sufficiently small resulting in negligible contribution to the relaxation parameters, the spectral density function can be described by

$$J(\omega) = \frac{2}{5} \left[\frac{S^2 \tau_c}{1 + (\omega\tau_c)^2} + \frac{(S_f^2 - S^2)\tau_s'}{1 + (\omega\tau_s')^2} \right] \tag{2.10}$$

with

$$\tau_s' = \frac{\tau_s \tau_c}{\tau_s + \tau C}. \tag{2.11}$$

The T_1 and T_2 relaxation times and the NOE enhancement of an amide ^{15}N spin, taking only dipolar and chemical shift anisotropy interactions as sources of relaxation into account, are related to the spectral density function by[16]

$$R_1 = T_1^{-1} = \frac{d^2}{4} \left[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega N) \right] + c^2 J(\omega_N) \tag{2.12}$$

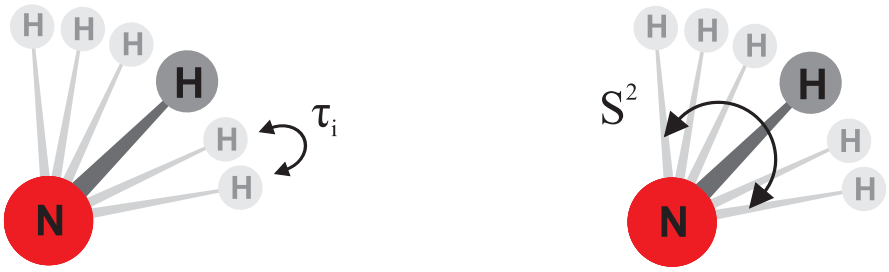


Figure 2.5: Schematic representation of the squared order parameter S^2 and the internal correlation time τ_i. According to the model free approach, internal dynamics of the molecule on the ps to ns time scale can be characterized by the spatial restriction (S^2) and the rate (τ_i) of the motion of the N–H vector.

$$R_2 = T_2^{-1} = \frac{d^2}{8}[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)] + \frac{c^2}{6}[4J(0) + 3J(\omega_N)] \tag{2.13}$$

$$NOE = 1 + \frac{d^2}{4R_1}\frac{\gamma_N}{\gamma_H}[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)] \tag{2.14}$$

in which $d = \mu_0 h \gamma_N \gamma_H r_{NH}^{-3}/(8\pi^2)$, $c = \omega_N \Delta\sigma/\sqrt{3}$, μ_0 is the permeability of vacuum; h is the Planck's constant; γ_H and γ_N are the gyromagnetic ratios of ^{1}H and the ^{15}N spin, respectively; r_{NH} is the N–H bond length; ω_H and ω_N are the Larmor frequencies of ^{1}H and the ^{15}N spin, respectively; $J(\omega_H)$ and $J(\omega_N)$ are the spectral density functions of ^{1}H and the ^{15}N spin, respectively; and $\Delta\sigma$ is the chemical shift anisotropy of the ^{15}N spin (assuming an axially symmetric chemical shift tensor).

Based on these equations, software packages like the program Modelfree by Arthur Palmer analyze the relaxation data by fitting the extended model free spectral density function to the experimental NMR spin relaxation data.[17,18] These simulated spectral density functions are optimized by varying the model free parameters, the generalized order parameter S^2 and the internal correlation time τ_i. These parameters provide information about the spatial restriction of the motion of the bond vector (S^2) and also about the rate of reorientation of this vector (τ_i) for each analyzed spin (Fig. 2.5). S^2 adopts values between 0 for isotropic internal motions according to equally probable orientations through the entire conformational space and 1 for completely restricted motions.

2.1.2 Dynamics in the μs to ms Time Scale

Molecular motions on time scales slower than the overall rotational correlation times (τ_c) for biomacromolecules are of particular interest because functionally important biological processes, including enzyme catalysis, signal transduction, ligand binding, and allosteric regulation occur in this time regime.[19] Motions on μs - ms time scales cause spin relaxation resulting from the modulation of isotropic chemical shifts by

altering the magnetic environments of the spins. The phenomenon of chemical exchange refers to any process in which the spins experience distinct magnetic environments.[20] Hence, chemical exchange provides information on conformational and kinetic processes in the time range slower than the overall molecular tumbling. The main experimental techniques for quantifying chemical exchange are longitudinal magnetization exchange, line shape analysis, and CPMG relaxation dispersion.

The observation of two or more resonance signals for a single nuclear spin in the NMR spectrum of a protein and subsequent confirmation with ZZ-exchange spectroscopy establish the existence of chemical exchange on the slow exchange time scale.[21] Conversely, studying slow chemical exchange processes by monitoring the exchange of longitudinal magnetization between distinct sites requires sufficient population of the minor sites of the exchange process in order to generate observable resonance signals. Furthermore, the exchange rate k_{ex} must not be much less than the longitudinal relaxation rate constant in the absence of chemical exchange, otherwise the signals decay due to relaxation faster than population transfer. However in many cases, owing to skewed population distributions and selective line broadening of resonances for the minor population, only one resonance corresponding to the major population is observed even for slow chemical exchange between multiple conformations or distinct magnetic environments.[22,23] Therefore, for such sites in slow exchange and all sites in fast exchange experimental techniques are based on determining the contribution of chemical exchange to the resonance linewidth or the transverse relaxation rate constant. Although resonance signals of sites experiencing chemical exchange display weakened intensity and broadened lineshapes, other factors such as amide proton exchange with the solvent and anisotropic global tumbling motion also reduce peak intensity and cause broadened lineshapes. Hence, chemical exchange is commonly characterized by measuring transverse relaxation rate constants rather than lineshapes. Sites in proteins that undergo chemical exchange can be identified by quantifying the chemical exchange contribution R_{ex} to the transverse relaxation rate constant R_2. This chemical exchange contribution R_{ex} is given by

$$R_{ex} = R_2 - R_2^0 \tag{2.15}$$

where R_2^0 represents the relaxation rate constant in the absence of chemical exchange, due to dipole-dipole, chemical shift anisotropy, and quadrupole relaxation mechanisms.

Exchange in the μs - ms time regime results in enhanced transverse relaxation rates. This increase in the observed R_2 is caused by altered precessional frequencies of the nuclear spin associated with the chemical exchange between distinct sites and as a consequence the spin coherence dephases. Relaxation compensated Carr-Purcell-Meiboom-Gill (CPMG) sequences have proven particularly suitable in characterizing protein dynamics on these time scales, as the implemented spin echo pulse trains can attenuate this kind of coherence loss.[24,25] In the case of a fast CPMG pulsing rate corresponding to a short time delay between the spin echo refocusing pulses, τ_{cp}, relative to the mean lifetime of the exchange event, the chemical exchange has only a negligible effect on the coherence dephasing and thus $R_2 \sim R_2^0$. Due to the different time regimes, these

relaxation mechanisms on the ps - ns time scale still proceed. However, in the case the exchange events occur on the time scale of or faster than the repetition rate of the 180° pulses realized by a slow CPMG pulsing rate, the induced dephasing significantly contributes to coherence loss which manifests in an elevated R_2.

In a two site exchange mechanism, a nuclear spin exchanges between sites a and b that differ in chemical shift by $\Delta\omega = \Omega_a - \Omega_b$. The kinetics of the exchange reaction are described by the apparent exchange rate constant $k_{ex} = k_1 + k_{-1}$, where k_1 and k_{-1} are the first order rate constants for the forward and reverse transitions, respectively. In the limit of fast exchange, $k_{ex} \gg \Delta\omega$, the dependence of the measured transverse relaxation rate on the time delay between the spin echo refocusing pulses, τ_{cp}, can be described by[26]

$$R_2(1/\tau_{cp}) = R_2^0 + \frac{\Phi_{ex}}{k_{ex}} \left[\frac{1 - 2 \tanh(k_{ex}\tau_{cp} \ / \ 2)}{k_{ex}\tau_{cp}} \right] \tag{2.16}$$

with $\Phi_{ex} = \Delta\omega^2 p_a p_b$ in which p_i is the fractional population of the spin in site i and $p_a + p_b = 1$. Alternatively to this fast limit equation, a completely general, albeit more complex expression for $R_2(1/\tau_{cp})$ is given by

$$R_2(1/\tau_{cp}) = \frac{1}{2} \left(R_{2a} + R_{2b} + k_{ex} - \frac{1}{\tau_{cp}} \cosh^{-1} \left[D_+ \cosh(\eta_+) - D_- \cos(\eta_-) \right] \right) \tag{2.17}$$

where

$$D_\pm = \frac{1}{2} \left[\pm 1 + \frac{(\Psi + 2\Delta\omega^2)}{(\Psi^2 + \zeta^2)^{\frac{1}{2}}} \right] \tag{2.18}$$

$$\eta_\pm = \frac{\tau_{cp}}{\sqrt{2}} \left[\pm\Psi + (\Psi^2 + \zeta^2)^{\frac{1}{2}} \right]^{\frac{1}{2}} \tag{2.19}$$

$$\Psi = (R_{2a} - R_{2b} - p_a k_{ex} + p_b k_{ex})^2 - \Delta\omega^2 + 4 p_a p_b k_{ex}^2 \tag{2.20}$$

$$\zeta = 2\Delta\omega(R_{2a} - R_{2b} - p_a k_{ex} + p_b k_{ex}). \tag{2.21}$$

Typically, these CPMG pulse sequences are implemented in a constant time manner, meaning a constant relaxation delay T_{cp} comprises a varying number of spin echo refocusing pulses equivalent to varying time delays, τ_{cp}, between these 180° pulses. The decay of $R_2(1/\tau_{cp})$ as a function of $1/\tau_{cp}$ can be fitted to equation 2.17 by various software packages like Cpmgfit by Arthur Palmer.[28] Nuclear spins that display no dispersive behavior experience no chemical exchange, whereas chemically exchanging residues exhibit well defined relaxation dispersion profiles which enable the extraction of the kinetic parameter k_{ex} in addition to the thermodynamic parameters p_a, p_b, and $\Delta\omega$. However, single field measurements of these relaxation dispersion profiles can yield a significant probability of erroneous parameters with a large deviation from physical reality.[27] As the relaxation curves depend on the static magnetic field, acquisition of a second data set at a different magnetic field strength increases the number of data for each nuclear spin and hence remarkably improves the analysis of the exchange process.

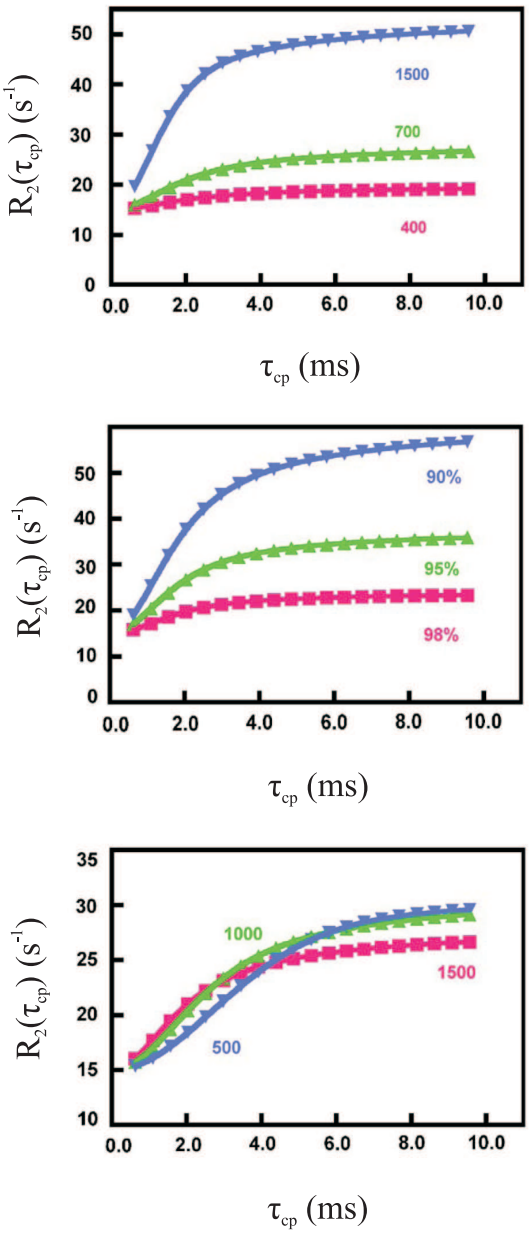


Figure 2.6: Parameter dependence of relaxation dispersion curves.[27] Top: Dependence of $R_2(\tau_{cp})$ on $\Delta\omega$, with $\Delta\omega = 400$ s^{-1} (red), $\Delta\omega = 700$ s^{-1} (green), $\Delta\omega = 1500$ s^{-1} (blue), and $k_{ex} = 1500$ s^{-1}, $p_a = 0.95$, and $R_2^0 = 15.0$ s^{-1}. Middle: Influence of fractional conformer population on $R_2(\tau_{cp})$, where $p_a = 0.90$ (blue), $p_a = 0.95$ (green), $p_a = 0.98$ (red), and $k_{ex} = 1500$ s^{-1}, $R_2^0 = 15.0$ s^{-1}, and $\Delta\omega = 1000$ s^{-1}. Bottom: Effect of k_{ex} on $R_2(\tau_{cp})$, with $k_{ex} = 500$ s^{-1} (blue), $k_{ex} = 1000$ s^{-1} (green), $k_{ex} = 1500$ s^{-1} (red), and $\Delta\omega = 700$ s^{-1}, $p_a = 0.95$, and $R_2^0 = 15.0$ s^{-1}. In this case of a two site exchange, the magnitude of $R_2(\tau_{cp})$ is primarily determined by $\Delta\omega$ and p_a, whereas the kinetics of exchange (k_{ex}) have a more subtle effect on the NMR data. However, since all of these parameters contribute to the relaxation dispersion profile of each spin and as equation 2.17 clearly indicates, analysis of these CPMG curves is quite complex.

This field dependence is predominantly caused by $\Delta\omega$ that can just be scaled to the distinct magnetic fields due to its linear field dependence. Additionally, R_2^0 shows a small field dependence owing to the ^{15}N chemical shift anisotropy.[27] Thus, fitting the various relaxation dispersion profiles acquired at two different static magnetic fields allows an unambiguous estimation of the kinetic and thermodynamic parameters. Although the majority of exchange processes justifies the assumption of a two site exchange mechanism, some exchange events require a three site mechanism. Therefore, established software packages also offer the possibility of analyzing the dispersion data by a linear three site exchange model.

In the case a nuclear spin exchanges between distinct sites, the contribution of the chemical exchange to the transverse relaxation rate constant is of course also present in the analysis of the dynamics in the ps - ns time scale. Hence, the software Modelfree accounts for the chemical exchange, but also for all other dynamical processes on the slow time scale, by including an additional term into the algorithm if the simulated relaxation data differ from the experimental ones. This exchange broadening parameter is scaled quadratically with respect to the static magnetic field provided that data for more than one field are available. Since relaxation of protonated ^{15}N heteronuclei is dominated by dipolar interactions with the directly bonded ^{1}H spins, fast internal motions monitored by ^{15}N–^{1}H relaxation experiments are approximately independent of the static magnetic field strengths, thus observed variations in the relaxation data on the fast time scale at different magnetic fields indicate slow internal motions.

2.2 Amide Proton Exchange

Characterization of conformational dynamics in biological macromolecules is not necessarily restricted to the investigation of spin relaxation, but can also be studied by amide proton exchange and residual dipolar coupling constants.

Measurement of the hydrogen exchange rates additionally provides information about the dynamical features of the protein. Whereas the classical H–D exchange cannot resolve rates faster than the time required for the acquisition of a two dimensional NMR spectrum, the so-called Mexico pulse sequence enables the determination of fast amide proton exchange rates in the range of milliseconds to seconds.[29] This experiment takes advantage of quantifying build-up rates of signals that are caused by the exchange process itself. For this purpose, all contributions from the natural Boltzmann magnetization of the amide protons have to be filtered out prior to the acquisition of the actual experiment. After quantitative suppression of all protons attached to nitrogen and carbon, the water protons are turned back onto the z axis and during the subsequent mixing time exchange of z magnetization from the water protons to the amide protons occurs. The build-up of z magnetization of the amide protons depends on their exchange rate with the water protons and also on the mixing time of the pulse sequence. Finally, the established z magnetization of the amide protons can be detected by a standard HSQC experiment. Analyzing the peak intensity as

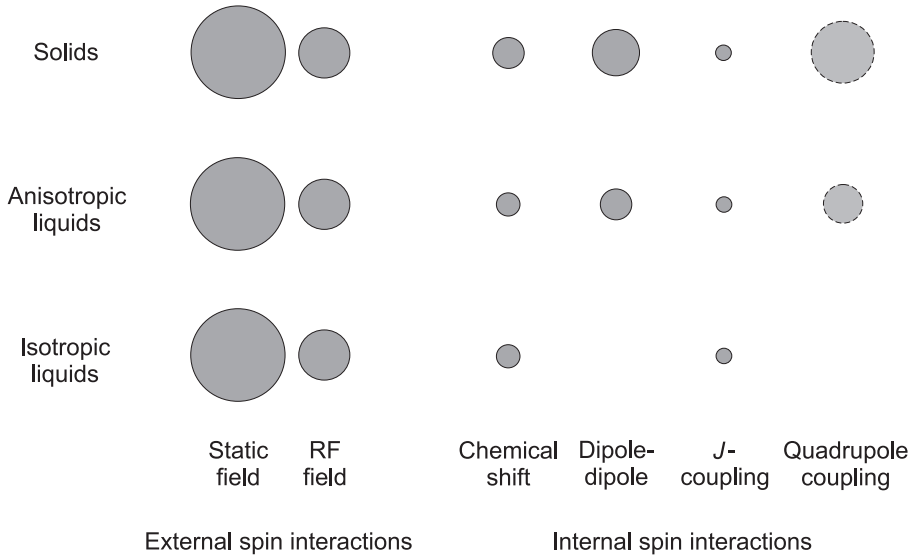


Figure 2.7: Summary of the motionally averaged spin interactions and their rough relative magnitudes in different phases of matter. In the special case of isotropic solution the largest internal terms average to zero. In addition, the quadrupolar coupling always vanishes for spin-$\frac{1}{2}$ nuclei independent of the phase of matter.

a function of the mixing time reveals the exchange rate k_{ex}. In order to minimize NOE effects that will also occur during the mixing time, both the nitrogen and the carbon bound protons have to be saturated prior to the mixing time. Appropriate pulse schemes are able to filter nitrogen and carbon bound protons,[30] however this clearly requires uniformly ^{15}N and ^{13}C labeled protein samples. For only ^{15}N labeled proteins, the NOE between amide protons and carbon bound protons also contributes to the resonance signal intensity and hence leads to inaccurate exchange rates.

2.3 Residual Dipolar Couplings

A nuclear spin experiences various electric and magnetic fields of distinct origin. Spin interactions of external origin correspond to the static magnetic field and the applied radio frequency field. Spin interactions of internal origin include indirect magnetic interaction of the external magnetic field and the nuclear spin through the involvement of electrons (chemical shift term), electric interaction of the nuclear spin with the surrounding electric fields (quadrupolar coupling), direct magnetic interactions of nuclear spins with each other (direct dipole-dipole coupling), indirect magnetic interactions of nuclear spins with each other through the involvement of electrons (J coupling) (Fig. 2.7). The dipole-dipole coupling does not involve any electrons and might be either intramolecular or intermolecular. The magnitude of this through space interaction

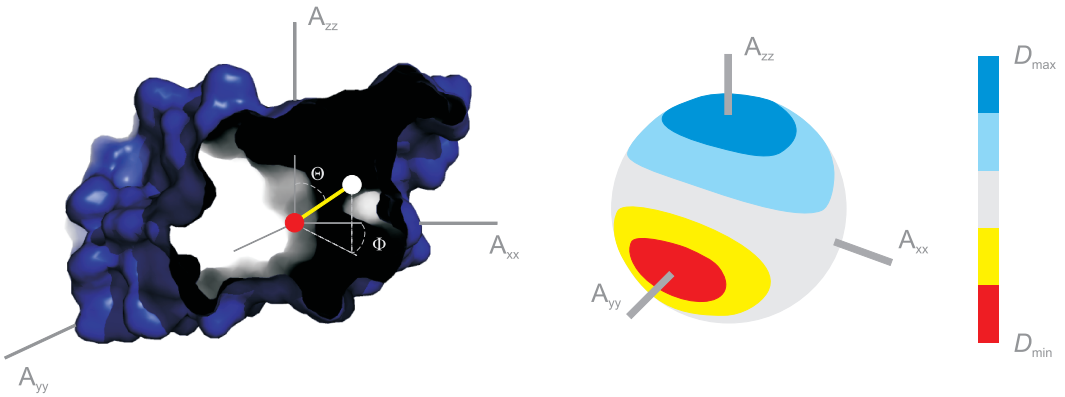


Figure 2.8: Schematic representation of the dependence of RDC values on the orientation of the internuclear vector and the alignment tensor. The measured coupling can be described in terms of the orientation of the interspin vector by the angles Θ and Φ in the frame of the alignment tensor, given by A_{xx}, A_{yy}, A_{zz}. These orientations of the interdipolar vector are plotted on the surface as a function of the range of couplings.

is given by the dipole-dipole coupling constant

$$D = -\frac{\gamma_I \gamma_S \hbar \mu_0}{8\pi^2 r^3} \left(3\cos^2\Theta - 1\right) \tag{2.22}$$

where Θ represents the angle between the vector joining the interacting spins and the external magnetic field and r the spin-spin distance. Apart from these parameters the coupling constant only depends on physical constants: the gyromagnetic ratios γ_I and γ_S of spin I and S, respectively, the Planck constant $\hbar$, and the permeability of vacuum μ_0. This dipole-dipole interaction decreases with the inverse cube of the internuclear distance and scales linearly with the gyromagnetic ratio of the interacting spins. Furthermore, it depends on the molecular orientation. According to equation 2.22 the dipole-dipole coupling constant has opposite signs for spin pairs aligned along the field ($\Theta = 0$) compared to spin pairs oriented perpendicular to the field ($\Theta = \pi/2$) and equals to zero at an angle of $\approx 54.74°$. In an isotropic liquid intramolecular dipole-dipole couplings average to zero, due to equal probabilities of all molecular orientations. Similarly, the short range intermolecular dipole-dipole couplings average to zero, because of the translational motion of the molecule. Although the long range intermolecular dipole-dipole couplings survive the motional averaging, these couplings are negligible. Even though the internuclear magnetic dipole couplings average out completely in isotropic solution, their effect on nuclear spin relaxation results in measurable NOEs. In contrast, in an anisotropic liquid the intramolecular dipole-dipole couplings do not average out due to the preferential molecular orientation. Whereas the intermolecular interactions behave the same, owing to the still existing translational motion of the molecules.

The non-isotropic orientational averaging of the molecule can be described in terms of an alignment tensor. The orientation of the alignment tensor with respect to the coordinate frame of the molecule can in return be defined simply via a three dimensional Euler rotation $R(\alpha, \beta, \gamma)$. The measured dipolar coupling corresponds to the orientation of the internuclear spin vector (specified by the angles Θ and Φ) in the frame of the alignment tensor, with its values A_{xx}, A_{yy}, and A_{zz} (Fig. 2.8). These values reflect the deviation from the isotropic probability distribution of the x, y, and z axes of the alignment tensor frame to be parallel to B_0 ($A_{xx} + A_{yy} + A_{zz} = 0$). For instance, for an isotropically reorienting molecule the probabilities would be $A_{xx} = A_{yy} = A_{zz} = 0$ and for a fully aligned molecule $A_{xx} = A_{yy} = -\frac{1}{3}$, $A_{zz} = \frac{2}{3}$; where by convention $| A_{xx} | \leq | A_{yy} | \leq | A_{zz} |$. Therefore, the equation for the dipolar coupling constant in a weakly anisotropic solution has to be rewritten as

$$D(\Theta, \Phi) = D_a \left[(3 \cos^2\Theta - 1) + R \sin^2\Theta \cos 2\Phi \right] \tag{2.23}$$

where $D_a = D_{max} A_{zz}$ with $D_{max} = -\frac{\gamma_I \gamma_S \hbar \mu_0}{8\pi^2 r^3}$ and $R = \frac{A_{xx} - A_{yy}}{A_{zz}}$. D_a refers to the magnitude of the dipolar coupling tensor, commonly normalized to the N–H dipolar interaction, and R represents the rhombicity. Altogether, five parameters characterize the alignment tensor (D_a, R, α, β, γ). These parameters can be determined directly via singular value decomposition from a set of at least five RDCs for any known structure.

In order to permit facile measurement of dipolar interactions it is essential that these interactions are still averaged to a very small fraction (typically $\sim 10^{-3}$) of their static value.[31] Several alignment media such as bicelles,[31] filamentous phage Pf1,[32] and polyacrylamide gels[33] enable the partial orientation of macromolecules to a relatively small degree. This weak alignment still allows the molecules to diffuse freely resulting in narrow resonances in the spectra, but on the other hand limits the measurement of dipolar couplings to the largest interactions, such as one bond ^{15}N–^{1}H, ^{13}C–^{1}H, ^{13}C–^{13}C. For directly bonded pairs of atoms, the internuclear distance is accurately known and hence the measured dipolar coupling provides information on the relative orientation of the internuclear bond vector.

Since the establishment of these alignment media that permit the measurement of dipolar couplings while retaining high quality spectra, residual dipolar couplings have evolved into an inherent part in structural and dynamic characterization of biomolecules in solution. The provided information on the biomolecular conformation is incorporated into structure refinement, fold determination, and quaternary structure determination of proteins. In addition, in terms of molecular dynamics residual dipolar couplings have also proven to be of particular benefit. RDCs are averages over all interspin orientations of the dipolar interaction that are sampled up to the time scale defined by the inverse of the alignment induced coupling and consequently report on averages over relatively long time scales (up to and beyond the millisecond range). Measurement of these couplings therefore encodes key information for understanding protein motions on time scales that are highly complementary to the dynamic picture derived from spin relaxation experiments. Although it is in principle possible to derive this valuable

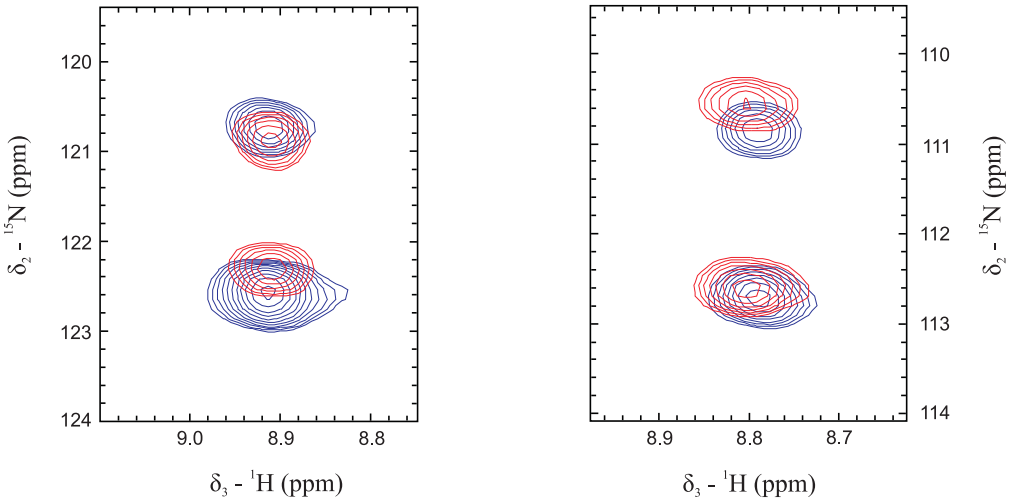


Figure 2.9: Extracting dipolar coupling constants by comparing line splittings in isotropic (blue) and anisotropic (red) solution. These splittings display representative examples for positive and negative residual dipolar couplings corresponding to an orientation rather along the z axis of the alignment tensor frame and respectively rather perpendicular to this axis. In this case, the partial alignment of the USH-I/LWEQ domain of talin was achieved by non-ionic liquid crystalline media and RDC values were extracted from IPAP-HNCO experiments without decoupling during acquisition.

information and several studies demonstrate its remarkable potential, the other side of the coin is the immense amount of data needed for this comprehensive analysis.[34–37] Figure 2.8 indicates that a certain dipolar coupling corresponds to an entire cone of Θ,Φ solutions. Whereas for applications like structure refinement the identification of this cone is already sufficient to validate the determined structure, the characterization of the motional averaging of each individual vector requires precise knowledge of its orientation. This detailed description of the vector orientation necessitates five independent RDC values, corresponding to five independent alignment media. Because of the impossible prediction of linearly independent alignment media, this dynamic analysis represents a quite extensive method to characterize backbone motions on the ms time scale, even though protein dynamics in this time regime are of particular importance for biologically relevant processes.

Chapter 3

The Antibody Domain C_L Explains Differences in Ig Amyloidogenicity

Prerequisite for the astonishing diversity and selectivity in biochemical processes of living systems is the capacity of even the most intricate molecular structures to self assemble with remarkable precision. During evolution proteins might be predominantly optimized for functional purposes rather than for most effective folding which compromises both the speed of folding and the native state stability.[38] This evolutionary pressure for function over folding increases the risk of off-pathway events such as protein misfolding and aggregation. In general, all proteins transiently populate partially folded species on their pathway to the native state suggesting that these intermediates enhance the folding efficiency considerably.[39] However, impaired speed of folding leads to persistently populated intermediates that are particularly susceptible to misfolding and misassembly reactions. Since the properties of the folding intermediate essentially determine whether a protein folds robustly or has the tendency to misfold, structural characterization of folding intermediates and thus understanding of protein folding in general is one of the great challenges of modern science. Yet, due to their transient nature, folding intermediates are poorly accessible to high resolution techniques.

The intrinsically slow folding reaction of the antibody domain C_L enables a detailed investigation of its major folding intermediate. Furthermore, trapping the intermediate in equilibrium by a single point mutation allowed its structural characterization at atomic resolution. NMR experiments in combination with molecular dynamics simulations provide a precise picture of the major C_L folding intermediate. For the core region of this protein, the typical β-barrel topology is well established, yet some edge strands are distorted and hence adopt a variety of conformations. Surprisingly, two short strand connecting helices conserved in constant antibody domains assume their completely native structure already in the intermediate. Thus, these helical elements might

provide a scaffold for adjacent strands and assist in the formation of the hydrophobic core. Since the constant antibody domains have never been directly associated with amyloidogenic diseases, these helices might prevent harmful misfolding reactions. Interestingly, the variable antibody domain (V_L) and β_2-microglobulin (β_2m) are highly homologous members of the same superfamily but their partially folded intermediate states are prone to amyloid formation. Remarkably, neither amyloidogenic protein possesses the short strand connecting helices. To verify the theory that these helical elements promote robust folding, the unstructured loop regions in β_2m were exchanged against the corresponding helical elements of the C_L domain. Indeed, this modification significantly reduces the propensity of β_2m to form amyloids. Thus, a high degree of local structuring in a protein folding intermediate can shape the folding landscape decisively to favor either folding or misfolding. Furthermore, minor structural differences between members of the same protein superfamily can lead to avoidance of pathogenic misfolding reactions while preserving an identical protein topology.

3.1 Introduction

3.1.1 The Antibody Molecule Immunoglobulin G

Already more than 100 years ago Paul Ehrlich proposed a remarkably modern model for an antibody molecule.[40] According to this model, antibody molecules can be described as highly flexible receptors located at the surface of the cell. In the case these receptors recognize any toxic molecule, they bind the pathogen and initiate the production of additional antibody molecules. The cell only generates the type of receptor that is specific for the bound toxic molecule and transports the antibodies to the extracellular matrix. As the free receptors still possess the capacity to bind the pathogen, they are able to protect the cell against any further attack. This model also includes the concept of complementarity in antibody-antigen recognition, consistent with the 'lock and key' fit proposed by Fischer for enzymes.[41] In the last century notable effort has been made to gain more precise insights into the molecular mechanism of the immune response resulting in a detailed knowledge of function and structure of antibody molecules.

To effectively fight the wide range of pathogens, lymphocytes have evolved to recognize a great variety of different antigens from bacteria, viruses, and other disease causing organisms. Two types of lymphocytes exist, B cells and T cells, each with quite different roles in the immune system and distinct antigen receptors. The antigen recognizing molecules of B cells are immunoglobulins (Ig). Each B cell produces immunoglobulin molecules of a single specificity which are bound to the membrane on the cell surface and serve as a receptor for antigen recognition. After encounter with an antigen, B cells differentiate into antibody secreting plasma cells. The antigen recognizing molecules of T cells are solely membrane bound proteins. These T cell receptors are related to immunoglobulins, however in contrast to the B cell receptors they recognize short peptide fragments of the antigens which are often buried within

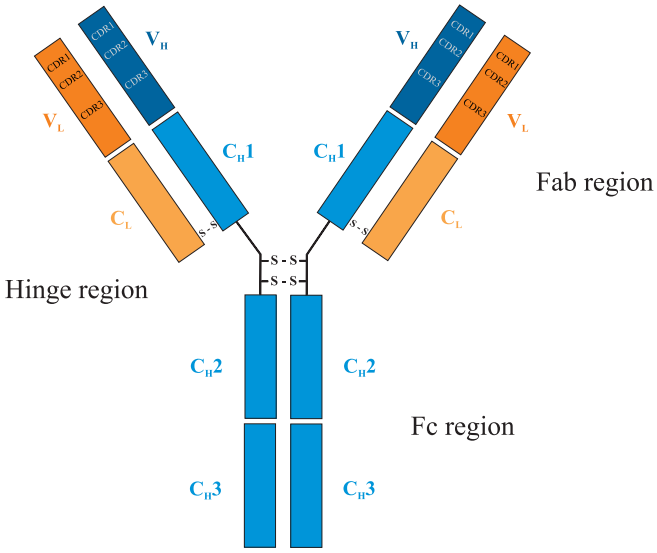


Figure 3.1: Schematic illustration of an antibody molecule. The Y shaped immunoglobulin G consists of two light chains (depicted in orange) and two heavy chains (colored in blue). Recognition of antigens takes place in the Fab region whereas the Fc region engages various elimination mechanisms.

the antigen molecule. Infected cells degrade foreign antigens and subsequently present the resulting peptides at the cell surface by specialized glycoproteins - the MHC (major histocompatibility complex) molecules. In this case it is a complex of the degraded peptide fragment and the MHC molecule on the surface of the cells that is recognized by the T cell receptors. After encounter with the antigen peptide fragment, T cells differentiate into one of several different functional types of effector T cells. These activated T cells mainly focus on killing of cells which are infected with viruses or other intracellular pathogens, activation of antigen stimulated B cells to differentiate and produce antibodies, and regulation of the activity of other lymphocytes to control immune response. The secreted antibody molecule from the B cells has to fulfill distinct functions, binding specifically to molecules from the pathogen and recruiting other cells and molecules that destroy the pathogen once the antibody is bound to it. Recognition and effector functions are structurally separated in the antibody molecule (Fig. 3.1). The antigen binding region varies extensively between antibody molecules and is therefore known as the variable region which allows each antibody to bind a different specific antigen. Altogether the total repertoire of antibodies of a single individual is large enough to ensure that virtually any structure can be recognized. The part of the antibody molecule that engages the various effector mechanisms is relatively conserved and hence termed the constant region. The different effector functions are defined by the structure of the heavy chains of the antibody molecule. According to the number of heavy chain domains, the amount and location of disulfide bonds, and the distribution of linked carbohydrate groups, antibodies can be divided into five major classes: immunoglobulin M (IgM), immunoglobulin D (IgD), immunoglobulin G

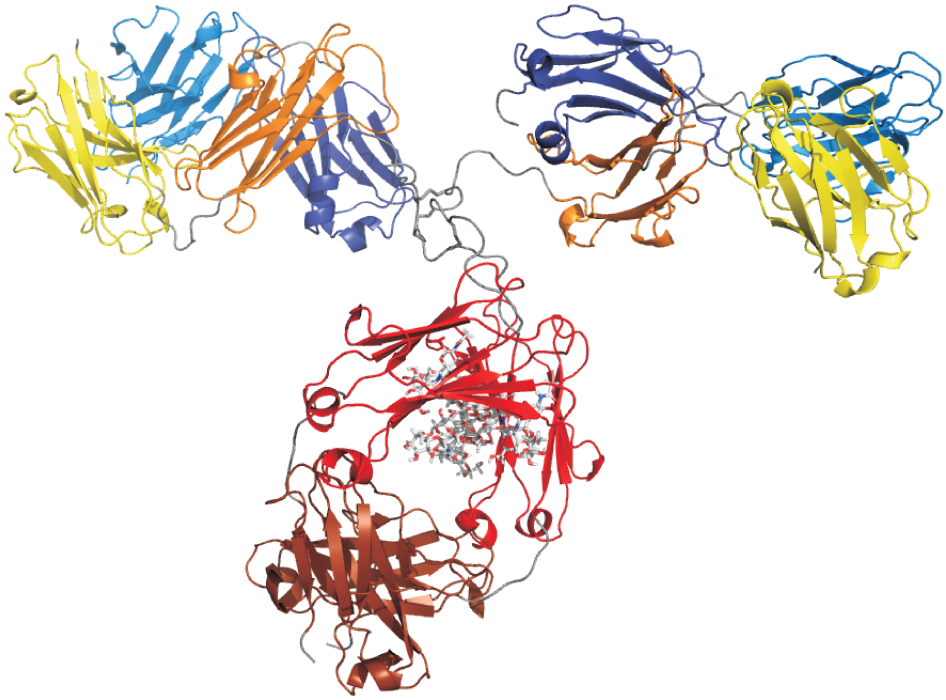


Figure 3.2: Cartoon representation of immunoglobulin G (pdb 1igt). This heterotetrameric glycoprotein is made up of a series of immunoglobulin domains. C_H3 (brown), C_H2 (red), C_H1 (orange), and V_H (yellow) form the heavy chain and C_L (dark blue) and V_L (light blue) represent the light chain.

(IgG), immunoglobulin A (IgA), and immunoglobulin E (IgE).[42] Immunoglobulin G is the predominant antibody in the human serum and the most abundant therapeutic antibody.

The immunoglobulin G molecule is a heterotetrameric glycoprotein assembled from two heavy chains and two light chains[43] which are covalently linked via disulfide bridges.[44,45] Each chain is composed of a series of structurally almost identical compact immunoglobulin domains. The immunoglobulin domains belonging to the constant region of the antibody molecule are named constant domain of the light chain C_L and constant domain of the heavy chain C_H1, C_H2, C_H3, respectively. Whereas the immunoglobulin domains which are part of the variable region of the antibody molecule are called variable domain of the light chain V_L and variable domain of the heavy chain V_H, respectively. The three dimensional structure of this antibody molecule adopts a Y shape and consists of three independent protein moieties which are connected through a flexible linker or hinge region (Fig. 3.2).[46–49] Two of these moieties are identical and include the antigen specific binding site, the third comprises interaction sites for ligands that activate the distinct effector mechanisms.[50,51] The hinge region enables segmental flexibility, hence each Fab region can bind an antigen while the Fc region remains accessible to effector ligands.[52]

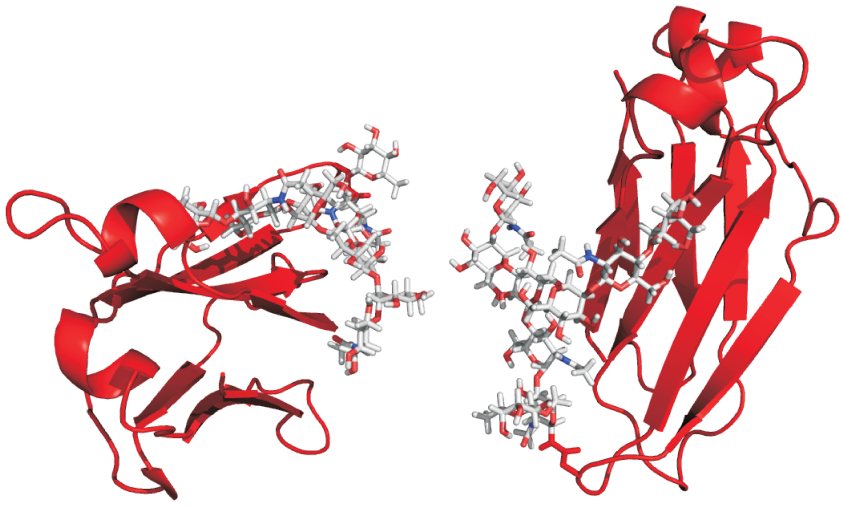


Figure 3.3: Cartoon representation of the glycosylated homodimer of the immunoglobulin domain C_H2. This illustration highlights the exclusive interaction of the C_H2 domains by its sugar moieties, consistent with deglycosylation of this domain leading to monomeric C_H2.[53]

The IgG-Fc region is a glycosylated homodimer of two immunoglobulin domains (C_H2, C_H3) that dimerizes via inter-heavy chain disulfide bridges at the N-terminal hinge region and due to non-covalent interactions between the C_H3 domains. In contrast to the C_H3 domain, the C_H2 domain does not show any interchain protein-protein contacts, but solely interacts with the second C_H2 domain by its sugar moieties (Fig. 3.3).[45] The structure of the attached 'core' heptasaccharide GlcNAc-GlcNAc-Man-(Man-GlcNAc)$_2$ reveals branching oligosaccharide arms at the first mannose residue. Besides, the oligosaccharide chain may be extended by an additional fucose residue at the primary GlcNAc and an additional galactose residue at the final GlcNAc of each arm.[54–57] The sugar molecules of the oligosaccharide trunk together with one branching arm make multiple non-covalent interactions with the surface of the C_H2 domain, predominantly through hydrophobic amino acid residues. Whereas the second arm of the oligosaccharide chain extends towards the opposite C_H2 domain and interacts with its corresponding sugar moiety.[44,45] Complete deglycosylation of the C_H2 domain abolishes the capacity of the Fc region to bind ligands[58–60] and as a consequence the antibody molecule fails to initiate the corresponding effector mechanisms.[61] Since direct contact between the oligosaccharide and the ligand is not essential for the formation of the complex that activates the effector mechanisms, the sugar moiety has to fulfill a spacer function to hold both C_H2 domains apart. Recent studies have shown that truncation of the oligosaccharide results in a closer approach of the C_H2 domains which then need to undergo conformational rearrangements prior to ligand binding.[53,62,63] As all glycoproteins consist of an ensemble of different oligosaccharides attached to a common peptide, immunoglobulins also possess a distinct array

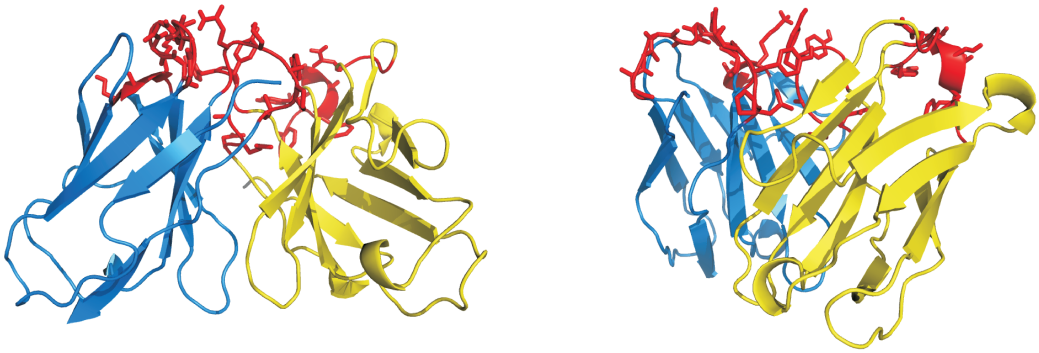


Figure 3.4: Cartoon representation of the heterodimer of the immunoglobulin variable domains V_H (yellow) and V_L (light blue) with their hypervariable regions (depicted in red).

of carbohydrates at a conserved position in the constant region of the heavy chain. Superposition of different IgG-Fc glycoform structures indicates that the C_H3 domains adopt a relatively rigid homodimeric structure whereas the position of the C_H2 domains varies with the distinct oligosaccharides and hence affects protein assembly, secretion and functional activity.[53,57] However, before the Fc region of the antibody molecule triggers any effector mechanism, foreign macromolecules have to be recognized and bound by the Fab region.

The Fab region is composed of the complete light chain (V_L and C_L) paired with the two N-terminal domains of the heavy chain (V_H and C_H1).[64,65] Association of the two polypeptide chains is based on non-covalent interactions between both the variable domains and the constant domains of each chain.[66] Furthermore, a disulfide bridge at the C-termini of the C_L domain and the C_H1 domain connects both chains covalently. For the antigen binding activity of the Fab region, effective assembly of the two chains is required,[66,67] even though crystallographic analyses of several antibody-antigen complexes have shown that antigen binding primarily takes place at the N-terminal tip of the Fab fragment.[68–70] The variable domains that are responsible for the antigen binding have the same overall topology and the same conserved framework structure like the constant domains, but in addition they possess hypervariable regions. On the basis of sequence variations, the residues in the variable domains are assigned to the framework regions or to hypervariable regions respectively complementarity-determining regions (CDRs).[71] Each of the variable domains has three of this complementarity-determining regions which cluster together and form a continuous hypervariable surface (Fig. 3.4). The variability in sequence and size of the CDRs causes a large diversity of the topography of this surface. Therefore, the specificity and affinity of the binding sites for the huge variety of distinct antigens are governed by the structure of the six CDRs each of varied size and shape which are embedded into a scaffold of conserved sequence and structure.[71,72] To understand the molecular mechanism of antigen binding, several structures of Fab regions with specifically bound ligands have been elucidated by X-ray diffraction.[68–70] Experimental evidence indicates that the relationship between amino

acid sequence and structure of the binding site can be described by a canonical model, at least for five of the six hypervariable loops.[73–76] In this model, antibodies have only a few main chain conformations or canonical structures for each hypervariable region. Most variations in sequence alter the CDR surface constituted by the side chains, while the canonical main chain structure remains. Only a few changes in sequence at specific positions result in a different main chain conformation.[77] However, this model is not applicable for the most variable CDR within the V_H domain. As this loop varies significantly in length with very different sequences and patterns of interaction, a clear relationship between sequence and structure was not detectable.[74]

Still, already the earliest studies of antibody structures illustrated that all antibody domains, whether constant or variable, form compact globular structures with a characteristic fold, named the immunoglobulin fold by Poljak and coworkers.[78]

3.1.2 The Immunoglobulin Fold

It is well known that three dimensional structures are much more conserved in evolution than sequences. A comprehensive structural and sequence comparison revealed that a common topology of fold can be achieved by fundamentally different sequences.[79–82] Even so, the occurrence in nature and diversity in terms of biological activity of members of the immunoglobulin fold is quite impressive.

The immunoglobulin like fold is one of the nine proposed superfolds that is adopted by distantly related proteins or evolutionary unrelated proteins with no sequence homology.[83] Domains with an immunoglobulin like fold can be divided into superfamilies in which the sequence and structural similarity between members suggests that a common evolutionary origin is possible. Within these superfamilies domains may be grouped again into families in which either sequence or functional and structural similarities imply that a common evolutionary origin is very likely.[84,85] The class of immunoglobulin like β-sandwich proteins includes several superfamilies, amongst others immunoglobulin, integrin domains, transcription factor, and hemocyanin. However, the immunoglobulin superfamily comprises by far the greatest number of members and is itself grouped again into families (Fig. 3.5). Originally, the immunoglobulin superfamily consisted of three subtypes, namely V set domains (antibody variable domain like), C1 set domains (antibody constant domain like), and C2 set domains (CD2, CD4, cell adhesion molecules).[84] In the meantime, the increasing number of structures of immunoglobulin like domains has identified new families, including subtype I, C3, C4, and Fn3.[86,87]

The topology of the immunoglobulin fold has been described as a greek key β-barrel[88] composed of 7-9 β-strands, distributed between two sheets with typical topology and connectivity. Structural alignment of domains adopting the immunoglobulin fold reveals a common structural core which consists of only four β-strands (*b*, *c*, *e*, and *f*).[89] These structurally conserved core strands are embedded in an antiparallel β-sheet sandwich with three to five additional strands (*a*, *c*', *c*", *d*, *g*). The position of the additional edge strands varies relative to the common core depending on the

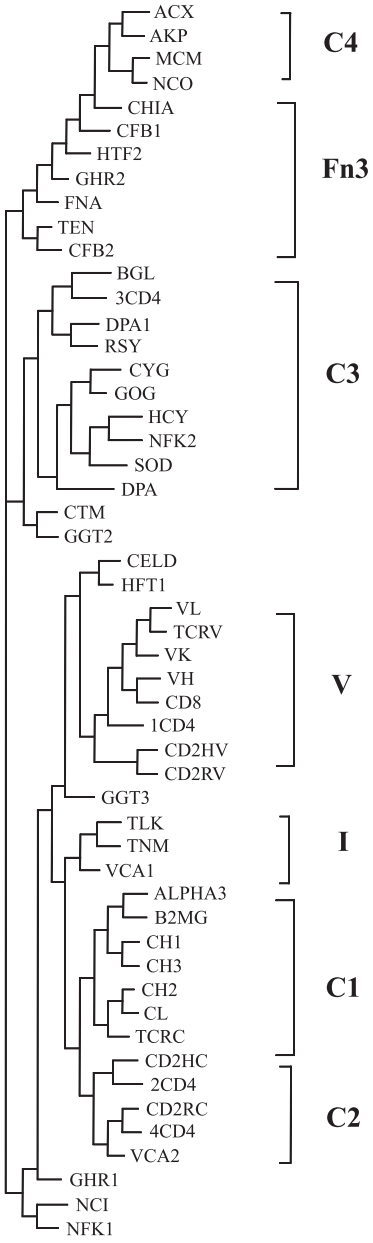


Figure 3.5: Structural tree of the immunoglobulin fold. Meanwhile, the immunoglobulin superfamily comprises several distinct families (V, C1, C2, I, C3, C4, and Fn3). Proteins investigated in this study all belong to the subtype C1 (antibody constant domain like).

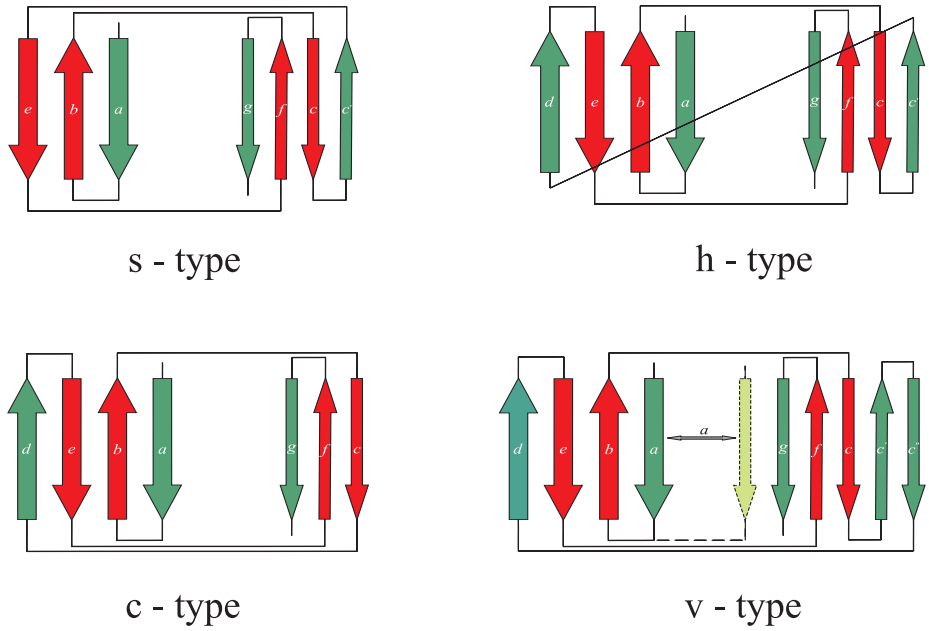


Figure 3.6: Topology diagrams of observed hydrogen patterns of the immunoglobulin fold. The 7-9 β-strands form a sandwich of 2 sheets (back sheet thin arrows; front sheet thick arrows). The common core is shown in red. Immunoglobulin constant domains consist of 7 β-strands in a topology displayed at the bottom left (c-type). Immunoglobulin variable domains have an additional hairpin between strands c and d, with a total of 9 β-strands (v-type). Strand a can adopt two alternative positions in the v-type domains. Other immunoglobulin like domains also consist of 7 β-strands, but differ from c-type in the 4th strand which has switched sheets (s-type). The last type represents an 8-stranded hybrid between c- and s-type that possesses both strands d and c, so each sheet is composed of 4 β-strands (h-type).

length of the intervening sequence between strands c and e, the most variable region in sequence. Based on the variation in the position of the edge strands, four different topological subtypes can be distinguished (Fig. 3.6).[89]

Generally, it is assumed that two polypeptide chains fold very similar if they share at least 30% sequence identity.[90,91] The sequence identity between most members of the immunoglobulin fold is far below this threshold, but still the fold is remarkably similar.[87,89] Several studies indicate that not only the sequence but also the overall amino acid composition influences the folding pattern of a protein.[92,93] Some of the positions of a particular fold are always occupied by hydrophobic amino acids which are key residues constituting the common core and responsible for maintenance of the fold.[94] In the case of the immunoglobulin fold only position $c3$ is hydrophobic within the complete superfamily, at positions $b1$ and $e5$ hydrophobic amino acids are present in more than 75% of the sequences, at position $f5$ only in 67% of the sequences (Fig. 3.7).[87] This low conservation of specific amino acids clearly illustrates the great diversity of this superfamily and the necessity for specific interactions that stabilize the folded domains. Besides the formation of a typical hydrophobic core coded by

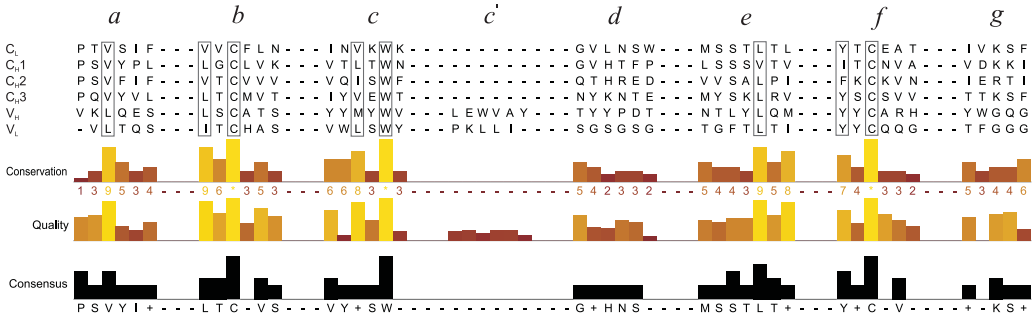


Figure 3.7: Sequence alignment of the distinct domains of immunoglobulin G. The alignment is based on the regular secondary structures and generated with the software Jalview.[95] This multiple alignment editor also calculates the conservation (an index reflecting the conservation of physiochemical properties), the quality (the quality score indicates the likelihood of observing mutations), and the consensus (an index representing the most common residue) of the analyzed sequences. For the immunoglobulin G domains, identical residues are found at positions $b3$, $c5$, and $f3$. In addition, positions $a3$, $b1$, $c3$, $e5$, and $f1$ are occupied by hydrophobic residues. The typical hydrophobic core of the immunoglobulin fold is mainly established by the inner β-strands b, c, e, and f, including the most buried positions $b3$, $c3$, and $f3$.

the sequence, in several subtypes disulfide bridges are introduced which influence the overall domain shape and also the symmetry between the two sheets.

Members of the immunoglobulin superfold are impressively widespread in nature. Immunoglobulin like domains have been identified in eukaryotes and prokaryotes, in vertebrates and invertebrates, and in fungi, parasites, bacteria, viruses, and plants.[82] Furthermore, members of the immunoglobulin like superfamilies exhibit a remarkable diversity in terms of biological activity.[82] Consequently, the question of the origin of the immunoglobulin fold rises, whether these proteins are homologous or analogous. If these proteins are homologous and originate from a common ancestor, the sequence has diverged during evolution to fulfill a substantial variety of distinct and independent functions. Otherwise, if these proteins are analogous, the immunoglobulin fold is a rather stable framework which many sequences have adopted during evolution. As most members of the immunoglobulin fold share such a low level of sequence identity, it is very difficult to distinguish between convergent and divergent mechanisms of evolution.[96] So still nowadays the question remains, is the immunoglobulin fold derived from a common ancestor or is it a stable structure to which many sequences have converged?

3.1.3 Immunoglobulin Amyloidogenicity

The enormous diversity of immunoglobulin like domains in terms of biological activity and specificity comes at a price. During diversification, structurally compromised immunoglobulin like domains are frequently generated. Normally, dysfunctional molecules are eliminated by distinct quality control mechanisms of the cell.[97,98] However, occasionally the expression of structurally compromised molecules has pathological consequences.[99,100] For instance, the variable domain of the immunoglobulin light chain is

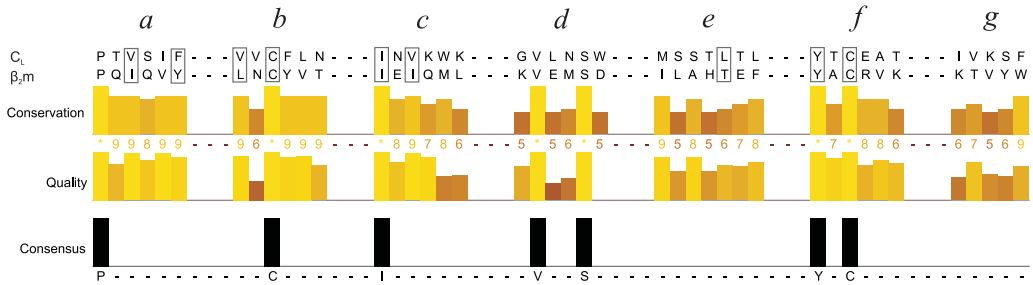


Figure 3.8: Sequence alignment of C_L and β_2-microglobulin. The alignment is based on the regular secondary structures and generated with the software Jalview[95] according to Fig. 3.7. Positions which are occupied by hydrophobic residues within the IgG domains are also conserved in β_2m. However, the overall conservation is significantly more pronounced between C_L and β_2m compared to C_L and the other IgG domains. Yet, C_L and β_2m differ fundamentally in respect of misfolding and aggregation.

involved in different human diseases, most of them associated with the aggregation of the polypeptide, like the light chain deposition disease (LCDD)[101] or the light chain amyloidosis.[102] Additionally, the MHC component β_2-microglobulin, also a member of the immunoglobulin superfamily, is responsible for the dialysis related amyloidosis.[103] Amyloidosis is the origin of various disorders which are caused by the extracellular deposition of insoluble amyloid fibrils in organs and/or tissues. These amyloids are long lived dissociation- and degradation-resistant structures, made up of β-strands that are arranged into sheets lying perpendicular to the long fiber axis and possess a core cross-β structure.[104] Despite the large variety of native folds adopted by amyloidogenic proteins, these structural features seem to be a recurring motif in amyloids suggesting a common assembly mechanism.[105] In addition, several studies indicate that folding intermediates are key precursors for the formation of amyloid fibrils.[106, 107]

Bearing in mind that intermediates are a rather general aspect of protein folding reactions and that most polypeptides are in principle susceptible to amyloid formation,[108] the question arises how proteins avoid aggregation in the majority of cases. To address this issue the folding pathway of the constant domain of the light chain (C_L) was investigated with high structural resolution. Characterization of the folding intermediate of the C_L domain might reveal insights into the essential prerequisites for a productive folding pathway. Moreover, the C_L domain is a particularly instructive model system because it has never been directly associated with amyloidogenic diseases even if present at much higher concentrations in the blood than the amyloidogenic protein β_2-microglobulin.[109] Interestingly, β_2m belongs to the same protein family and, like C_L, forms a β-sandwich composed of seven strands stabilized by a single disulfide bond between strands b and f (Fig. 3.8).[110, 111] The amyloidogenic properties of β_2m are exceptionally well studied. As β_2m readily forms amyloid fibrils if partially unfolded (e.g. at acidic pH),[112] the reaction is thought to be initiated by the population of a partially folded intermediate.[112, 113] When such intermediates become long lived rather than transient species, they especially run the risk of misfolding and misassembly. Productive folding of β_2m from a native like intermediate to the native state is limited by

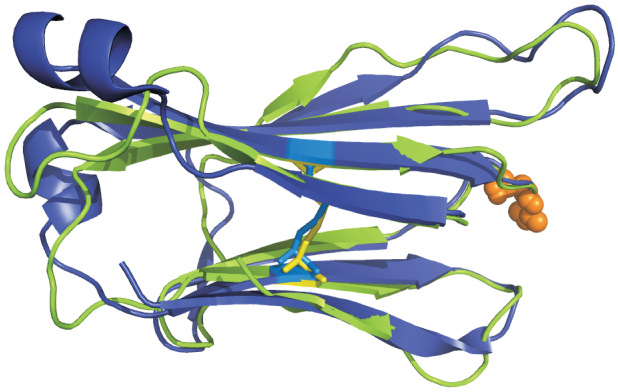


Figure 3.9: Cartoon representation of the C_L domain compared with β_2-microglobulin. The C_L domain (shown in blue) and β_2m (illustrated in green) are homologous members of the same protein family and adopt a highly conserved three dimensional structure, including the single disulfide bridge and the *cis* proline residue (depicted in orange).

an intrinsically slow *trans*-to-*cis* peptidyl-prolyl isomerization reaction.[113,114] Experiments in which the critical proline residue was held in the *trans* state confirmed that this intermediate is the major determinant in amyloid formation.[113,115] In addition, further studies revealed that the most probable amyloidogenic precursor already possesses large parts of the native β-sheet topology with only the outer strands and loop regions being distorted.[113,114,116] Remarkably, the *cis* proline residue associated with the amyloidogenic potential of β_2m is conserved in the C_L domain (Fig. 3.9).[110] Furthermore, the overall folding mechanisms of the two proteins are highly similar,[111,113] each populating an intermediate state en route to the native state. However, despite their sequence and structural similarities the folding landscapes of C_L and β_2m must exhibit considerable differences to either favor or prevent misfolding and aggregation.

3.2 Folding Pathway of the C_L Domain

In the current view, almost all proteins are believed to populate partially folded species, so called folding intermediates, along their pathways to the native state.[117–119] The characteristics of folding intermediates determine whether a protein is able to fold robustly or has the tendency to misfold.[120] Thus, a detailed structural characterization of folding intermediates is essential for the understanding of protein folding in general. Because they are transient, however, only very few folding intermediates have been described in atomic detail so far.[121–124]

3.2.1 Backbone Assignment

Backbone resonance assignment is a basic requirement for the structure determination of proteins as well as for the analysis of protein dynamics and kinetics by NMR spectroscopy. Nowadays NMR studies on biomolecules typically use multidimensional techniques and isotopic labeling to resolve and identify the large number of peaks present in the spectrum. Already a simple two dimensional ^{1}H-^{15}N HSQC experiment reveals insights into structural elements and dynamics of the biomolecule. Based on the dispersion of the proton dimension, it is possible to distinguish between an unfolded (Fig. 4.6) and a well structured (Fig. 3.10) protein. In the special case of all α-helical (Fig. 5.4) and all β-sheet (Fig. 3.10) proteins it is even feasible to identify these secondary structure elements due to their characteristic proton dispersion. Furthermore, the line width in the proton dimension enables conclusions about the dynamical properties of each single peak. To transfer this information onto an atomic level, each peak appearing in the HSQC spectrum has to be assigned according to its amino acid type and position within the protein sequence.

Therefore, the triple-resonance experiments HNCO, HNcaCO, HNCA, CBCAcoNH, and HNCACB were recorded on a uniformly ^{15}N, ^{13}C labeled sample of the murine C_L domain (103 AA). These spectra provide the C', C$^{\alpha}$, and C$^{\beta}$ chemical shifts of residue i and $i-1$. The chemical shifts are assembled into arrays termed pseudo-residues, each of them associated with a single resonance in the HSQC spectrum. According to the very sensitive relationship between the carbon chemical shifts and the amino acid type, the pseudo-residues can be linked to appropriate amino acids. Furthermore, each pseudo-residue holds the information on the carbon chemical shifts of its preceding residue. In the assignment process, the possibilities of suitable amino acids for each pseudo-residue and its adjacent residue are adjusted to the sequence of the protein. This assignment process was carried out semi-automatically using the program Pasta.[125] For the C_L domain, this strategy resulted in the assignment of 73 residues out of 98 non-proline residues (Fig. 3.10). Besides very few isolated residues that could not be assigned, most of the unassigned residues concentrate to one region of the protein, the outer strand d and parts of its flanking strand e including the connecting loop. This indicates that the residues in this region probably exhibit high water exchange rates of the amide protons or chemical exchange.

3.2.2 Secondary Structure

On the way to the three dimensional structure determination, a central step is the identification of secondary structure elements. Several indicators exist which present hints on the secondary structure elements α-helix, β-sheet or loop region. However, an unambiguous characterization of the secondary structure of a protein requires the combination of distinct methods. The most reliable of these methods is the identification of helical or β-sheet secondary structure elements by their specific sequential ^{1}H-^{1}H NOE pattern.[126] Whereas within an α-helix the short $^{1}H^{N}_{i}$-$^{1}H^{N}_{i+1}$ distance gives rise

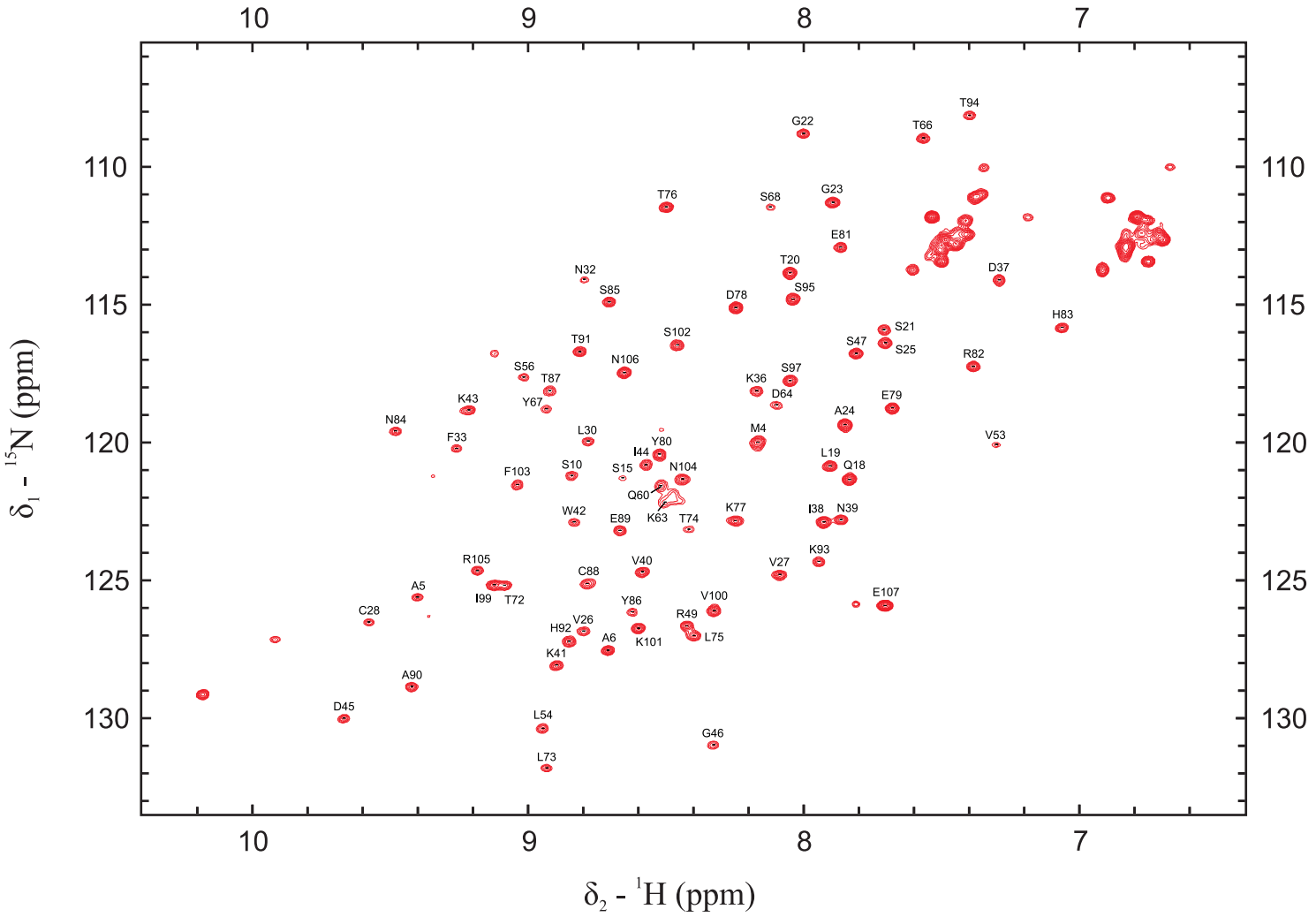


Figure 3.10: Assigned ^{1}H-^{15}N HSQC spectrum of the C_L domain (103 AA) recorded at a temperature of 25 °C and a proton frequency of 750 MHz.

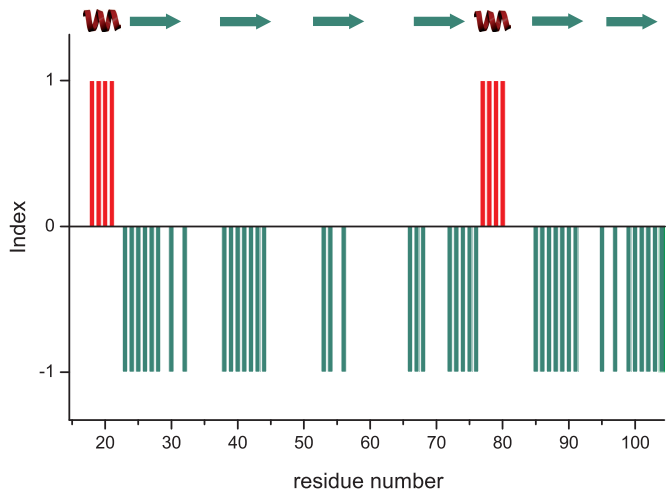


Figure 3.11: Consensus chemical shift index of the C_L domain. These consensus predictions are based on a simple majority rule, applied whenever at least two out of three different chemical shift indices argue for one particular secondary structure element. Analysis of the carbon chemical shifts C', C^{α}, and C^{β} of the C_L domain indicates the expected 7 β-strands as well as the two short helical elements.

to an intense cross peak, the characteristic pattern within the β-sheet shows an intense $^{1}H^{\alpha}_{i}$-$^{1}H^{N}_{i+1}$ cross peak. Another indicator for the different structural elements are three bond J couplings which are correlated with the intervening torsion angles Φ and Ψ by the Karplus equation.[127] In natively folded proteins the backbone Φ and Ψ conformations depend on steric effects, both within individual amino acids and between side chains of different residues, and on secondary structure interactions such as hydrogen bond formation. The well established Karplus relationship between the backbone torsion angle Φ and the $^{3}J_{H^{N}H^{\alpha}}$ coupling constant enables conclusions about the secondary structure elements as the coupling constant ranges from 3.9 Hz for α helices to 8.9 Hz for anti-parallel β-sheets and 9.7 Hz for parallel β-sheets.[128,129] And finally the probably simplest method for identifying protein secondary structure is the analysis of backbone ^{13}C chemical shifts. Due to the strong correlation between carbon chemical shift and secondary structure, a chemical shift index (CSI) can be assigned to each residue based on the deviation of its chemical shifts from the random coil values.[130,131] Secondary structure elements are defined according to the values and local densities of these chemical shift indices.

As the study of the C_L domain mainly focuses on the folding pathway and in addition various crystal structures of the C_L domain in the context of the Fab region or even the whole antibody molecule have already been elucidated, there is no need for a complete structure determination of the C_L domain in solution. However, since the backbone resonance assignment is the basic prerequisite for the characterization of the folding pathway, the carbon chemical shifts obtained for the assignment process can be further analyzed in respect of secondary structure elements. This analysis allows conclusions

about the secondary structure of the protein in solution without any additional experiments. Comparing the measured carbon chemical shifts with the random coil values results in chemical shift indices for each backbone carbon nucleus. The CSI takes the value 0 for chemical shifts within the range of the random coil values, 1 for chemical shifts greater than the range and −1 for shifts less than the range. For the carbon nuclei C' and C^{α} a group of four or more CSI values of 1 not interrupted by a CSI value of −1 indicates an α-helix, whereas a group of three or more continuous CSI values of −1 suggests a β-strand. All other regions are designated as coil. In contrast, the C^{β} chemical shift index only enables the identification of β-sheet structures. In this case, four or more consecutive CSI values of 1 support the assumption of a β-strand being present. For the C_L domain, this approximate evaluation suggests an identical secondary structure in solution as compared to the crystal. Remarkably, even though this method only provides hints the residues constituting the secondary structure elements are completely consistent. Furthermore, this approach clearly indicates the existence of the two short strand connecting helical elements also in solution (Fig. 3.11).

3.2.3 Folding Kinetics

The C_L domain folds via an intermediate on two parallel pathways to its native state.[111] The slower pathway is limited by the isomerization of the Tyr-34–Pro-35 bond to the native *cis* conformation which predominantly adopts the *trans* conformation in the unfolded state. In general, proline residues are exclusively synthesized in the *trans* form by the ribosome. Since the *cis* and *trans* conformation of peptide bonds preceding a proline residue are almost isoenergetic, with the *trans* form being slightly more favorable, unfolded polypeptides exhibit a mixture of *cis* and *trans* states.[132] Due to the thermodynamic equilibrium, 10-30% of these peptide bonds N-terminal to a proline residue occur in the *cis* conformation.[133] However, native proteins exist essentially completely in either the *trans* or the *cis* form. As a consequence, molecules with a native *cis* conformation have to undergo the intrinsically slow isomerization reaction. Within the cell enzymes such as peptidyl-prolyl isomerases catalyze the *trans*-*cis* isomerization. *In vitro* this isomerization decisively determines the rate of protein folding allowing the characterization of the folding pathway by real time NMR spectroscopy. In the case of the C_L domain only ∼10% of the molecules adopt the native *cis* conformation in the unfolded state at equilibrium and are able to fold to the native state within a few seconds. Whereas ∼90% of the molecules exist in the non-native *trans* conformation and need to undergo the isomerization reaction before complete folding to the native state.[111,134] At 2 °C this reaction takes several hours and therefore populates the major kinetic intermediate for a considerable amount of time.

In order to characterize the intermediate state and also the folding pathway of the C_L domain, real time ^{1}H-^{15}N HSQC spectra were measured during refolding from the chemically denatured state. For this purpose, ^{15}N-labeled unfolded C_L in PBS buffer containing 2 M GdmCl was diluted 10-fold by adding ice cold PBS buffer without GdmCl and HSQC spectra were recorded immediately after mixing at 2 °C every

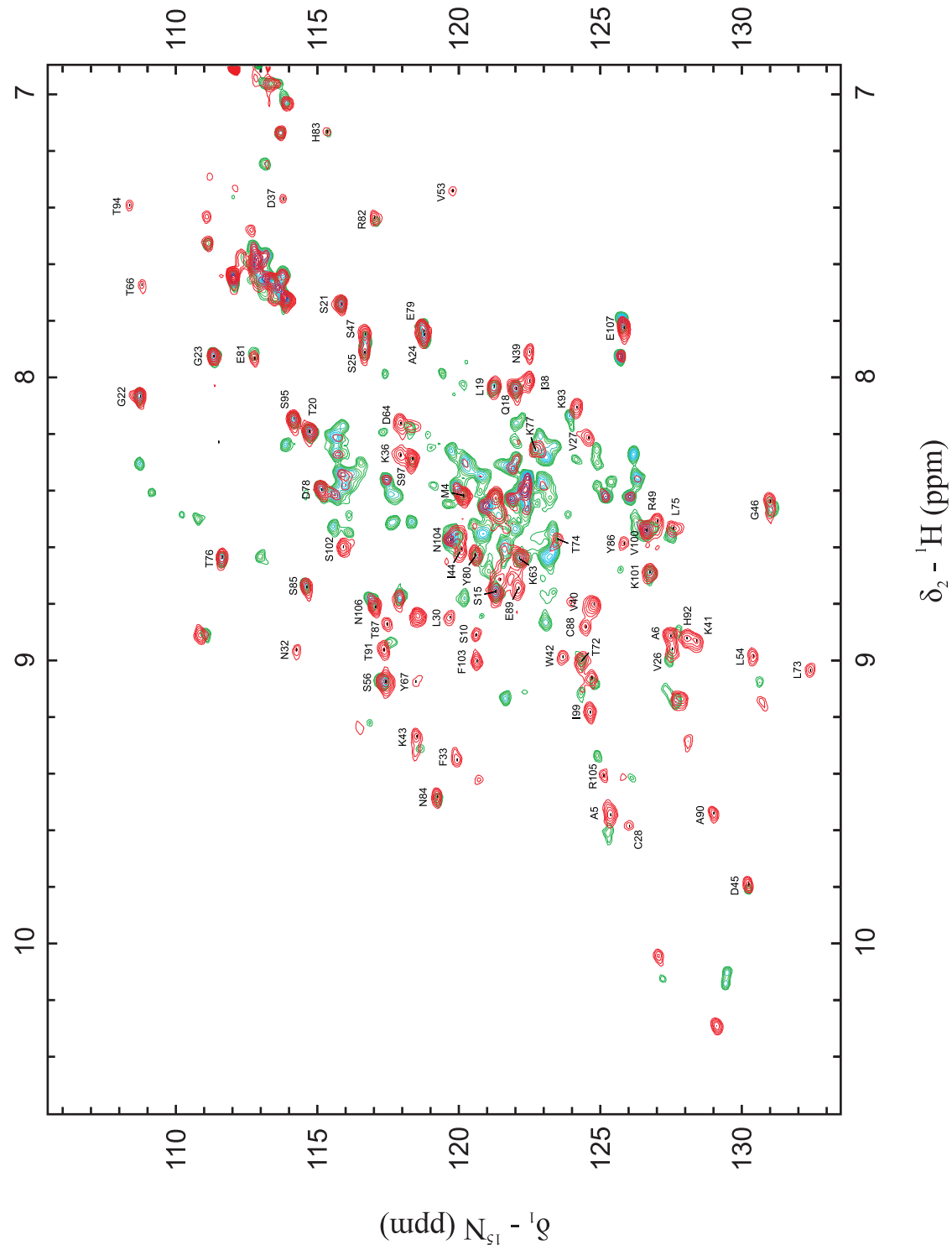


Figure 3.12: Superimposed ^{1}H-^{15}N HSQC spectra of the native state (shown in red) and the intermediate state (depicted in green) of the C_L domain. These spectra were recorded at 2 °C with a residual GdmCl concentration of 0.2 M during refolding from the chemically denatured state.

14 min. The first spectrum reflects almost exclusively the kinetic intermediate and had only to be corrected for 10% of the C_L molecules possessing the correct Tyr-34–Pro-35 isomerization state. This correction was achieved by subtracting 10% of the final spectrum after 7 h from the first HSQC spectrum with the software Topspin 1.3 (Bruker BioSpin). Because the chemical shifts of the amide protons strongly depend on their molecular environment, overlaying the HSQC spectra of the intermediate and the native state reveals similarities and changes in their environment during the folding process (Fig. 3.12). For the intermediate state of the C_L domain some residues already seem to sense their native environment whereas other residues indicate substantial differences between the native and the intermediate state. To gain more insights into the structural properties of the intermediate, the change in peak intensity at the native chemical shift position was followed over time for each assigned residue. Consistent with the first spectrum representing solely the intermediate state and likewise studying only the slow folding pathway to the native state, all spectra during refolding were corrected for the fast folding pathway analogous to the first spectrum. Except for the residues with already native like intensity after the dead time of the experiment, in every case the change in intensity during the folding process could be well described by a single exponential function (Fig. 3.13). The time constants of the folding of the individual residues show stochastic behavior around a mean value of $\tau = 199$ min at 2 °C without any significant systematic deviations for any part of the protein. In contrast, extrapolating the exponential functions to obtain the initial amplitudes at time zero presents considerable discrepancies for distinct parts of the protein. Even though every residue folds cooperatively with the same rate constant to the native state, the initial amplitudes indicate a varying degree of local structuring within the intermediate state. As the amide resonances represent very sensitive indicators for changes in their local environment, it has to be kept in mind that low initial amplitudes are not mandatory equal to missing secondary structure, but might just as well indicate the requirement for local rearrangements such as side chain packing or hydrogen bond formation. In Fig. 3.13 regions of high or low local structuring are mapped on the crystal structure of the C_L domain revealing an interesting pattern: almost native initial amplitudes are found in vicinity of the two short helices connecting strand *a* and *b* as well as *e* and *f* and adjacent β-sheet termini suggesting that these parts already adopt a native like conformation in the intermediate state, whereas low initial amplitudes are observed for some of the β-strands, in particular strands *c* and *d* which argues for a lack of native interactions. Interestingly, the two helices and their local environment seem to be already completely structured in the major folding intermediate of the C_L domain.

Certainly, the initial HSQC amplitudes only provide hints on the structural properties of the intermediate. Equilibrium spectroscopic data would give information more directly related to structure. Mutating the Pro-35 residue against another amino acid that preferentially adopts a *trans* peptide bond,[135] might trap the kinetic intermediate. Populating the folding intermediate at equilibrium would enable a detailed investigation of its structural features.

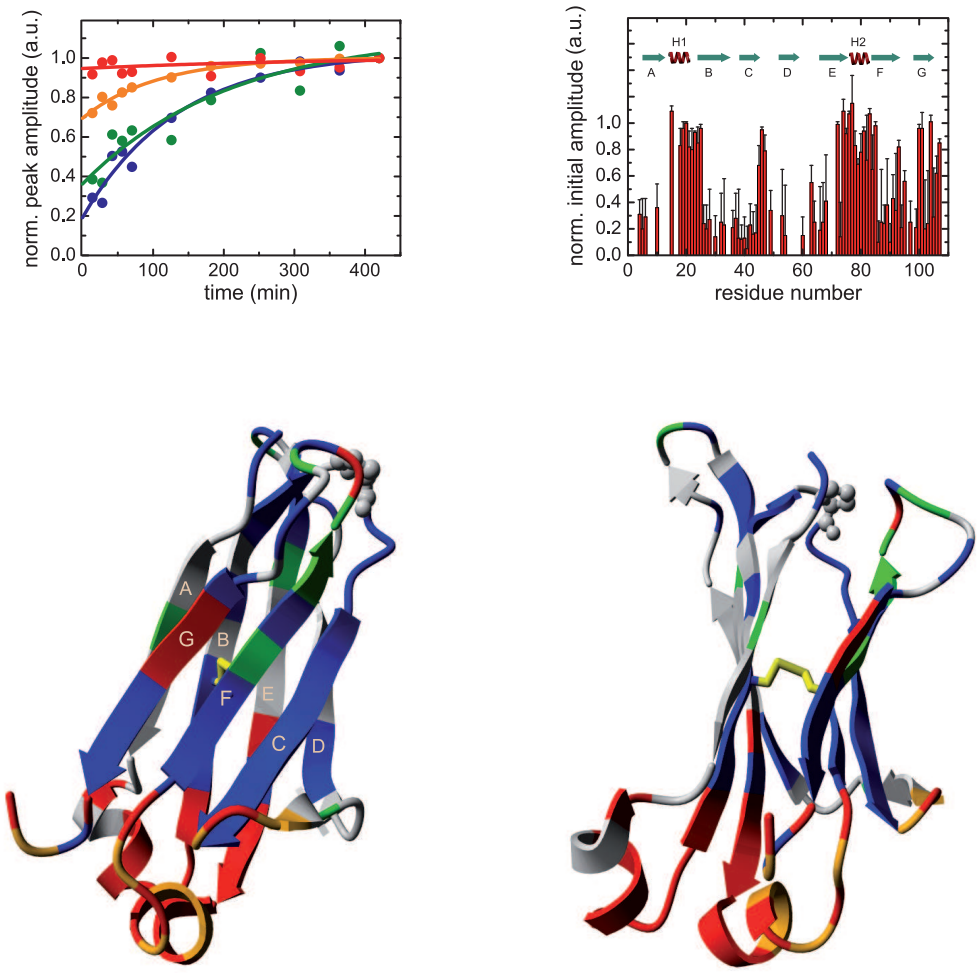


Figure 3.13: Folding kinetics of the C_L domain. Top left: For each assigned residue the change in the peak intensity was followed over time, fitted by a single exponential function and extrapolated to time zero (blue: Lys-36, green: Glu-89, orange: Glu-79, and red: Leu-19 selected as representative residues). Top right: The initial HSQC amplitudes indicate significantly different degrees of local structuring for the individual residues in the intermediate state. Bottom: Mapping the initial HSQC intensities on the crystal structure of the C_L domain. The colors represent the initial amplitudes relative to the native ones (red: $> 80\%$, orange: $\leq 80\%$, green: $\leq 60\%$, blue: $\leq 30\%$).

3.3 Structural Characterization of the Trapped Folding Intermediate

By mutating the Pro-35 residue against Ala (C_L^{P35A}), the kinetic intermediate might be trapped at equilibrium. [1] Indeed, far-UV and near-UV CD spectra of C_L^{P35A} agree quite well with the respective spectra of the kinetic intermediate. In addition, denaturant induced unfolding transitions show that the Pro35Ala mutation leads to a stability reduction of the C_L domain by >50% which is consistent with a partially folded species representing the major equilibrium state. Furthermore, so called chevron plots enable the determination of folding and unfolding rate constants. For this purpose, the protein sample is rapidly diluted to a particular denaturant concentration and the establishment of the new equilibrium is directly monitored by fluorescence spectroscopy. Plotting the logarithm of the calculated relaxation rates as a function of the denaturant concentration results in the chevron plot. The analysis of these chevron plots indicates that the wild type (C_L^{wt}) folds to its intermediate state with a very similar rate constant as C_L^{P35A} folds to its final state. These data argue for the C_L^{P35A} mutant populating the kinetic intermediate at equilibrium which is trapped by the *trans* state of the bond preceding Ala-35. Based on these experiments, the structure of C_L^{P35A} was further characterized by NMR spectroscopy.

3.3.1 Backbone and Aliphatic Side Chain Assignment

The ^{1}H-^{15}N HSQC spectra of the C_L^{wt} kinetic intermediate and the C_L^{P35A} mutant are almost completely superimposable (Fig. 3.14), demonstrating equivalent secondary and tertiary structure in both species. Crucially, trapping of the intermediate state allows an assignment to be carried out which is not feasible for the transiently populated kinetic intermediate observed in the folding pathway of C_L^{wt}. Backbone resonance assignment for the C_L^{P35A} mutant was achieved by recording the same standard backbone experiments and following the same assignment procedure as for the wild type C_L^{wt}.

To highlight similarities and dissimilarities regarding the structural properties of the native and the intermediate state, further experiments for both, the wild type and the mutant, are required. In general, folding intermediates often adopt a near native topology with incompletely folded or partially misfolded structural elements such as side chain interactions.[121–124, 136, 137] To verify the side chain packing and interactions in the wild type C_L^{wt} and the C_L^{P35A} mutant, in addition to the backbone resonances the side chain resonances also have to be assigned. Therefore, the triple resonance experiments (H)CCH-TOCSY which correlates the carbon nuclei of a complete spin system and (H)CCH-COSY which gives information about adjacent carbon nuclei were measured. The combination of these spectra that take advantage of the higher chemical

1 Since this project was conducted in collaboration with the group of Prof. Dr. Johannes Buchner, Lehrstuhl Biotechnologie, Department Chemie, Universität München, the biochemical experiments and also the protein sample preparation were performed by Matthias Feige.

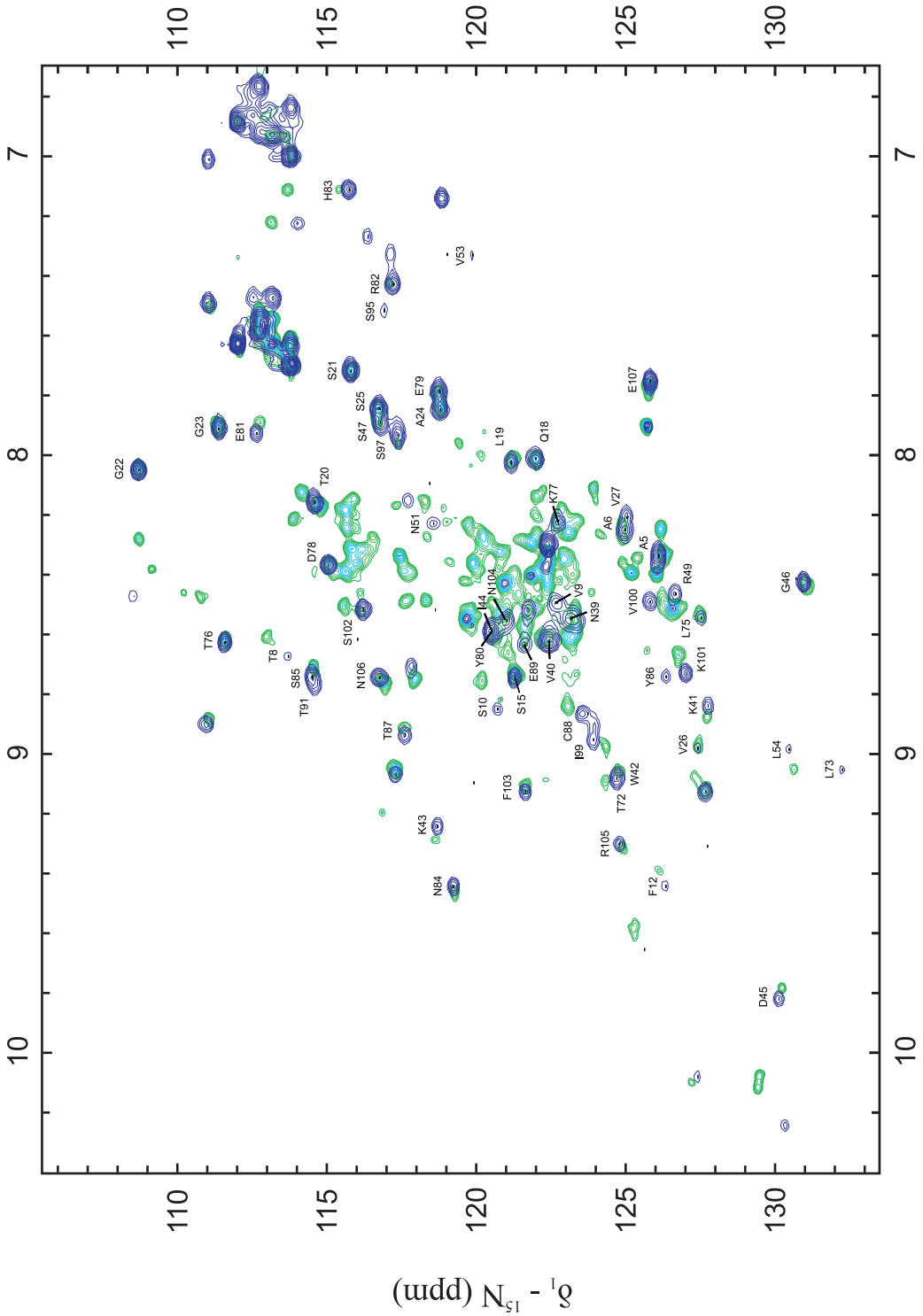


Figure 3.14: Superimposed ^{1}H-^{15}N HSQC spectra of the kinetic (depicted in green) and the trapped (shown in blue) folding intermediate of the C_L domain. Both spectra were recorded at 2 °C with a residual GdmCl concentration of 0.2 M.

shift dispersion in the carbon dimension resulted in the assignment of the aliphatic side chains. Aromatic resonances generally escape this assignment process. As the connection of the aromatic protons with the aliphatic side chains necessitates several additional experiments, the assignment of the aromatic resonances was not taken into account. Still, based on the assignment of the backbone resonances and the aliphatic side chains of both species, C_L^{wt} and C_L^{P35A}, a set of NOESY spectra can identify differences in the structure of the native and the intermediate state.

3.3.2 Comparison of Secondary and Tertiary Structure of the Wild Type and the Mutant

It is well known that the chemical shifts of backbone atoms in proteins are extremely sensitive to local conformations. Based on this strong correlation between chemical shifts and protein secondary structure, Wishart defined the chemical shift index (CSI) which enables the characterization of secondary structure elements subsequently to the protein backbone assignment.[130,131] This primarily idea of the chemical shift index was further expanded by the software Talos.[138] Besides the chemical shift itself, Talos takes advantage of homologous proteins which display quite similar patterns of secondary chemical shifts. Meanwhile the database of the program contains C^{α}, C^{β}, C', H^{α}, and N^{H} chemical shifts of the structured parts of 186 proteins for which a high resolution ($\leq$ 2.2 Å) X-ray crystal structure is available. Talos searches this database for strings of three adjacent residues with the closest similarity in secondary chemical shift and amino acid sequence to those of the query triplet. Averaging the backbone angles of the ten most consistent triplets from the database, provides a reliable prediction for the Φ and Ψ dihedral angles of the central residue within the query sequence. The ten Φ/Ψ pairs with the highest similarities are graphed in the Ramachandran plot which also highlights the most populated areas of the entire database (Fig. 3.15). Despite the relatively small database Talos searches through, the approach to combine both chemical shift and residue type information and also to consider a string of three residues instead of a single one, enables a reliable prediction of the Φ and Ψ torsion angles and consequently for the backbone secondary structure.

A successful search for chemical shift similarities requires an identical referencing of all chemical shifts. Before crystal structures and corresponding chemical shifts are included into the Talos database, an algorithm is applied to the deposited chemical shifts which corrects all secondary shifts according to the appropriate reference standard for each nucleus and hence ensures that all of the shifts are defined in the same manner. To allow a reasonable similarity search, the same procedure applies for all experimental chemical shifts. Typically, ^{1}H and ^{13}C chemical shifts are referenced to the internal standard trimethylsilyl propionate (TSP). In the case of C_L^{wt} and C_L^{P35A} referencing of the H^{N} dimension was achieved by the external standard TSP and subsequently all other dimensions were referenced indirectly. To be consistent with the direct referencing to the carbon chemical shift of the TSP methyl groups, 2.6 ppm had to be added to all

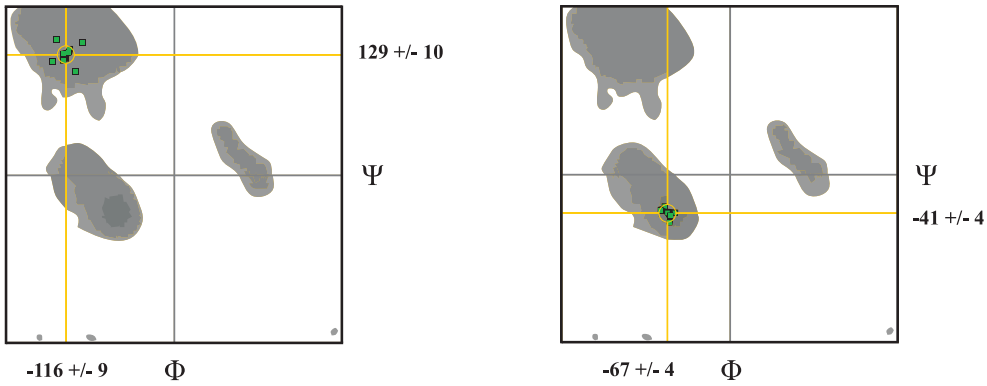


Figure 3.15: Prediction of the Φ and Ψ dihedral angles by the software Talos. The ten Φ/Ψ pairs from the database with the highest similarity in secondary chemical shift and amino acid sequence to the query triplet are indicated as small green squares in the Ramachandran plot. Highlighted in gray are the areas which correspond to the torsion angles characteristic for the secondary structure elements β-sheet, right handed α-helix, and left handed α-helix. Left: Talos output for the residue Val-27 of the C_L domain. Due to the consistent matches with triplets from the database, the predicted Φ and Ψ angles argue for a β-sheet structure at position 27. Right: Talos output for residue Glu-79 of the C_L domain. In this case, the predicted torsion angles clearly suggest that residue 79 is part of an α-helical structure. It should be noted that approximately 2% of the predictions made by Talos are found to be in error.

carbon dimensions.

Prediction of the Φ and Ψ dihedral angles by the software Talos reverts to the chemical shifts of C^{α}, C^{β}, C', H^{α}, and N^{H}. However, the standard backbone resonance assignment is restricted to the carbon chemical shifts C', C^{α}, and C^{β} and the nitrogen chemical shift N^{H}. Although the H^{α} chemical shift requires additional experiments, it should be included in the characterization of the protein secondary structure as this proton chemical shift is one of the most sensitive indicators for the different backbone conformations.[131] Therefore, the side chain assignment of the wild type C_L^{wt} and the mutant C_L^{P35A} is not only prerequisite for the interpretation of the NOESY spectra, but also allows the determination of the backbone Φ and Ψ angles. For nearly all of the assigned residues, Talos was able to find highly consistent matches within its database, arguing for a reliable prediction of the dihedral angles. Fig. 3.15 illustrates the graphical output of the ten Φ/Ψ pairs with the highest similarities and the corresponding averaged Φ and Ψ values for two representative residues of the C_L domain. The backbone torsion angles of C_L^{wt} and C_L^{P35A} indicate an almost identical secondary structure in the intermediate and the native state. This particularly holds for the central β-strands *b*, *c*, *f*, and partly strand *e* that constitute the hydrophobic core of the protein and also for the outer strand *g*. Notably, the predicted Φ and Ψ dihedral angles for the C_L^{P35A} mutant clearly identify the two short helices, suggesting that these secondary structure elements are already completely evolved in the intermediate state. Since it is generally assumed that folding intermediates adopt a native like topology,[136,137] C_L^{wt} and C_L^{P35A} are examined more precisely in respect of their tertiary structure.

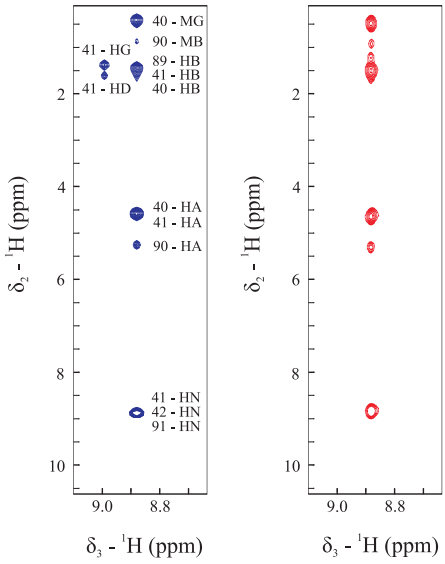


Figure 3.16: Comparison of backcalculated (shown in blue) and experimental (shown in red) NOESY-strip for Lys-41 as representative residue of the C_L domain. The simulated NOESY strip for Lys-41 indicates the assignment for expected cross peaks within a radius of 5 Å. Additionally the software requires a threshold for the peak intensity to account for the experimental noise. To simplify matters, adjacent NOESY-strips from residues other than Lys-41 within the pictured experimental strip have been omitted.

Thus, a set of 3D-NOESY spectra was recorded, including NNH- and CNH-NOESY spectra[139] in addition to the conventional ^{15}N-HSQC-NOESY spectrum. Analysis of these NOESY spectra was performed in combination with a backcalculation software (Fig. 3.16). This in-house software simulates the NOESY strip of a certain residue according to the adjusted parameters, like for instance the NOESY mixing time, the correlation time of the protein, and the reachable radius around the query residue. For C_L^{wt} the backcalculation of the NOESY spectra was based on the crystal structure of the C_L domain (pdb 1ors). The notable agreement between the backcalculated NOESY strips from the crystal structure and the experimental NMR data for all of the recorded spectra evidence an identical three dimensional structure for the C_L domain in the crystal and in solution. Furthermore, the simulated NOESY strip indicates the assignment and also the expected intensity for each cross peak. This facilitated the interpretation of the NOESY spectra of C_L^{wt} and the identification of structural varieties between the native state and the intermediate. The comparison of the experimental NOESY strips of corresponding residues of the wild type C_L^{wt} and the mutant C_L^{P35A} revealed that large parts of the proteins give rise to cross peaks within the same range. Regions in C_L^{P35A} with significant differences in the carbon chemical shifts and the NOESY pattern from C_L^{wt} are exclusively located around Ala-35 and strands *d* and *e*. Remarkably, the two helices connecting strands *a* and *b* as well as *e* and *f* are fully formed in the C_L^{P35A} mutant, as judged from the characteristic NOE pattern for helical structural elements.[126]

3.3.3 NMR-restrained MD Simulations

To gain further dynamic and structural information that cannot be deduced from the NMR experiments alone, molecular dynamics (MD) simulations with NMR-derived restraints were performed. [2] The C_L^{P35A} mutant was created from the crystal structure of the C_L domain (pdb 1fh5) by replacing Pro-35 against Ala and setting the bond preceding Ala-35 to *trans*. This mutant was energy-minimized with the steepest-descent and the adopted Newton-Rhapson methods.[140] The restraints for the MD simulations were derived from a comparison of the experimental NMR data of C_L^{wt} and C_L^{P35A}. Residues with similar chemical shifts and NOESY cross peaks within the same range in both proteins were restrained by minimizing the RMSD of the interatomic distances with respect to the crystal structure. Residues that differ in chemical shifts and the NOESY pattern or are unassigned in the C_L^{P35A} mutant alone were restrained toward zero native contacts. Furthermore, Φ and Ψ dihedral angles were restrained according to the values determined by Talos. To minimize the number of restraints, only residues located in secondary structure elements were taken into account. These restraints were applied in conjunction with a simulated annealing like protocol to enhance the sampling of the conformational space. This approach resulted in an ensemble of structures which best represent the partially folded state (Fig. 3.17).

The simulations verify that the overall topology of C_L^{P35A} is on average well retained. Additionally, both helices are fully structured in the simulations and display only minor fluctuations. Strikingly, the edge strands *a* and *g* adopt native conformations in all simulated structures. However, the ensemble of structures reveals considerable heterogeneity at the ends of some strands which is especially pronounced around Ala-35, arguing for a partial distortion of one edge of the protein by the *trans* state of the peptide bond preceding residue 35. The only strand that experiences significant fluctuations is strand *d*. Just residues Val-53 and Leu-54 of this strand form native like interactions in the ensemble of structures which is in good agreement with NOE signals observed exclusively for these residues within strand *d*. Furthermore, the N termini of strand *e* and strand *g* show some flexibility as indicated by a mixture of structures with full or partial ordering of the respective strands in the MD simulations. Interestingly, the partial disordering of strand *e* and one of its ends provides a rationale for the missing NMR assignment of some residues from its flanking strand *b* which itself is highly structured in the simulations.

Taken together, the combination of NMR experiments and MD simulations presents a powerful approach to obtain a detailed picture of the major C_L folding intermediate. The two small helices and their local environment are completely folded and the intermediate exhibits a native like core structure despite the presence of flexible regions that are able to adopt a variety of conformations.

2 The MD simulations themselves were run by Zu Thur Yew, a PhD student in the group of Dr. Emanuele Paci, Astbury Center for Structural Molecular Biology, University of Leeds, UK.

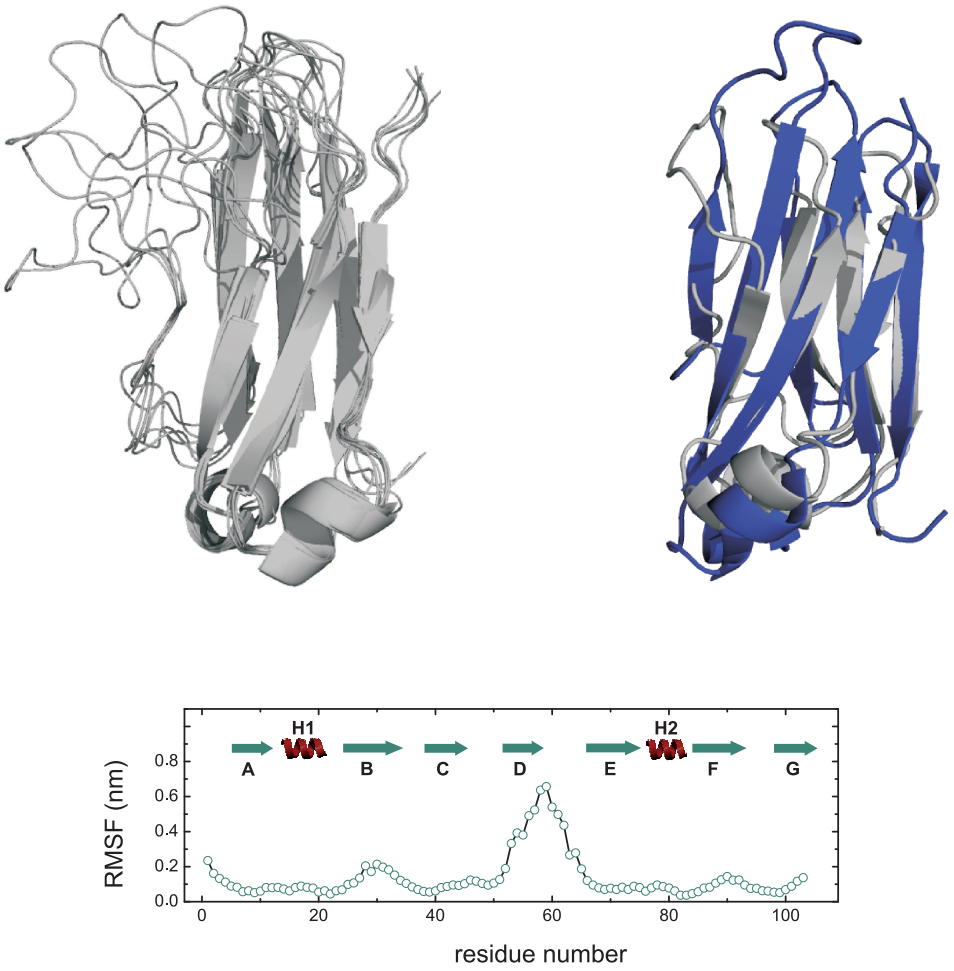


Figure 3.17: Structural characterization of the folding intermediate of the C_L domain by NMR-restrained MD simulations. Top left: Overlay of seven structures from the restrained C_L^{P35A} simulations to highlight flexible parts. Top right: Overlay of the native C_L structure (shown in blue) with the average of 30 structures from the NMR-restrained MD simulations of C_L^{P35A}. Bottom: Root mean square fluctuations (RMSF) for the last 30 C_L^{P35A} structures derived from the simulations.

Figure 3.18: Sequence of β_2m, C_L, and the β_2m$^{\text{toCL}}$ exchange mutant. The unstructured loop regions of human β_2m are depicted in blue, the corresponding helical elements in C_L are colored in red. The exchange mutant β_2m$^{\text{toCL}}$ was synthesized and the altered sequences between strands a and b as well as strands e and f are shown.

3.4 Amyloidogenic Properties of the C_L Domain and β_2-microglobulin

No indication exists that the constant domain of the antibody light chain is in any manner directly responsible for amyloidogenic processes even though it adopts the same topology as the amyloidogenic variable domain (V_L) or β_2-microglobulin. In both cases, amyloid formation is assumed to proceed from a partially folded intermediate state.[116,141] Interestingly, non of these proteins possesses the short strand connecting helices of the C_L domain. As these helical elements are already fully evolved in the C_L folding intermediate, their sequence or structure might play a role in the inhibition of amyloid formation. Consequently, the exchange of the unstructured loop regions connecting strands a and b as well as strands e and f in β_2m against the corresponding helical elements of the C_L domain (β_2m$^{\text{toCL}}$) could provide insights into the differences in amyloidogenicity (Fig. 3.18).

The β_2m$^{\text{toCL}}$ exchange mutant folds to a well defined structure with similar far-UV CD-spectroscopic properties like wild type β_2m. The transplanted helical elements destabilize β_2m$^{\text{toCL}}$ against thermal denaturation, but have only a minor effect on its pH stability as compared to β_2m. According to the Tango algorithm[142] that identifies β-aggregating regions of a protein sequence, the aggregation propensity of the primary sequence is left unaffected by the mutations. To study the amyloidogenic properties of the distinct proteins, fibrillization experiments were performed analogous to established reaction conditions for β_2m.[143] Each protein solution of C_L^{wt}, C_L^{P35A}, β_2m, and β_2m$^{\text{toCL}}$ respectively was incubated at pH 1.5 and pH 3.0 for 7 days under slight shaking at 37 °C. Additionally, all proteins were incubated under physiological conditions at pH 7.4 and 37 °C either seeded with β_2m fibrils or not. Formation of amyloid fibrils was monitored by atomic force microscopy (AFM).

As expected, C_L^{wt} and the C_L^{P35A} mutant are not at all prone to fibril formation. Importantly, an obvious difference in amyloidogenicity is observed for wild type β_2m and the β_2m$^{\text{toCL}}$ mutant (Fig. 3.19). Whereas β_2m readily forms fibrils under all conditions

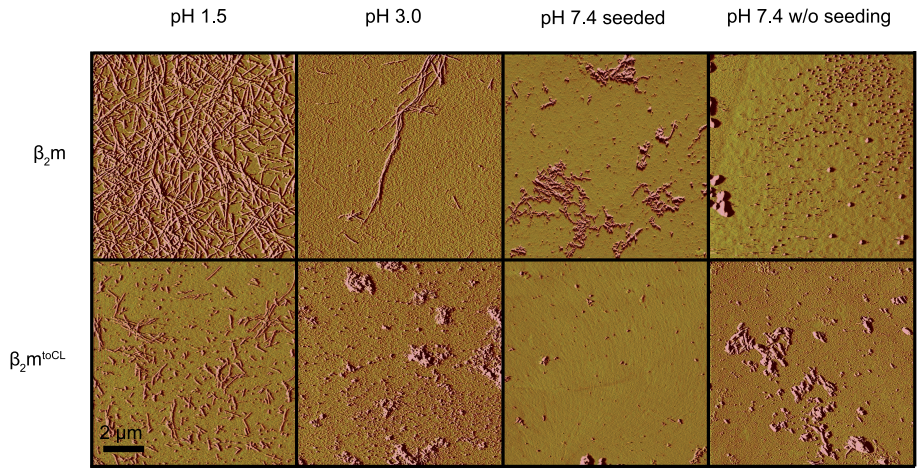


Figure 3.19: Amyloidogenic properties of β_2m and the β_2m$^{\text{toCL}}$ exchange mutant. The exchange mutant β_2m$^{\text{toCL}}$ clearly differs from wild type β_2m in respect of its amyloidogenicity.

tested, in the case of β_2m$^{\text{toCL}}$ just pH 1.5 induced the formation of fibrils, however these fibrils are considerably shorter and less pronounced as compared to wild type β_2m. Already at pH 3.0, amyloid fibrils were detectable only in very few experiments, and under physiological conditions fibril formation was unverifiable, even in cross-seeding experiments with sonicated β_2m fibrils. These data clearly demonstrate that transplanting the sequences corresponding to the C_L helices into the β_2m framework significantly reduces its amyloidogenicity.

3.5 Discussion

The structural characteristics of a folding intermediate decisively determine the folding landscape of a protein to either favor productive folding like for the C_L domain or to run the risk of harmful misfolding and aggregation like for β_2m. Therefore, a detailed picture of the folding intermediate of the C_L domain can provide insights into the determinants that distinguish the productive folding pathway of C_L from the aggregation prone folding pathway of the structurally similar β_2m. A combination of NMR experiments together with MD simulations reveals an ensemble of structures that represent the major kinetic folding intermediate of the C_L domain. These structures indicate that only minor variations in the sequences of highly homologous proteins can remarkably influence the folding landscapes and thus result in vital or pathogenic biomolecules.

The major kinetic folding intermediate of the C_L domain already adopts the typical immunoglobulin β-sheet topology for large parts of the protein. The only strand that seems to pass through a variety of distinct conformations and thus appears highly distorted in the ensemble of structures is the edge strand *d*. All other strands and in particular strands *b*, *c*, *e*, and *f* which constitute the folding nucleus of Ig proteins[144, 145] exhibit some dynamics but are already well structured. The most remarkable structural

feature of the intermediate are the two completely folded small helices. Although these helical elements are strongly conserved in constant antibody domains, their role in the folding process and their influence on the structure of the intermediate has not been recognized yet. These helices seem to fulfill a spacer and orienting function between strands *a*-*b* and *e*-*f* and provide hydrogen bond donors and acceptors for adjacent strands and loops. In addition, the helices appear to position hydrophobic residues (e.g. Tyr-80 in helix 2) that enables their participation in the formation of the hydrophobic core. Hence, these two helices can be regarded as a scaffold within the C_L intermediate favoring the establishment of a native like topology by correctly positioning important parts of the molecule.

A widespread assumption for the underlying mechanism of protein aggregation assigns the keystone to intermolecular β-sheet formation. β-strands buried within the hydrophobic core of the protein are naturally prone to aggregation as these strands always attempt to satisfy their hydrogen bonding potential. Certainly, the edge strands of β-sheet proteins must differ from the central β-strands in order to promote solubility over aggregation for natively folded proteins. Characteristically, these edge strands are irregular or short β-strands or otherwise unsuitable for further β interactions. Thus, the edge strands of β-sheet proteins are the decisive factor for prevention of amyloid like aggregation.[146,147] Interestingly, the edge strands *a* and *g* on one side of the β-sheet of the C_L intermediate are native like in all simulated structures. In contrast, in the amyloidogenic intermediate of β_2m these strands *a* and *g* and also strand *d* experience significant fluctuations resulting in disordered β-strands on both sides of the protein.[113–116,148] Distorsion of the edge strands of a β-sheet protein leaves the core β-strands unprotected leading to intermolecular β-sheet formation to satisfy the hydrogen bond potential which turns a soluble monomer into an aggregating oligomer. Therefore, the marked difference in amyloidogenicity between C_L and β_2m might be based on the distinct structuring of their edge strands. Whereas in the amyloidogenic intermediate of β_2m the central β-strands are unprotected on both sides of the protein molecule enabling a linear arrangement of monomers into fibrils, within the C_L intermediate just the edge strand *d* on one side of the protein is partially distorted, however on the other side of the β-sheet the edge strands *a* and *g* fulfill their protective function and thus might prevent oligomerization.

Since the helical elements in the C_L intermediate inhere a crucial role in productive folding, their function as scaffold elements might be responsible for the higher degree of local structuring particularly pronounced for the edge strand *a* as compared to the intermediate of β_2m. Indeed, grafting the sequences that correspond to the C_L helices onto the corresponding positions in β_2m significantly reduces the propensity of β_2m to form amyloids. As these helices represent robust folding elements, their structural characteristics seem to cause the marked difference in amyloidogenicity. However, sequence effects by the transplanted elements themselves cannot be completely ruled out. Besides, β_2m readily forms fibrils even though it is considerably more stable than C_L.[111,113] Accordingly, the most important factor that determines the amyloidogenic propensity of a protein is not the stability of the native state per se, but rather the

sequence of the protein[149] and structural characteristics of partially folded species that may be populated along the folding pathway.[143]

In conclusion, these data clearly demonstrate how a high degree of local structuring in a protein folding intermediate can substantially influence the folding landscape and favor robust folding over harmful misfolding. The distinct properties of C_L and β_2m can be understood in evolutionary terms. Selection of antibodies took place under harsh extracellular conditions with high concentrations of the multimeric protein present,[150] whereas β_2m is found at much lower concentrations and usually associated with the MHC complex.[151] Thus, small differences, acquired over the course of evolution, between members of the same protein superfamily can lead to the avoidance of pathogenic misfolding reactions while preserving an identical protein topology.

Chapter 4

The Antibody Domain C_H1 Controls the Secretion of IgG Molecules

Extracellular proteins are essential for many processes underlying multicellular life. Prerequisite for their biological functions is proper folding into the native tertiary structure and also assembly into a defined quaternary structure.[152] A particularly important example in this respect are IgG antibodies in which two heavy and two light chains have to associate prior to secretion from the endoplasmic reticulum (ER).[153,154] While isolated antibody light chains can be secreted from the ER, unassembled Ig heavy chains are actively and efficiently retained in the ER by interaction with the molecular chaperone BiP.[155,156] Since the heavy chains comprise most antibody effector functions, a tight quality control of their assembly prior to secretion is vital.[152] It is known that the first constant domain of the heavy chain, the C_H1 domain, plays a crucial role in this retention process.[153,155] If deleted or replaced with another antibody domain, isolated heavy chains can be secreted as it occurs in the case of the rare heavy chain diseases[157,158] or naturally in camelid antibodies.[159] In the context of the whole IgG molecule, the C_H1 domain adopts the typical immunoglobulin fold and is associated with the constant domain of the light chain, the C_L domain.

As the basic principles underlying the unusual behavior of the C_H1 domain have remained enigmatic, the role of the C_H1/C_L association for correct antibody assembly was studied in more detail. In general, the individual domains of an immunoglobulin molecule exhibit highly similar three dimensional structures and fold to their native state autonomously.[134,160,161] In contrast, the isolated C_H1 domain of the heavy chain is a natively unfolded protein. However, the C_L domain, the cognate association partner of C_H1 in the antibody, is necessary and sufficient to induce structure formation in C_H1. Structural characterization of the C_H1 domain and its association coupled folding reaction by NMR spectroscopy revealed an atomic level description of the folding

pathway that proceeds via a partially structured encounter complex. Specific recognition of a few key interface residues between C_L and C_H1 results in the formation of a folding nucleus in the C_H1 domain. Subsequently, the isomerization of peptidyl-prolyl bonds within C_H1 which possesses an unusually high number of three *cis* prolines in the native state[110] paves the path to the completely structured C_H1 domain.

The interaction of C_H1 with the molecular chaperone BiP could be reconstituted *in vitro* which identified the folding status of the C_H1 domain as the key determinant for chaperone binding and allowed to establish an overall mechanism for antibody secretion control *in vitro*. Furthermore, *in vivo* experiments in which the folding characteristics of the C_H1 domain were transplanted into the IgG light chain proof that its natively unfolded nature plays indeed the decisive role in antibody secretion control in the cell. These data provide insights into the folding pathway and the assembly of natively disordered proteins and highlight the structural plasticity within a highly conserved protein topology to fulfill specific biological requirements.

4.1 Introduction

4.1.1 Secretion of Antibody Molecules

Proteins destined for secretion, for the plasma membrane, and for the secretory and endocytic organelles are synthesized by ribosomes associated with the membrane of the endoplasmic reticulum (ER).[152] These ribosomes release the nascent protein chains into the ER which provides an environment optimized for protein folding and maturation. Like the lumen of other organelles of the secretory pathway, the ER lumen is extracytosolic and therefore topologically equivalent to the extracellular space. To ensure that only native conformers reach their target organelles and compartments, the ER possesses a variety of quality control mechanisms for 'proof-reading' newly synthesized proteins.[162–164] The primary quality control mechanism is based on general biophysical and structural properties such as the exposure of hydrophobic regions or the tendency to aggregate that distinguish native from non-native protein conformations. During the folding and assembly process into the native conformer, these features are naturally transient resulting in a temporary interaction of newly synthesized proteins with the quality control factors. These molecular chaperones and folding enzymes that reside in the ER lumen in high concentrations[165–167] stabilize incompletely folded polypeptide chains, protect them from aggregation, and also in many cases promote folding. Once folded, the cargo molecules are no longer retained in the ER and consequently enter the exit sites. This export of individual proteins is regulated by various selective mechanisms which constitute the secondary quality control mechanism.[163,168] Each molecule involved in the secondary quality control has its own specific recognition mechanism for certain protein species or protein families. These molecules assist in assembling of oligomeric proteins, accompany native conformers out of the ER, and provide signals for intracellular transport. After passing through this quality control

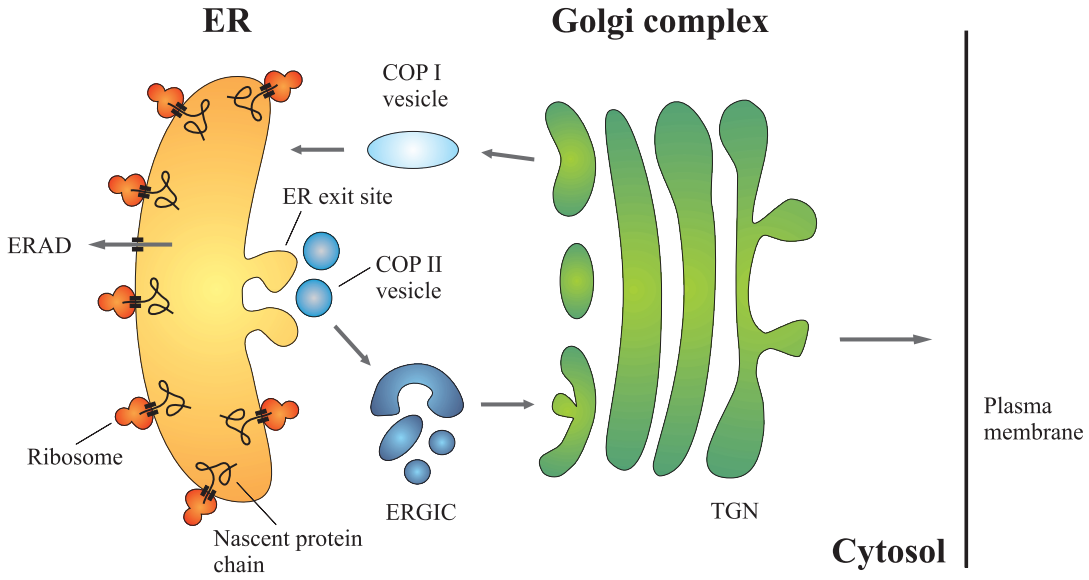


Figure 4.1: Organelles involved in the secretory pathway. The endoplasmic reticulum (ER) is the site of synthesis and maturation of proteins entering the secretory pathway. Once folded and assembled into the native conformer, the cargo molecules enter the ER exit sites. Vesicles that are coated with the coatomer protein (COP)II bud off and traffic through the ER-Golgi intermediate compartment (ERGIC) to the *cis*-face of the Golgi complex. In certain cases, the retrieval of misfolded proteins from the Golgi complex by COPI vesicles occurs. On their way through the Golgi complex the proteins become further modified, like phosphorylated or glycosylated and finally sorted according to their function and destination. After the proteins passed through the *trans*-Golgi network (TGN), they proceed to the plasma membrane or beyond.

natively folded and assembled proteins leave the ER for the Golgi complex (Fig. 4.1). In contrast, partially folded and incompletely assembled proteins are not transported to their final destination in the cell, but retained in the ER. Persistently misfolded proteins and unassembled oligomers are retro-translocated from the ER into the cytosol and subsequently degraded by proteasomes, a process named ER-associated degradation (ERAD).[169–172]

Recent studies indicated that 30-75% of newly synthesized proteins and in some cases even growing nascent chains are degraded within 20 min.[173–175] The reason for this apparently wasteful behavior of the cell is still unknown. Most of the peptides bound to the MHC class I molecules on the surface of T cells originate from the ongoing degradation of newly synthesized and nascent proteins.[176] Therefore this apparent wastefulness might actually be an essential function, as in the case of an infection of the cell by a virus and production of viral protein, this infection can rapidly be detected by the T cells.[175] Nevertheless, this immense production of degradation prone proteins also results in various diseases. If polypeptides in non-native conformations are not recognized efficiently or the degradation system fails, these proteins aggregate and accumulate in the cell and tissues leading to the formation of degradation resistant amyloids.[177]

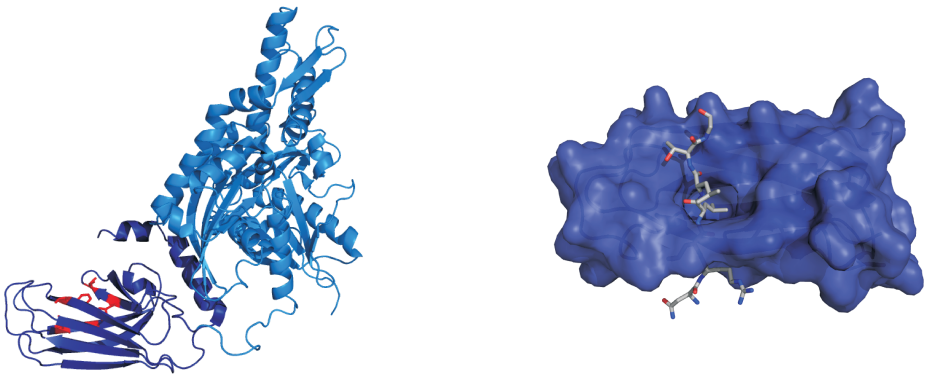


Figure 4.2: Immunoglobulin heavy chain binding protein (BiP) is one of the most abundant proteins in the ER lumen. Left: Cartoon representation of the crystal structure of bovine Hsc70 (pdb 1yuw),[178] a well studied representative of the Hsp70 class. Like all members of this chaperone family, the molecule possesses two separable functional units: the N-terminal (light blue) ATPase activity and the C-terminal (dark blue) substrate binding that is almost completely determined by the interaction with a five residue core (depicted in red). Right: Surface representation of the substrate binding domain of DnaK, another representative of the Hsp70 family, bound to a heptapeptide (pdb 1dkz).[179] This peptide interacts with DnaK in an extended conformation through a channel defined by loops from the β sandwich. An increased flexibility in the outer loops of the binding domain in the ATP state frees up the binding site for substrate release and rebinding.

In addition to the reliable recognition of non-native protein conformations, the ER has to fulfill a second equally important function, the assistance in productive folding and assembly. Although some proteins fold quite efficiently, many of them fold inefficiently.[180, 181] Moreover, especially hetero-oligomeric proteins show a poor success rate of maturation and single subunits as well as incompletely assembled oligomers have to be retained in the ER and eventually retro-translocated for degradation.[182] The ER specific Hsp70 chaperone family member BiP plays an essential role in directing protein folding and assembly as well as targeting misfolded proteins for retrograde translocation. This molecular chaperone identifies non-native protein conformations by recognizing a short heptapeptide of aliphatic residues in alternating positions.[183–185] Even though this is a common motif present in most proteins, the interaction of BiP with this binding site is specific and limited to a few secretory pathway proteins.[167] Besides the fact that naturally these recognition motifs are exposed only transiently and buried rapidly while the protein folds to its native state, the binding to the hydrophobic peptide sequence depends not only on the exposure but also on the dynamic properties and the sterically accessibility.[186] Several studies indicate that BiP undergoes cycles of binding and release from unfolded or partially folded proteins with folding occurring during the release cycle.[187] Like all members of the Hsp70 family, BiP also binds both ADP and ATP. Due to differences in conformation and mobility between the ATP state and the ADP state, substrate binding activity can be modulated by the ATPase unit (Fig. 4.2).[188]

This molecular chaperone BiP also controls the folding of individual immunoglobulin domains and the assembly of mature antibody molecules into the correct quaternary structure.[155] It has been shown that BiP binds temporarily to the variable domain of certain light chains to assist in folding and to remove improperly folded chains from the secretory pathway.[167] Properly folded immunoglobulin light chains can be secreted from the ER without association to the heavy chain.[189] In contrast, BiP retains unassembled heavy chains inside the ER and prevents their secretion or transport to the cell surface.[157] While the interaction with the constant region domains of heavy chains (C_H2, C_H3) is also transient,[153] BiP remains bound to the first constant domain of the heavy chain (C_H1) in the absence of light chain synthesis.[157] Maybe for this reason the light chains are expressed in excess of the heavy chains, to ensure high rates of heavy chain incorporation into the oligomeric complex.

Despite the highly similar three dimensional structure of all antibody domains investigated so far, the C_H1 domain varies significantly from other immunoglobulin domains in respect of stable BiP binding. Due to the nature of BiP to recognize hydrophobic peptide sequences, the C_H1 domain must be either intrinsically unfolded or incompletely folded or the folded C_H1 domain possesses exposed hydrophobic residues before assembly with the light chain.

4.1.2 Intrinsically Disordered Proteins

Protein folding has attracted enormous research effort for many years now and discovering the underlying mechanism is one of the major challenges of modern science. In the early 1960s Anfinsen suggested, based on the experimental experience that proteins fold reversibly, that native structures of globular proteins are thermodynamically stable states and therefore conformations at the global minima of their accessible free energies.[190,191] However, already a few years later Levinthal remarked that there are too many possible conformations for proteins to find the native structure within the whole conformational space by random searching.[192] This argumentation, called 'Levinthal's paradox', led to the assumption that proteins must fold by specific 'folding pathways'.[193] These two approaches differ fundamentally, as Anfinsen proposed a thermodynamically controlled folding mechanism which means that a protein reaches its global minimum in energy and that folding is pathway independent but time consuming due to the extensive search. In contrast, Levinthal favored a kinetically controlled mechanism depending on the pathway and hence resulting in a fast folding process but as a consequence the proteins may only reach local minima and display variable final structures.

In recent years a new model arose which has revolutionized our understanding of the protein folding process. As folding of a polypeptide chain does not involve starting from one specific conformation, the new perspective replaces the pathway concept of sequential events with the funnel concept of parallel events.[194–196] In this case, individual chains move asynchronous, each adopting different conformations. However, as the folding chains progress toward lower intrachain free energies their conformational

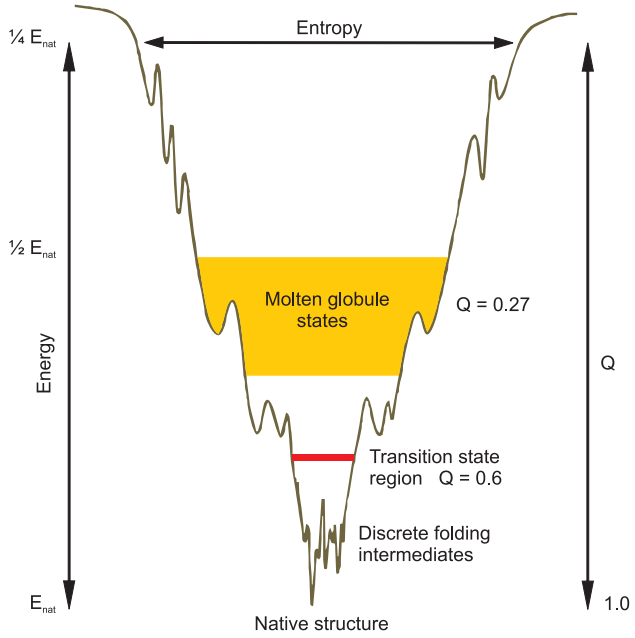


Figure 4.3: Schematic folding funnel with indicated regions for molten globules and transition states. The width of the funnel represents the entropy, the depth of the funnel relates to the native energy and the slope of the folding funnel provides an indication of the velocity of the folding process. Q corresponds to the fraction of native contacts correctly made. On its way to the native structure, the protein may be transiently trapped by any local minimum. This effect of ruggedness of the funnel landscape slows down the diffusion through the conformational space. The rate of finding the folded state depends on the conformational diffusion rate as well as the free energy barriers arising from the tradeoff between energy and entropy.

options increasingly narrow and finally result in one native structure. Its most important point, namely the prerequisite that protein folding proceeds via multiple routes going downhill rather than a single pathway, has immediately elegantly shown a way out of the long standing Levinthal paradox.

This new model of 'folding funnels' or 'energy landscapes' (Fig. 4.3) assumes that the natural fluctuations in the conformation of an unfolded or incompletely folded polypeptide chain enable even residues that are highly separated in the amino acid sequence to come into contact to one other.[105, 117, 197] Because in the majority of cases native like interactions are more stable than non-native ones, each polypeptide chain is able to find its lowest energy structure by a process of trial and error. Moreover, many studies indicate that the fundamental mechanism of protein folding involves the interaction of a relatively small number of residues to form a folding nucleus.[197–199] Interactions of these key residues force the polypeptide chain to adopt an overall topology comparable to the native fold. Once the correct topology has been achieved, the native structure will almost invariably emerge during the final stages of folding (Fig. 4.4). Conversely, if these key interactions are not established, the protein cannot fold to a stable globular structure.

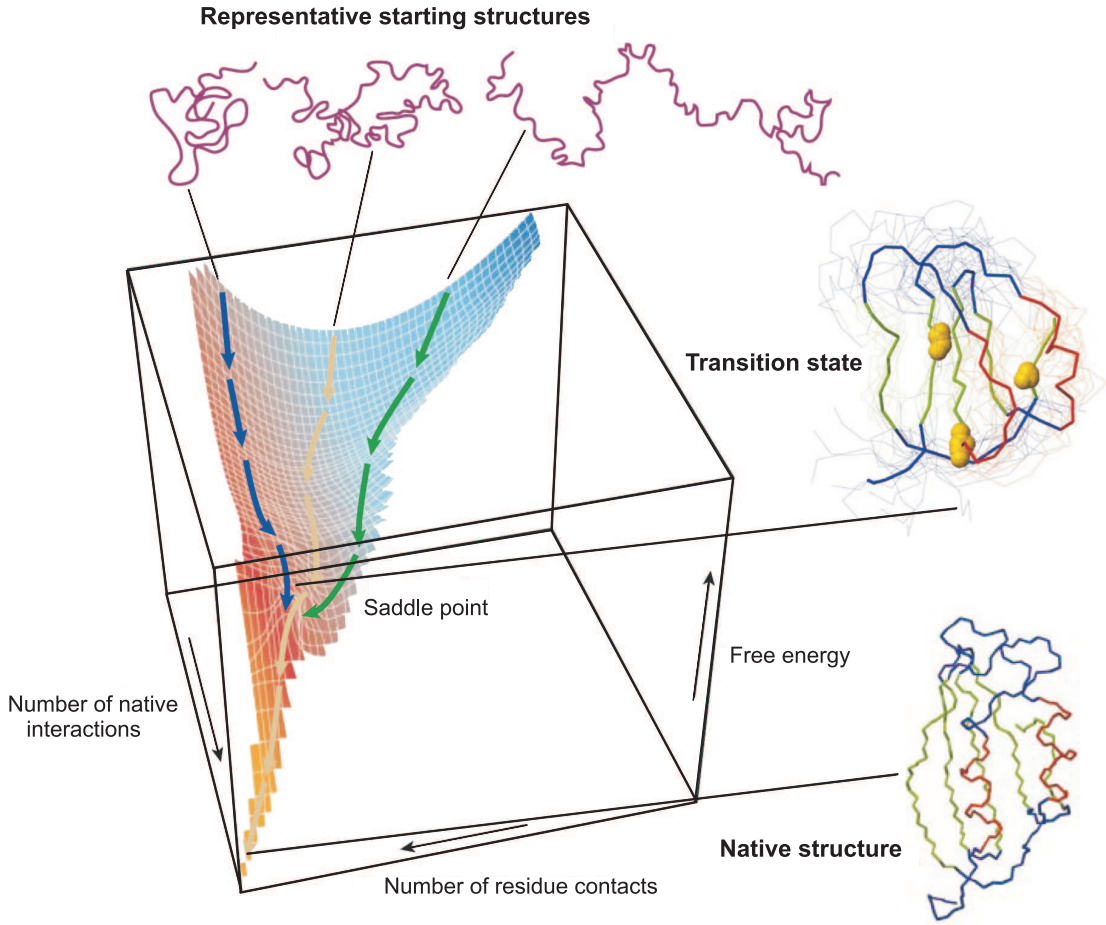


Figure 4.4: Schematic energy landscape for protein folding.[105] An ensemble of distinct conformations of unfolded polypeptide chains finds its lowest energy structure by a process of trial and error. Depending on the shape of the landscape, only a small number of all possible conformations needs to be sampled, enabling a few key residues to come into contact and forming a folding nucleus. When these residues (depicted in yellow in the transition state) form their native like contacts, the overall topology of the native state is established. This transition state still consists of an ensemble of structures reflecting the multiple pathways of protein folding. After reaching this saddle point, the native structure will almost invariably be generated during the final stages of folding.

Hence the question raises, what distinguishes a protein that navigates with remarkable ease through a complicated energy landscape to its native structure from a polypeptide chain that gets lost in this landscape unable to reach its native fold in a reasonable period of time. The most realistic model of a protein is a minimally frustrated polypeptide with a rugged funnel like landscape biased toward the native structure.[200] This folding landscape is necessarily rugged because biomolecules can sample many conformations during their motions and have the possibility of making inappropriate contacts between residues. For a unique native state to be thermodynamically stable, conformations structurally similar to this global minimum must be thermodynamically favored over dissimilar ones. For the global minimum to be kinetically accessible, the states that are low in energy must be dissimilar in structure. These states act as traps during the folding process. Heteropolymers with randomly chosen amino acid sequences have rough energy landscapes because of conflicts between various interactions between the residues.[200] It is believed that sequences of foldable proteins must be non-random to the extent that in the sampled conformations these conflicts are locally minimized. This so called 'principle of minimal frustration' suggests that folded proteins have an energy landscape that is not too rough and in which, on average, the energy decreases the more the molecule resembles the native structure.[201] Polypeptides acting on this principle of minimal frustration have a considerably smoother energy landscape than a random heteropolymer, and in addition there are progressive forces which funnel the protein toward the native state. However, not all of the individual interactions in a protein need to be minimally frustrated for the landscape as a whole to be funneled. Even in folded proteins some frustrations from conflicting interactions may be present locally without affecting the protein's foldability nor its mutational robustness.[202] Yet, these frustrated interactions come along with the possibility of kinetically trapping the polypeptide chain on its way downhill the funnel to the native structure. Rapid search through the folding landscape aimed toward the specific native structure requires the landscape to have an overall gradient that is large compared to the local ruggedness.[203] In the case the local ruggedness exceeds the slope of the funnel, forces guiding the protein toward the native state diminish. Accordingly, the protein is stuck in the rough landscape and not able to reach its native fold in a reasonable period of time. This highly rugged landscape corresponds to the energy landscape of an intrinsically unfolded protein.[204] As the shape of the energy landscape is encoded by the amino acid sequence, the ability of a protein to navigate successfully through the complicated energy landscape is determined by its primary amino acid sequence.

For many years now there has been a common conviction in structural biology that a folded protein structure is essential for biological function. Of course protein structure and function are intimately connected, but the awareness that not all biologically functional proteins fold spontaneously into stable globular structures gains in importance. An extensive bioinformatic survey revealed that intrinsically disordered proteins are rather a general feature than a rare exception.[205,206] Due to the persistence of sequences encoding natively unfolded proteins throughout evolution, these non-globular

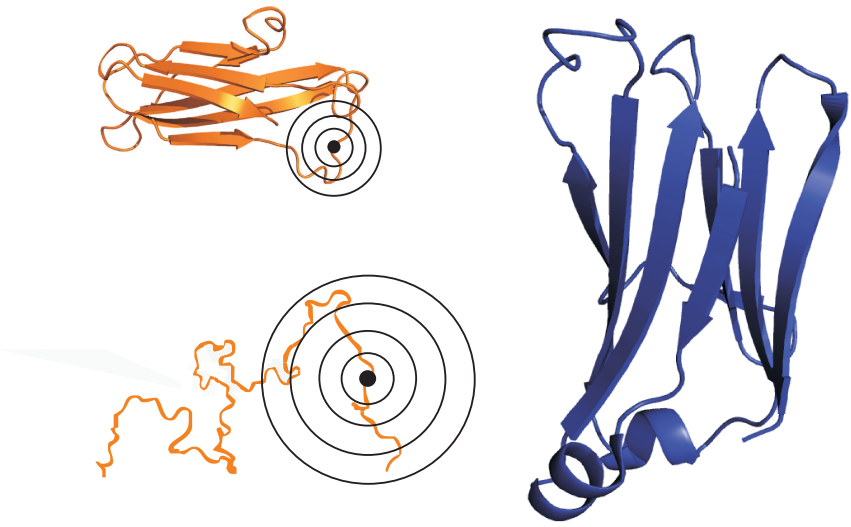


Figure 4.5: Schematic illustration of the 'fly casting' mechanism. A partially structured or unstructured protein has a greater capture radius than a folded protein with its limited flexibility. Therefore, the partially folded ensemble is already able to form a few initial contacts to the binding site while the folded structure remains out of range because of the smaller fluctuations in the folded state. Although these initial contacts are weak, if they are native like, the disordered chain can complete folding and binding simultaneously. Accordingly, the speed of association is enhanced for a natively unstructured protein.

structures seem to play a pivotal role in cellular biology.[207] Originally, biomolecular recognition processes were illustrated as the association of folded proteins that dock as rigid bodies, known as the 'lock and key' mechanism.[41] Another more complex model proposed for biomolecular recognition is based on the formation of an encounter complex that undergoes small local conformational changes to optimize the initial interactions, named the 'induced fit' mechanism.[208] As it turns out that many cellular proteins appear to be partially or even completely unstructured under native conditions, but form a perfectly ordered structure in the presence of the appropriate ligand, the mechanism of 'folding upon binding' describes a further possibility for biomolecular recognition.[209,210] In this case the disordered polypeptide chain forms an encounter complex with a rigid protein and generates random binding contacts with the surface of the folded protein while staying unfolded itself. If the energy landscape for binding is strongly funneled and consequently correlated with the folding landscape, the disordered polypeptide chain folds upon binding. Otherwise, if the binding interactions are non-native and therefore not strong enough to induce a funnel like folding landscape, the complex dissociates.[211]

Since intrinsically disordered proteins always run the risk of being degraded within the cell, there must be physiological advantages for the use of unstructured proteins that only fold upon reaching their targets as biomolecular recognition process. On the one hand, natively unfolded proteins are more adaptive thus possessing the capability

to bind to several distinct targets as well as to overcome steric clashes.[212,213] On the other hand, these extended polypeptide chains are able to form complexes with large interfaces. To achieve a comparable extensive interface, a stably folded protein would have to increase in size immensely.[214] As a consequence, disordered proteins provide a simple solution for having large intermolecular interfaces while maintaining a small genome. Furthermore, the 'fly casting' mechanism postulates a kinetic advantage for being initially unfolded before binding.[215] A partially structured or unstructured protein has a greater capture radius than a folded protein with its limited flexibility for a specific binding site, thereby enhancing the speed of association (Fig. 4.5).

4.2 Investigation of the Structural Characteristics of the C_H1 Domain

The biological activity of IgG molecules depends on proper assembly into the correct quaternary structure. For this reason, the endoplasmic reticulum (ER) possesses an elaborate assembly control mechanism which ensures the association of two heavy and two light chains into a functional antibody molecule prior to secretion from the ER.[153,154] An essential process within this complex control mechanism is the retention of unassembled heavy chains in the ER by interaction with the molecular chaperone BiP.[153,156] It is known that the first constant domain of the heavy chain, the C_H1 domain, plays a crucial role in this retention process.[153,155] Hence, investigation of the structural characteristics of the C_H1 domain might reveal further insights into the secretion control mechanism.

4.2.1 The Unassembled C_H1 Domain is Natively Unfolded

In the context of the whole antibody molecule, the C_H1 domain adopts the typical immunoglobulin fold and associates with the constant domain of the light chain C_L.[46] Surprisingly, the isolated C_H1 domain from a murine IgG1, despite its ability to form its intrinsic disulfide bridge, is a natively unfolded protein. This is in marked contrast to all antibody domains studied so far.[160,161,216,217]

Besides the typical random coil CD spectrum, the isolated C_H1 domain also displays a ^{1}H-^{15}N HSQC spectrum characteristic of an unstructured protein (Fig. 4.6). Moreover, the intensities of the HSQC peaks differ significantly, thus indicating distinct dynamical properties within the protein. In the case of the C_H1 domain, decreasing the temperature resulted in altered backbone dynamics in this respect that the amide resonances show less line broadening and consequently the spectral quality increased considerably. As the carbon chemical shifts predominantly depend on the amino acid type itself, the dispersion in the carbon dimensions is sufficient for the assignment of the backbone resonances even in unstructured proteins. The backbone resonances of the isolated C_H1 domain (98 AA) were assigned at 12 °C, due to the favorable dynamical characteristics at lower temperatures and consistent with the real time measurements

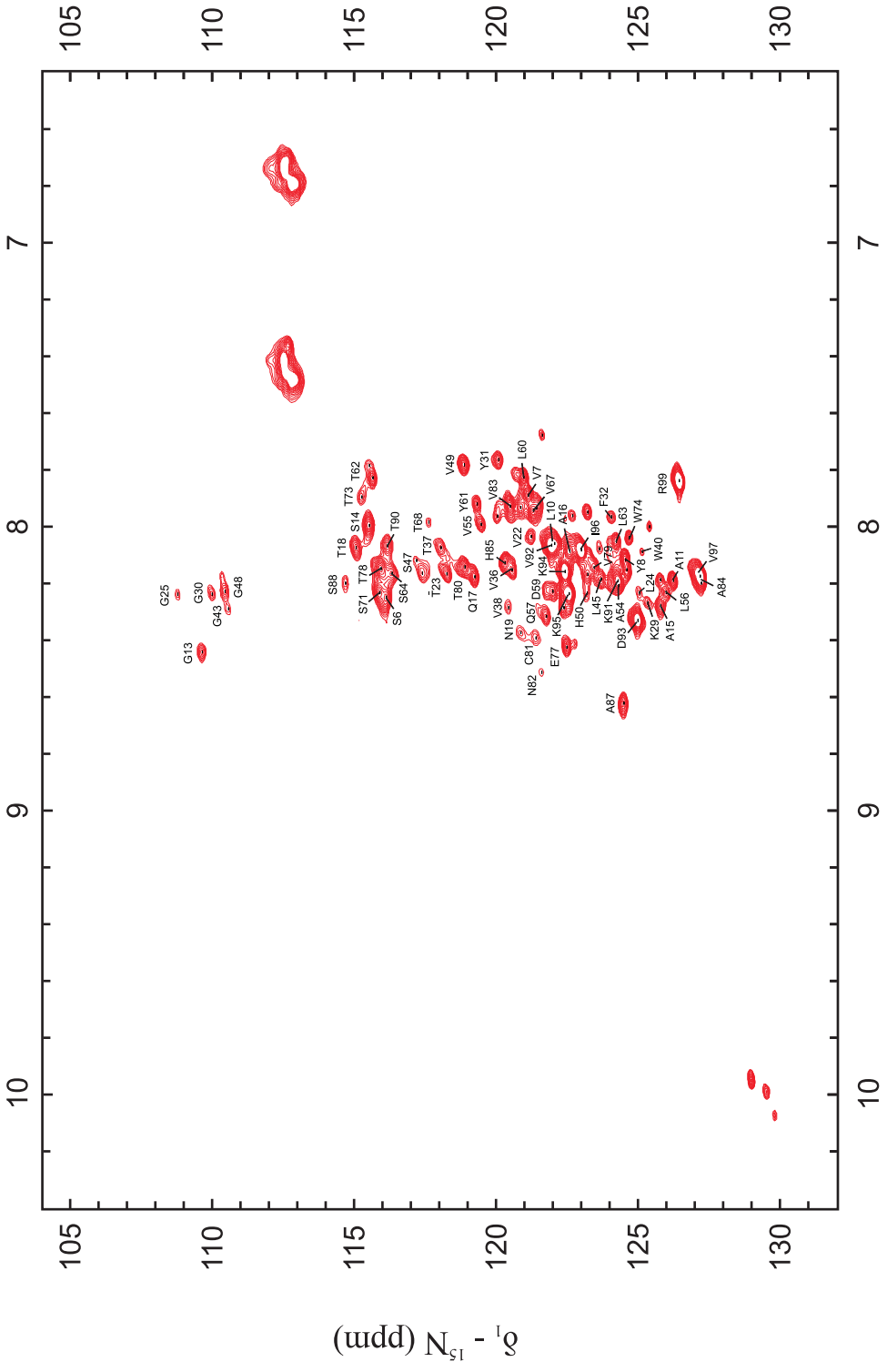


Figure 4.6: ^{1}H-^{15}N HSQC spectrum of the isolated C_H1 domain (98 AA) recorded at a temperature of 12 °C and a proton frequency of 900 MHz. The dispersion of the peaks in the proton dimension is characteristic of an unstructured protein. Besides, the intensities of the peaks differ significantly, thus indicating distinct dynamical properties within the unfolded C_H1 domain.

during folding, recording standard triple resonance experiments and performing the semi-automatic assignment procedure as described for the C_L domain (Section 3.2.1). Despite the unfolded nature, an assignment of 77% of the backbone resonances was achieved (67 residues out of 87 non-proline residues). Interestingly, the HSQC spectrum of the reduced C_H1 domain is identical to the one of the oxidized C_H1 domain, except for slight changes in the chemical shifts of the cysteines themselves and the directly adjacent residues. Hence, the unassembled C_H1 domain is natively unfolded irrespective of the formation of the internal disulfide bridge.

Analysis of the carbon chemical shifts according to the chemical shift index (CSI, section 3.2.2) indicates no existence of secondary structure elements within the isolated C_H1 domain. Nevertheless, it is assumed that even unstructured proteins are structured in terms of formation of a hydrophobic core or transient interactions. HetNOE data of the unfolded C_H1 domain suggest that the unassembled C_H1 exists as an ensemble of relatively flexible structures experiencing fluctuations comparable to those of disordered regions in folded proteins. To gain information about any preferential conformations present in the disordered C_H1 domain, NOESY spectra on a highly deuterated sample were recorded. Deuteration substantially lengthens the relaxation times of carbon and proton spins in proximity to the substituted deuterons,[218] due to the lower gyromagnetic ratio of deuterons relative to that of protons ($\gamma_H/\gamma_D \approx 6.5$). This slower decay of the diagonal peak intensity as a function of mixing times enables an increase in the sensitivity of NOESY experiments through the use of longer mixing times.[219] However, the drawback of complete deuteration for NOESY experiments is that only NOEs between exchangeable protons which means predominantly amide protons can be observed. Still, long mixing time NOESY spectra recorded on perdeuterated unfolded proteins allow observation of longer range H^N-H^N NOEs belonging to only a subset of conformers even though these cross peaks are relatively weak due to the low population and the dynamic properties.[220] Yet, measurement of NNH- and HNH-NOESY spectra with a mixing time of 600 ms on the highly deuterated disordered C_H1 domain in aqueous PBS buffer could not detect any long range NOEs. Instead, an exchange with protons from the aqueous buffer is observed for all residues. Consequently, the unfolded state of C_H1 seems to be an ensemble of relatively flexible conformations lacking specific amide-amide contacts or hydrogen bonds. As the disulfide bridge present in the unfolded state brings at least some residues in close proximity, these might just interact by their hydrophobic side chains. Study of possible hydrophobic interactions and the required assignment of the aliphatic and aromatic side chains is still under progress.

Despite a not yet available detailed picture of the unfolded C_H1 domain, the intrinsically disordered nature already provides a rationale for the strong interaction with the chaperone BiP and the pivotal role of this domain in retaining the unassembled heavy chain in the ER.

4.2.2 Association coupled Folding of the C_H1 Domain

In contrast to the unassembled C_H1 domain, in the presence of the C_L domain, the cognate association partner of C_H1 in the complete antibody molecule, the C_H1 domain folds to its characteristic greek key β-barrel topology. However, this association coupled folding process depends on the redox status of the C_H1 domain [1]. Reduced C_H1 is not able to assemble with C_L, as evidenced by anisotropy measurements of Lucifer yellow labeled C_L where no change in anisotropy was observed upon addition of reduced C_H1 as compared to oxidized C_H1. Therefore, the internal disulfide bridge is prerequisite for the association of the C_H1 domain with the C_L domain and consequently for inducing structure formation in C_H1.

Backbone Assignment

Since the unfolded state and the folded state of the C_H1 domain are in slow exchange, the peaks of the backbone resonances of the unfolded state disappear during the folding process while the peaks corresponding to the folded state appear at different positions. Therefore, the backbone resonances of the structured C_H1 domain have to be newly assigned. To ensure complete structure formation in C_H1, twofold excess of unlabeled C_L was added to ^{15}N respectively ^{15}N, ^{13}C labeled C_H1 and incubated for at least 6 h at room temperature prior to the NMR measurements. For the backbone assignment the standard triple resonance experiments (Section 3.2.1) were recorded at 25 °C in order to increase the overall tumbling of the heterodimer and thus slow down the decay of magnetization. However, the carbon chemical shifts were limited to the C' and C^α chemical shifts due to the relaxation properties of the protein complex. Although the molecular weight of around 20 kDa is not really critical for NMR measurements, the shape of the C_L/C_H1 complex seems to be disadvantageous in regard to the correlation time. The increased correlation time results in an increased loss of magnetization via the H^α nuclei making the magnetization transfer from the C^α nuclei to the C^β nuclei inefficient. As the amino acid type and sequential information just based on the C' and C^α chemical shifts does not allow an unambiguous assignment of the C_H1 domain, an additional parameter is required to verify and expand the backbone resonance assignment. Due to the low stability of the heterodimer neither increasing the temperature in a range that considerably improves the relaxation properties nor deuteration of the complex provides an appropriate solution. However, the analysis of residual dipolar couplings shows an elegant way out of this challenging task. Since in the case of the C_H1 domain the crystal structure in complex with the C_L domain is already available, dipolar couplings calculated from the known structure can be compared with experimental values. The program Mars[221, 222] routinely includes RDCs into the backbone assignment of proteins with known structure. To achieve the necessary partial alignment of the assembled C_H1 domain with respect to the magnetic field for the measurement of the

1 Since this project was conducted in collaboration with the group of Prof. Dr. Johannes Buchner, Lehrstuhl Biotechnologie, Department Chemie, Universität München, the biochemical experiments and also the protein sample preparation were performed by Matthias Feige.

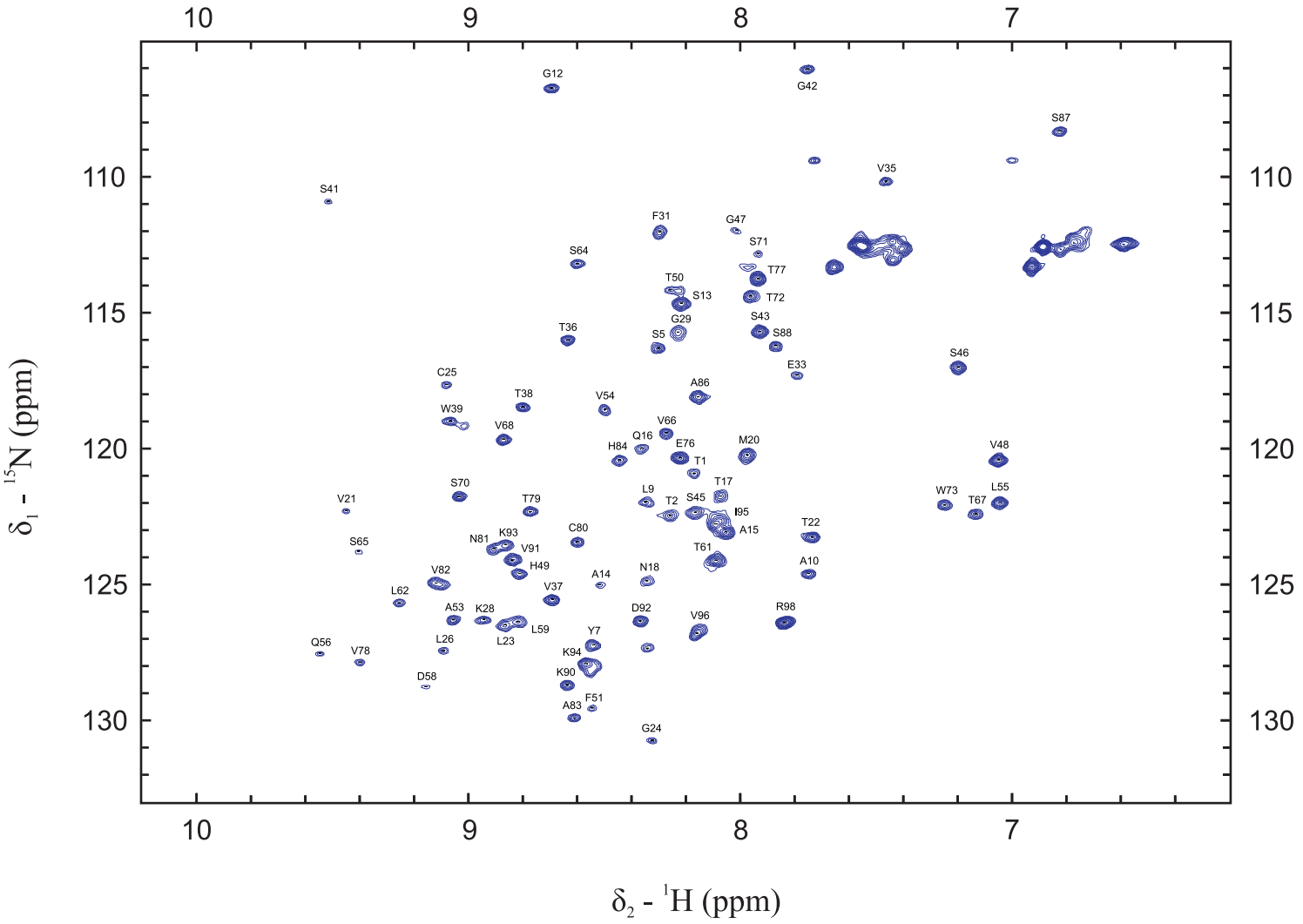


Figure 4.7: ^{1}H-^{15}N HSQC spectrum of the C_H1 domain (98 AA) in complex with the C_L domain recorded at a temperature of 25 °C and a proton frequency of 750 MHz. To ensure complete structure formation in C_H1, twofold excess of unlabeled C_L was added to ^{15}N respectively ^{15}N, ^{13}C labeled C_H1 and incubated for at least 6 h at room temperature prior to the NMR measurements.

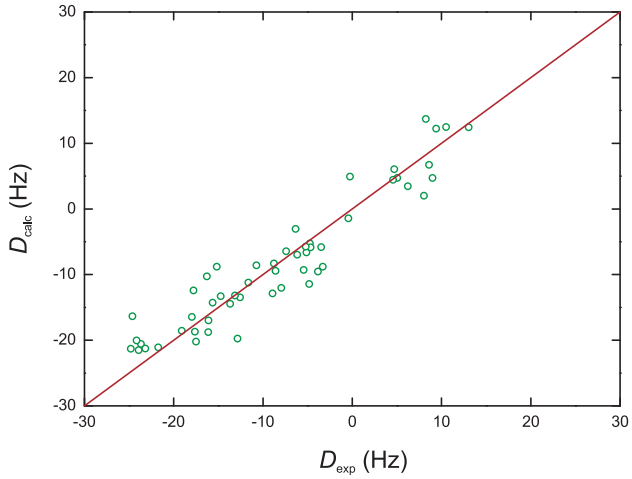


Figure 4.8: Correlation between the backcalculated and the experimental $^1D_{NH}$ couplings of the folded C_H1 domain. The backcalculated values for the dipolar couplings were derived from the RDC-based backbone resonance assignment by the program Mars. These values based on the known crystal structure of the assembled C_H1 domain agree quite well with the experimental N-H RDCs (quality factor $Q = 0.24$; correlation factor $R = 0.95$).

residual dipolar couplings, non-ionic liquid crystalline media were applied to the NMR sample as published by Otting and coworkers.[223] Since the C_H1 domain associated with the C_L domain displays well dispersed spectra (Fig. 4.7), RDC values could be extracted from 2D IPAP-HSQC spectra. These experimental N-H dipolar couplings in combination with the amino acid type and sequential information of the C' and C^α chemical shifts as well as the crystal structure of the assembled C_H1 domain (pdb 1ors) serve as input for the software Mars. This approach extended the backbone resonance assignment to 76 residues out of 87 non-proline residues. In addition, the considerable correlation between the backcalculated and the experimental $^1D_{NH}$ couplings (Fig. 4.8) indicates an identical structure of the C_H1 domain in complex with the C_L domain in solution and in the crystal.

Folding Kinetics

The folded C_H1 domain possesses an unusually high number of three *cis* prolines.[110] The energetically almost equal *trans* and *cis* conformation of peptide bonds preceding a proline residue result in an equilibrium distribution of around 10-30% *cis* conformation per proline residue in an unfolded polypeptide chain (Section 3.2.3). For proteins with *cis* prolines in the native state the required isomerization reaction from the mainly populated *trans* to the *cis* conformation is associated with a high activation energy. Whereas *in vivo* this peptidyl-prolyl isomerization is enzymatically catalyzed, *in vitro* this reaction is an intrinsically slow process. Consequently, the time frame necessitated to complete folding *in vitro* is shifted into an appropriate range for real time NMR spectroscopy. To resolve the specific recognition between C_L and C_H1 as well as the folding pathway of the natively disordered C_H1 domain at the level of atomic resolution,

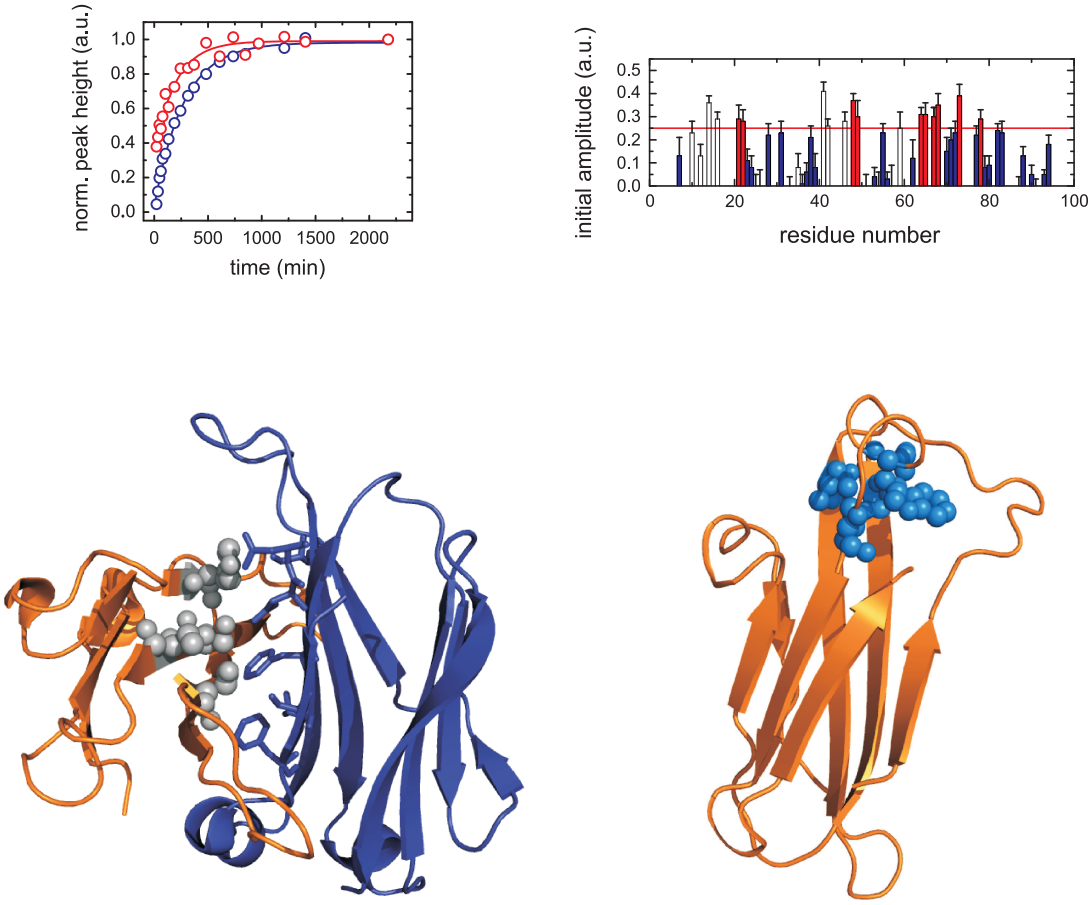


Figure 4.9: Folding kinetics of the C_H1 domain. Top left: The time dependent peak intensities of each assigned residue could be fitted by a single exponential function and extrapolated to time zero (blue: Lys-90 and red: Val-68 selected as representative residues). Top right: Initial amplitudes for each assigned C_H1 residue were derived from the single exponential fit. Residues in loop regions are shown as open bars, residues with an initial HSQC amplitude below the threshold of 25% native intensity are colored blue and residues above the threshold red. Bottom left: Mapping the residues which already exhibit significant intensities after several minutes on the crystal structure of the C_H1/C_L dimer reveals key residues involved in the interface recognition and the hydrophobic folding nucleus. Residues in the C_H1 domain (orange) with initial amplitudes above the threshold which constitute the C_L interface are indicated as gray spheres, the corresponding residues in the C_L domain (blue) that are involved in this interaction are shown as sticks. Bottom right: The interaction with C_L apparently initiates the formation of a hydrophobic cluster in the C_H1 domain. Residues with initial amplitudes above the threshold that are not part of the interface between C_H1 and C_L but form the hydrophobic folding nucleus are depicted as blue spheres.

twofold excess of unlabeled C_L was added to ^{15}N labeled C_H1 and HSQC spectra were recorded immediately after mixing every 14 min at 12.5 °C.

For each assigned residue, changes of the amplitudes over time could be described by a single exponential function. Notably, some residues already exhibit significant intensities in the first spectrum recorded after 20 min. These residues can be assumed to adopt a native like conformation prior to the slow peptidyl-prolyl isomerization reaction. Mapping the residues which are part of the β-sheets that form the mature structure and are found to be in a native like environment, on the crystal structure of the C_H1 domain reveals how the association coupled folding reaction of this antibody domain might proceed (Fig. 4.9). Residues Thr-22, His-49, Ser-65, and Thr-67 in the C_H1 domain which constitute the C_L interface seem to be already correctly positioned in the encounter complex. Importantly, His-49 and Ser-65 are involved in hydrogen bonds with the C_L domain. The interaction with C_L apparently initiates the formation of a hydrophobic cluster in the C_H1 domain comprised of Val-21, Val-68, Trp-73, and Val-78. Thus, a few key interactions between C_L and C_H1 establish an interface between the two domains which allows the formation of a hydrophobic folding nucleus in the C_H1 domain and subsequently prolyl isomerization paves the path to the native state.

4.3 Insights into the Secretion Control Mechanism

Although association coupled folding of the C_H1 domain is accordingly an intrinsic feature of this natively unfolded protein, in the cell folding, assembly, and subsequent secretion of IgG molecules involves additional factors.[224] The molecular chaperone BiP that is present in high concentrations in the ER plays an important role in retaining the unassembled heavy chain in the ER.[154, 155] As BiP recognizes specific hydrophobic heptapeptides in extended conformations which are characteristic of partially structured or completely unstructured proteins, this process of retention is mainly based on the interaction of the intrinsically unfolded C_H1 domain with the chaperone. After synthesis of the polypeptide chain by the ribosome, the C_H1 domain of the unassembled heavy chain remains in a reduced state.[225] According to analytical HPLC experiments, BiP forms stable complexes with the reduced C_H1 domain as well as with the oxidized C_H1 domain, although with slightly less affinity. This indicates that the oxidation of the C_H1 domain could in principle take place in the BiP bound state. Since reduced C_H1 does not bind to C_L (Section 4.2.2) nor could association of C_L to BiP bound C_H1 be detected (Fig. 4.10), the data allow to establish an order of events for the reaction between BiP, C_H1, and C_L. It is most likely that *in vivo* the C_H1 domain early binds to BiP as it enters the ER cotranslationally in a reduced state. Then, triggered by association with the light chain, and presumably controlled by additional components of the Ig assembly machinery,[154, 224, 226] the oxidation of the internal disulfide bridge between Cys-25 and Cys-80 takes place which brings at least two of the residues (Val-21 and Val-78) that are involved in the formation of the hydrophobic folding nucleus in close proximity to each other.[111] Only after release from the chaperone, oxidized C_H1

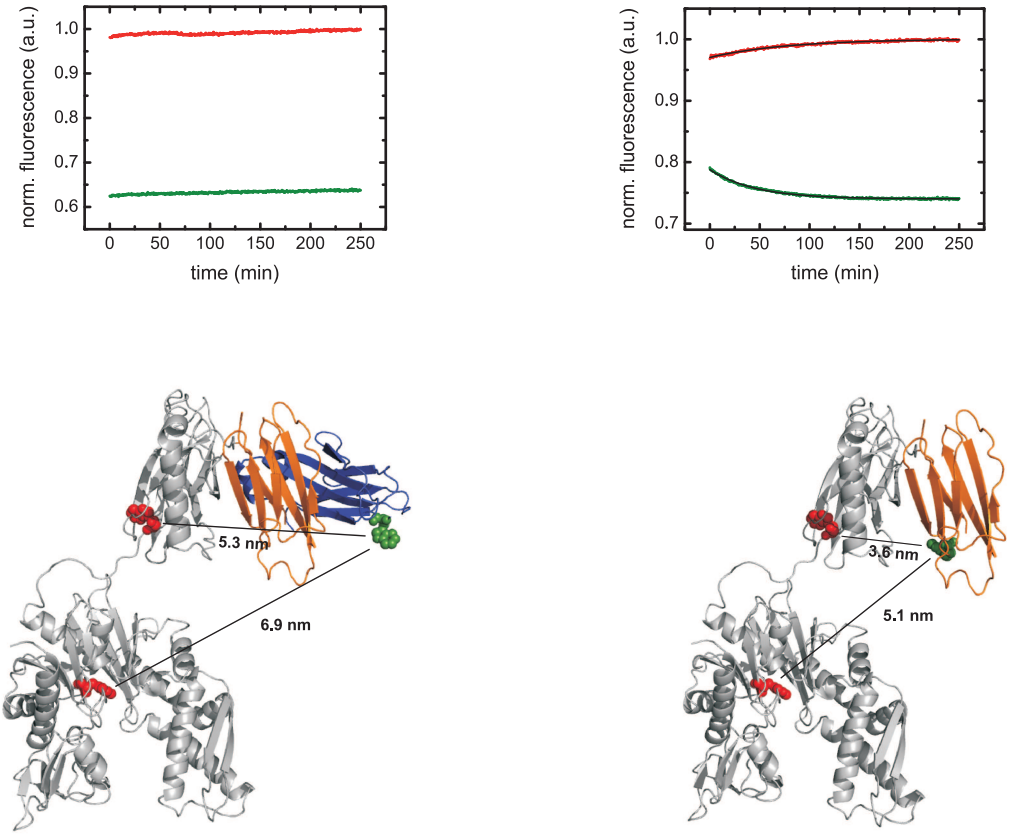


Figure 4.10: Verification of the formation of triple complexes between BiP, C_H1, and C_L. Top left: To investigate whether C_L can associate with BiP bound C_H1 or the C_L/C_H1 complex binds to BiP, ATTO594 labeled BiP and unlabeled C_H1 were added to ATTO532 labeled C_L. The donor, ATTO532, was excited at 500 nm and the donor fluorescence was recorded at 550 nm (green line), the fluorescence of the acceptor, ATTO594, was recorded at 625 nm (red line). Since no change in fluorescence can be detected, the formation of triple complexes does not occur. Top right: As a control of the Fret system, the association between ATTO594 labeled BiP and ATTO532 labeled C_H1 was measured. Bottom left: Cartoon representation of a putative triple complex between BiP (gray, model based on pdb 3c7n), C_H1 (orange), and C_L (blue). The labels are schematically shown in green (ATTO532) respectively in red (ATTO594) and distances are indicated. Bottom right: Illustration of the complex between BiP and C_H1 including the schematically depicted labels and the corresponding distances. The Förster radius of the dye pair is 6.6 nm according to the manufacturer.

can complete its folding in association with C_L which is in agreement with putative BiP binding sequences in the C_H1 domain[227] being also involved in the interaction with C_L. *In vivo* studies found that the C_H1 domain stays in the reduced state prior to light chain association.[154,225] It is conceivable that in the ER light chain association, oxidation, and folding of C_H1 are tightly coupled by additional components of the assembly machinery,[224] as the cell based experiments did not allow a temporal dissection of the events.[225] In the readily assembled antibody molecule, the C_L domain and the C_H1 domain are covalently cross-linked via an intermolecular disulfide bridge. Once this disulfide bridge is formed, it abolishes binding of C_H1 to BiP as evidenced *in vitro* by analytical HPLC experiments and thereby renders the association process between C_L and C_H1 irreversible. In addition, since the formation of this disulfide bridge depends on proper folding of C_H1, only incorrectly or incompletely folded C_H1 domains re-associate with BiP.

This comprehensive model puts the unexpected intrinsically unfolded nature of the C_H1 domain at the center of the secretion control mechanism of murine IgG antibodies. Therefore, the assumption that the general folding status of an antibody domain is governing retention in the ER and strong binding to BiP was further analyzed *in vivo* [2]. For this purpose, various immunoglobulin chain constructs were transiently expressed in the established eukaryotic COS cell line. This african green monkey kidney fibroblast line was metabolically labeled with ^{35}S methionine and cysteine respectively. Secreted immunoglobulin chains were immunoprecipitated from the expression medium, whereas immunoglobulin chains retained in the ER and also the ER specific chaperone BiP were immunoprecipitated from the cell lysate. Subsequently, analyzed SDS-polyacrylamide gels were exposed to XAR films to detect the ^{35}S labeled proteins.

To support the theory that the intrinsically unfolded nature of C_H1 is regulatory for secretion control, the C_L domain in the light chain was replaced by the C_H1 domain. This light chain now behaved like a heavy chain in terms of retention in the ER and interaction with BiP. This shows that the structural characteristics of the C_H1 domain and its role in antibody retention are intrinsic, context independent features. To more directly address the structural prerequisites for antibody retention, the small helical elements of the C_L domain which have been reported to play a crucial role in the folding of this domain (Chapter 3), were exchanged against the corresponding elements of the C_H1 domain which themselves have not been classified as BiP binding sequences.[227] This exchange transformed the C_L domain into a natively unfolded protein. An antibody light chain containing the altered C_L domain strongly interacted with BiP *in vivo* and was no longer secreted from the ER, arguing that the folding status of an antibody domain is key for its retention and that the transplanted residues might play a role in the folding properties of the C_H1 domain.

2 *In vivo* experiments were carried out by Dr. Yuichiro Shimizu, a postdoctoral fellow in the group of Prof. Dr. Linda Hendershot, Department of Genetics and Tumor Cell Biology, St. Jude Children's Research Hospital, Memphis, USA.

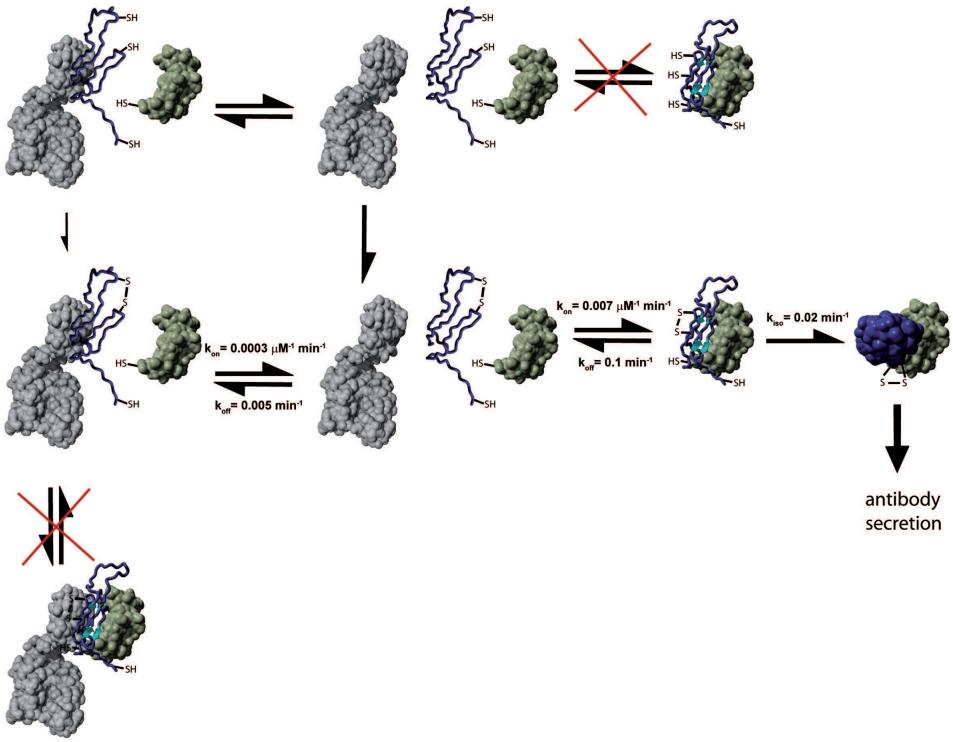


Figure 4.11: Secretion control mechanism of IgG antibodies. After synthesis of the polypeptide chain by the ribosome, the C_H1 domain of the heavy chain (blue) enters the ER in a reduced state. Due to its intrinsically unfolded nature, the C_H1 domain stably binds to the ER specific chaperone BiP (gray). This interaction involves cycles of binding and release from the chaperone which enables the assembly with the C_L domain of the light chain (green) during the release cycle. Since reduced C_H1 does not associate with C_L nor does C_L interact with BiP bound C_H1, the oxidation of the internal disulfide bridge is most likely controlled by additional components of the Ig assembly machinery. After release from the chaperone BiP, the oxidized C_H1 domain can complete its folding in association with the C_L domain. Once C_H1 is properly folded, formation of the intermolecular disulfide bridge which covalently cross-links the C_L domain and the C_H1 domain takes place. This renders the assembly of the heavy chain and the light chain irreversible and allows the secretion of the antibody molecule from the ER.

4.4 Discussion

Investigation of the structural properties of the C_H1 domain discovered unexpected characteristics. In general, the individual domains of an immunoglobulin molecule adopt highly similar three dimensional structures and fold to their native state autonomously.[160,161,216] Interestingly, the isolated C_H1 domain of the murine IgG1 antibody is an intrinsically unfolded protein, despite its capability to form its internal disulfide bridge. Since oxidation of this disulfide bridge could still take place in the BiP bound state, BiP seems to preserve C_H1 in an association competent state rather than directly promoting the association reaction with C_L. After release from BiP, C_H1 gains the typical immunoglobulin fold upon assembly with C_L, a reaction that turns irreversible by the formation of the interchain disulfide bridge. This association coupled folding reaction proceeds via a partially structured encounter complex. A detailed NMR spectroscopic analysis of the underlying pathway suggests that the reaction is initiated by the recognition of a few key interface residues between the C_L domain and the C_H1 domain which then promotes the formation of a hydrophobic folding nucleus in the C_H1 domain. Both reactions render each other energetically more favorable and thereby allow the entropically demanding structuring of an unfolded polypeptide chain. As for most antibodies secretion of isolated light chains as well as heavy chains devoid of the C_H1 domain can take place, the essential control step for the overall IgG assembly seems to be realized in the C_H1/C_L interaction. Taken together, these data provide a detailed molecular mechanism by which BiP and a natively unfolded antibody domain control the secretion of murine IgG1 antibodies.

A rigorous assembly control mechanism is of particular importance for antibody molecules which pass through asynchronous expression rates of heavy and light chains depending on the distinct developmental states of the lymphocytes. B cells arise from precursors in the bone marrow, so called pre-B cells. These pre-B cells rearrange heavy chain genes first which are largely retained in the cell and undergo rapid degradation[228] except for a limited number that assembles with the surrogate light chain.[229] The developmentally more mature B cells also produce light chains which assemble with the heavy chains and allow their transport to the cell surface. Finally, the terminally differentiated plasma cells produce enormous quantities of Ig molecules[230] that must be appropriately assembled to bind specifically to antigens and fulfill their effector functions. *In vivo* studies indicate that the control mechanism for all Ig isotypes focuses on the C_H1 domain which counteracts immature secretion.[157,231,232] For secretion of mature antibody molecules the C_H1 domain has to assemble with the C_L domain and adopt the characteristic immunoglobulin fold. This association coupled folding reaction of C_H1 in complex with C_L additionally requires disulfide bridge formation and peptidyl-prolyl isomerization, reactions that are readily accelerated by the ER folding machinery.[152,224,233,234] Association of C_H1 with either BiP or C_L are concentration dependent reactions and consequently will not become rate limiting at the high antibody and chaperone concentrations in the ER. Hence, the nature of the reactions governing C_H1 folding and assembly with C_L allow efficient yet accurate assembly of

antibodies prior to secretion and hint towards the co-evolution of substrates and folding helpers in the ER. These data not only provide novel insights into antibody assembly control but furthermore extend the growing class of natively unfolded proteins[209, 235] by a prominent member of the immunoglobulin superfamily. Since C_H1 does not possess the typical sequence signature of an intrinsically disordered protein,[236, 237] it might be the first representative of a novel type of natively unfolded proteins. The large number of proline residues in C_H1, e.g. in a helical part of the molecule which was identified as important for antibody domain folding might be one determinant of its unfolded nature. From the evolutionary point of view, the C_H1 domain has experienced a divergent evolutionary fate dependent on its folding and homodimerization properties. One of the oldest adaptive immune systems based on antibody molecules is found in sharks. Their antibody molecules assemble only from heavy chains that consist of single C_H1 homologous domains.[238] In contrast, the heavy-chain-only antibody molecules of camels are devoid of the C_H1 domain.[159] How the unexpected biophysical properties of this protein have evolved in different species to fulfill specific biological functions remains a challenging subject of current research.

Chapter 5

pH Dependent Talin Binding to Actin Promotes FA Remodeling

Vitally important for most anti-pathogen immune responses is the rapid production of antibody molecules. Yet, B cells specific for any given antigen are exceedingly rare, sometimes as few as one cell in a million.[239] For this reason, B cells have to ensure an efficient survey for their specific antigen. Naive B lymphocytes, B cells that have not yet encountered an antigen, continually enter the secondary lymphoid organs from the bloodstream. These secondary lymphoid organs are strategically positioned to sample antigens that enter through almost any body surface, the spleen filters antigens from the blood, lymph nodes filter lymph draining from skin or mucosal surfaces. After randomly migrating through a lymphoid organ without detecting a cognate antigen, B cells return to the circulation and move on to survey further lymphoid organs.[240] In contrast, upon antigen encounter cell migration undergoes rapid changes. Rather than migrating randomly, B cells now migrate with directional preference, pair with antigen specific helper T cells, and become antibody secreting plasma cells.[241] This whole process of an effective immune response decisively depends on the migratory properties of the cells. As in general cell migration is a central process in the development and maintenance of multicellular organisms, elucidation of the underlying mechanism has attracted great interest.

It is assumed that the rate limiting step in cell migration is the binding of actin filaments to the focal adhesion (FA) associated proteins that stabilize the adhesions between the cell and the substrate. At the leading edge of migrating cells, focal adhesions undergo rapid cycles of assembly and turnover, creating and disrupting, respectively, sites of traction necessary for forward movement of the cell body. Force generation for traction requires linkage among the extracellular matrix, integrin receptors, and the actin cytoskeleton. Actin filaments do not directly bind to the cytoplasmic domain

of integrins but bind to integrin associated focal adhesion proteins such as talin.[242] Although F-actin binding by talin is known to be pH sensitive *in vitro*, with lower affinity at higher pH,[243,244] the functional significance of this pH dependence remains unclear. Because increased intracellular pH (pH_i) promotes cell migration[245,246] and is a hallmark of metastatic carcinomas,[247] the question arises whether it increases FA remodeling through lower affinity talin actin binding. Talin contains several actin binding sites,[248] however only the C-terminal USH-I/LWEQ module binds actin in a pH dependent manner, with lower affinity and decreased maximal binding at higher pH. NMR spectroscopic investigations of this talin module revealed a structural mechanism for pH dependent actin binding. A cluster of titratable amino acids with upshifted pK_a values, including His-2418, was identified at one end of the five helix bundle distal from the actin binding site. Protonation of this histidine residue induces changes in the conformation and dynamics of the remote actin binding site. Furthermore in motile fibroblasts, increasing pH_i decreased FA lifetime and increased the migratory rate. These data suggest a molecular mechanism for pH sensitive actin binding by talin and indicate that FA turnover is pH dependent and in part mediated by pH dependent affinity of talin for binding actin.

5.1 Introduction

5.1.1 Cell Migration

Cell migration is an essential prerequisite for many physiological and pathophysiological processes. Already the origin of life requires cell migration shortly after conception to enable the morphogenesis throughout embryonic development. Also in adult life, cell movement continues to play an extremely important role in processes such as tissue repair and immune surveillance. For an efficient immune response, the leukocytes have to migrate from the circulation into the surrounding tissue to destroy invading microorganisms and infected cells.[239] After accompanying us throughout the whole life, cell migration even contributes to our death. Migration is involved in several pathological processes, including vascular diseases, chronic inflammatory diseases, and cancer.[249] Hence, understanding the fundamental mechanism underlying cell migration holds the promise of effective therapeutic approaches for treating diseases.

From the present point of view cell migration can be illustrated as a cyclic process.[250] The initial response of a cell to a migration promoting agent is to polarize and extend protrusions in the direction of migration (Fig. 5.1). These protrusions are stabilized by adhering to the extracellular matrix (ECM) or adjacent cells via transmembrane receptors linked to the actin cytoskeleton. The adhesions at the cell front serve as traction sites for migration, whereas the adhesions at the cell rear disassemble, allowing the cell to move forward.

The basic requirement for cell migration is the polarization of the cell, which means that the distribution of signaling molecules and the cytoskeleton is asymmetrical be-

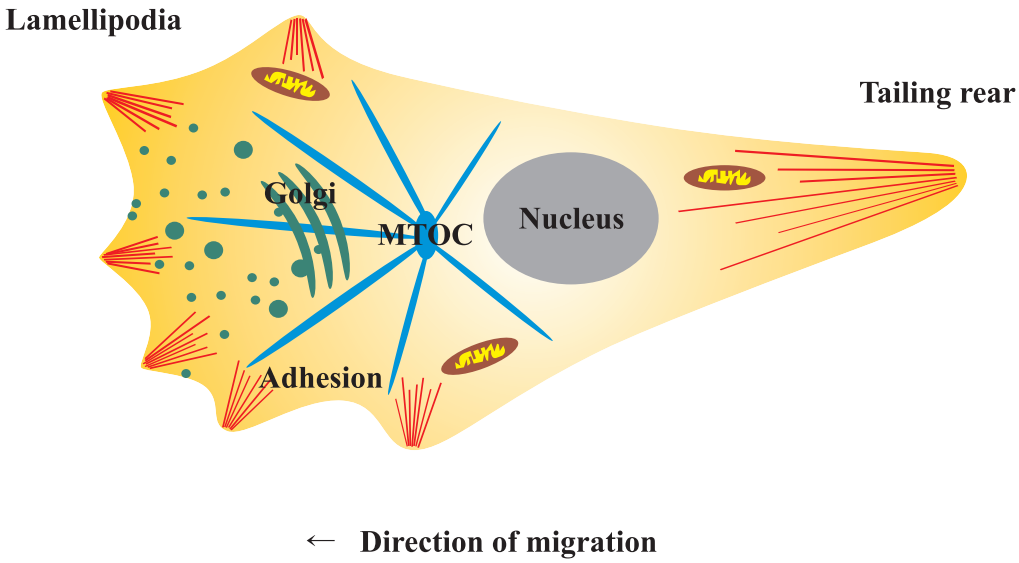


Figure 5.1: Schematic illustration of a migrating cell. The initial response of a cell to an extracellular stimuli is to polarize and extend protrusions in the direction of migration. Polarization of the cell means asymmetrical distribution of signaling molecules and the cytoskeleton between the front and the rear of a moving cell. In addition, polarity is also affected by localizing the microtubule-organizing center (MTOC) and the Golgi apparatus in front of the nucleus which facilitates microtubule growth into the lamellipodia and delivery of Golgi derived vesicles to the leading edge providing membrane and associated proteins required for forward protrusion. These protrusions are stabilized by adhering to the extracellular matrix (ECM) or adjacent cells via transmembrane receptors linked to the actin cytoskeleton. The adhesions at the cell front serve as traction sites for migration, whereas the adhesions at the cell rear disassemble, allowing the cell to move forward.

tween the front and the back of a moving cell. Establishing and maintaining cell polarity in response to extracellular stimuli seems to be arranged among other components by integrins, microtubules, and vesicular transport. Cells are able to respond directionally to marginal differences in concentrations of migration promoting agents.[251] This shallow gradient between the front and the rear of the cell needs to be intensified into steeper intracellular signaling gradients in order to initiate a cellular response. Several signaling molecules exist that accumulate at the leading edge of the cell stimulated by a extracellular chemoattractant and thereby activate further molecules involved in this complex regulatory machinery.[251,252] Molecules downstream the regulatory pathway can in return activate the signaling molecules, and additionally stabilize microtubules as well as induce recruitment and clustering of activated integrins at the leading edge of the cell to form new adhesions.[253–255] To ensure a directional migration rather than a random movement the cell possesses mutual antagonists that suppress each others activity.[256] Molecules activated by the extracellular chemoattractant accumulate at the leading edge of the cell suppressing their antagonists, whereas the antagonists localize at the sites and the rear of the cell thus preventing protrusion at sites other than the leading edge.[257,258]

The actual protrusive event is assumed to occur not just by elongation of the actin filaments but by a mechanism, in which thermal energy bends the nascent short filaments. Unbending of an elongated filament against the leading edge of the cell provides the driving force for pushing the membrane forward.[259] In this protrusion process distinct actin binding proteins govern the actin polymerization. The pool of available monomers and free ends is regulated by proteins that prevent self-nucleation of actin monomers, restrict polymerization of new filaments close to the membrane, and disassemble older filaments into actin monomers which are required for the polymerization at the front end.[259, 260]

After a protrusion has formed, it has to be stabilized by attaching to the surroundings. Many different receptors contribute to the adhesion, however integrins play the leading part. Integrins are heterodimeric receptors consisting of α and β chains with large ligand binding extracellular domains and short cytoplasmic domains, thus providing a linkage between the extracellular matrix and via adaptor proteins the actin filaments on the inside of the cell (Fig. 5.2). Interactions of the cytoplasmic tail with the cytoskeletal linker protein talin alters the conformation of the integrin extracellular domains thereby activating the integrins which preferentially localize to the leading edge of the cell to form new adhesions.[261, 262] Migrating cells can be distinguished according to the extent of their integrin clusters ('focal adhesions') between rapidly migrating ones, such as leukocytes which possess few integrin clusters and cells with large integrin clusters that are either non-migratory or move very slowly. Besides the density of adhesion receptors on the cell, the migration velocity and hence the strength of cell attachment is determined by the density of adhesive ligands on the substrate and the affinity of the receptors for the adhesive ligands.[250] In addition to connecting the extracellular matrix to the intracellular cytoskeleton, integrins also transfer information about the physical state of the ECM into the cell and correspondingly vary the cytoskeletal dynamics.[263, 264]

At the rear of migrating cells, adhesions must disassemble to promote retraction of the tail. In fibroblasts, the rearmost adhesions often tether the cell strongly to the substratum, resulting in a long tail. At any time, the tension is sufficient to physically break the linkage between the integrin and the actin cytoskeleton, allowing the cell to move on.[250] Integrins detached from the cytoskeleton and the ECM are, at least partially, recycled by endocytosis to the front of the cell. Furthermore, intracellular calcium levels also favor adhesion turnover at the rear. The tension generated by the strong adhesions at the back of the migrating cell can open certain calcium channels.[265] Increased intracellular calcium concentration activates the protease calpain, which has the potential to cleave several focal adhesion proteins, including integrins, talin, and vinculin.[266, 267]

Although research into the molecular basis of cell migration has progressed rapidly over the past few years, there are still many unresolved questions regarding how cells establish and maintain their polarity, how adhesions form and break, how cells migrate *in vivo*, and how cells recognize their targets.

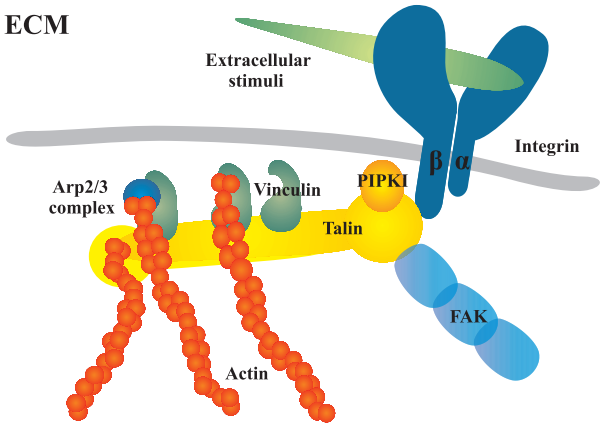


Figure 5.2: Schematic illustration of the linkage between integrin receptors and the actin cytoskeleton. Talin connects actin filaments to focal adhesions directly through the C-terminal I/LWEQ domain and the N-terminal FERM domain and indirectly through the vinculin binding sites. Furthermore, talin functions not only as dynamical linkage between the cytoplasmic tail of β-integrins and the actin cytoskeleton but also increases the affinity of the extracellular domain of integrins for stimuli of the ECM.[268, 269]

5.1.2 The Focal Adhesion Protein Talin

Adhesions regulate cell motility in response to extracellular cues by assembly and disassembly of molecular interactions between the cell and the substrate. Their structures can be classified into focal complexes, focal adhesions, and fibrillar adhesions. This classification depends on their size, shape, intracellular localization, molecular composition, and dynamics.[270–272] Focal complexes disassemble and reassemble at the leading edge during protrusion in a process called focal adhesion turnover. Thus, these focal complexes are highly dynamic nascent adhesions and in the early state mainly established by β_3-integrin, talin, and paxillin.[273, 274] In slow moving cells, some focal complexes mature into focal adhesions. Focal adhesions localize at the cell periphery and more centrally in less motile regions. Connecting the cell body to the extracellular matrix requires more stable adhesions with slower turnover rates. These focal adhesions contain high levels of vinculin, talin, paxillin, focal adhesion kinase (FAK), and integrin $\alpha_v\beta_3$.[274] However, the molecular mechanism underlying the decision of a focal complex to mature into a more stable focal adhesion or to disassemble and turnover remains unclear. Fibrillar adhesions can arise from focal adhesions and characteristically associate with the extracellular fibronectin fibrils.[275]

The focal adhesion protein talin emerges as key modulator of adhesion dynamics (Fig. 5.3).[276–278] Talin is an antiparallel homodimer[279] made of a small globular head domain (∼50 kDa) and a large elongated rod domain (∼220 kDa). The N-terminal FERM domain directly binds integrin α_{IIb}, β_1, and β_3 cytoplasmic tails,[280, 281] FAK,[282] PIP2,[283] and weakly to actin filaments. The rod domain contains an I/LWEQ actin filament binding site[248] as well as multiple vinculin binding sites[284] and also mediates dimerization via a C-terminal coiled coil segment. Hence, talin connects actin filaments

to focal adhesions directly through the I/LWEQ and the FERM domains and indirectly through the vinculin binding sites. Furthermore, talin functions not only as linkage between the cytoplasmic tail of β-integrins and the actin cytoskeleton but also increases the affinity of the extracellular domain of integrins.[268] Talin binding to integrins disrupts an intracellular salt bridge between the α and β integrin subunits, leading to increased integrin affinity, which strengthens the interaction with the ECM - a phenomenon known as 'inside out' integrin activation.[285–287]

In vivo studies with talin deficient flies revealed that in the absence of talin integrins are still able to bind the ECM, but can neither accumulate into clusters nor connect to the cytoskeleton.[288] Another approach verified that talin is the major integrin effector of inside out integrin activation, since the knockdown of talin reduces integrin activation irrespective of the many distinct signaling pathways and cannot be rescued by other integrin regulators but only by the expression of an integrin activating fragment of talin.[262] Maybe due to the vital role of talin in adhesion formation and integrin activation, mammals have two highly similar isoforms of talin (talin1 and talin2 with $\sim$74% identity) suggesting redundant functions.

During migration, adhesions assemble and disassemble. One way to regulate disassembly is the cleavage of adhesion components by proteases, such as calpain.[289] Various mechanisms activate calpain *in vivo*, though the best known mechanism is the activation by calcium.[290] Except under pathological conditions associated with cell death, the levels of calcium required to activate calpains maximally *in vitro* do not exist within living cells.[291] This apparent paradox led to the assumption that other regulatory mechanisms must be able to reduce the amount of calcium needed *in vivo*. Diverse modes of regulation have been identified *in vitro*, for instance binding of phospholipids[292] or protein-protein interactions[293] which decrease the calcium requirement for calpains, but their relevance *in vivo* still has to be corroborated. Friedrich postulated another explanation for this paradox.[294] During evolution calpain proteases developed this need for high amounts of calcium as a safety device to prevent potentially destructive hyperactivity and therefore it is preferable for calpains to work at much less than half maximal activity. Furthermore, multiple phosphorylation sites additionally control the activity of calpain proteases.[295] Interestingly, so far no single consensus sequence could be identified for predicting whether a protein can be proteolyzed by calpains or even where calpains cleave a known substrate. However, most substrates are cleaved in a limited fashion in disordered regions between structured domains. Several indications exist, that the proteolysis of talin regulates focal adhesion turnover, which might present the principle function of calpain mediated adhesion disassembly.[296]

Besides the formation of focal adhesion contacts to the extracellular matrix,[263, 297] cell migration necessitates reorganization of the cytoskeleton,[298, 299] activation of ion channels and transporters,[245, 300, 301] and membrane recycling by endo- and exocytosis.[302, 303] Activity of the plasma membrane isoform of the Na^+/H^+ exchanger, NHE1, is a prerequisite for cell polarity, directed migration, and also contributes to cell adhesion.[304] Predominantly, NHE1 regulates the intracellular pH homeostasis by electroneutral substitution of intracellular H^+ for extracellular Na^+ thereby protecting the

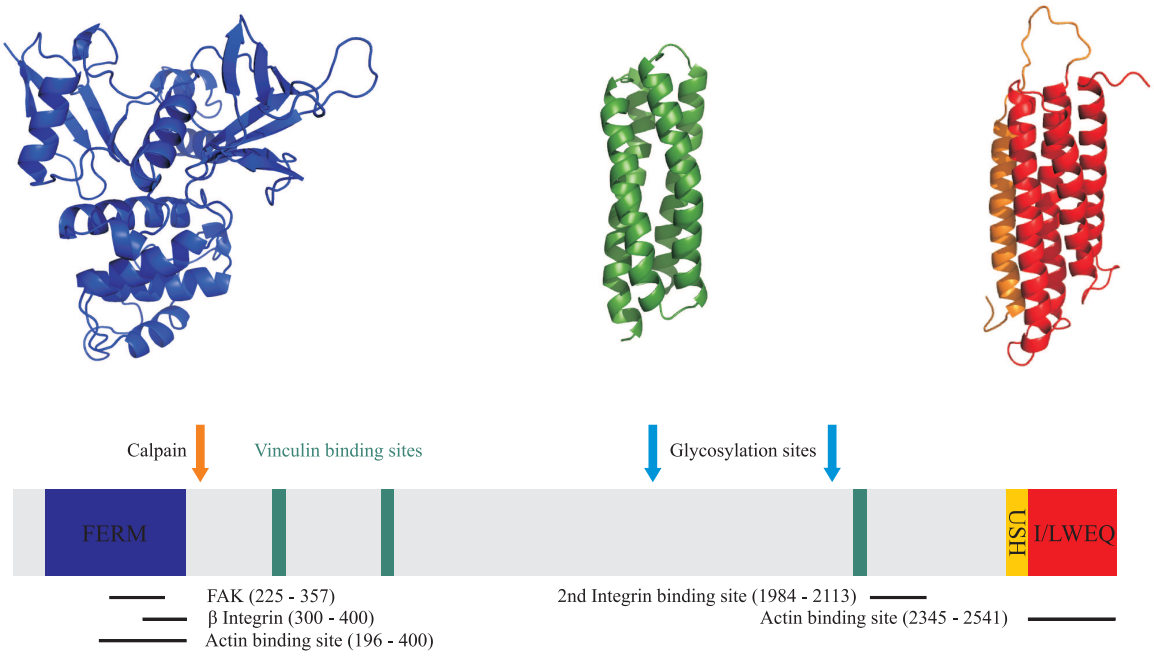


Figure 5.3: The focal adhesion protein talin. Talin (2541 residues) is a flexible antiparallel dimer that plays a decisive role in integrin mediated adhesion. The N-terminal FERM domain (residues 86-400, depicted in blue, model based on pdb 1mix) directly binds integrin α_{IIb}, β_1, and β_3 cytoplasmic tails,[280, 281] FAK,[282] PIP2,[283] and weakly to actin filaments. The rod domain contains a highly conserved I/LWEQ actin binding site[248] (residues 2341-2541, illustrated in red, pdb 2jsw) as well as multiple vinculin binding sites[284] (residues 607-636, 852-876, and 1944-1969, shown in green, pdb 2b0h) and also mediates dimerization via a C-terminal coiled coil segment (residues 2501-2541). In addition, the rod domain comprises a second integrin binding site.[242] Calpain II, an isoform of the calcium activated protease, cleaves talin between residues 433-434 generating a small globular head domain (∼50 kDa) and a large elongated rod domain (∼220 kDa).

cell from intracellular acidification.[305] In fibroblasts, NHE1 mainly localizes in lamellipodia regulating the intracellular pH (pH_i) according to the stimulation by membrane receptors, such as integrins that respond to extracellular migratory cues.[306,307] Increased pH_i promotes the recruitment of the actin severing protein cofilin to the leading edge of migrating cells, where it generates an increased availability of actin monomers and consequently enables enhanced actin polymerization.[308–310] In addition to its role as an ion transport protein, NHE1 anchors the actin cytoskeleton to the plasma membrane by direct interaction with actin binding proteins.[245,311] Moreover, NHE1 favors the dynamic reorganization of the actin based cytoskeleton.[245] Hence, through its localization and its function as cytoskeletal anchor and ion exchanger, NHE1 establishes cell polarity and directed cell migration. Remarkably, the lack of ion translocation inhibits de-adhesion resulting in trailing edges that fail to retract.

Because dynamic turnover of adhesions to the ECM is essential for cell migration, the mechanisms by which adhesion complexes form and subsequently disassemble are key to cell motility. Although several mechanisms contribute to focal adhesion remodeling in migrating cells,[312] emerging evidence indicates that talin plays a central role in the dynamic linkage between integrins and actin filaments.[242,313] Talin functions in distinct albeit complementary mechanisms that promote focal adhesion turnover. Namely the cleavage of talin by the protease calpain, which also modulates adhesion complex composition and likely signaling functions of talin.[296] The second mechanism is regulated talin binding to actin filaments, which is proposed to act as a clutch to control focal adhesion turnover and membrane protrusion dynamics.[314,315] How actin binding by talin is dynamically regulated during cell migration, however, remains undetermined. Previous studies suggest that actin binding by talin *in vitro* is pH sensitive, with lower affinity binding at higher pH,[243,244] although the functional significance of this regulation is unknown. In motile cells, an increase in intracellular pH favors focal adhesion remodeling[307,316] and velocity.[245,246] Accordingly, it is tempting to speculate that increased pH_i in motile cells might promote dynamic turnover of focal adhesions by lowering the affinity of actin binding by talin.

5.2 Investigation of the Dynamics of the Actin Binding Site of Talin

Since the N-terminal FERM domain and the C-terminal I/LWEQ domain of talin both contain actin binding sites that could mediate a clutch like action for focal adhesion remodeling,[248] the pH dependence of actin binding by each domain was analyzed by F-actin cosedimentation. This approach identified the FERM domain (1-433) and the I/LWEQ domain including the coiled coil dimerization segment (2341-2541) as pH insensitive within the physiological range of pH 6.5-7.5. However, a module containing the I/LWEQ domain and the adjacent upstream helical segment (USH) (2300-2541) shows pH sensitive F-actin binding with higher affinity at pH 6.5 compared to pH 7.5.

5.2.1 Assignment of the pH Sensitive Actin Binding Domain of Talin

As the pH sensitive actin binding module includes the coiled coil dimerization segment, this talin construct reaches a molecular weight of around 50 kDa, far beyond the optimal size for NMR studies. Therefore, initial NMR measurements were carried out on the talin USH-I/LWEQ domain lacking the C-terminal dimerization helix (2300-2501). During the course of this study, Critchley and coworkers determined an NMR structure of the talin I/LWEQ domain[317] including the USH segment and excluding the C-terminal dimerization domain (pdb 2jsw). However, the construct and the temperature at which which the structure was determined differ from the one investigated *in vitro* and also *in vivo* in this study. Since the poorly dispersed ^{1}H-^{15}N HSQC spectrum of this relatively large, purely helical talin module (Fig. 5.4) does not allow an unambiguous transfer of the backbone resonances, the pH sensitive actin binding domain was newly assigned.

In order to figure out the most suitable conditions for the NMR measurements at which the immense peak overlap decreases and thus the spectral quality improves, at least as much as possible, a temperature series from 22 °C to 37 °C and a pH series within the range of pH 6.0 to pH 8.0 was recorded. Accordingly, the backbone resonances of the talin USH-I/LWEQ domain (2300-2501) were assigned at a temperature of 32 °C and a pH value of 6.0 recording standard triple resonance experiments and performing the semi-automatic assignment procedure as described in section 3.2.1. Due to the low sequence complexity of this talin module, in some cases even the analysis of the carbon chemical shifts C', C^{α}, and C^{β} was not sufficient enough to achieve an unambiguous assignment. Purely helical proteins characteristically display a poor proton dispersion, but on the other hand the helical secondary structure possesses an advantageous ^{1}H-^{1}H NOE pattern.[126] Typically, within an α-helix the short $^{1}H^{N}_{i}$-$^{1}H^{N}_{i+1}$ distance gives rise to an intense cross peak. Consequently, recording an NNH-NOESY spectrum[139] in combination with the ^{15}N-HSQC-NOESY spectrum enables the identification of adjacent amino acids within the ^{1}H-^{15}N HSQC spectrum. This approach of combining the conventional assignment procedure with the specific sequential ^{1}H-^{1}H NOE pattern of α-helices resulted in the backbone resonance assignment of 156 residues out of 198 non-proline residues (Fig. 5.4).

5.2.2 Titration Studies of the Monomeric and Dimeric Talin Construct

The talin USH-I/LWEQ domain contains just a single histidine residue, the only amino acid with a pK_a close to the physiological range. Near this histidine there are also a number of negatively charged amino acids located which could have upshifted pK_a values. These residues are conserved in mammalian talin1 but not in *Caenorhabditis elegans* or *Drosophila* and might form a pH sensor distal from the putative actin binding site as defined by mutagenesis.[318] To corroborate the hypothesis of a pH sensor

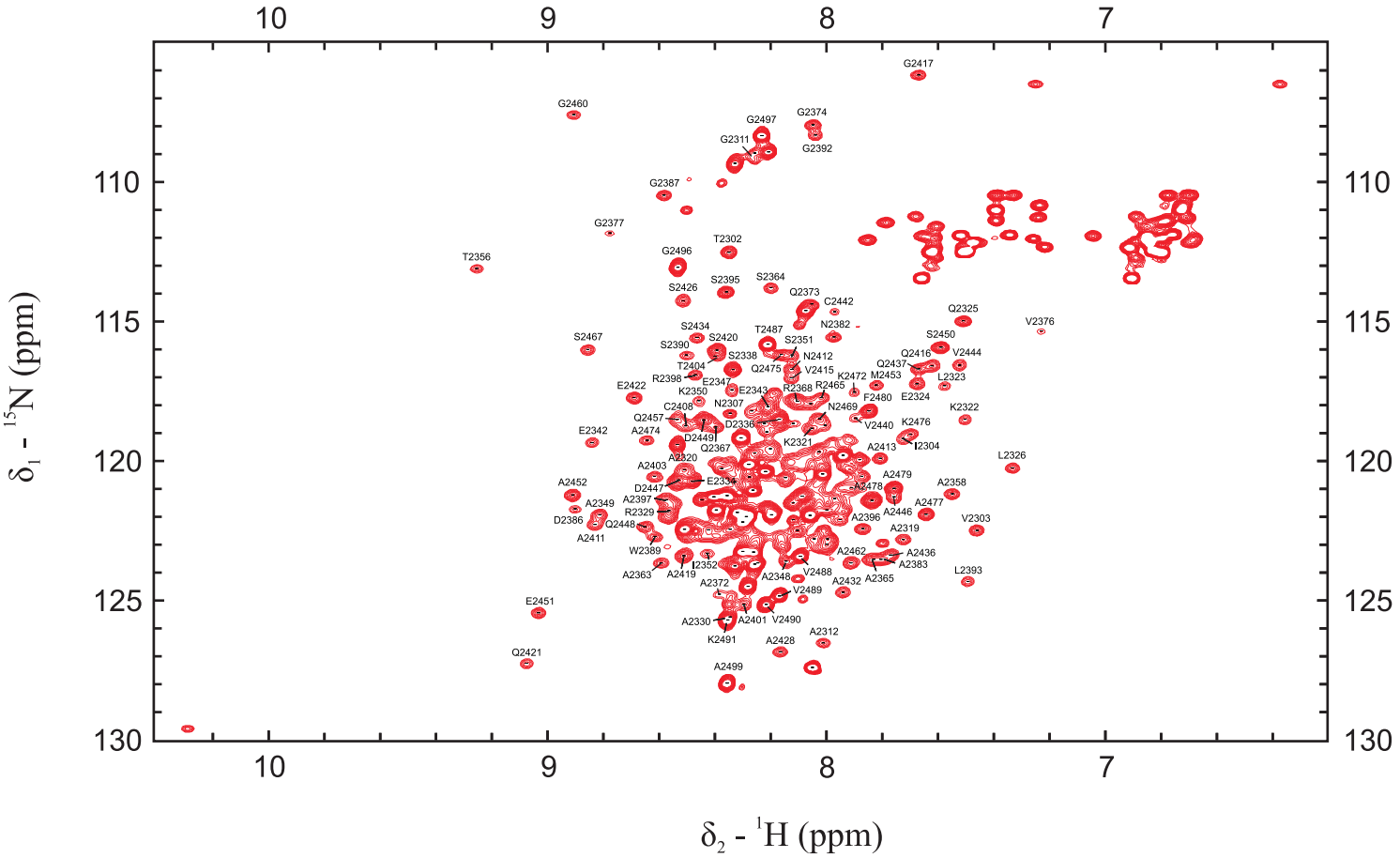


Figure 5.4: Assigned ^{1}H-^{15}N HSQC spectrum of the pH sensitive actin binding domain of talin (202 AA) recorded at a temperature of 32 °C and a proton frequency of 800 MHz. For simplicity, just the assignment of isolated peaks is shown in this spectrum, the complete backbone resonance assignment is listed in the appendix.

sensitive to small variations in intracellular pH which initiate conformational or dynamical changes at the remote actin binding site to either favor or diminish the interaction with actin, titration studies were carried out on the talin USH-I/LWEQ module (2300-2501). Increasing solvent pH from pH 6.5 to pH 7.5 according to the *in vitro* and *in vivo* experiments was associated with spectral perturbations for a subset of backbone amide resonances (Fig. 5.5). The residues with the largest chemical shift changes cluster to one area of the USH-I/LWEQ domain, the region predicted to comprise the pH sensor. Analysis of the titration curve of the $H^{\delta 2}$ proton of the single histidine residue indicates a pK_a of 7.2 which is upshifted from the pK_a of ∼6.5 for isolated histidine in aqueous buffer.[319] Due to this upshifted pK_a of His-2418, physiological changes in the intracellular pH result in significant differences in the fractional protonation of the histidine side chain (∼85% protonation at pH 6.5 and ∼20% protonation at pH 7.5). Additionally, Asp-2482 also shows a significantly upshifted pK_a and consequently titrates in the range from pH 6.5 to pH 7.5 as well. Even though the pK_a values of the glutamic acid residues 2334, 2342, and 2481 likewise deviate from the one for the isolated amino acid, these shifts are not sufficient enough to cause considerable alterations in the fractional protonation within the physiological range. Still, this clearly indicates the existence of a pH sensor composed of residues with upshifted pK_a values that cluster together at one end of the helical bundle and adopt distinct protonation states in response to small changes in pH_i influencing the conformation of the actin binding site, located ∼ 40 Å away.

However, the truncated 40 C-terminal residues of the I/LWEQ domain are not only proposed to form a coiled coil responsible for dimerization, but are also required for actin binding.[318,320] Actin cosedimentation assays confirm a highly restricted affinity of the monomeric talin module lacking the dimerization helix for F-actin. Therefore, to reveal possible mechanisms by which the talin USH-I/LWEQ domain binds actin in a pH dependent manner, the dimeric talin construct (2300-2541) was further investigated in respect of residues titratable within the physiological range of the intracellular pH. Remarkably, despite the considerable molecular weight of ∼50 kDa the intensity and amount of the backbone amide resonances in the ^{1}H-^{15}N HSQC spectrum is consistent with the monomeric talin module. Although the particularly beneficial shape of the homodimer provides further insights into the *in vivo* relevant putative pH sensor, yet the NMR experiments are limited to the amide resonances. Comparison of the amide chemical shift perturbations with pH for monomeric and dimeric USH-I/LWEQ shows almost identical changes for constructs with and without the C-terminal dimerization helix. This result suggests that the model of pH allosterically regulating the actin binding site, although based on NMR experiments with the monomeric construct, may also explain the pH dependent binding of full length, dimeric talin to actin.

These titration studies point towards a mechanism in which protonation of residues that constitute the pH sensor, including His-2418, may modulate conformation and dynamics of distal residues in the actin binding site.

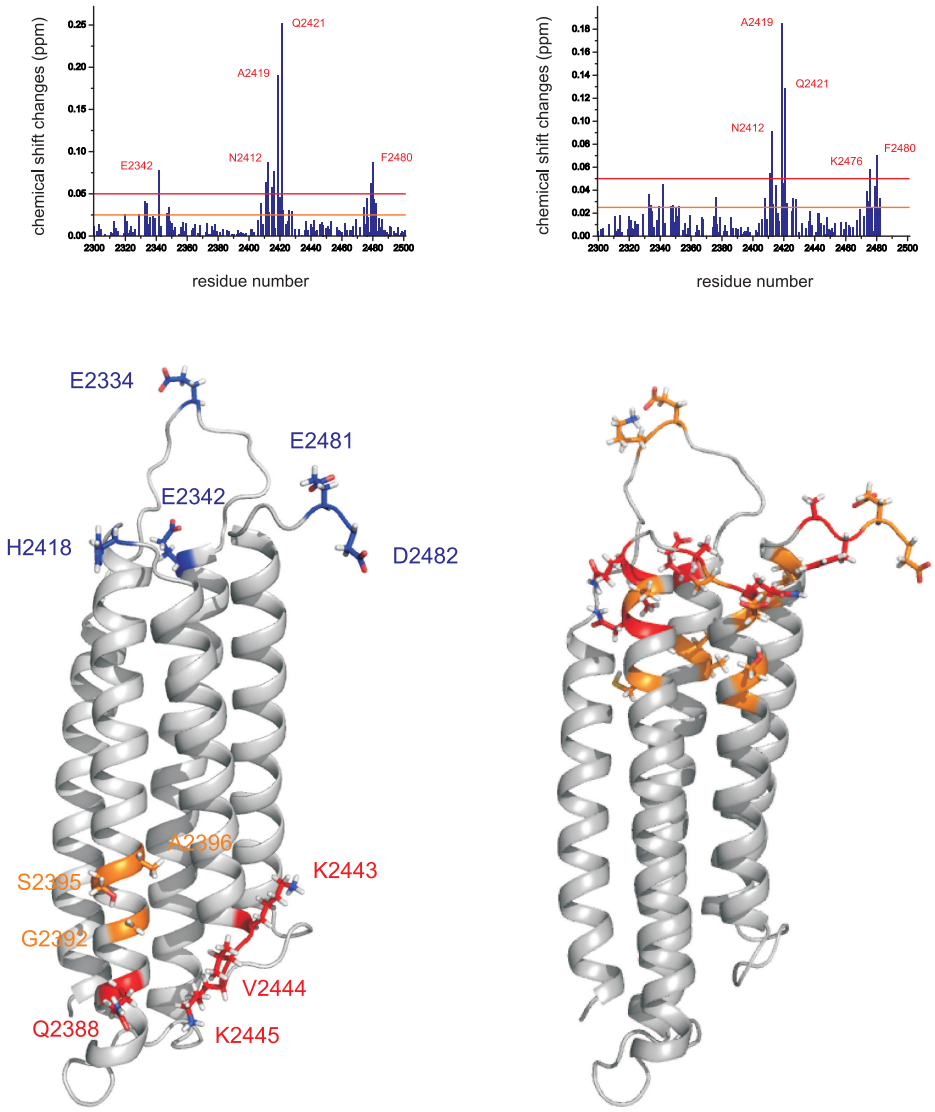


Figure 5.5: pH dependent conformational changes in talin USH-I/LWEQ. Top left: Chemical shift changes according to $\Delta\delta = |\ \Delta^1\mathrm{H}\ | + |\ \Delta^{15}\mathrm{N}\ | \cdot 0.2$ for the monomeric talin construct lacking the dimerization helix caused by an increase in pH from 6.5 to 7.5. Top right: Since the truncated 40 C-terminal residues of the I/LWEQ domain are required for actin binding, chemical shift changes for the dimeric talin module were determined analogous. Comparison of the amide chemical shift perturbations with pH for monomeric and dimeric USH-I/LWEQ shows almost identical changes. Bottom left: In response to physiological pH changes the protonation states for His-2418 and Asp-2482, which constitute the pH sensor along with Glu-2334, Glu-2342, and Glu-2481, vary and hence influence the conformation of the distal actin binding site (residues whose mutation significantly reduced the affinity for actin are depicted in orange and respectively in red for the ones with the most pronounced effects on the binding capacity[318]). Bottom right: Increasing solvent pH from 6.5 to 7.5 led to chemical shift perturbations for a subset of backbone amide resonances (classified as medium with $\Delta\delta > 0.025$ ppm, colored in orange and large with $\Delta\delta > 0.05$ ppm, shown in red). Remarkably, all residues with considerable alterations in the chemical shifts cluster to one area of the five helical bundle around the residues predicted to form the pH sensor.

5.2.3 Conformational Coupling between the pH Sensor and the Actin Binding Site

Assuming that talin possesses a pH sensor sensitive to the changes in intracellular pH which demonstrably occur during cell migration, the question remains how does this influence the actin binding site, located $\sim$ 40 Å away at the other end of the helical bundle and how or are the residues in the pH sensor at all coupled to the residues in the actin binding site through the helices of the I/LWEQ module. To get any idea about possible mechanisms underlying this signal transduction, Critchley and coworkers kindly recorded in addition to the NOESY spectra used for the structure determination another set of NOESY spectra at increased solvent pH and subsequently excluded significant structural changes as cause for the altered affinity of talin to actin at different pH values. This leaves conformational and/or dynamical modulations as possible mechanisms of the coupling between the pH sensor and the actin binding site.

Residual Dipolar Couplings

Molecular dynamics, apparent in backbone and side chain mobilities, play a crucial role in protein stability and function. Although local backbone fluctuations on the picosecond to nanosecond time scale can be characterized by NMR spectroscopy and also molecular simulation dynamics, slower motions in the submicrosecond to millisecond range remain poorly accessible. However, these slow time scales are of particular interest as functionally important biological processes, including enzyme catalysis,[321] signal transduction,[322] ligand binding, and allosteric regulation,[323] are supposed to occur in this time range. Because of the overall tumbling of the protein in aqueous solution, molecular dynamics in the time scale beyond the typical tumbling time of around $5 - 10$ ns are averaged out within a small number of molecular rotations. Due to the anisotropic environment and thus the partial alignment of the protein, residual dipolar couplings provide key information for understanding protein motions in the submicrosecond to millisecond time range.[31,324] Even though an accurate analysis of the backbone dynamics via residual dipolar couplings requires several independent alignment media,[325] in the case of talin USH-I/LWEQ RDC measurements with one alignment medium but at different pH values might already reveal insights into any feasible mechanism for the pH dependent actin binding.

To achieve the necessary partial alignment for the measurement of the residual dipolar couplings, non-ionic liquid crystalline media were added to the ^{15}N, ^{13}C labeled monomeric talin USH-I/LWEQ domain (2300-2501) as published by Otting and coworkers.[223] Since the 2D ^{1}H-^{15}N HSQC spectrum is dominated by immense peak overlap, N-H dipolar couplings have to be extracted from 3D IPAP-HNCO spectra. According to the backbone assignment, RDC measurements were carried out at 32 °C and in order to characterize the coupling between the pH sensor and the actin binding site at pH 6.5 and pH 7.5, respectively. These experimental $^{1}D_{\mathrm{NH}}$ couplings along with the structure of the talin USH-I/LWEQ domain (pdb 2jsw) serve as input for the software Pales that determines the alignment tensor by singular value decomposition.[326] Further-

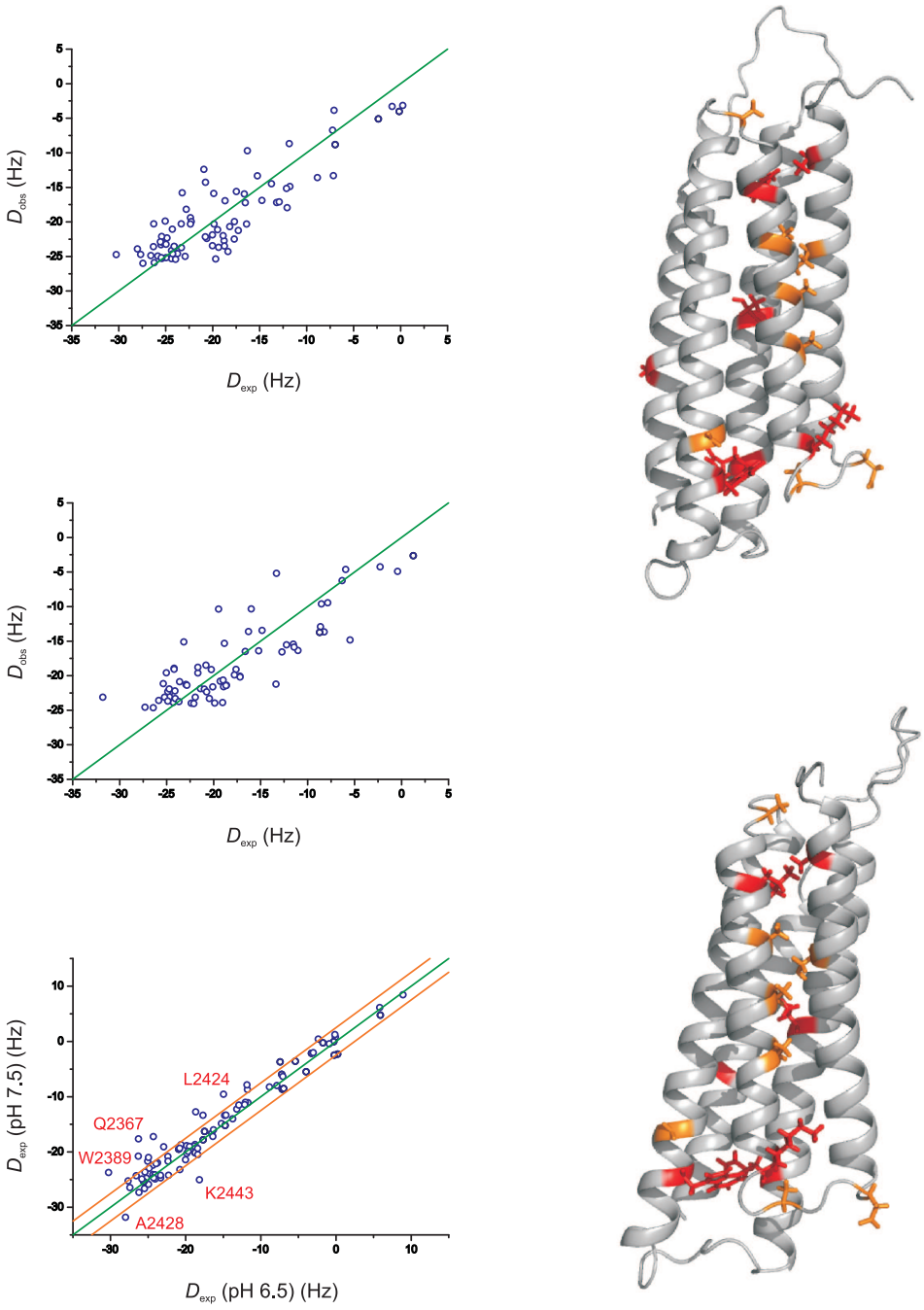


Figure 5.6: pH dependent residual dipolar couplings of talin USH-I/LWEQ. Partial alignment for the measurement of $^1D_{NH}$ couplings was achieved with non-ionic liquid crystalline media and RDC values were extracted from IPAP-HNCO spectra. Left top: Correlation between the backcalculated and the experimental dipolar couplings at pH 6.5 ($R = 0.87$, $Q = 0.18$) Left middle: Correlation between the backcalculated and the experimental dipolar couplings at pH 7.5 ($R = 0.84$, $Q = 0.23$) Left bottom: Experimental residual dipolar couplings at pH 6.5 compared with the experimental $^1D_{NH}$ couplings at pH 7.5 revealed on average just 1.1 ± 0.7 Hz deviation for the vast majority of the residues but also few amino acids that differ significantly. Right: Mapping residues associated with pH dependent couplings according to deviations >2.5 Hz (orange) and >5 Hz (red) onto the solution structure of talin USH-I/LWEQ.

more, knowledge of the alignment tensor based on the experimental couplings and the three dimensional structure enables the backcalculation of the N-H dipolar couplings. For both pH values the experimental $^1D_{\mathrm{NH}}$ couplings agree quite well with the back-calculated ones (correlation factor $R = 0.87$, respectively $R = 0.84$ and quality factor $Q = 0.18$, respectively $Q = 0.23$ for pH 6.5 and pH 7.5 respectively, Fig. 5.6).

For the vast majority of the USH-I/LWEQ residues on average the experimental residual dipolar couplings at pH 6.5 deviate from the experimental $^1D_{\mathrm{NH}}$ couplings at pH 7.5 just by 1.1 ± 0.7 Hz. However, a few amino acids show considerable differences of >2.5 Hz or even >5 Hz with variation of the solvent pH (depicted in orange and red, respectively in the cartoon representation of the talin USH-I/LWEQ domain in Fig. 5.6). Interestingly, mapping these residues on the solution structure of the talin module reveals an exciting pattern (Fig. 5.7). The amino acids associated with pH dependent couplings are either supposed to be involved in actin binding or located within the helix which connects the pH sensor with the actin binding site. Furthermore, the residues within this helix are all linked via hydrogen bonds. This amazing pattern tempts to speculate about a potential mechanism for the regulation of actin binding by talin. Since the single histidine is the decisive modulator of the pH sensor, signal transduction depends on the distinct protonation states of His-2418. Deprotonation of the side chain provides electrons for the formation of a hydrogen bond with the side chain of Ser-2420. This serine residue initiates the signal transfer through the helix to the remote actin binding site as its carbonyl is hydrogen bonded to the amide of Leu-2424. The pathway further downstream incorporates Ala-2428, Ala-2432, Ala-2436, Leu-2439, and finally reaches Lys-2443, one of the residues identified as fundamental for actin binding. In contrast, if the side chain of His-2418 is protonated at pH 6.5 the first hydrogen bond of this cascade is not formed and hence might result in altered modulations of the distal actin binding site. The participation of Leu-2439 and Ser-2420 in this signaling pathway could not be confirmed by the RDC measurement, because of the missing assignment for Leu-2439 and the typically averaged coupling due to the location of Ser-2420 in the loop region. However, the principle mechanism of dynamically transmitting information across hydrogen bonds has been extensively investigated by Blackledge and coworkers.[325] Although this long range network of dynamic correlations between amino acids connected via hydrogen bonds transfers information across a four stranded β-sheet, there is no plausible reason why this mechanism should not account for signal transduction through helices as well.

In order to obtain another indication for this mechanism, a second RDC data set was measured and analyzed. The partial alignment was achieved the same way as described above, but in this case the $^1D_{\mathrm{C^\alpha C'}}$ couplings were extracted from 3D IPAP-HNCO spectra. However, these RDC values show no significant differences associated with increasing solvent pH from 6.5 to 7.5. According to the study by Blackledge and coworkers[325] this is consistent with a signal transfer via hydrogen bonds assuming that motions about the C^α_{i-1}–C^α_i axis connecting sequential amino acids are dominant and that reorientation about orthogonal axes are less pronounced, as illustrated below.

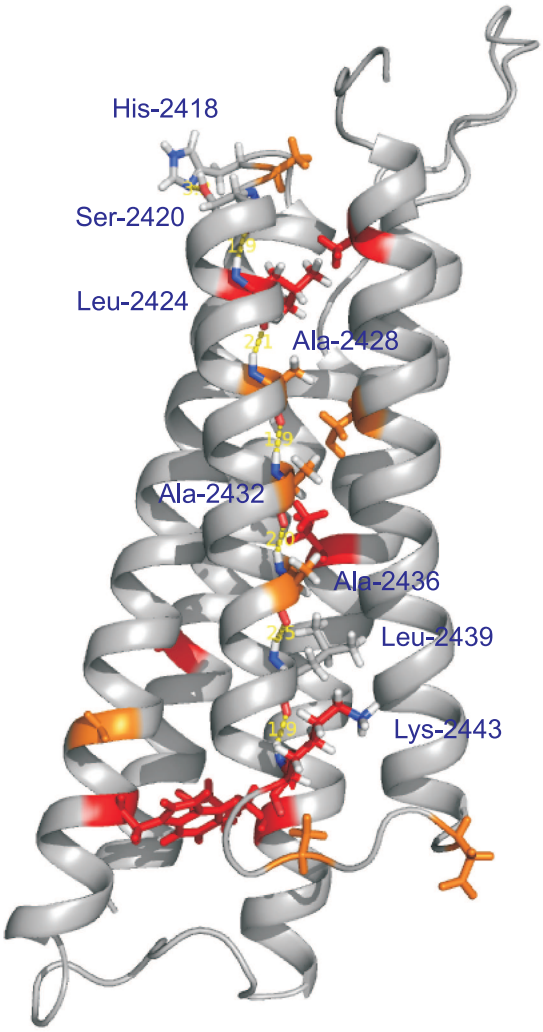
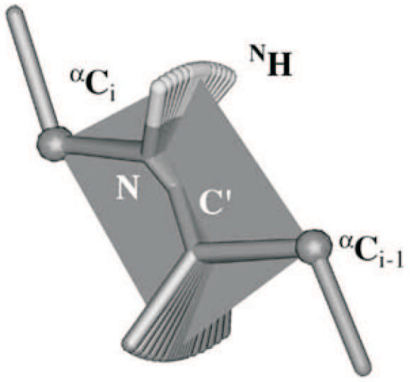


Figure 5.7: Speculation about a possible mechanism for the regulation of actin binding by talin. At pH 7.5 the deprotonated state of the side chain of His-2418 enables the formation of a hydrogen bond to the side chain of Ser-2420. From this serine residue the information is dynamically transfered further downstream across the hydrogen bonds of Leu-2424, Ala-2428, Ala-2432, Ala-2436, Leu-2439, and finally reaches Lys-2443, one of the residues identified as fundamental for actin binding. In contrast, the protonated state of the side chain of His-2418 at pH 6.5 prevents the formation of the first hydrogen bond of this cascade and hence might result in altered modulations of the distal actin binding site. Therefore, the residues of the pH sensor might be coupled with the remote actin binding site by a long range network of dynamic correlations between amino acids connected via hydrogen bonds. The amino acids associated with pH dependent couplings are colored according to their deviations in orange for >2.5 Hz and red for >5 Hz (on average the experimental $^{1}D_{\mathrm{NH}}$ couplings differ just by 1.1 ± 0.7 Hz between pH 6.5 and pH 7.5).



Therefore, $^1D_{\mathrm{NH}}$ RDCs from another alignment medium might verify these backbone motions. Though, this approach is still under progress.

CPMG Relaxation Dispersion Experiments

Protein conformational dynamics on microsecond to millisecond time scales are also evident as chemical exchange line broadening. Increasing solvent pH from 6.5 to 7.5 is accompanied with significant line broadening for a subset of backbone amide resonances. Remarkably, this decrease in peak intensity caused by conformational exchange is most pronounced for the loop region constituting the pH sensor as well as for the loop region associated with the actin binding.

Another indication for motions in the slow time range comes from CPMG relaxation dispersion experiments. Conformational or chemical exchange can alter the transverse relaxation rate at sites that exhibit a change in chemical shift with the dynamic event. Consistent with the observation of just a single set of cross peaks derived from the major conformer in the ^{1}H-^{15}N correlation spectra, conformational rearrangements from the predominant state might only occur at low levels. Typically, measurements provide information just for the extensively populated conformer. In contrast, Carr-Purcell-Meiboom-Gill (CPMG) based spin relaxation rates can be sensitive to the presence of minor conformers as long as the rates of interconversion are on the millisecond time scale and large chemical shift differences between the states are present. Conformational exchange on the microsecond to millisecond time scale leads to a decay of transverse signal, which however can be suppressed by application of radio frequency B_1 fields. Thus, the effective decay of transverse magnetization, $\mathrm{R_2}^{\mathrm{eff}}$, decreases as a function of increasing B_1 field strength, ν_{CPMG}. Analysis of the relaxation dispersion profiles, $\mathrm{R_2}^{\mathrm{eff}}$ versus ν_{CPMG}, provides information on the kinetic (rates of interconversion) and the thermodynamic (populations) parameters describing the exchange process.

To identify the sites of conformational or chemical exchange in the talin USH-I/LWEQ domain, CPMG relaxation dispersion experiments were carried out at 32 °C and both pH values (6.5 and 7.5). Dispersion profiles for the backbone amides were recorded at two different magnetic field strengths corresponding to ^{1}H frequencies of

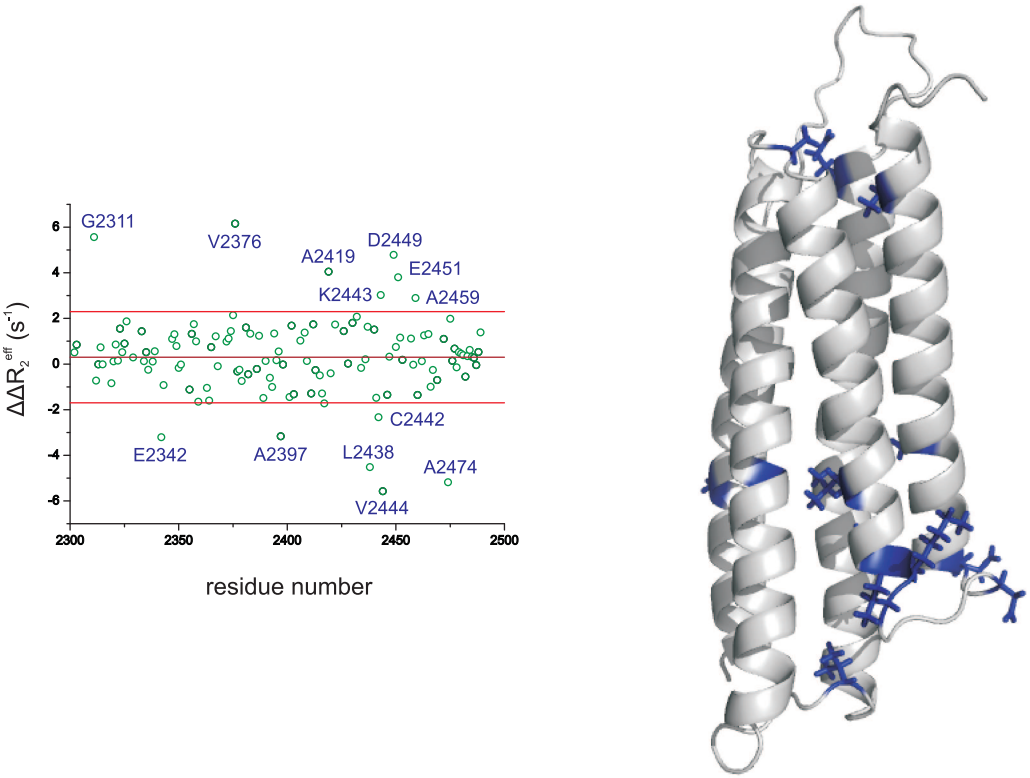


Figure 5.8: pH dependent CPMG relaxation dispersion experiments of talin USH-I/LWEQ. An effective method to map allostery is the approach of equilibrium perturbation. Differences in the backbone amide relaxation dispersion at CPMG fields of 25 Hz and 1000 Hz (ΔR_2^{eff}) can be compared for the high affinity (pH 6.5) and the low affinity (pH 7.5) state of actin binding. Significant deviations from the average $\Delta\Delta R_2^{eff}$ value indicate residues that experience conformational fluctuations. Variations in the intracellular pH lead to conformational exchange in the actin binding site of the talin USH-I/LWEQ domain.

500 and 800 MHz. A fully interleaved pulse sequence and also a software package for analyzing the obtained CPMG curves was kindly provided by Lewis Kay. In the case of the monomeric USH-I/LWEQ talin most of the measured profiles display a gentle slope, indicating that these backbone amides do not sense an exchange process. Yet, some of the CPMG curves suggest the existence of conformationally exchanging regions. However, all attempts to fit these relaxation dispersion profiles properly were without success, hence neither the exchange rate constant nor the population of the distinct conformers could be derived from the experimental data. Ascertaining the reason for this failure and improving the pulse sequence to obtain interpretable curves is still subject to current research. Even though the data do not allow a detailed description of the exchange process yet, they still provide information in terms of pH dependent equilibrium perturbation on the backbone amide relaxation dispersion.[327] Sites experiencing conformational fluctuations can be mapped by the difference in ΔR_2^{eff} between the high affinity and the low affinity state for actin binding. ΔR_2^{eff} corresponds to the change in R_2^{eff} at the CPMG field strength of 25 Hz compared to 1000 Hz for pH 6.5 and pH 7.5, respectively. The pH dependent $\Delta\Delta R_2^{\text{eff}}$ presents a sensitive indicator for residues that are affected by the distinct protonation states within the pH sensor. Residues with $\Delta\Delta R_2^{\text{eff}}$ values remarkably deviating from the average value cluster mainly to the region supposed to bind F-actin (Fig. 5.8). Although complete analysis of the CPMG relaxation dispersion experiments is still in progress the approach of equilibrium perturbation already indicates conformational exchange in the actin binding site associated with variation in the intracellular pH.

5.3 Insights into the pH Dependent Mechanism of Focal Adhesion Remodeling

This model of pH regulated focal adhesion turnover in migrating cells based on *in vitro* and NMR experiments was further investigated *in vivo* [1]. To study the effect of physiological changes in the intracellular pH on the lifetime of focal adhesions, experiments were carried out with fibroblasts deficient for the Na^+/H^+ exchanger NHE1 but stably expressing wild type NHE1 (WT cells) with a pH_i of ∼7.5 or a mutant NHE1 with an E266I substitution that lacks proton translocation (E266I cells) resulting in a pH_i of ∼7.0.[311,328] Focal adhesion turnover in cells migrating at the edge of a wounded monolayer was determined by real time imaging of GFP-paxillin, which localizes to focal adhesions and had similar expression rates in WT and E266I cells. The decrease in intracellular pH decreases focal adhesion turnover in extending lamellipodia and consequently increases the lifetime of focal adhesions in E266I cells (37.1 ± 3.2 min) compared with WT cells (15.5 ± 1.1 min).

1 *In vivo* experiments were performed by Dr. Jyoti Srivastava, a postdoctoral fellow in the group of Prof. Dr. Diane Barber, Department of Cell and Tissue Biology, University of California, San Francisco, USA.

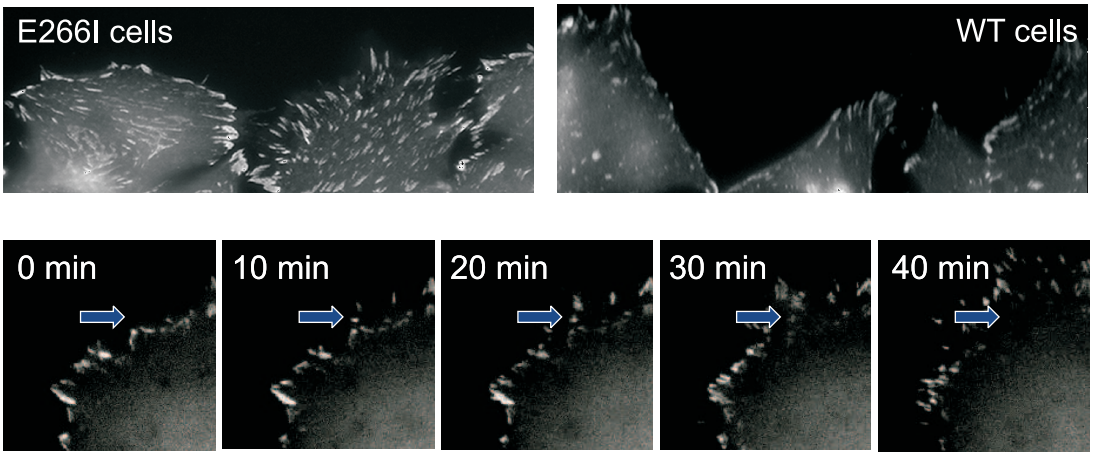


Figure 5.9: pH dependent focal adhesion stability. Images from movies of fibroblasts expressing GFP-paxillin which localizes to focal adhesions at the edge of a wounded monolayer. These fibroblasts are deficient for the Na^+/H^+ exchanger NHE1 but either stably express wild type NHE1 (WT cells) and hence regulate the pH_i to the physiological value of ∼7.5 or a mutant NHE1 with an E266I substitution that lacks proton translocation (E266I cells) resulting in a pH_i of ∼7.0.[311, 328] Decreased pH_i increases the lifetime of focal adhesions in E266I cells (Left, 37.1 ± 3.2 min) compared with WT cells (Right, 15.5 ± 1.1 min). These *in vivo* studies clearly evidence that the velocity of migrating cells depends on the intracellular pH, as focal adhesion stability is significantly increased in cells with defective H^+ efflux (Bottom).

To verify the assumption that the single histidine (His-2418) in the talin USH-I/LWEQ domain is the key residue in pH sensing, this histidine was substituted against a phenylalanine (H2418F). Indeed, actin binding by the USH-I/LWEQ-H2418F mutant was relatively pH insensitive as confirmed by F-actin cosedimentation. However, this mutant had an intermediate affinity and maximal binding compared with the wild type talin construct at low and high pH. Consistent with the intermediate actin binding affinity of the mutant talin, the focal adhesion lifetime increased in WT cells but decreased in E266I cells. As the substitution of H2418F in talin changes focal adhesion stability, attenuating turnover in WT cells with pH_i of 7.5 but increasing turnover in E266I cells with pH_i of 7.0, these data indicate that His-2418 plays a central role in sensing intracellular pH changes to regulate focal adhesion stability and cell migration rate.

5.4 Discussion

Increased pH_i was previously assumed to promote focal adhesion remodeling.[245, 316] The focal adhesion protein talin senses variations in the intracellular pH within the physiological range and regulates the stability of focal adhesions by altered affinity for the cytoskeletal protein actin. A structural model for pH dependent actin binding by monomeric USH-I/LWEQ talin suggests that protonation of His-2418, which has

an upshifted pK_a value, modulates conformation and dynamics at the remote actin binding site. Signal transduction from the pH sensor at one end of the five helical bundle to the distal actin binding site at the other end might occur via a long range network of amino acids connected by hydrogen bonds. In addition, the C-terminus of talin seems to be the essential determinant in focal adhesion remodeling, since impaired force generation in talin null cells is restored by expression of full length talin but not by a truncated talin lacking the C-terminus but retaining the FERM domain[329] which also has an actin binding site although this binding is pH insensitive. In contrast, a mutant talin lacking the FERM domain but retaining the C-terminus targets to focal adhesions.[330] Although the FERM domain binds the cytoplasmic tail of β integrins,[281] the C-terminus of talin1 contains a second integrin binding site.[242] Furthermore, this pH dependent focal adhesion remodeling could be initiated by different migratory cues, since integrin activation,[307,331] growth factors,[328,332] and monolayer wounding[328] stimulate NHE1 activity and increase pH_i. Therefore, pH regulated focal adhesion stability based on dynamic changes in talin affinity for F-actin may be one of several complementary mechanisms controlling focal adhesion turnover.

Changes in intracellular pH regulate a number of normal and pathological cell processes. Increases in pH_i are permissive for growth factor induced cell proliferation,[311] cell cycle progression,[333] and differentiation[334] and are necessary for cell migration.[335] On the contrary, increased cytosolic pH is a hallmark of transformed cells from different tissue origins and genetic backgrounds, making it a common characteristic of distinct cancers and possibly a common critical driving force for tumor progression.[336] Because cytosolic pH homeostasis is tightly regulated,[337] dramatic differences in cell behavior are driven by relatively small changes in pH_i. The increased pH_i in transformed cells differs only by 0.3 – 0.5 pH units as compared to normal cells that generally maintain their pH_i at $\sim$ 7.2 and variations of 0.3 – 0.4 pH units below the normal value already trigger apoptosis. Despite established effects of small changes in pH_i on diverse cell functions, the underlying mechanisms by which these changes affect proteins and macromolecular assemblies driving specific cell processes remains relatively unknown. One advantage of protons to serve for signaling and regulation as opposed to post-translational modifications or cofactors is the potential for extremely rapid temporal responses. Furthermore, another intriguing aspect of pH sensitivity is the ability to regulate multiple proteins in unison. Increased pH_i is an evolutionarily conserved signal necessary for several stages of cell migration in addition to focal adhesion remodeling, including polarity[245,246] and the assembly of cytoskeletal filaments.[246,338] Moreover, besides talin a number of proteins bind actin in a pH dependent manner, including HIP1,[339] cofilin,[310] and villin.[340] However, increased pH_i also plays a decisive role in most metastatic cancers, regardless of the tissue origin or genetic background.[247] Several studies have highlighted that intracellular alkalinization is an essential determinant in tumor transformation, development, cell growth, and in the maintenance of the metastatic process.[336,341,342] Oncogenes are able to induce the necessary and permissive cytosolic alkalinization that precedes cell transformation.[307,336] In transformed and cancer cells the NHE1 is usually hyperactive, affecting the behavior of the cells

in this respect that the cells persistently maintain in a status of permanent and uncontrolled proliferation.[333,343] Further, NHE1 directly coordinates tumor cell motility and invasion by selectively regulating cytoskeletal events such as focal adhesion assembly.[336,344] Hence, the higher pH_i of metastatic cells compared with normal cells could increase focal adhesion turnover to promote migratory capacity. Besides, this study underscores the importance of bridging structural and cellular biology to elucidate how physiological changes in pH_i regulate cell behaviors.

Chapter 6

Summary

The present thesis highlights the advantageous feature of solution state NMR spectroscopy. Whereas for the pure structure determination X-ray crystallography still represents the method of choice, NMR spectroscopy has the unique capacity to investigate dynamic properties of molecules over a range of different time scales with atomic resolution. Since most proteins function through transitions from ground states to sparsely populated excited states, which inevitably involves conformational changes, a thorough understanding of protein function is not possible without understanding protein dynamics. These dynamic processes cover a large time regime, including very fast fluctuations of individual atoms on the ps time scale, loop and domain motions on the ns time scale, and conformational rearrangements on the ms time scale. Many critical biological activities occur at the μs to ms time scale, such as protein folding and allosteric regulation, and manifest as the chemical exchange phenomena in NMR spectroscopy.

Folding and unfolding of proteins are crucial in regulating biological activity and targeting proteins to distinct cellular locations. Aggregation of misfolded proteins that escape the cellular quality control mechanisms is a common feature of a wide range of diseases. Since it is assumed that the shape of the folding landscape and hence the tendency to fold or misfold is encoded in the intrinsic properties of the amino acid sequence, investigation of highly homologous members of the same protein family but with remarkably different success rates in reaching the native state within a reasonable amount of time might reveal hints on the determinants of proper folding or harmful misfolding. Furthermore, it is generally accepted that potentially all proteins fold via an intermediate state that acts as a stepping stone to the native state. Therefore, these intermediate states play a key role in defining protein folding and assembly pathways as well as those of misfolding and aggregation. Yet, due to their transient nature, folding intermediates are poorly accessible to high resolution techniques. Chapter 3 describes

the intrinsically slow folding reaction of the constant antibody domain of the light chain (C_L) by real time NMR spectroscopy and specifies its major folding intermediate in detail. Moreover, a single point mutation enabled trapping of the intermediate in equilibrium and hence its characterization at atomic resolution based on a comprehensive analysis of chemical shifts, dihedral angle restraints, and NOESY cross peaks for the native state and the intermediate state of the C_L domain in conjunction with a simulated annealing like protocol. The NMR experiments in combination with the simulations provide a detailed picture of the major C_L folding intermediate. Thus, the intermediate state of the C_L domain exhibits the basic β-barrel topology typical for the immunoglobulin fold, however some flexible regions adopt a variety of conformations that especially holds for the flanking edge strands at one side of the protein. Surprisingly, two short strand connecting helices assume their completely native structure already in the intermediate state and therefore might fulfill a spacer and orienting function for adjacent strands and loops. Hence, these two helices can be regarded as a scaffold within the C_L intermediate favoring the formation of a native like topology by correctly positioning important parts of the molecule. This assumption was further verified by comparing the major folding intermediate of the C_L domain that favors productive folding with the intermediate state of the well studied β_2-microglobulin that has the tendency to misfold and form amyloid fibrils. Although C_L and β_2m are highly homologous members of the widespread immunoglobulin superfamily, C_L has never been directly associated with amyloidogenic processes, whereas β_2m is known to cause amyloid diseases. Transplanting the sequences corresponding to the C_L helices, which are conserved in constant antibody domains but lacking in β_2-microglobulin, into the β_2m framework significantly reduces its amyloidogenicity. This study clearly demonstrates that minor structural differences in an intermediate can shape the folding landscape decisively to favor either folding or misfolding.

Besides productive folding to the native state, the biological functions of proteins also depend on the assembly into the correct quaternary structure. A particularly important example in this respect are IgG antibodies in which two heavy and two light chains have to associate prior to secretion from the endoplasmic reticulum (ER). Whereas isolated antibody light chains are in most cases readily secreted, unassembled Ig heavy chains are retained in the ER by interaction with the molecular chaperone BiP until association with the light chain. The first constant domain of the heavy chain, the C_H1 domain, plays a crucial role in this retention process. Since the individual domains of an immunoglobulin exhibit highly similar three dimensional structures and generally fold to their native state autonomously, the underlying principles for this special behavior of the C_H1 domain remained enigmatic. In Chapter 4 the structural properties of this antibody domain are analyzed in more detail. Surprisingly, in marked contrast to all isolated antibody domains studied so far, the C_H1 domain is a natively unfolded protein. It gains structure only upon interaction with its native partner, the C_L domain. Characterization of the folding pathway of the intrinsically disordered C_H1 domain in complex with the C_L domain by real time NMR spectroscopy revealed that this association coupled folding reaction proceeds via a partially structured encounter complex.

Based on the atomic level description of the folding pathway, key residues that allow specific recognition of the two proteins and thereby establish a folding nucleus in the C_H1 domain could be identified. In order to verify the theory that the intrinsically unfolded nature of the C_H1 domain controls the secretion of antibody molecules in a cellular context, the structural characteristics of C_H1 were transplanted into the IgG light chain. *In vivo*, this altered light chain strongly interacts with the molecular chaperone BiP and is no longer secreted from the ER. Hence, these structural characteristics of the C_H1 domain and its role in antibody retention are intrinsic, context independent features. These data provide insight into the folding pathway of natively disordered proteins and highlight the plasticity of folding properties within one protein topology to fulfill specific biological requirements.

Another essential process in nature to regulate biological activity of proteins involves the coupling of conformational changes between two widely separated binding sites. Allostery is crucial to living cells and controls metabolism either through positive feedback regulation or negative inhibition. The propagation of conformational changes over long distances or through a closely packed protein is remarkably subtle and effective. Yet, how these conformational changes are generated and transmitted remains a challenging subject of current research. Chapter 5 deals with the fundamental process of cell migration. At the leading edge of a migrating cell, focal adhesions undergo rapid cycles of assembly and turnover, creating and disrupting, respectively, sites of traction necessary for forward movement of the cell. Force generation for traction requires linkage among the extracellular matrix, integrin receptors, and actin filaments. Actin filament binding by the integrin associated focal adhesion protein talin is pH sensitive with lower affinity at higher pH, although the functional significance of this pH dependence remains unknown. Because increased intracellular pH (pH_i) promotes cell migration and is a hallmark of metastatic carcinomas, focal adhesion remodeling might be increased through lower affinity talin-actin binding. NMR spectroscopic investigations on the pH sensitive USH-I/LWEQ domain of talin reveals a possible structural mechanism of pH dependent actin binding. A cluster of titratable amino acids with upshifted pK_a values, including the solely histidine residue, was identified at one end of the five helix bundle distal from the actin binding site. Protonation of the histidine modulates conformation and dynamics at the remote actin binding site. CPMG relaxation dispersion experiments indicate conformational changes in the actin binding site associated with physiological changes in the intracellular pH. Furthermore, analysis of RDCs at two different pH values suggests that signal transduction from the pH sensor to the distal actin binding site might occur via a long range network of amino acids connected by hydrogen bonds. This model of pH regulated focal adhesion turnover in migrating cells based on NMR experiments could be confirmed *in vivo*. In motile fibroblasts, increasing pH_i decreases focal adhesion lifetime and increases the migratory rate. These data identify a molecular mechanism for pH sensitive actin binding by talin and suggest that focal adhesion is in part allosterically regulated by pH dependent affinity of talin for actin binding.

Appendix A

The C_L Domain of the Immunoglobulin G Antibody

A.1 Sample Preparation

The antibody domain C_L was expressed, refolded, and purified as published.[111,345] The C_L gene was obtained by PCR amplification using the cDNA of the murine MAK33 κ light chain as a template. The PCR product was cloned into the pET28a expression vector (Novagen, Darmstadt, Germany) via the NdeI and HindIII restriction sites and transformed into the E. coli strain BL21(DE3). Cells were grown at 37 °C in selective LB medium, and expression was started by addition of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at an OD_{600} of 0.8. After expression overnight at 37 °C, cells were harvested by centrifugation (4000 g). The cell pellet was resuspended in 100 mM Tris-HCl, 100 mM NaCl, 5 mM EDTA, pH 7.5 at 4 °C. Cells were cracked with a cell disruptor and DNA was removed by the addition of 10 μg/mL DNAseI, 3 mM $MgCl_2$ and incubation for 30 min at room temperature. Thereafter, 2.5% (v/v) Triton X-100 was added and the mixture was stirred for further 30 min at 4 °C. Inclusion bodies were isolated by centrifugation (40000 g, 30 min, 4 °C) and washed two times with 100 mM Tris-HCl, 100 mM NaCl, 5 mM EDTA, pH 7.5. Subsequently, inclusion bodies were solubilized in 100 mM sodium phosphate (pH 7.5), 6 M GdmCl, 10 mM β-mercaptoethanol for 2 h at room temperature. Insoluble components were removed by centrifugation (40000 g, 25 min, 20 °C). The supernatant was applied to a Ni-chelating column equilibrated in 100 mM sodium phosphate, pH 7.5, 3 M GdmCl. After washing for five column volumes, elution was performed with the same buffer at pH 4.0. Refolding was carried out via dialysis in 250 mM Tris-HCl (pH 8.0), 100 mM L-Arg, 10 mM EDTA, 1 mM GSSG, 0.5 mM GSH at 4 °C overnight. Aggregates were removed by centrifugation (40000 g, 25 min, 4 °C) and 0.25 units Thrombin (re-

striction grade, Novagen, Darmstadt, Germany) for each mg protein were added to remove the N-terminal His-tag. The reaction was allowed to proceed for 16 h at 4 °C. After an additional centrifugation step to remove aggregates (40000 g, 25 min, 4 °C), the concentrated supernatant was applied to a Superdex 75 26/60 gel filtration column (Amersham Biosciences, Upsala, Sweden) equilibrated in PBS buffer. This protocol results in a yield of around 10 mg/mL pure protein.

The C_L^{P35A} mutant was generated by site directed mutagenesis (Stratagene) and cloned, expressed, refolded, and purified according to the protocol for the C_L wild type. Isotope labeled proteins were expressed in M9 minimal medium (6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 1 g/L NH_4Cl, 2 mL of 1 M $MgSO_4$, 2 g/L glucose, 0.2 mL of 0.5 M $CaCl_2$, antibiotics, 1 mg biotin, 0.5 mL of 2 mg/mL thiamine hydrochloride, 1 mL of 15 mg/mL $FeCl_2$ in 1 M HCl, 1 mL of 15 mg/mL $ZnCl_2$ in slightly acidic H_2O, pH 7.4) containing either ^{15}N ammonium chloride as the only nitrogen source or additionally ^{13}C glucose as the only carbon source.

The β_2m gene was obtained by PCR amplification from human cDNA (RZPD), the gene for $\beta_2 m^{toCL}$ was synthesized (GATC Biotech). Both proteins were cloned into the pET28a expression vector (Novagen) without a His-tag and expressed overnight at 37 °C in E. coli BL21(DE3) cells, and the inclusion bodies were prepared as described above. IBs were solubilized in 50 mM Tris-HCl (pH 7.5), 8 M urea, 50 mM β-mercaptoethanol, 10 mM EDTA and applied to a Q-sepharose column equilibrated in 50 mM Tris-HCl (pH 7.5), 5 M urea, 10 mM EDTA. The protein of interest in each case did not bind to the column, and the flow through was refolded as described for the C_L domain. A Superdex 75 26/60 column (GE Healthcare) equilibrated in PBS buffer was used as a final purification step.

All plasmids were sequenced and the mass of each protein was confirmed by MALDI-TOF MS.

For the NMR analysis the concentration of the distinct proteins in PBS buffer with 5% (v/v) D_2O was 0.5 - 1.0 mM. All spectra were recorded at 25 °C on Bruker DMX600, DMX750, and AVANCE900 spectrometers. Processing of the spectra was performed using the program Topspin 1.3 (Bruker BioSpin) and analysis of the spectra was carried out with the program Sparky (www.cgl.ucsf.edu/home/sparky). The ^{15}N labeled unfolded C_L for the folding studies was obtained via the same protocol as described above except the refolding step and in addition the gel filtration column in the final purification step was equilibrated in PBS buffer containing 2 M GdmCl. This unfolded C_L in PBS containing 2 M GdmCl was diluted tenfold by adding ice cold PBS buffer without GdmCl to start the folding process and spectra were measured immediately after mixing. The folding reaction was monitored by a series of real time ^1H-^{15}N HSQC spectra acquired every 14 minutes at 2 °C.

A.2 Experiments and Assignment of C_L^{wt} and C_L^{P35A}

Table A.1: Experiments on the wild type C_L domain

Experiment	Nuclei	NS	Time domain	Data matrix
NHSQC	$^{1}H\times^{15}N$	4	1024×128	1024×256
CHSQC	$^{1}H\times^{13}C$	8	1024×256	1024×512
HNCO	$^{1}H\times^{15}N\times^{13}C$	8	1024×76× 82	1024×128×256
HNcaCO	$^{1}H\times^{15}N\times^{13}C$	16	1024×64× 82	1024×128×256
HNCA	$^{1}H\times^{15}N\times^{13}C$	12	1024×76× 92	1024×128×128
CBCAcoNH	$^{1}H\times^{15}N\times^{13}C$	8	1024×76× 96	1024×128×128
HNCACB	$^{1}H\times^{15}N\times^{13}C$	32	1024×64× 92	1024×128×128
CCHcosy	$^{1}H\times^{13}C\times^{13}C$	16	1024×72× 92	1024×128×128
CCHtocsy	$^{1}H\times^{13}C\times^{13}C$	16	1024×72× 92	1024×128×128
CNHnoesy	$^{1}H\times^{15}N\times^{13}C$	16	1024×64× 96	1024×128×128
HNHnoesy	$^{1}H\times^{15}N\times\ ^{1}H$	16	1024×64×128	1024×128×256
NNHnoesy	$^{1}H\times^{15}N\times^{15}N$	16	1024×64× 96	1024×128×128
MEXICO	$^{1}H\times^{15}N$	8	1024×128	1024×256
HetNOE	$^{1}H\times^{15}N$	128	1024×128	1024×256

Table A.2: Experiments on the C_L^{P35A} mutant

Experiment	Nuclei	NS	Time domain	Data matrix
NHSQC	$^{1}H\times^{15}N$	4	1024×128	1024×256
CHSQC	$^{1}H\times^{13}C$	4	1024×256	1024×512
HNCO	$^{1}H\times^{15}N\times^{13}C$	8	1024×64× 64	1024×128×256
HNcaCO	$^{1}H\times^{15}N\times^{13}C$	64	1024×48× 64	1024×128×128
HNCA	$^{1}H\times^{15}N\times^{13}C$	24	1024×56× 78	1024×128×128
CBCAcoNH	$^{1}H\times^{15}N\times^{13}C$	16	1024×48× 64	1024×128×128
HNCACB	$^{1}H\times^{15}N\times^{13}C$	32	1024×52× 72	1024×128×128
CCHcosy	$^{1}H\times^{13}C\times^{13}C$	16	1024×72× 92	1024×128×128
CCHtocsy	$^{1}H\times^{13}C\times^{13}C$	16	1024×72× 92	1024×128×128
CNHnoesy	$^{1}H\times^{15}N\times^{13}C$	24	1024×64× 92	1024×128×128
HNHnoesy	$^{1}H\times^{15}N\times\ ^{1}H$	16	1024×56×102	1024×128×256
NNHnoesy	$^{1}H\times^{15}N\times^{15}N$	16	1024×64× 82	1024×128×128
MEXICO	$^{1}H\times^{15}N$	8	1024×128	1024×256
HetNOE	$^{1}H\times^{15}N$	32	1024×128	1024×256

Table A.3: Backbone and side chain assignments of the wild type C_L domain

Residue	H^N (N)	C'	H^α (C^α)	H^β (C^β)	H^γ (C^γ)	H^δ (C^δ)
G1	–	–	–	–	–	–
S2	–	–	–	–	–	–
H3	–	174.46	4.53 (55.97	2.83, 2.90 (30.94)	–	–
M4	8.12 (119.87)	174.99	4.45 (56.16)	1.99 (35.13)	2.43, 2.59 (32.33)	–
A5	9.37 (125.64)	175.23	4.61 (52.09)	1.32 (21.77)	–	–
A6	8.69 (127.42)	174.97	4.43 (49.89)	1.24 (17.51)	–	–
P7	–	–	–	–	–	–
T8	–	–	–	–	–	–
V9	–	174.71	5.14 (61.19)	1.56 (33.51)	0.58 (21.56)	–
S10	8.84 (121.25)	–	4.65 (57.04)	3.57, 3.66 (66.15)	–	–
I11	–	–	–	–	–	–
F12	–	–	–	–	–	–
P13	–	–	–	–	–	–
P14	–	–	–	–	–	–
S15	8.66 (121.15)	178.92	4.65 (59.65)	3.92 (63.32)	–	–
S16	–	–	–	–	–	–
E17	–	178.91	3.92 (59.63)	1.80, 1.91 (29.05)	2.15, 2.19 (36.40)	–
Q18	7.83 (121.38)	179.13	3.82 (58.70)	1.87, 2.00 (27.79)	2.17 (33.39)	–
L19	7.89 (120.81)	180.47	4.29 (57.80)	1.30, 1.80 (41.92)	1.53 (26.89)	0.64 (23.31), 0.75 (25.30)
T20	8.03 (113.86)	175.51	3.91 (65.43)	4.17 (69.09)	1.16 (21.57)	–
S21	7.68 (115.82)	174.69	4.42 (59.26)	3.87, 3.96 (64.19)	–	–
G22	7.98 (108.67)	174.04	3.54, 4.32 (45.74)	–	–	–
G23	7.87 (111.30)	170.19	3.46, 4.31 (44.23)	–	–	–
A24	7.82 (119.11)	174.30	3.82 (51.33)	0.42 (21.41)	–	–
S25	7.70 (116.50)	172.36	4.99 (57.12)	3.26, 3.39 (63.67)	–	–
V26	8.79 (126.92)	174.11	4.00 (60.96)	1.60 (32.58)	0.42 (20.77)	–
V27	8.09 (124.83)	174.00	5.00 (61.08)	1.71 (35.58)	0.60 (21.17)	–
C28	9.57 (126.43)	173.98	5.40 (53.24)	2.81, 2.85 (43.70)	–	–
F29	–	174.37	4.77 (54.38)	2.42, 2.56 (41.61)	–	–
L30	8.76 (119.89)	–	5.22 (52.67)	– (41.89)	–	–
N31	–	172.04	3.99 (54.94)	2.65, 2.77 (37.82)	–	–
N32	8.79 (114.08)	174.66	5.09 (54.27)	1.86, 2.42 (42.93)	–	–

Table A.3: continued

Residue	H^N (N)	C'	H^α (C^α)	H^β (C^β)	H^γ (C^γ)	H^δ (C^δ)
F33	9.23 (120.10)	175.10	4.33 (58.91)	2.87, 3.07 (42.31)	–	–
Y34	–	–	–	–	–	–
P35	–	–	– (62.07)	– (35.00)	–	–
K36	8.16 (118.07)	175.32	3.42 (58.98)	0.83 (33.36)	0.03, 0.83 (24.90)	0.77, 0.91 (29.16)
D37	7.29 (113.86)	174.09	4.24 (55.12)	2.24, 2.30 (40.75)	–	–
I38	7.92 (122.77)	172.03	4.48 (60.38)	1.48 (40.90)	0.52, 1.24 (28.48), 0.63 (16.59)	−0.41 (13.52)
N39	7.85 (122.74)	173.27	4.96 (52.35)	2.44, 2.49 (42.23)	–	–
V40	8.58 (124.66)	174.14	4.59 (60.94)	1.57 (34.43)	0.40 (20.82), 0.48 (22.23)	–
K41	8.88 (127.98)	173.76	4.57 (54.44)	1.44 (36.36)	– (24.47)	–
W42	8.82 (122.84)	176.94	5.21 (56.17)	2.77, 2.97 (32.15)	–	–
K43	9.19 (118.73)	175.11	4.77 (54.32)	0.99, 1.25 (37.17)	0.73, 0.91 (25.37)	1.15, 1.21 (29.43)
I44	8.57 (121.01)	176.61	4.37 (59.49)	1.19 (39.49)	0.77, 1.20 (25.85), −0.17 (16.87)	0.60 (14.11)
D45	9.67 (129.85)	176.42	4.23 (55.97)	2.56, 2.97 (39.78)	–	–
G46	8.29 (129.73)	173.84	3.35, 4.08 (45.41)	–	–	–
S47	7.79 (116.71)	174.03	4.6, (57.18)	3.61, 3.72 (64.51)	–	–
E48	–	175.75	4.55 (57.26)	1.80 (30.30)	1.95 (36.47)	–
R49	8.41 (126.70)	174.26	4.52 (53.82)	1.43, 1.59 (32.04)	1.41 (27.23)	3.01, 3.17 (42.10)
Q50	–	–	–	–	–	–
N51	–	–	–	–	–	–
G52	–	173.27	3.52, 3.95 (45.83)	–	–	–
V53	7.29 (119.98)	175.90	4.48 (61.61)	1.79 (33.42)	0.50 (22.11), 0.78 (21.09)	–
L54	8.94 (130.26)	–	4.62 (53.95)	1.34, 1.47 (44.98)	1.30 (26.81)	0.76 (24.22)
N55	–	174.89	5.71 (51.82)	2.38, 2.70 (39.91)	–	–
S56	9.00 (117.57)	174.91	5.27 (57.10)	3.65 (65.32)	–	–
W57	–	–	–	–	–	–
T58	–	–	–	–	–	–
D59	–	176.18	4.61 (54.41)	2.48, 2.62 (41.44)	–	–
Q60	8.51 (121.69)	175.76	4.41 (56.05)	1.85, 2.06 (29.70)	2.25, 2.29 (33.88)	–
D61	–	–	–	–	–	–
S62	–	175.10	4.08 (59.60)	3.72, 3.78 (63.63)	–	–
K63	8.50 (122.20)	177.06	4.14 (57.28)	1.66 (32.18)	1.21, 1.27 (24.36)	1.42 (28.81)
D64	8.09 (118.49)	175.81	4.46 (53.56)	2.38, 2.57 (41.51)	–	–
S65	–	173.32	3.95 (59.92)	3.57, 3.65 (62.65)	–	–

Table A.3: continued

Residue	H^N (N)	C'	H^α (C^α)	H^β (C^β)	H^γ (C^γ)	H^δ (C^δ)
T66	7.55 (108.86)	172.91	4.27 (60.80)	3.88 (71.53)	0.77 (21.29)	–
Y67	8.90 (118.56)	174.31	4.47 (57.85)	1.57, 2.70 (42.15)	–	–
S68	8.09 (111.35)	172.40	5.30 (58.06)	3.68, 4.04 (67.11)	–	–
M69	–	–	–	–	–	–
S70	–	–	–	–	–	–
S71	–	174.33	4.37 (56.56)	3.26 (65.29)	–	–
T72	9.08 (125.15)	172.55	4.91 (61.94)	3.79 (70.27)	0.78 (20.75)	–
L73	8.93 (131.76)	174.66	4.17 (52.95)	−0.88, 0.48 (41.09)	0.71 (26.31)	0.20 (23.08), 0.29 (25.79)
T74	8.40 (123.01)	173.22	4.61 (62.30)	3.75 (69.94)	0.87 (21.17)	–
L75	8.40 (126.93)	176.98	4.72 (53.19)	1.57, 1.61 (46.01)	1.38 (27.11)	0.71 (23.26), 0.69 (23.99)
T76	8.49 (111.36)	175.26	4.44 (61.39)	4.59 (70.64)	1.23 (21.98)	–
K77	8.20 (122.78)	177.01	3.59 (60.49)	1.46 (32.56)	1.25 (24.48)	1.53, 1.70 (29.42)
D78	8.23 (114.97)	178.11	4.15 (57.00)	2.42 (40.77)	–	–
E79	7.66 (118.68)	178.54	3.80 (59.42)	1.92, 2.02 (30.28)	2.29 (36.95)	–
Y80	8.50 (120.35)	178.65	4.62 (59.10)	2.74, 2.87 (39.27)	–	–
E81	7.86 (112.78)	176.00	3.62 (57.42)	1.88, 1.98 (29.18)	2.03, 2.66 (37.95)	–
R82	7.37 (117.23)	175.24	4.05 (56.36)	1.42, 1.45 (31.00)	1.37, 1.55 (27.53)	2.90, 2.97 (43.46)
H83	7.04 (115.95)	173.25	4.78 (54.20)	2.50, 2.96 (31.09)	–	–
N84	9.44 (119.46)	174.16	4.98 (55.28)	2.63, 2.73 (41.92)	–	–
S85	8.70 (114.75)	172.64	5.01 (57.02)	3.50, 3.56 (65.13)	–	–
Y86	8.63 (126.39)	174.28	4.57 (57.99)	2.32, 2.54 (41.61)	–	–
T87	8.89 (118.00)	171.38	4.84 (61.44)	3.61 (72.04)	0.85 (22.12)	–
C88	8.77 (125.08)	171.24	4.05 (52.63)	1.90 (43.96)	–	–
E89	8.65 (123.14)	175.06	4.96 (53.55)	1.45, 1.55 (32.69)	1.64, 1.72 (36.32)	–
A90	9.41 (128.85)	176.03	5.26 (49.67)	0.87 (22.16)	–	–
T91	8.81 (116.69)	172.78	4.49 (61.46)	3.76 (70.07)	0.90 (21.15)	–
H92	8.85 (127.28)	175.13	4.55 (56.61)	2.50, 2.62 (36.25)	–	–
K93	7.92 (124.17)	176.12	3.79 (58.63)	1.22, 1.50 (32.48)	1.15 (24.27)	1.51 (28.70)
T94	7.38 (108.10)	174.97	3.93 (63.05)	4.33 (70.35)	1.38 (24.15)	–
S95	8.02 (114.70)	174.87	4.75 (57.41)	3.73, 3.81 (64.82)	–	–
T96	–	174.94	4.05 (63.59)	4.23 (68.68)	1.12 (21.76)	–
S97	8.04 (117.73)	172.10	4.77 (55.94)	3.58, 3.70 (63.71)	–	–
P98	–	–	– (62.83)	– (31.92)	–	–

Table A.3: continued

Residue	H^N (N)	C'	H^α (C^α)	H^β (C^β)	H^γ (C^γ)	H^δ (C^δ)
I99	9.11 (125.13)	175.08	3.97 (61.39)	1.50 (39.30)	0.97, 1.52 (27.28), 0.70 (17.90)	0.74 (14.37)
V100	8.32 (125.98)	175.91	4.65 (61.06)	1.69 (34.50)	0.59 (20.52), 0.48 (21.72)	–
K101	8.59 (126.81)	173.91	4.59 (53.71)	1.42, 1.60 (35.69)	0.99, 1.20 (24.71)	1.67, 1.80 (28.14)
S102	8.45 (116.44)	173.67	5.67 (56.78)	3.37, 3.45 (66.84)	–	–
F103	9.02 (121.46)	171.82	4.77 (56.95)	2.87, 3.10 (42.32)	–	–
N104	8.43 (121.26)	175.16	5.41 (50.69)	2.41, 2.52 (40.30)	–	–
R105	9.17 (124.51)	176.35	4.07 (58.39)	1.73, 1.87 (31.10)	1.62 (27.77)	2.58, 2.77 (42.62)
N106	8.62 (117.26)	173.95	4.63 (53.01)	2.56, 2.64 (39.15)	–	–
E107	7.69 (125.87)	173.97	3.97 (58.13)	1.77, 1.92 (31.26)	2.06, 2.09 (36.45)	–

Table A.4: Backbone and side chain assignments of the C_L^{P35A} mutant

Residue	H^N (N)	C'	H^α (C^α)	H^β (C^β)	H^γ (C^γ)	H^δ (C^δ)
G1	–	–	–	–	–	–
S2	–	–	–	–	–	–
H3	–	174.99	4.44 (56.12)	2.87, 2.91 (30.31)	–	–
M4	8.11 (122.15)	175.20	4.24 (55.08)	1.72, 1.85 (32.63)	2.21, 2.28 (31.79)	–
A5	8.15 (125.91)	176.13	4.12 (52.02)	1.15 (19.08)	–	–
A6	8.06 (124.90)	174.49	4.47 (50.05)	1.17 (18.45)	–	–
P7	–	–	– (62.41)	– (30.77)	–	–
T8	8.53 (113.73)	172.35	4.56 (60.49)	3.99 (72.90)	1.04 (21.21)	–
V9	8.32 (122.58)	174.75	4.99 (60.62)	1.55 (34.87)	0.65 (21.98), 0.79 (22.72)	–
S10	8.72 (120.85)	–	4.78 (56.95)	3.59, 3.64 (66.14)	–	–
I11	–	173.03	5.37 (59.07)	1.38 (42.06)	0.90, 1.66 (29.37), 0.71 (16.14)	0.84 (14.94)
F12	9.31 (126.43)	–	4.84 (54.27)	2.72, 2.94 (41.56)	–	–
P13	–	–	–	–	–	–
P14	–	–	– (62.06)	–	–	–
S15	8.60 (121.19)	–	4.68 (59.63)	3.88 (62.73)	–	–
S16	–	–	–	–	–	–
E17	–	178.90	3.91 (59.48)	1.78, 1.90 (28.68)	2.10, 2.19 (36.66)	–
Q18	7.81 (121.56)	179.15	3.82 (58.59)	1.87, 2.00 (27.58)	2.17 (33.72)	–
L19	7.86 (120.88)	180.46	4.29 (57.87)	1.29, 1.78 (41.80)	1.53 (27.59)	0.64 (23.47), 0.74 (25.56)
T20	7.99 (114.08)	175.42	3.91 (65.36)	4.17 (68.91)	1.16 (21.80)	–
S21	7.61 (115.78)	174.59	4.42 (59.09)	3.87, 3.95 (64.20)	–	–
G22	7.93 (108.66)	173.94	3.54, 4.32 (45.61)	–	–	–
G23	7.80 (111.32)	170.06	3.45, 4.31 (44.37)	–	–	–
A24	7.75 (118.96)	174.14	3.83 (51.28)	0.42 (21.36)	–	–
S25	7.69 (116.62)	172.25	5.02 (57.23)	3.27, 3.42 (63.42)	–	–
V26	8.80 (127.02)	173.83	4.01 (60.78)	1.63 (32.31)	0.42 (21.08)	–
V27	8.02 (125.02)	174.17	5.02 (61.13)	1.78 (35.37)	0.50 (21.47)	–
C28	9.54 (125.81)	–	5.43 (53.56)	2.81 (44.07)	–	–
F29	–	174.60	–	–	–	–
L30	8.94 (119.92)	–	5.11 (52.10)	– (41.96)	–	–
N31	–	–	–	–	–	–
N32	–	172.98	3.85 (54.58)	2.52, 2.71 (36.64)	–	–

Table A.4: continued

Residue	H^N (N)	C'	H^α (C^α)	H^β (C^β)	H^γ (C^γ)	H^δ (C^δ)
F33	8.43 (115.97)	–	4.14 (58.62)	2.76, 2.80 (40.60)	–	–
Y34	–	–	–	–	–	–
A35	–	–	–	–	–	–
K36	–	176.21	4.16 (56.57)	– (32.06)	–	–
D37	7.94 (118.63)	174.58	4.46 (53.41)	2.41, 2.50 (39.53)	–	–
I38	7.24 (119.28)	174.73	4.20 (60.44)	1.31 (39.80)	0.58, 1.03 (26.98), 0.25 (18.10)	0.16 (14.04)
N39	8.41 (123.27)	173.12	4.77 (52.67)	2.46, 2.50 (41.34)	–	–
V40	8.41 (122.79)	174.56	4.76 (60.77)	1.59 (34.42)	0.40 (21.39), 0.54 (23.24)	–
K41	8.73 (127.72)	174.16	4.56 (54.48)	1.46 (36.00)	– (24.83)	–
W42	8.90 (124.25)	176.91	5.25 (56.13)	2.77, 3.02 (31.99)	–	–
K43	9.13 (118.79)	175.03	4.80 (54.34)	1.01, 1.27 (36.98)	0.72, 0.95 (25.55)	1.17, 1.22 (29.80)
I44	8.49 (120.83)	176.53	4.39 (59.48)	1.19 (39.31)	0.79, 1.19 (26.20), −0.17 (17.12)	0.59 (14.23)
D45	9.66 (129.95)	176.17	4.24 (55.80)	2.56, 2.97 (39.64)	–	–
G46	8.26 (101.75)	173.75	3.36, 4.08 (45.34)	–	–	–
S47	7.74 (116.69)	173.89	4.62 (57.22)	3.61, 3.71 (64.33)	–	–
E48	–	175.66	4.54 (57.64)	1.81 (29.95)	1.96 (36.62)	–
R49	8.35 (126.47)	174.13	4.51 (53.98)	1.44, 1.58 (32.22)	1.84 (27.27)	3.00, 3.14 (42.47)
Q50	–	–	–	–	–	–
N51	–	–	–	–	–	–
G52	–	173.16	3.53, 3.93 (45.74)	–	–	–
V53	7.24 (119.95)	–	4.42 (61.84)	1.77 (32.96)	0.78 (21.79)	–
L54	8.86 (130.44)	–	4.59 (53.68)	1.31, 1.45 (45.07)	1.26 (27.17)	0.69 (24.44)
N55	–	–	–	–	–	–
S56	–	–	–	–	–	–
W57	–	–	–	–	–	–
T58	–	–	–	–	–	–
D59	–	–	–	–	–	–
Q60	–	–	–	–	–	–
D61	–	–	–	–	–	–
S62	–	–	–	–	–	–
K63	–	–	–	–	–	–
D64	–	–	–	–	–	–
S65	–	–	–	–	–	–

Table A.4: continued

Residue	H^N (N)	C'	H^α (C^α)	H^β (C^β)	H^γ (C^γ)	H^δ (C^δ)
T66	–	–	–	–	–	–
Y67	–	–	–	–	–	–
S68	–	–	–	–	–	–
M69	–	–	–	–	–	–
S70	–	–	–	–	–	–
S71	–	174.06	5.25 (56.61)	3.68, 3.87 (65.12)	–	–
T72	8.98 (124.86)	172.38	4.96 (61.91)	4.59 (70.61)	0.77 (21.76)	–
L73	8.90 (131.86)	–	4.17 (53.15)	−0.92, 0.48 (41.18)	0.28 (26.71)	0.17 (23.56)
T74	–	173.09	4.60 (62.39)	3.73 (69.20)	0.86 (21.31)	–
L75	8.37 (127.12)	176.95	4.73 (53.36)	1.58, 1.63 (45.82)	0.81 (27.42)	0.69 (24.47)
T76	8.46 (111.41)	175.56	4.44 (61.39)	3.79 (70.55)	0.79 (21.20)	–
K77	8.13 (122.72)	176.94	3.58 (60.67)	1.45 (32.17)	1.24, 1.29 (24.70)	1.52 (29.41)
D78	8.20 (114.96)	178.03	4.15 (56.98)	2.42 (40.58)	–	–
E79	7.62 (118.70)	178.47	3.79 (59.19)	1.92, 2.02 (30.08)	2.29 (37.17)	–
Y80	8.45 (120.32)	178.59	4.62 (59.02)	2.75, 2.86 (39.05)	–	–
E81	7.80 (112.65)	175.90	3.62 (57.30)	1.88, 1.97 (28.98)	2.02, 2.67 (38.24)	–
R82	7.31 (117.24)	175.12	4.03 (56.27)	1.41, 1.47 (30.73)	1.35, 1.53 (27.84)	2.90, 2.97 (43.71)
H83	6.99 (115.88)	173.12	4.77 (54.36)	2.50, 2.92 (30.92)	–	–
N84	9.36 (119.37)	174.03	4.97 (55.27)	2.62, 2.73 (41.80)	–	–
S85	8.63 (114.54)	172.62	5.04 (56.97)	3.51, 3.57 (64.90)	–	–
Y86	8.64 (126.33)	173.89	4.62 (58.06)	2.42, 2.61 (41.55)	–	–
T87	8.85 (117.90)	171.54	4.94 (61.23)	3.63 (72.14)	0.86 (22.08)	–
C88	8.73 (123.79)	171.24	4.27 (52.66)	1.63, 1.92 (44.36)	–	–
E89	8.52 (121.92)	174.83	4.94 (53.54)	1.51, 1.66 (32.470	1.72, 1.75 (36.44)	–
A90	9.21 (127.88)	176.37	5.22 (50.31)	1.02 (22.55)	–	–
T91	8.63 (114.45)	172.62	4.59 (60.59)	3.94 (71.63)	0.91 (22.01)	–
H92	8.41 (119.21)	175.81	4.78 (55.72)	2.95, 3.13 (32.84)	–	–
K93	–	177.63	3.96 (59.03)	1.75, 1.79 (32.43)	–	–
T94	8.29 (108.72)	174.59	4.27 (61.50)	4.39 (68.72)	1.09 (22.13)	–
S95	7.45 (116.77)	174.79	4.30 (58.17)	3.43, 3.55 (63.71)	–	–
T96	–	174.52	4.19 (62.24)	4.20 (68.92)	1.09 (21.84)	–
S97	7.83 (117.36)	171.77	4.66 (56.14)	3.62, 3.75 (63.75)	–	–
P98	–	–	– (61.73)	– (31.98)	–	–

Table A.4: continued

Residue	H^N (N)	C'	H^α (C^α)	H^β (C^β)	H^γ (C^γ)	H^δ (C^δ)
I99	8.79 (123.71)	174.77	4.12 (61.89)	1.69 (38.82)	1.02, 1.59 (27.53), 0.79 (18.91)	0.72 (13.98)
V100	8.31 (125.72)	175.72	4.80 (60.72)	1.70 (34.51)	0.63 (21.38)	–
K101	8.58 (126.89)	173.69	4.57 (54.02)	1.61 (35.96)	1.12, 1.26 (24.99)	1.19, 1.40 (29.15)
S102	8.35 (116.32)	173.36	5.69 (56.77)	3.40, 3.49 (66.51)	–	–
F103	9.03 (121.89)	171.76	4.78 (56.91)	2.87, 3.11 (42.19)	–	–
N104	8.39 (121.39)	175.05	5.39 (50.73)	2.41, 2.51 (40.04)	–	–
R105	9.11 (124.54)	176.31	4.02 (58.38)	1.72, 1.86 (30.83)	1.61 (28.03)	2.54, 2.77 (42.98)
N106	8.58 (116.98)	173.93	4.62 (53.05)	2.57, 2.62 (38.97)	–	–
E107	7.62 (125.81)	180.87	3.96 (58.19)	1.77, 1.92 (31.14)	2.09 (36.65)	–

Table A.5: Talos dihedral angle restraints of the wild type C_L domain

Residue	$\phi \pm \Delta\phi$ [°]	$\psi \pm \Delta\psi$ [°]
G1	–	–
S2	–	–
H3	-86 ± 11	-15 ± 14
M4	-113 ± 32	145 ± 12
A5	-109 ± 23	126 ± 11
A6	-95 ± 27	125 ± 15
P7	–	–
T8	–	–
V9	-111 ± 15	131 ± 10
S10	-139 ± 20	154 ± 15
I11	–	–
F12	–	–
P13	–	–
P14	-64 ± 5	139 ± 10
S15	-60 ± 9	-32 ± 8
S16	–	–
E17	-67 ± 5	-41 ± 6
Q18	-66 ± 3	-43 ± 8
L19	-67 ± 7	-36 ± 6
T20	-63 ± 8	-39 ± 11
S21	-90 ± 15	-1 ± 14
G22	84 ± 12	15 ± 16
G23	-94 ± 26	154 ± 17
A24	-116 ± 31	150 ± 16
S25	-105 ± 24	117 ± 13
V26	-109 ± 17	125 ± 6
V27	-116 ± 9	129 ± 10
C28	-117 ± 10	136 ± 10
F29	-130 ± 13	139 ± 13
L30	-128 ± 30	139 ± 16
N31	-84 ± 28	150 ± 20
N32	-128 ± 19	167 ± 17
F33	-90 ± 18	127 ± 31
Y34	–	–
P35	-62 ± 12	149 ± 14
K36	-62 ± 3	-19 ± 13
D37	-88 ± 10	-10 ± 10
I38	-106 ± 15	137 ± 14
N39	-123 ± 14	143 ± 10
V40	-95 ± 14	131 ± 5
K41	-133 ± 17	152 ± 17
W42	-110 ± 17	132 ± 10
K43	-139 ± 15	151 ± 11
I44	-101 ± 24	133 ± 16
D45	-65 ± 16	134 ± 9
G46	94 ± 11	-10 ± 20
S47	-99 ± 24	146 ± 23

Table A.5: continued .

Residue	$\phi \pm \Delta\phi$ [°]	$\psi \pm \Delta\psi$ [°]
E48	-84 ± 12	132 ± 14
R49	-116 ± 21	144 ± 19
Q50	–	–
N51	–	–
G52	81 ± 6	-1 ± 10
V53	-102 ± 15	136 ± 21
L54	-129 ± 18	138 ± 14
N55	-111 ± 19	129 ± 13
S56	-108 ± 13	139 ± 22
W57	–	–
T58	–	–
D59	-87 ± 30	146 ± 9
Q60	-115 ± 29	138 ± 28
D61	–	–
S62	-69 ± 16	-34 ± 15
K63	-76 ± 15	-20 ± 30
D64	-101 ± 7	0 ± 14
S65	57 ± 3	46 ± 11
T66	-135 ± 27	167 ± 14
Y67	-121 ± 26	148 ± 8
S68	-147 ± 20	153 ± 14
M69	–	–
S70	–	–
S71	-104 ± 10	119 ± 12
T72	-95 ± 14	110 ± 12
L73	-101 ± 20	125 ± 10
T74	-102 ± 14	128 ± 10
L75	-120 ± 17	153 ± 12
T76	-79 ± 14	163 ± 10
K77	-59 ± 4	-41 ± 10
D78	-60 ± 7	-39 ± 6
E79	-67 ± 4	-41 ± 4
Y80	-67 ± 6	-37 ± 5
E81	-70 ± 5	-24 ± 10
R82	-91 ± 26	-32 ± 19
H83	-102 ± 18	133 ± 30
N84	-106 ± 31	145 ± 15
S85	-100 ± 18	134 ± 11
Y86	-107 ± 17	130 ± 9
T87	-115 ± 30	133 ± 11
C88	-123 ± 14	144 ± 12
E89	-124 ± 20	135 ± 18
A90	-117 ± 16	133 ± 19
T91	-117 ± 19	119 ± 10
H92	-117 ± 36	133 ± 26
K93	-62 ± 6	-36 ± 10
T94	-79 ± 10	-16 ± 12
S95	-99 ± 24	150 ± 29

Table A.5: continued .

Residue	$\phi \pm \Delta\phi$ [°]	$\psi \pm \Delta\psi$ [°]
T96	-91 ± 26	147 ± 24
S97	-102 ± 21	133 ± 31
P98	-63 ± 7	144 ± 7
I99	-91 ± 20	130 ± 12
V100	-109 ± 22	127 ± 11
K101	-126 ± 17	141 ± 20
S102	-123 ± 18	150 ± 9
F103	-138 ± 21	146 ± 15
N104	-113 ± 16	132 ± 17
R105	-63 ± 6	-23 ± 9
N106	-94 ± 11	1 ± 4
E107	–	–

Table A.6: Talos dihedral angle restraints of the C_L^{P35A} mutant

Residue	$\phi \pm \Delta\phi$ [°]	$\psi \pm \Delta\psi$ [°]
G1	–	–
S2	–	–
H3	-77 ± 6	-18 ± 16
M4	-91 ± 29	137 ± 32
A5	-89 ± 27	125 ± 13
A6	-78 ± 18	138 ± 22
P7	-59 ± 7	140 ± 10
T8	-124 ± 33	146 ± 19
V9	-116 ± 23	145 ± 13
S10	-130 ± 19	137 ± 15
I11	-120 ± 11	142 ± 13
F12	-136 ± 12	145 ± 14
P13	–	–
P14	-72 ± 18	130 ± 26
S15	-105 ± 29	115 ± 27
S16	–	–
E17	-65 ± 6	-42 ± 5
Q18	-66 ± 3	-42 ± 9
L19	-68 ± 7	-33 ± 10
T20	-65 ± 7	-36 ± 14
S21	-91 ± 14	-3 ± 15
G22	84 ± 12	15 ± 16
G23	-94 ± 26	154 ± 17
A24	-123 ± 29	150 ± 16
S25	-111 ± 27	119 ± 16
V26	-109 ± 23	124 ± 6
V27	-116 ± 8	129 ± 7
C28	-125 ± 12	136 ± 15
F29	-117 ± 24	143 ± 12

Table A.6: continued .

Residue	$\phi \pm \Delta\phi$ [°]	$\psi \pm \Delta\psi$ [°]
L30	-133 ± 27	153 ± 21
N31	–	–
N32	66 ± 9	25 ± 19
F33	-116 ± 41	140 ± 27
Y34	–	–
A35	–	–
K36	-74 ± 17	-15 ± 19
D37	-90 ± 12	-5 ± 21
I38	-122 ± 25	135 ± 26
N39	-117 ± 20	137 ± 14
V40	-108 ± 22	131 ± 5
K41	-130 ± 16	140 ± 17
W42	-106 ± 19	134 ± 9
K43	-138 ± 11	153 ± 9
I44	-98 ± 21	132 ± 18
D45	52 ± 2	42 ± 7
G46	81 ± 9	0 ± 15
S47	-111 ± 19	140 ± 14
E48	-80 ± 14	131 ± 12
R49	-110 ± 17	144 ± 21
Q50	–	–
N51	–	–
G52	86 ± 8	-1 ± 12
V53	-119 ± 18	129 ± 10
L54	-119 ± 12	139 ± 23
N55	–	–
S56	–	–
W57	–	–
T58	–	–
D59	–	–
Q60	–	–
D61	–	–
S62	–	–
K63	–	–
D64	–	–
S65	–	–
T66	–	–
Y67	–	–
S68	–	–
M69	–	–
S70	–	–
S71	-111 ± 13	127 ± 16
T72	-124 ± 17	127 ± 8
L73	-102 ± 26	124 ± 27
T74	-91 ± 16	127 ± 12
L75	-116 ± 18	150 ± 15
T76	-78 ± 13	163 ± 10
K77	-59 ± 4	-41 ± 10

Table A.6: continued .

Residue	$\phi \pm \Delta\phi$ [°]	$\psi \pm \Delta\psi$ [°]
D78	-61 ± 6	-38 ± 6
E79	-66 ± 4	-41 ± 4
Y80	-70 ± 15	-30 ± 10
E81	-69 ± 6	-25 ± 11
R82	-90 ± 25	-25 ± 27
H83	-102 ± 18	133 ± 30
N84	-106 ± 29	147 ± 15
S85	-100 ± 18	132 ± 14
Y86	-108 ± 15	130 ± 9
T87	-109 ± 26	134 ± 11
C88	-133 ± 16	153 ± 7
E89	-118 ± 20	132 ± 15
A90	-126 ± 13	134 ± 10
T91	-123 ± 22	142 ± 19
H92	-90 ± 32	139 ± 20
K93	-65 ± 3	-25 ± 10
T94	-92 ± 24	-10 ± 19
S95	-77 ± 20	145 ± 23
T96	-93 ± 19	-10 ± 22
S97	-97 ± 19	137 ± 29
P98	-65 ± 7	146 ± 9
I99	-103 ± 22	125 ± 14
V100	-112 ± 22	126 ± 10
K101	-127 ± 18	142 ± 18
S102	-118 ± 17	150 ± 10
F103	-138 ± 21	146 ± 15
N104	-113 ± 23	132 ± 17
R105	-64 ± 7	-23 ± 10
N106	-81 ± 27	-7 ± 29
E107	–	–

Appendix B

The C_H1 Domain of the Immunoglobulin G Antibody

B.1 Sample Preparation

The murine IgG1 MAK33 cDNA served as a template to amplify the gene of the antibody domain C_H1. The PCR product was cloned into the pET28a vector (Novagen, Gibbstown, NJ, USA) without any attached tag. The C_H1 domain was expressed as inclusion bodies overnight at 37 °C in E. coli BL21(DE3) cells and selective LB medium. Isotope labeled protein for the NMR experiments was expressed in M9 medium with ^{15}N ammonium chloride as the only nitrogen source or additionally ^{13}C as the only carbon source. Inclusion bodies were isolated as published.[345] After expression overnight at 37 °C, cells were harvested by centrifugation (4000 g). The cell pellet was resuspended in 100 mM Tris-HCl, 100 mM NaCl, 5 mM EDTA, pH 7.5 at 4 °C. Cells were cracked with a cell disruptor and DNA was removed by the addition of 10 μg/mL DNAseI, 3 mM $MgCl_2$ and incubation for 30 min at room temperature. Thereafter, 2.5% (v/v) Triton X-100 was added and the mixture was stirred for further 30 min at 4 °C. Inclusion bodies were isolated by centrifugation (40000 g, 30 min, 4 °C) and washed two times with 100 mM Tris-HCl, 100 mM NaCl, 5 mM EDTA, pH 7.5. Subsequently, inclusion bodies were solubilized in 50 mM Tris-HCl, pH 8.0, 10 mM β-mercaptoethanol, 10 mM EDTA, 8 M urea and insoluble components were removed by centrifugation (40000 g). The supernatant was applied to a Q-sepharose column equilibrated in the same buffer with 5 M urea. Under these conditions, C_H1 did not bind to the column. Refolding was carried out via dialysis in 250 mM Tris-HCl, pH 8.2, 100 mM L-Arg, 10 mM EDTA, 1 mM GSSG, 0.5 mM GSH overnight at 4 °C. After refolding and an additional centrifugation step to remove aggregates, the protein was applied to a Superdex 75 26/60 gel filtration column (GE Healthcare, München, Germany) equilibrated in PBS

buffer. This protocol results in a yield of around 5 mg/mL pure protein. The vector with the gene of the C_H1 domain was sequenced and the protein mass was verified by MALDI-TOF MS.

For the NMR analysis the concentration of either the assembled or the unassembled C_H1 domain in PBS buffer with 5% (v/v) D_2O was 0.5 - 1.0 mM. Spectra of the C_H1 domain in complex with the C_L domain were recorded at 25 °C on Bruker DMX600, and DMX750 spectrometers, whereas spectra of the unfolded C_H1 domain were measured at 12 °C on a Bruker AVANCE900 spectrometer. Processing of all spectra was performed using the program Topspin 1.3 (Bruker BioSpin) and analysis of the spectra was carried out with the program Sparky (www.cgl.ucsf.edu/home/sparky). In order to ensure complete folding of the C_H1 domain for the spectra of the assembled C_H1, unlabeled C_L was added to labeled C_H1 at a molar ratio of $\sim 2:1$ and the folding reaction was allowed to proceed for at least 6 h at room temperature. For the RDC measurement, this complex was subsequently partially aligned with non-ionic liquid crystalline media composed of *n*-alkyl-polyethylene glycol and *n*-alkyl alcohol. According to the paper by Otting and coworkers,[223] for the given salt concentration, pH value, and temperature range 3 wt% C12E5/hexanol (molar ratio r = 0.96) was chosen as alignment media. In the case of the assembled C_H1 domain, the sample volume of 400 μL resulted in 12.5 μL C12E5 and 2.5 μL hexanol to obtain a stable alignment. The association coupled folding reaction of the C_H1 domain was followed by real time ^{1}H-^{15}N HSQC spectra recorded every 14 minutes directly after adding unlabeled C_L to the labeled C_H1 at 12.5 °C due to the reduced reaction rate at lower temperature.

B.2 Experiments and Assignment of the Folded and Unfolded C_H1 Domain

Table B.1: Experiments on the assembled C_H1 domain

Experiment	Nuclei	NS	Time domain	Data matrix
NHSQC	$^{1}H\times^{15}N$	8	1024×128	1024×256
HNCO	$^{1}H\times^{15}N\times^{13}C$	8	$1024\times72\times 60$	$1024\times128\times256$
HNcaCO	$^{1}H\times^{15}N\times^{13}C$	32	$1024\times52\times 72$	$1024\times128\times256$
HNCA	$^{1}H\times^{15}N\times^{13}C$	16	$1024\times64\times 92$	$1024\times128\times256$
HNcoCA	$^{1}H\times^{15}N\times^{13}C$	48	$1024\times52\times 72$	$1024\times128\times256$
IPAP-HSQC	$^{1}H\times^{15}N$	128	1024×256	1024×512
MEXICO	$^{1}H\times^{15}N$	128	1024×128	1024×256

Table B.2: Experiments on the unassembled C_H1 domain

Experiment	Nuclei	NS	Time domain	Data matrix
NHSQC	$^{1}H\times^{15}N$	8	1024×256	1024×512
HNCO	$^{1}H\times^{15}N\times^{13}C$	8	$1024\times72\times\ 96$	$1024\times128\times256$
HNcaCO	$^{1}H\times^{15}N\times^{13}C$	16	$1024\times72\times\ 96$	$1024\times128\times256$
CBCAcoNH	$^{1}H\times^{15}N\times^{13}C$	24	$1024\times72\times\ 92$	$1024\times128\times256$
HNCACB	$^{1}H\times^{15}N\times^{13}C$	32	$1024\times70\times\ 90$	$1024\times128\times256$
HNHnoesy	$^{1}H\times^{15}N\times\ ^{1}H$	16	$1024\times72\times128$	$1024\times128\times256$
NNHnoesy	$^{1}H\times^{15}N\times^{15}N$	16	$1024\times96\times\ 96$	$1024\times128\times128$
HetNOE	$^{1}H\times^{15}N$	128	1024×220	1024×512
MEXICO	$^{1}H\times^{15}N$	64	1024×128	1024×256

Table B.3: Backbone assignments of the assembled C_H1 domain

Residue	H^N (N)	C'	C^α
T1	8.17 (120.92)	176.25	62.50
T2	8.26 (122.47)	176.19	56.34
P3	–	–	–
P4	–	–	62.00
S5	8.30 (116.32)	176.00	58.38
V6	–	176.29	63.45
Y7	8.54 (127.25)	168.73	58.07
P8	–	–	62.17
L9	8.35 (121.97)	175.91	53.02
A10	7.75 (124.62)	175.53	48.54
P11	–	–	62.46
G12	8.69 (106.75)	174.25	45.23
S13	8.22 (114.67)	174.67	58.83
A14	8.52 (125.03)	177.51	52.41
A15	8.05 (123.08)	177.88	52.48
Q16	8.36 (120.00)	176.24	55.74
T17	8.07 (121.75)	175.57	61.97
N18	8.34 (124.89)	176.07	53.89
S19	–	174.46	53.15
M20	7.97 (120.24)	174.94	53.58
V21	9.45 (122.29)	173.30	60.52
T22	7.74 (123.26)	172.00	62.38
L23	8.86 (126.51)	175.60	52.11
G24	8.32 (130.75)	171.71	46.81
C25	9.08 (117.66)	171.76	54.63
L26	9.09 (127.44)	169.77	53.91
V27	–	173.64	61.51
K28	8.94 (126.30)	176.53	55.65
G29	8.23 (115.73)	176.34	48.12
Y30	–	173.10	59.27
F31	8.30 (112.05)	175.68	55.56
P32	–	–	63.32
E33	7.79 (117.31)	174.91	56.87
P34	–	–	62.36
V35	7.46 (110.17)	175.36	60.05
T36	8.63 (116.01)	173.56	60.69
V37	8.69 (125.57)	175.66	60.82
T38	8.80 (118.48)	171.08	59.56
W39	9.07 (118.99)	177.78	55.63
N40	–	176.77	53.96
S41	9.52 (110.91)	174.27	58.95
G42	7.75 (106.04)	174.97	43.98
S43	7.93 (115.70)	174.43	60.58
L44	–	175.93	52.47
S45	8.17 (122.36)	173.18	55.87
S46	7.20 (117.03)	176.04	53.17
G47	8.01 (111.98)	173.75	46.09

Table B.3: continued .

Residue	H^N (N)	C'	C^α
V48	7.05 (120.43)	176.63	60.75
H49	8.81 (124.61)	172.44	54.79
T50	8.26 (114.17)	172.94	62.16
F51	8.55 (129.56)	172.97	57.65
P52	–	–	62.85
A53	9.06 (126.32)	177.79	52.30
V54	8.50 (118.57)	174.43	60.54
L55	7.04 (122.00)	173.90	53.90
Q56	9.55 (127.55)	–	52.36
S57	–	173.21	53.82
D58	9.16 (128.77)	176.05	54.20
L59	8.82 (126.38)	175.16	55.61
Y60	–	171.79	58.37
T61	8.09 (124.12)	173.82	60.69
L62	9.26 (125.69)	173.17	52.36
S63	–	172.48	56.87
S64	8.60 (113.20)	172.39	56.11
S65	9.40 (123.80)	171.58	56.00
V66	8.27 (119.43)	171.70	59.49
T67	7.13 (122.42)	173.94	60.44
V68	8.87 (119.67)	172.22	57.07
P69	–	–	63.14
S70	9.03 (121.78)	179.50	60.73
S71	7.93 (112.83)	175.53	59.23
T72	7.96 (114.40)	171.88	63.33
W73	7.25 (122.08)	174.82	55.10
P74	–	–	–
S75	–	174.48	62.91
E76	8.22 (120.36)	175.04	54.27
T77	7.93 (113.75)	175.41	62.80
V78	9.40 (127.87)	173.89	62.67
T79	8.77 (122.33)	172.91	60.45
C80	8.60 (123.44)	171.69	52.95
N81	8.91 (123.69)	173.92	51.33
V82	9.12 (124.96)	174.05	60.84
A83	8.61 (129.92)	175.37	50.07
H84	8.44 (120.43)	172.51	52.12
P85	–	–	62.08
A86	8.15 (118.10)	178.16	54.47
S87	6.82 (108.33)	173.96	57.12
S88	7.87 (116.23)	173.99	58.91
T89	–	174.66	61.90
K90	8.64 (128.70)	174.63	56.16
V91	8.84 (124.12)	172.73	60.59
D92	8.37 (126.35)	175.22	52.53
K93	8.86 (123.55)	173.73	52.62
K94	8.57 (127.92)	176.28	56.36
I95	8.09 (122.79)	175.92	60.62

Table B.3: continued .

Residue	H^N (N)	C'	C^α
V96	8.16 (126.80)	174.16	59.34
P97	–	–	63.27
R98	7.84 (126.41)	180.98	57.35

Table B.4: Backbone assignments of the unassembled C_H1 domain

Residue	H^N (N)	C'	C^α	C^β
T1	–	–	–	–
T2	–	–	–	–
P3	–	–	–	–
P4	–	–	63.05	31.88
S5	8.25 (116.13)	174.02	58.22	63.78
V6	7.89 (121.13)	175.11	61.89	32.70
Y7	8.12 (124.56)	173.70	55.55	38.05
P8	–	176.31	62.99	–
L9	8.06 (122.07)	176.68	54.75	42.31
A10	8.19 (126.23)	175.40	50.31	17.87
P11	–	177.57	63.51	31.89
G12	8.44 (109.63)	174.28	45.18	–
S13	7.99 (115.52)	174.31	58.60	63.91
A14	8.28 (125.79)	177.40	52.55	19.07
A15	8.09 (122.58)	177.75	52.59	18.96
Q16	8.18 (119.24)	176.19	55.82	29.30
T17	8.07 (115.07)	174.22	62.07	69.61
N18	8.37 (120.87)	175.15	53.02	38.82
S19	–	–	–	–
M20	–	175.11	55.63	26.60
V21	8.03 (121.22)	176.01	62.47	32.66
T22	8.17 (118.25)	174.20	61.92	69.61
L23	8.23 (125.04)	177.67	55.72	42.24
G24	8.24 (108.78)	–	45.39	–
C25	–	–	–	–
L26	–	–	–	–
V27	–	175.83	62.02	–
K28	8.27 (125.34)	176.64	56.72	32.61
G29	8.24 (109.99)	172.88	44.86	–
Y30	7.76 (120.08)	174.41	57.86	39.20
F31	7.97 (124.06)	172.70	54.83	39.17
P32	–	–	–	–
E33	–	–	–	–
P34	–	176.66	63.01	31.92
V35	8.15 (120.56)	176.25	62.66	32.39
T36	8.07 (118.05)	174.31	62.05	69.69
V37	8.28 (120.43)	172.03	59.95	–
T38	–	173.07	61.85	69.82

Table B.4: continued .

Residue	H^N (N)	C'	C^α	C^β
W39	8.09 (125.14)	174.51	55.49	30.58
N40	–	–	–	–
S41	–	174.99	–	–
G42	8.29 (110.56)	178.53	45.14	–
S43	–	174.54	58.32	63.82
L44	8.19 (123.69)	177.51	55.38	41.97
S45	–	174.43	58.26	63.79
S46	8.16 (117.40)	174.74	58.53	63.66
G47	8.23 (110.44)	173.28	45.12	–
V48	7.78 (118.86)	175.67	62.27	32.42
H49	8.23 (123.19)	–	56.09	30.90
T50	–	–	–	–
F51	–	–	–	–
P52	–	176.14	56.02	28.72
A53	8.20 (124.32)	177.61	52.29	19.00
V54	7.99 (119.47)	175.84	62.16	32.48
L55	8.23 (125.98)	177.09	54.90	42.09
Q56	8.31 (121.76)	175.93	56.13	28.99
S57	–	174.34	58.61	63.75
D58	8.23 (122.00)	176.18	54.61	40.81
L59	7.83 (120.96)	177.26	55.75	41.98
Y60	7.92 (119.32)	175.95	57.95	38.50
T61	7.83 (115.65)	174.29	62.44	69.69
L62	8.05 (124.22)	177.44	55.56	42.20
S63	8.16 (116.32)	–	58.59	63.55
S64	–	–	–	–
S65	–	174.12	58.24	63.92
V66	7.94 (121.30)	175.90	62.20	32.48
T67	7.98 (117.63)	–	61.47	–
V68	–	–	–	–
P69	–	176.63	63.05	32.01
S70	8.23 (115.89)	175.60	58.30	63.77
S71	8.23 (117.62)	174.22	58.38	63.72
T72	7.89 (115.25)	173.68	61.73	69.64
W73	8.04 (124.69)	174.17	54.70	28.86
P74	–	176.70	63.23	32.05
S75	8.35 (116.18)	174.54	58.46	63.81
E76	8.42 (122.49)	176.37	56.64	30.21
T77	8.15 (115.93)	174.25	62.15	69.65
V78	8.14 (123.45)	175.97	62.26	32.60
T79	8.14 (118.88)	174.03	61.83	69.64
C80	8.39 (121.39)	173.94	55.28	40.59
N81	8.51 (121.60)	174.62	53.19	38.73
V82	7.93 (120.50)	175.32	62.18	32.61
A83	8.20 (127.21)	176.86	52.09	19.09
H84	8.13 (120.31)	173.46	54.09	30.14
P85	–	176.73	63.30	31.96
A86	8.62 (124.48)	177.91	52.56	19.04

Table B.4: continued .

Residue	H^N (N)	C'	C^α	C^β
S87	8.20 (114.69)	177.89	58.34	63.67
S88	–	174.55	58.33	63.88
T89	8.07 (116.17)	174.18	62.02	69.56
K90	8.19 (124.29)	176.13	56.22	32.80
V91	8.07 (121.94)	175.47	62.06	32.71
D92	8.33 (124.99)	175.86	54.20	41.16
K93	8.16 (122.43)	176.12	56.16	32.81
K94	8.24 (122.56)	176.16	56.34	32.61
I95	8.08 (122.99)	175.86	60.75	38.20
V96	8.17 (127.19)	174.05	59.66	32.44
P97	–	175.74	63.31	31.98
R98	7.84 (126.44)	168.96	57.36	31.41

Table B.5: $^1J_{\mathrm{NH}}$-, $^1J_{\mathrm{NH}}$+$^1D_{\mathrm{NH}}$-couplings and $^1D_{\mathrm{NH}}$ restraints for the assembled C_H1 domain

Residue	$^1J_{\mathrm{NH}}$ [Hz]	$^1J_{\mathrm{NH}}$+$^1D_{\mathrm{NH}}$ [Hz]	$^1D_{\mathrm{NH}}$ restraint
A10	93.40	84.80	−8.60
V21	94.17	99.20	5.03
T22	93.23	92.98	−0.25
L23	93.63	90.15	−3.48
G24	95.17	80.42	−14.75
C25	91.79	73.99	−17.80
L26	92.62	83.84	−8.78
K28	93.67	88.54	−5.13
E33	91.74	104.78	13.04
V35	90.98	75.79	−15.19
T36	92.42	76.28	−16.14
V37	91.84	67.22	−24.62
T38	93.48	68.66	−24.82
W39	89.91	76.22	−13.69
S41	94.73	102.76	8.03
G42	94.27	98.83	4.56
S43	93.22	71.47	−21.75
S45	93.52	102.13	8.61
S46	94.47	104.98	10.51
V48	93.70	88.94	−4.76
H49	94.83	82.26	−12.57
F51	93.42	90.12	−3.30
A53	95.09	101.31	6.22
V54	91.32	86.65	−4.67
L55	92.18	86.74	−5.44
Q56	99.22	108.21	8.99
D58	92.71	75.07	−17.64
L59	94.58	99.29	4.71
L62	94.38	78.09	−16.29

Table B.5: continued .

Residue	$^1J_{\mathrm{NH}}$ [Hz]	$^1J_{\mathrm{NH}}+^1D_{\mathrm{NH}}$ [Hz]	$^1D_{\mathrm{NH}}$ restraint
S64	92.17	79.04	−13.13
S65	95.78	87.83	−7.95
T67	93.86	88.64	−5.22
V68	94.80	88.46	−6.34
S70	94.23	78.11	−16.12
S71	93.25	74.14	−19.11
T72	92.68	92.24	−0.44
W73	94.52	89.69	−4.83
T77	93.69	101.92	8.23
V78	93.40	82.67	−10.73
T79	93.11	86.94	−6.17
C80	93.57	81.92	−11.65
N81	95.22	71.30	−23.92
V82	91.51	73.55	−17.96
A83	92.57	79.71	−12.86
H84	92.44	88.60	−3.84
S87	92.35	84.93	−7.42
S88	95.28	79.64	−15.64
K90	94.27	70.60	−23.67
V91	93.00	84.06	−8.94
D92	93.12	68.96	−24.16
K93	93.76	76.25	−17.51
K94	93.40	70.21	−23.19
V96	94.82	104.22	9.40

Appendix C

The pH Sensitive USH-I/LWEQ Domain of Talin

C.1 Sample Preparation

The gene of the USH-I/LWEQ domain of talin was obtained by PCR amplification using the cDNA of the murine talin1 as a template. The PCR product was cloned into the pGEX6P2 expression vector (GE Healthcare, Piscataway, NJ, USA) via the BamHI and EcoRI restriction sites resulting in a GST fusion protein and transformed into the E. coli strain BL21(DE3). Cells were grown at 37 °C in selective LB medium till an OD_{600} of 0.5 and subsequently adapted to a temperature of 30 °C for approximately half an hour. When the OD_{600} reached 0.8 the expression was started by addition of 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG). After expression for 6 h at 30 °C, cells were harvested by centrifugation (4000 g). The cell pellet was resuspended in PBS buffer (10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 140 mM NaCl, 2.7 mM KCl, pH 7.2). Disruption of the cells was carried out by high-pressure treatment and DNA was removed by the addition of 10 μg/mL DNAseI, 3 mM $MgCl_2$ and incubation for 30 min at room temperature. Thereafter, 1% Triton X-100 was added and the mixture was incubated for further 60 min at 4 °C. Finally, removal of the insoluble components was achieved by centrifugation (40000 g). For the batch purification, 5 mL glutathione sepharose per L cell culture was washed three times with PBS buffer and subsequently added to the supernatant. After incubation for 2 h under slight shaking at room temperature and washing three times with protease buffer (44 mM K_2HPO_4, 5.8 mM KH_2PO_4, 1 mM EDTA, 1 mM DTT, pH 8.0), the glutathione sepharose beads were resuspended in 8 mL/L protease buffer and 50 μL/L prescission protease (GE Healthcare, Piscataway, NJ, USA) were added to remove the N-terminal GST-tag. The cleavage was allowed to proceed overnight at 4 °C. After an additional centrifugation step to spin down

the sepharose beads (500 g, 5 min, 4 °C), the concentrated supernatant contained the cleaved talin construct. This protocol results in a yield of around 20 mg/mL pure protein. Isotope labeled protein was expressed in M9 minimal medium that additionally included 1 g/L Isogro (Sigma-Aldrich, St. Louis, MO, USA) and 10 mL/L 100x MEM vitamins (Invitrogen, Carlsbad, CA, USA) containing either ^{15}N ammonium chloride as the only nitrogen source or additionally ^{13}C glucose as the only carbon source. The plasmid was sequenced and the protein mass was confirmed by MALDI-TOF MS.

For the NMR analysis the concentration of the USH-I/LWEQ domain of talin in the phosphate buffer (44 mM K_2HPO_4, 5.8 mM KH_2PO_4, 1 mM EDTA, 1 mM DTT) with 5% (v/v) D_2O at either pH 6.5 or pH 7.5 was around 1.0 mM. Spectra of this talin construct were recorded at 32 °C on Bruker DRX500, DMX750, and AVANCE800 spectrometers. Processing of all spectra was performed using the program Topspin 1.3 (Bruker BioSpin) and analysis of the spectra was carried out with the program Sparky (www.cgl.ucsf.edu/home/sparky). For the RDC measurement, the protein was partially aligned with non-ionic liquid crystalline media composed of *n*-alkyl-polyethylene glycol and *n*-alkyl alcohol. According to the paper by Otting and coworkers,[223] for the given salt concentration, pH value, and temperature range 3 wt% C12E5/hexanol (molar ratio r = 0.96) was chosen as alignment media. In the case of the talin construct, the sample volume of 400 μL resulted in 12.5 μL C12E5 and 2.5 μL hexanol to obtain a stable alignment.

C.2 Experiments and Assignment of the Talin USH-I/LWEQ Domain

Table C.1: Experiments on the talin USH-I/LWEQ domain

Experiment	Nuclei	NS	Time domain	Data matrix
NHSQC	$^{1}H\times^{15}N$	8	1024×256	1024×512
HNCO	$^{1}H\times^{15}N\times^{13}C$	8	$1024\times\ 64\times128$	$1024\times128\times256$
HNcaCO	$^{1}H\times^{15}N\times^{13}C$	16	$1024\times\ 64\times128$	$1024\times128\times256$
CBCAcoNH	$^{1}H\times^{15}N\times^{13}C$	24	$1024\times\ 64\times\ 96$	$1024\times128\times256$
HNCACB	$^{1}H\times^{15}N\times^{13}C$	24	$1024\times\ 64\times\ 90$	$1024\times128\times256$
HNHnoesy	$^{1}H\times^{15}N\times\ ^{1}H$	8	$1024\times128\times256$	$1024\times256\times512$
NNHnoesy	$^{1}H\times^{15}N\times^{15}N$	16	$1024\times112\times112$	$1024\times256\times256$
IPAP-HNCO	$^{1}H\times^{15}N\times^{13}C$	16	$1024\times128\times\ 92$	$1024\times256\times256$
IPAP-HNCO	$^{1}H\times^{15}N\times^{13}C$	16	$1024\times\ 64\times192$	$1024\times256\times256$
CPMG	$^{1}H\times^{15}N$	16	1024×256	1024×512
MEXICO	$^{1}H\times^{15}N$	8	1024×300	1024×512

Table C.2: Backbone assignments of the talin USH-I/LWEQ domain

Residue	H^N (N)	C'	C^α	C^β
D1	8.28 (123.71)	177.81	53.06	41.41
P2	–	180.56	64.60	32.21
T3	8.36 (112.53)	178.87	64.65	68.43
V4	7.47 (122.50)	180.30	64.52	32.14
I5	7.73 (119.21)	180.32	63.16	37.65
A6	–	–	–	–
E7	–	179.38	57.35	29.79
N8	8.36 (118.30)	178.43	53.60	38.46
E9	–	–	–	–
L10	–	–	–	–
L11	–	–	55.70	41.99
G12	8.28 (109.09)	179.62	46.81	–
A13	8.02 (126.49)	181.62	54.83	17.12
A14	8.04 (119.70)	182.28	55.47	17.54
A15	8.16 (120.57)	183.38	55.05	17.54
A16	8.02 (123.04)	–	55.21	17.73
I17	–	–	–	–
E18	–	–	–	–
A19	–	183.22	54.88	17.68
A20	7.74 (122.79)	182.63	54.92	17.56
A21	8.52 (120.27)	183.36	54.89	17.95
K22	8.06 (118.80)	181.98	58.78	31.89
K23	7.52 (118.54)	181.97	59.03	31.93
L24	7.59 (117.36)	181.44	56.61	41.66
E25	7.69 (117.25)	179.72	58.03	29.59
Q26	7.52 (114.98)	178.79	55.67	28.98
L27	7.35 (120.25)	178.84	55.22	42.40
K28	–	–	–	–
P29	–	179.12	62.74	32.11
R30	8.57 (121.80)	178.68	56.01	30.91
A31	8.37 (125.61)	–	52.30	19.05
K32	–	–	–	–
P33	–	179.59	63.08	32.03
K34	8.46 (121.32)	179.53	57.03	32.52
E35	8.49 (120.71)	179.05	56.32	29.73
A36	8.26 (123.67)	180.05	52.83	18.91
D37	8.18 (118.52)	179.15	54.23	40.83
E38	–	179.56	57.17	29.83
S39	8.34 (116.69)	177.23	59.28	63.45
L40	8.01 (122.83)	180.04	55.21	42.03
N41	–	–	–	–
F42	–	179.39	60.87	38.94
E43	8.84 (119.35)	181.01	61.12	28.41
E44	8.22 (118.05)	181.67	59.72	29.22
Q45	–	178.28	55.59	29.41
I46	8.13 (122.09)	178.34	60.94	38.65
L47	–	181.28	58.24	41.17

Table C.2: continued .

Residue	H^N (N)	C'	C^α	C^β
E48	8.36 (117.48)	182.19	59.27	29.48
A49	8.16 (123.59)	182.85	54.73	17.41
A50	8.83 (121.96)	181.62	55.46	17.87
K51	8.47 (117.84)	182.73	60.59	32.33
S52	8.14 (116.20)	180.24	62.69	–
I53	8.45 (123.32)	178.03	59.58	32.10
A54	–	–	–	–
A55	–	183.32	54.63	18.37
A56	7.99 (121.30)	183.09	54.42	18.44
T57	9.26 (113.09)	–	68.51	70.48
S58	–	179.16	62.52	–
A59	7.56 (121.16)	183.37	55.16	17.56
L60	8.34 (120.64)	–	58.59	32.50
V61	–	–	–	–
K62	–	–	–	–
A63	–	182.76	54.74	16.92
A64	8.61 (123.64)	181.39	55.05	17.70
S65	8.21 (113.77)	180.47	62.48	–
A66	7.85 (123.49)	182.85	56.56	–
A67	–	181.97	54.85	17.36
Q68	8.41 (118.73)	179.99	60.26	26.79
R69	8.11 (117.86)	181.92	59.87	30.06
E70	–	–	–	–
L71	–	–	–	–
V72	–	182.62	61.05	38.69
A73	8.39 (124.80)	180.36	55.31	18.25
Q74	8.07 (114.47)	179.61	56.21	29.26
G75	8.06 (107.95)	177.85	45.82	–
K76	8.23 (118.95)	179.17	56.95	33.48
V77	7.25 (115.42)	178.53	61.06	34.48
G78	8.79 (111.85)	176.30	45.77	–
A79	8.36 (122.48)	180.00	52.38	20.22
I80	7.89 (119.94)	178.43	58.34	39.07
P81	–	181.07	65.20	31.88
A82	8.36 (120.21)	180.74	53.94	18.48
N83	7.99 (115.55)	177.86	52.22	38.27
A84	7.82 (123.51)	182.53	55.46	18.59
L85	–	–	–	–
D86	–	181.08	56.51	41.12
D87	8.91 (121.71)	181.69	56.98	40.27
G88	8.60 (110.45)	178.71	47.03	–
Q89	–	181.78	58.67	28.25
W90	8.63 (122.70)	181.43	61.40	27.63
S91	8.51 (116.18)	–	62.40	–
Q92	–	181.42	58.66	28.03
G93	8.06 (108.29)	179.12	46.61	–
L94	7.51 (124.32)	–	58.13	–
I95	–	179.90	63.52	31.97

Table C.2: continued .

Residue	H^N (N)	C'	C^α	C^β
S96	8.38 (114.01)	178.52	62.85	70.44
A97	7.88 (122.41)	182.33	54.97	17.55
A98	8.59 (121.39)	182.01	55.75	18.07
R99	8.48 (116.91)	182.71	59.87	29.81
M100	–	–	–	–
V101	–	182.30	66.21	–
A102	8.31 (125.15)	181.95	55.75	17.69
A103	8.44 (122.45)	182.94	55.33	17.91
A104	8.63 (120.55)	183.02	54.77	17.88
T105	8.41 (116.11)	177.79	–	–
N106	–	179.82	57.90	38.63
N107	8.24 (118.63)	180.68	56.44	37.68
L108	–	181.16	–	–
C109	8.51 (118.75)	179.58	64.91	26.56
E110	8.27 (119.71)	182.03	59.44	–
A111	–	182.87	54.76	17.44
A112	8.84 (122.22)	180.77	55.13	17.84
N113	8.14 (116.71)	179.58	56.34	38.94
A114	7.82 (119.89)	182.55	55.01	17.89
A115	–	183.19	–	–
V116	8.14 (117.00)	179.36	65.06	31.12
Q117	7.64 (116.59)	178.60	56.00	28.98
G118	7.68 (106.16)	176.89	45.41	–
H119	–	176.58	53.57	28.72
A120	8.53 (123.40)	179.78	52.57	19.65
S121	8.40 (116.01)	177.80	56.15	63.54
Q122	9.09 (127.25)	180.03	60.20	28.11
E123	8.70 (117.72)	181.89	59.96	28.84
K124	–	181.60	56.30	30.30
L125	8.13 (121.50)	176.92	53.51	41.04
I126	–	180.72	66.13	37.81
S127	8.53 (114.28)	–	62.42	–
S128	–	179.30	62.56	–
A129	8.18 (126.82)	182.95	55.86	18.10
K130	–	176.23	59.48	31.55
Q131	8.29 (120.57)	178.80	58.22	30.62
V132	–	181.25	67.07	31.16
A133	7.96 (124.72)	182.40	55.97	17.24
A134	–	183.63	55.19	18.18
S135	8.47 (115.56)	183.60	62.32	–
T136	–	178.28	67.99	–
A137	7.78 (123.36)	182.84	55.50	17.54
Q138	7.68 (116.67)	181.12	57.91	27.91
L139	–	–	–	–
L140	–	181.97	59.32	41.51
V141	7.91 (118.51)	180.86	66.75	32.05
A142	7.92 (121.01)	183.38	54.84	17.45
C143	7.98 (114.59)	178.56	–	–

Table C.2: continued .

Residue	H^N (N)	C'	C^α	C^β
K144	–	179.64	58.53	32.63
V145	7.54 (116.60)	179.22	64.87	32.10
K146	–	178.01	55.21	35.00
A147	7.77 (121.34)	179.19	51.95	20.12
D148	8.53 (120.68)	180.10	53.85	41.34
Q149	8.66 (122.32)	178.19	57.53	28.90
D150	8.45 (118.54)	178.44	53.83	40.73
S151	7.60 (115.91)	178.14	57.87	64.57
E152	9.04 (125.39)	181.92	59.63	29.02
A153	8.92 (121.19)	183.56	55.24	17.48
M154	7.83 (117.28)	181.88	55.83	30.76
K155	–	–	–	–
R156	–	–	–	–
L157	–	180.76	58.39	41.87
Q158	8.55 (118.55)	181.19	58.74	28.48
A159	8.01 (121.81)	183.72	55.41	18.00
A160	8.12 (122.49)	183.18	55.07	17.71
G161	8.92 (107.57)	178.87	47.45	–
N162	–	180.29	55.76	37.40
A163	7.92 (123.62)	183.78	55.36	17.57
V164	8.22 (119.72)	180.13	67.54	40.86
K165	–	180.75	60.04	32.03
R166	8.04 (117.76)	181.16	59.16	30.14
A167	7.89 (120.53)	183.05	54.95	18.34
S168	8.87 (115.99)	183.05	62.60	–
D169	–	181.83	57.41	39.83
N170	8.04 (118.50)	180.75	55.81	37.51
L171	8.04 (123.08)	–	55.65	37.31
V172	–	180.25	67.63	31.52
K173	7.91 (117.56)	181.94	59.46	32.11
A174	–	182.37	54.83	17.88
A175	8.66 (119.25)	181.71	54.48	18.55
Q176	8.17 (116.19)	180.62	58.03	28.43
K177	7.72 (119.07)	179.90	57.59	32.08
A178	7.66 (121.91)	180.78	53.33	18.56
A179	7.85 (121.39)	180.62	53.20	18.63
A180	7.77 (120.93)	180.56	52.94	18.55
F181	7.86 (118.18)	178.74	58.06	39.38
E182	8.10 (121.24)	178.90	56.76	30.02
E183	8.28 (121.01)	179.08	54.52	40.84
Q184	8.23 (120.34)	179.01	56.09	29.30
E185	8.41 (121.24)	179.05	56.73	29.66
N186	8.32 (119.13)	177.87	53.44	38.99
E187	8.37 (121.23)	179.21	56.78	29.93
T188	8.22 (115.79)	177.07	62.39	69.55
V189	8.11 (123.39)	178.50	62.49	32.42
V190	8.18 (124.83)	178.63	62.40	32.44
V191	8.23 (125.15)	178.72	62.39	32.33

Table C.2: continued .

Residue	H^N (N)	C'	C^α	C^β
K192	8.37 (125.71)	179.19	56.39	32.88
E193	8.41 (121.76)	178.93	56.60	29.98
K194	–	178.96	56.11	32.90
M195	8.33 (121.88)	178.76	55.55	32.92
V196	8.21 (121.93)	179.25	62.38	32.62
G197	8.54 (113.05)	177.22	45.22	–
G198	8.24 (108.32)	176.63	45.11	–
I199	7.95 (119.80)	178.57	60.83	38.76
A200	8.37 (127.98)	180.03	52.37	18.93
Q201	8.29 (120.10)	178.28	55.62	29.33
I202	8.07 (121.95)	178.31	60.99	38.62

Bibliography

[1] Hoult, D.; Ginsberg, N. The quantum origins of the free induction decay signal and spin noise, *J. Magn. Reson.* **2001**, *148*, 182 – 199.

[2] Karplus, M.; McCammon, J. Dynamics of proteins: elements and function, *Annu. Rev. Biochem.* **1983**, *52*, 263 – 300.

[3] Dayie, K.; Wagner, G.; Lefèvre, J. Theory and practice of nuclear spin relaxation in proteins, *Annu. Rev. Phys. Chem.* **1996**, *47*, 243 – 282.

[4] Bloch, F.; Hansen, W.; Packard, M. The nuclear induction experiment, *Phys. Rev.* **1946**, *70*, 474 – 485.

[5] Bloch, F. Nuclear Induction, *Phys. Rev.* **1946**, *70*, 460 – 474.

[6] Bloembergen, N.; Purcell, E.; Pound, R. Relaxation effects in nuclear magnetic resonance absorption, *Phys. Rev.* **1948**, *73*, 679 – 712.

[7] Wangsness, R.; Bloch, F. The dynamical theory of nuclear induction, *Phys. Rev.* **1953**, *89*, 728 – 739.

[8] Redfield, A. On the theory of relaxation processes, *IBM J. Res. Dev.* **1957**, *1*, 19 – 32.

[9] Levitt, M. *Spin dynamics - Basics of Nuclear Magnetic Resonance*. John Wiley, Chichester, 1961.

[10] Allerhand, A.; Doddrell, D.; Komoroski, R. Natural abundance carbon-13 partially relaxed fourier transform nuclear magnetic resonance spectra of complex molecules, *J. Chem. Phys.* **1971**, *55*, 189 – 198.

[11] Millet, O.; Muhandiram, D.; Skrynnikov, N.; Kay, L. Deuterium spin probes of side-chain dynamics in proteins. 1. Measurement of five relaxation rates per deuteron in ^{13}C-labeled and fractionally ^{2}H-enriched proteins in solution, *J. Am. Chem. Soc.* **2002**, *124*, 6439 – 6448.

[12] Kay, L.; Torchia, D.; Bax, A. Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease, *Biochemistry* **1989**, *28*, 8972 – 8979.

[13] Lipari, G.; Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity, *J. Am. Chem. Soc.* **1982**, *104*, 4546 – 4559.

[14] Lipari, G.; Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results, *J. Am. Chem. Soc.* **1982**, *104*, 4559 – 4570.

[15] Clore, G.; Szabo, A.; Bax, A.; Kay, L.; Driscoll, P.; Gronenborn, A. Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins, *J. Am. Chem. Soc.* **1990**, *112*, 4989 – 4991.

[16] Abragam, A. *The Principles of Nuclear Magnetism.* Clarendon Press, Oxford, 1961.

[17] Palmer, A.; Rance, M.; Wright, P. Intramolecular motions of a zinc finger DNA-binding domain from Xfin characterized by proton-detected natural abundance carbon-13 heteronuclear NMR spectroscopy, *J. Am. Chem. Soc.* **1991**, *113*, 4371 – 4380.

[18] Mandel, A.; Akke, M.; Palmer, A. Backbone dynamics of escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme, *J. Mol. Biol.* **1995**, *246*, 144 – 163.

[19] Wang, C.; Palmer, A. Solution NMR methods for quantitative identification of chemical exchange in ^{15}N-labeled proteins, *Magn. Reson. Chem.* **2003**, *41*, 866 – 876.

[20] Palmer, A. NMR characterization of the dynamics of biomacromolecules, *Chem. Rev.* **2004**, *104*, 3623 – 3640.

[21] Montelione, G.; Wagner, G. 2D chemical exchange NMR spectroscopy by proton-detected heteronuclear correlation, *J. Am. Chem. Soc.* **1989**, *111*, 3096 – 3098.

[22] Ishima, R.; Torchia, D. Estimating the time scale of chemical exchange of proteins from measurements of transverse relaxation rates in solution, *J. Biomol. NMR* **1999**, *14*, 369 – 372.

[23] Millet, O.; Loria, J.; Kroenke, C.; Pons, M.; Palmer, A. The static magnetic field dependence of chemical exchange linebroadening defines the NMR chemical shift time scale, *J. Am. Chem. Soc.* **2000**, *122*, 2867 – 2877.

[24] Allerhand, A.; Gutowsky, H. Spin-echo studies of chemical exchange. II. Closed formulas for two sites, *J. Chem. Phys.* **1965**, *42*, 1587 – 1599.

[25] Allerhand, A.; Thiele, E. Analysis of Carr-Purcell spin-echo NMR experiments on multiple-spin systems. II. The effect of chemical exchange, *J. Chem. Phys.* **1966**, *45*, 902 – 916.

[26] Davis, D.; Perlman, M.; London, R. Direct measurements of the dissociation-rate constant for inhibitor-enzyme complexes via the $T_{1\rho}$ and T_2 (CPMG) methods, *J. Magn. Reson.* **1994**, *104*, 266 – 275.

[27] Kovrigin, E.; Kempf, J.; Grey, M.; Loria, J. Faithful estimation of dynamics parameters from CPMG relaxation dispersion measurements, *J. Magn. Reson.* **2006**, *180*, 93 – 104.

[28] Palmer, A.; Kroenke, C.; Loria, J. Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules, *Methods Enzymol.* **2001**, *339*, 204 – 238.

[29] Gemmecker, G.; Jahnke, W.; Kessler, H. Measurement of fast proton exchange rates in isotopically labeled compounds, *J. Am. Chem. Soc.* **1993**, *115*, 11620 – 11621.

[30] Gemmecker, G.; Olejniczak, E.; Fesik, S. An improved method for selectively observing protons attached to ^{12}C in the presence of ^{1}H–^{13}C spin pairs, *J. Magn. Reson.* **1992**, *96*, 199 – 204.

[31] Tjandra, N.; Bax, A. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium, *Science* **1997**, *278*, 1111 – 1114.

[32] Hansen, M.; Mueller, L.; Pardi, A. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions, *Nat. Struct. Biol.* **1998**, *5*, 1065 – 1074.

[33] Tycko, R.; Blanco, F.; Ishii, Y. Alignment of biopolymers in strained gels: A new way to create detectable dipole-dipole couplings in high-resolution biomolecular NMR, *J. Am. Chem. Soc.* **2000**, *122*, 9340 – 9341.

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

[35] Peti, W.; Meiler, J.; Brüschweiler, R.; Griesinger, C. Model-free analysis of protein backbone motion from residual dipolar couplings, *J. Am. Chem. Soc.* **2002**, *124*, 5822 – 5833.

[36] Tolman, J. A novel approach to the retrieval of structural and dynamic information from residual dipolar couplings using several oriented media in biomolecular NMR spectroscopy, *J. Am. Chem. Soc.* **2002**, *124*, 12020 – 12030.

[37] Briggman, K.; Tolman, J. De novo determination of bond orientations and order parameters from residual dipolar couplings with high accuracy, *J. Am. Chem. Soc.* **2003**, *125*, 10164 – 10165.

[38] Brockwell, D.; Radford, S. Intermediates: ubiquitous species on folding energy landscapes?, *Curr. Opin. Struct. Biol.* **2007**, *17*, 30 – 37.

[39] Neuweiler, H.; Doose, S.; Sauer, M. A microscopic view of miniprotein folding: enhanced folding efficiency through formation of an intermediate, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 16650 – 16655.

[40] Ehrlich, P.; Morgenroth, J. Wirkung und Entstehung der aktiven Stoffe im Serum nach der Seitenkettentheorie, *Handbuch der pathogenen Mikroorganismen* **1904**, *1*, 430 – 451.

[41] Fischer, E. Einfluss der Configuration auf die Wirkung der Enzyme, *Ber. Dtsch. Chem. Ges.* **1894**, *27*, 2985 – 2991.

[42] Murphy, K.; Travers, P.; Walport, M. *Janeway's Immunobiology.* Garland Science, New York, 2008.

[43] Edelman, G.; Benacerraf, B. On structural and functional relations between antibodies and proteins of the gamma-system, *Proc. Natl. Acad. Sci. USA* **1962**, *48*, 1035 – 1042.

[44] Huber, R.; Deisenhofer, J.; Colman, P.; Matsushima, M.; Palm, W. Crystallographic structure studies of an IgG molecule and an Fc fragment, *Nature* **1976**, *264*, 415 – 420.

[45] Deisenhofer, J. Crystallographic refinement and atomic models of a human Fc fragment and its complex with fragment B of protein A from Staphylococcus aureus at 2.9 and 2.8 Å resolution, *Biochemistry* **1981**, *20*, 2361 – 2370.

[46] Harris, L.; Larson, S.; Hasel, K.; Day, J.; Greenwood, A.; McPherson, A. The three-dimensional structure of an intact monoclonal antibody for canine lymphoma, *Nature* **1992**, *360*, 369 – 372.

[47] Davies, D.; Chacko, S. Antibody structure, *Acc. Chem. Res.* **1993**, *26*, 421 – 427.

[48] Harris, L.; Larson, S.; Hasel, K.; McPherson, A. Refined structure of an intact IgG2a monoclonal antibody, *Biochemistry* **1997**, *36*, 1581 – 1597.

[49] Harris, L.; Skaletsky, E.; McPherson, A. Crystallographic structure of an intact IgG1 monoclonal antibody, *J. Mol. Biol.* **1998**, *275*, 861 – 872.

[50] Shields, R.; Namenuk, A.; Hong, K.; Meng, Y.; Rae, J.; Briggs, J.; Xie, D.; Lai, J.; Stadlen, A.; Li, B.; Fox, J.; Presta, L. High resolution mapping of the binding site on human IgG1 for FcγRI, FcγRII, FcγRIII, and FcRn and design of IgG1 variants with improved binding to the FcγR, *J. Biol. Chem.* **2001**, *276*, 6591 – 6604.

[51] Sondermann, P.; Huber, R.; Oosthuizen, V.; Jacob, U. The 3.2 Å crystal structure of the human IgG1 Fc fragment-FcγRIII complex, *Nature* **2000**, *406*, 267 – 273.

[52] Saphire, E.; Stanfield, R.; Crispin, M.; Parren, P.; Rudd, P.; Dwek, R.; Burton, D.; Wilson, I. Contrasting IgG structures reveal extreme asymmetry and flexibility, *J. Mol. Biol.* **2002**, *319*, 9 – 18.

[53] Krapp, S.; Mimura, Y.; Jefferis, R.; Huber, R.; Sondermann, P. Structural analysis of human IgG-Fc glycoforms reveals a correlation between glycosylation and structural integrity, *J. Mol. Biol.* **2003**, *325*, 979 – 989.

[54] Jefferis, R.; Lund, J.; Mizutani, H.; Nakagawa, H.; Kawazoe, Y.; Arata, Y.; Takahashi, N. A comparative study of the N-linked oligosaccharide structures of human IgG subclass proteins, *Biochem. J.* **1990**, *268*, 529 – 537.

[55] Lund, J.; Takahashi, N.; Pound, J.; Goodall, M.; Jefferis, R. Multiple interactions of IgG with its core oligosaccharide can modulate recognition by complement and human Fcγ Receptor I and influence the synthesis of its oligosaccharide chains, *J. Immunol.* **1996**, *157*, 4963 – 4969.

[56] Wormald, M.; Rudd, P.; Harvey, D.; Chang, S.; Scragg, I.; Dwek, R. Variations in oligosaccharide-protein interactions in immunoglobulin G determine the site-specific glycosylation profiles and modulate the dynamic motion of the Fc oligosaccharides, *Biochemistry* **1997**, *36*, 1370 – 1380.

[57] Wright, A.; Morrison, S. Effect of glycosylation on antibody function: implications for genetic engineering, *Trends Biotechnol.* **1997**, *15*, 26 – 32.

[58] Walker, M.; Lund, J.; Thompson, K.; Jefferis, R. Aglycosylation of human IgG1 and IgG3 monoclonal antibodies can eliminate recognition by human cells expressing FcγRI and/or Fcγ RII receptors, *Biochem. J.* **1989**, *259*, 347 – 353.

[59] Tao, M.; Morrison, S. Studies of aglycosylated chimeric mouse-human IgG. Role of carbohydrate in the structure and effector functions mediated by the human IgG constant region, *J. Immunol.* **1989**, *143*, 2595 – 2601.

[60] Jefferis, R.; Lund, J.; Pound, J. IgG-Fc-mediated effector functions: molecular definition of interaction sites for effector ligands and the role of glycosylation, *Immunol. Rev.* **1998**, *163*, 59 – 76.

[61] Burton, D.; Woof, J. Human antibody effector function, *Advan. Immunol.* **1992**, *51*, 1 – 84.

[62] Umana, P.; Jean-Mairet, J.; Moudry, R.; Amstutz, H.; Bailey, J. Engineered glycoforms of an antineuroblastoma IgG1 with optimized antibody-dependent cellular cytotoxic activity, *Nat. Biotechnol.* **1999**, *17*, 176 – 180.

[63] Mimura, Y.; Sondermann, P.; Ghirlando, R.; Lund, J.; Young, S.; Goodall, M.; Jefferis, R. Role of oligosaccharide residues of IgG1-Fc in FcγRIIb binding, *J. Biol. Chem.* **2001**, *276*, 45539 – 45547.

[64] Padlan, E.; Segal, D.; Spande, T.; Davies, D.; Rudikoff, S.; Potter, M. Stucture at 4.5 Å resolution of a phosphorylcholine-binding Fab, *Nat. New Biol.* **1973**, *245*, 165 – 167.

[65] Poljak, R.; Amzel, L.; Chen, B.; Phizackerley, R.; Saul, F. The three-dimensional structure of the Fab fragment of a human myeloma immunoglobulin at 2.0 Å resolution, *Proc. Natl. Acad. Sci. USA* **1974**, *71*, 3440 – 3444.

[66] Padlan, E.; Cohen, G.; Davies, D. Antibody Fab assembly: the interface residues between CH_1 and C_L, *Mol. Immunol.* **1986**, *23*, 951 – 960.

[67] Chothia, C.; Novotny, J.; Bruccoleri, R.; Karplus, M. Domain association in immunoglobulin molecules: the packing of variable domains, *J. Mol. Biol.* **1985**, *186*, 651 – 663.

[68] Sheriff, S.; Silverton, E.; Padlan, E.; Cohen, G.; Smith-Gill, S.; Finzel, B.; Davies, D. Three-dimensional structure of an antibody-antigen complex, *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 8075 – 8079.

[69] Colman, P.; Laver, W.; Varghese, J.; Baker, A.; Tulloch, P.; Air, G.; Webster, R. Three-dimensional structure of a complex of antibody with influenza virus neuraminidase, *Nature* **1987**, *326*, 358 – 362.

[70] Padlan, E.; Silverton, E.; Sheriff, S.; Cohen, G.; Smith-Gill, S.; Davies, D. Structure of an antibody-antigen complex: crystal structure of the HyHEL-10 Fab-lysozyme complex, *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 5938 – 5942.

[71] Wu, T.; Kabat, E. An analysis of the sequences of the variable regions of bence jones proteins and myeloma light chains and their implications for antibody complementarity, *J. Exp. Med.* **1970**, *132*, 211 – 250.

[72] Jones, P.; Dear, P.; Foote, J.; Neuberger, M.; Winter, G. Replacing the complementarity-determining regions in a human antibody with those from a mouse, *Nature* **1986**, *321*, 522 – 525.

[73] Kabat, E.; Wu, T.; Bilofsky, H. Unusual distributions of amino acids in complementarity-determining (hypervariable) segments of heavy and light chains of immunoglobulins and their possible roles in specificity of antibody-combining sites, *J. Biol. Chem.* **1977**, *252*, 6609 – 6616.

[74] Padlan, E.; Davies, D. Variability of three-dimensional structure in immunoglobulins, *Proc. Natl. Acad. Sci. USA* **1975**, *72*, 819 – 823.

[75] Chothia, C.; Lesk, A. Canonical structures for the hypervariable regions of immunoglobulins, *J. Mol. Biol.* **1987**, *196*, 901 – 917.

[76] Chothia, C.; Lesk, A.; Levitt, M.; Amit, A.; Mariuzza, R.; Phillips, S.; Poljak, R. The predicted structure of immunoglobulin D1.3 and its comparison with the crystal structure, *Science* **1986**, *233*, 755 – 758.

[77] Chothia, C.; Lesk, A.; Tramontano, A.; Levitt, M.; Smith-Gill, S.; Air, G.; Sheriff, S.; Padlan, E.; Davies, D.; Tulip, W.; Colman, P.; Spinelli, S.; Alzari, P.; Poljak, R. Conformations of immunoglobulin hypervariable regions, *Nature* **1989**, *342*, 877 – 883.

[78] Poljak, R.; Amzel, L.; Avey, H.; Chen, B.; Phizackerley, R.; Saul, F. Three-dimensional structure of the Fab fragment of a human immunoglobulin at 2.8 Å resolution, *Proc. Natl. Acad. Sci. USA* **1973**, *70*, 3305 – 3310.

[79] Ptitsyn, O. Protein folding and protein evolution: common folding nucleus in different subfamilies of c-type cytochromes?, *J. Mol. Biol.* **1998**, *278*, 655 – 666.

[80] Bashford, D.; Chothia, C.; Lesk, A. Determinants of a protein fold : unique features of the globin amino acid sequences, *J. Mol. Biol.* **1987**, *196*, 199 – 216.

[81] Clarke, J.; Cota, E.; Fowler, S.; Hamill, S. Folding studies of immunoglobulin-like β-sandwich proteins suggest that they share a common folding pathway, *Structure* **1999**, *7*, 1145 – 1153.

[82] Halaby, D.; Mornon, J. The immunoglobulin superfamily: an insight on its tissular, species, and functional diversity, *J. Mol. Evol.* **1998**, *46*, 389 – 400.

[83] Orengo, C.; Jones, D.; Thornton, J. Protein superfamilies and domain superfolds, *Nature* **1994**, *372*, 631 – 634.

[84] Williams, A.; Barclay, A. The immunoglobulin superfamily - domains for cell surface recognition, *Annu. Rev. Immunol.* **1988**, *6*, 381 – 405.

[85] Chothia, C.; Jones, E. The molecular structure of cell adhesion molecules, *Annu. Rev. Biochem.* **1997**, *66*, 823 – 862.

[86] Harpaz, Y.; Chothia, C. Many of the immunoglobulin superfamily domains in cell adhesion molecules and surface receptors belong to a new structural set which is close to that containing variable domains, *J. Mol. Biol.* **1994**, *238*, 528 – 539.

[87] Halaby, D.; Poupon, A.; Mornon, J. The immunoglobulin fold family: sequence analysis and 3D structure comparisons, *Protein Eng.* **1999**, *12*, 563 – 571.

[88] Richardson, J. β-sheet topology and the relatedness of proteins, *Nature* **1977**, *268*, 495 – 501.

[89] Bork, P.; Holm, L.; Sander, C. The immunoglobulin fold: structural classification, sequence patterns and common core, *J. Mol. Biol.* **1994**, *242*, 309 – 320.

[90] Chothia, C.; Lesk, A. The relation between the divergence of sequence and structure in proteins, *EMBO J.* **1986**, *5*, 823 – 826.

[91] Sander, C.; Schneider, R. Database of homology-derived protein structures and the structural meaning of sequence alignment, *Proteins* **1991**, *9*, 56 – 68.

[92] Nakashima, H.; Nishikawa, K.; Ooi, T. The folding type of a protein is relevant to the amino acid composition, *J. Biochem.* **1986**, *99*, 153 – 162.

[93] Dubchak, I.; Holbrook, S.; Kim, S. Prediction of protein folding class from amino acid composition, *Proteins* **1993**, *16*, 79 – 91.

[94] Poupon, A.; Mornon, J. Populations of hydrophobic amino acids within protein globular domains: identification of conserved "topohydrophobic" positions, *Proteins* **1998**, *33*, 329 – 342.

[95] Clamp, M.; Cuff, J.; Searle, S.; Barton, G. The Jalview java alignment editor, *Bioinformatics* **2004**, *20*, 426 – 427.

[96] Rost, B. Protein structures sustain evolutionary drift, *Fold. Des.* **1997**, *2*, S19 – S24.

[97] Hendershot, L. Immunoglobulin heavy chain and binding protein complexes are dissociated in vivo by light chain addition, *J. Cell Biol.* **1990**, *111*, 829 – 837.

[98] Gardner, A.; Aviel, S.; Argon, Y. Rapid degradation of an unassembled immunoglobulin light chain is mediated by a serine protease and occurs in a pre-Golgi compartment, *J. Biol. Chem.* **1993**, *268*, 25940 – 25947.

[99] Gallo, G.; Goni, F.; Boctor, F.; Vidal, R.; Kumar, A.; Stevens, F.; Frangione, B.; Ghiso, J. Light chain cardiomyopathy. Structural analysis of the light chain tissue deposits, *Am. J. Pathol.* **1996**, *148*, 1397 – 1406.

[100] Stevens, F.; Argon, Y. Pathogenic light chains and the B-cell repertoire, *Immunol. Today* **1999**, *20*, 451 – 457.

[101] Kyle, R. Monoclonal proteins and renal disease, *Annu. Rev. Med.* **1994**, *45*, 71 – 77.

[102] Bellotti, V.; Mangione, P.; Merlini, G. Immunoglobulin light chain amyloidosis - the archetype of structural and pathogenic variability, *J. Struct. Biol.* **2000**, *130*, 280 – 289.

[103] Eakin, C.; Miranker, A. From chance to frequent encounters: origins of β_2-microglobulin fibrillogenesis, *Biochim. Biophys. Acta Proteins Proteomics* **2005**, *1753*, 92 – 99.

[104] Fandrich, M. On the structural definition of amyloid fibrils and other polypeptide aggregates, *Cell. Mol. Life Sci.* **2007**, *64*, 2066 – 2078.

[105] Dobson, C. Protein folding and misfolding, *Nature* **2003**, *426*, 884 – 890.

[106] Kelly, J. The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways, *Curr. Opin. Struct. Biol.* **1998**, *8*, 101 – 106.

[107] Chiti, F.; Dobson, C. Protein misfolding, functional amyloid, and human disease, *Annu. Rev. Biochem.* **2006**, *75*, 333 – 366.

[108] Fandrich, M.; Fletcher, M.; Dobson, C. Amyloid fibrils from muscle myoglobin - even an ordinary globular protein can assume a rogue guise if conditions are right, *Nature* **2001**, *410*, 165 – 166.

[109] van Rhee, F.; Bolejack, V.; Hollmig, K.; Pineda-Roman, M.; Anaissie, E.; Epstein, J.; Shaughnessy, J.; Zangari, M.; Tricot, G.; Mohiuddin, A.; Alsayed, Y.; Woods, G.; Crowley, J.; Barlogie, B. High serum-free light chain levels and their rapid reduction in response to therapy define an aggressive multiple myeloma subtype with poor prognosis, *Blood* **2007**, *110*, 827 – 832.

[110] Augustine, J.; de la Calle, A.; Knarr, G.; Buchner, J.; Frederick, C. The crystal structure of the Fab fragment of the monoclonal antibody MAK33 - implications for folding and interaction with the chaperone BiP, *J. Biol. Chem.* **2001**, *276*, 3287 – 3294.

[111] Feige, M.; Hagn, F.; Esser, J.; Kessler, H.; Buchner, J. Influence of the internal disulfide bridge on the folding pathway of the C_L antibody domain, *J. Mol. Biol.* **2007**, *365*, 1232 – 1244.

[112] McParland, V.; Kad, N.; Kalverda, A.; Brown, A.; Kirwin-Jones, P.; Hunter, M.; Sunde, M.; Radford, S. Partially unfolded states of β_2-microglobulin and amyloid formation in vitro, *Biochemistry* **2000**, *39*, 8735 – 8746.

[113] Jahn, T.; Parker, M.; Homans, S.; Radford, S. Amyloid formation under physiological conditions proceeds via a native-like folding intermediate, *Nat. Struct. Mol. Biol.* **2006**, *13*, 195 – 201.

[114] Kameda, A.; Hoshino, M.; Higurashi, T.; Takahashi, S.; Naiki, H.; Goto, Y. Nuclear magnetic resonance characterization of the refolding intermediate of β_2-microglobulin trapped by non-native prolyl peptide bond, *J. Mol. Biol.* **2005**, *348*, 383 – 397.

[115] Eakin, C.; Berman, A.; Miranker, A. A native to amyloidogenic transition regulated by a backbone trigger, *Nat. Struct. Mol. Biol.* **2006**, *13*, 202 – 208.

[116] McParland, V.; Kalverda, A.; Homans, S.; Radford, S. Structural properties of an amyloid precursor of β_2-microglobulin, *Nat. Struct. Mol. Biol.* **2002**, *9*, 326 – 331.

[117] Dill, K.; Chan, H. From Levinthal to pathways to funnels, *Nat. Struct. Biol.* **1997**, *4*, 10 – 19.

[118] Sanchez, I.; Kiefhaber, T. Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding, *J. Mol. Biol.* **2003**, *325*, 367 – 376.

[119] Bai, Y. Energy barriers, cooperativity, and hidden intermediates in the folding of small proteins, *Biochem. Biophys. Res. Commun.* **2006**, *340*, 976 – 983.

[120] Jahn, T.; Radford, S. Folding versus aggregation: polypeptide conformations on competing pathways, *Arch. Biochem. Biophys.* **2008**, *469*, 100 – 117.

[121] Mayor, U.; Guydosh, N.; Johnson, C.; Grossmann, J.; Sato, S.; Jas, G.; Freund, S.; Alonso, D.; Daggett, V.; Fersht, A. The complete folding pathway of a protein from nanoseconds to microseconds, *Nature* **2003**, *421*, 863 – 867.

[122] Feng, H.; Zhou, Z.; Bai, Y. A protein folding pathway with multiple folding intermediates at atomic resolution, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 5026 – 5031.

[123] Religa, T.; Markson, J.; Mayor, U.; Freund, S.; Fersht, A. Solution structure of a protein denatured state and folding intermediate, *Nature* **2005**, *437*, 1053 – 1056.

[124] Nishimura, C.; Dyson, H.; Wright, P. Identification of native and non-native structure in kinetic folding intermediates of apomyoglobin, *J. Mol. Biol.* **2006**, *355*, 139 – 156.

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

[126] Wüthrich, K. *NMR of proteins and nucleic acids.* John Wiley, New York, 1986.

[127] Karplus, M. Contact electron-spin coupling of nuclear magnetic moments, *J. Chem. Phys.* **1959**, *30*, 11 – 15.

[128] Pardi, A.; Billeter, M.; Wüthrich, K. Calibration of the angular dependence of the amide proton-C^{α} proton coupling constants, ${}^{3}J_{HN\alpha}$, in a globular protein : use of ${}^{3}J_{HN\alpha}$ for identification of helical secondary structure, *J. Mol. Biol.* **1984**, *180*, 741 – 751.

[129] Wang, A.; Bax, A. Reparametrization of the Karplus relation for ${}^{3}J_{H^{\alpha}H^{N}}$ and ${}^{3}J_{H^{N}C'}$ in peptides from uniformly ${}^{13}C/{}^{15}N$-enriched human ubiquitin, *J. Am. Chem. Soc.* **1995**, *117*, 1810 – 1813.

[130] Wishart, D.; Sykes, B.; Richards, F. The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy, *Biochemistry* **1992**, *31*, 1647 – 1651.

[131] Wishart, D.; Sykes, B. The ${}^{13}C$ chemical shift index: a simple method for the identification of protein secondary structure using ${}^{13}C$ chemical shift data, *J. Biomol. NMR* **1994**, *4*, 171 – 180.

[132] Wedemeyer, W.; Welker, E.; Scheraga, H. Proline *cis-trans* isomerization and protein folding, *Biochemistry* **2002**, *41*, 14637 – 14644.

[133] Weiss, M.; Jabs, A.; Hilgenfeld, R. Peptide bonds revisited, *Nat. Struct. Biol.* **1998**, *5*, 676.

[134] Goto, Y.; Hamaguchi, K. Unfolding and refolding of the reduced constant fragment of the immunoglobulin light chain: kinetic role of the intrachain disulfide bond, *J. Mol. Biol.* **1982**, *156*, 911 – 926.

[135] Pappenberger, G.; Aygün, H.; Engels, J.; Reimer, U.; Fischer, G.; Kiefhaber, T. Nonprolyl *cis* peptide bonds in unfolded proteins cause complex folding kinetics, *Nat. Struct. Biol.* **2001**, *8*, 452 – 458.

[136] Balbach, J.; Forge, V.; van Nuland, N.; Winder, S.; Hore, P.; Dobson, C. Following protein folding in real time using NMR spectroscopy, *Nat. Struct. Biol.* **1995**, *2*, 865 – 870.

[137] Mizuguchi, M.; Kroon, G.; Wright, P.; Dyson, H. Folding of a β-sheet protein monitored by real-time NMR spectroscopy, *J. Mol. Biol.* **2003**, *328*, 1161 – 1171.

[138] Cornilescu, G.; Delaglio, F.; Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology, *J. Biomol. NMR* **1999**, *13*, 289 – 302.

[139] Diercks, T.; Coles, M.; Kessler, H. An efficient strategy for assignment of cross-peaks in 3D heteronuclear NOESY experiments, *J. Biomol. NMR* **1999**, *15*, 177 – 180.

[140] Brooks, B.; Bruccoleri, R.; Olafson, B.; States, D.; Swaminathan, S.; Karplus, M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations, *J. Comput. Chem.* **1983**, *4*, 187 – 217.

[141] Qin, Z.; Hu, D.; Zhu, M.; Fink, A. Structural characterization of the partially folded intermediates of an immunoglobulin light chain leading to amyloid fibrillation and amorphous aggregation, *Biochemistry* **2007**, *46*, 3521 – 3531.

[142] Fernandez-Escamilla, A.; Rousseau, F.; Schymkowitz, J.; Serrano, L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins, *Nat. Biotechnol.* **2004**, *22*, 1302 – 1306.

[143] Smith, D.; Jones, S.; Serpell, L.; Sunde, M.; Radford, S. A systematic investigation into the effect of protein destabilisation on β_2-microglobulin amyloid formation, *J. Mol. Biol.* **2003**, *330*, 943 – 954.

[144] Freund, C.; Honegger, A.; Hunziker, P.; Holak, T.; Plückthun, A. Folding nuclei of the scFv fragment of an antibody, *Biochemistry* **1996**, *35*, 8457 – 8464.

[145] Hamill, S.; Steward, A.; Clarke, J. The folding of an immunoglobulin-like greek key protein is defined by a common core nucleus and regions constrained by topology, *J. Mol. Biol.* **2000**, *297*, 165 – 178.

[146] Richardson, J.; Richardson, D. Natural β-sheet proteins use negative design to avoid edge-to-edge aggregation, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 2754 – 2759.

[147] Monsellier, E.; Chiti, F. Prevention of amyloid-like aggregation as a driving force of protein evolution, *EMBO Rep.* **2007**, *8*, 737 – 742.

[148] Hoshino, M.; Katou, H.; Hagihara, Y.; Hasegawa, K.; Naiki, H.; Goto, Y. Mapping the core of the β_2-microglobulin amyloid fibril by H/D exchange, *Nat. Struct. Biol.* **2002**, *9*, 332 – 336.

[149] Chiti, F.; Stefani, M.; Taddei, N.; Ramponi, G.; Dobson, C. Rationalization of the effects of mutations on peptide and protein aggregation rates, *Nature* **2003**, *424*, 805 – 808.

[150] Han, J.; Batey, S.; Nickson, A.; Teichmann, S.; Clarke, J. The folding and evolution of multidomain proteins, *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 319 – 330.

[151] Guo, H.; Jardetzky, T.; Garrett, T.; Lane, W.; Strominger, J.; Wiley, D. Different length peptides bind to HLA-Aw68 similarly at their ends but bulge out in the middle, *Nature* **1992**, *360*, 364 – 366.

[152] Ellgaard, L.; Helenius, A. Quality control in the endoplasmic reticulum, *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 181 – 191.

[153] Kaloff, C.; Haas, I. Coordination of immunoglobulin chain folding and immunoglobulin chain assembly is essential for the formation of functional IgG, *Immunity* **1995**, *2*, 629 – 637.

[154] Lee, Y.; Brewer, J.; Hellman, R.; Hendershot, L. BiP and immunoglobulin light chain cooperate to control the folding of heavy chain and ensure the fidelity of immunoglobulin assembly, *Mol. Biol. Cell* **1999**, *10*, 2209 – 2219.

[155] Haas, I.; Wabl, M. Immunoglobulin heavy chain binding protein, *Nature* **1983**, *306*, 387 – 389.

[156] Bole, D.; Hendershot, L.; Kearney, J. Posttranslational association of immunoglobulin heavy chain binding protein with nascent heavy chains in nonsecreting and secreting hybridomas, *J. Cell Biol.* **1986**, *102*, 1558 – 1566.

[157] Hendershot, L.; Bole, D.; Kohler, G.; Kearney, J. Assembly and secretion of heavy chains that do not associate posttranslationally with immunoglobulin heavy chain-binding protein, *J. Cell Biol.* **1987**, *104*, 761 – 767.

[158] Wahner-Roedler, D.; Kyle, R. Heavy chain diseases, *Best Pract. Res. Clin. Haematol.* **2005**, *18*, 729 – 746.

[159] Hamers-Casterman, C.; Atarhouch, T.; Muyldermans, S.; Robinson, G.; Hamers, C.; Songa, E.; Bendahman, N.; Hamers, R. Naturally occurring antibodies devoid of light chains, *Nature* **1993**, *363*, 446 – 448.

[160] Thies, M.; Mayer, J.; Augustine, J.; Frederick, C.; Lilie, H.; Buchner, J. Folding and association of the antibody domain C_H3: prolyl isomerization precedes dimerization, *J. Mol. Biol.* **1999**, *293*, 67 – 79.

[161] Feige, M.; Walter, S.; Buchner, J. Folding mechanism of the C_H2 antibody domain, *J. Mol. Biol.* **2004**, *344*, 107 – 118.

[162] Hurtley, S.; Helenius, A. Protein oligomerization in the endoplasmic reticulum, *Annu. Rev. Cell Biol.* **1989**, *5*, 277 – 307.

[163] Ellgaard, L.; Molinari, M.; Helenius, A. Setting the standards: quality control in the secretory pathway, *Science* **1999**, *286*, 1882 – 1888.

[164] Wickner, S.; Maurizi, M.; Gottesman, S. Posttranslational quality control: folding, refolding, and degrading proteins, *Science* **1999**, *286*, 1888 – 1893.

[165] Fra, A.; Fagioli, C.; Finazzi, D.; Sitia, R.; Alberini, C. Quality control of ER synthesized proteins: an exposed thiol group as a three-way switch mediating assembly, retention and degradation, *EMBO J.* **1993**, *12*, 4755 – 4761.

[166] Helenius, A.; Trombetta, E.; Hebert, D.; Simons, J. Calnexin, calreticulin and the folding of glycoproteins, *Trends Cell Biol.* **1997**, *7*, 193 – 200.

[167] Hellman, R.; Vanhove, M.; Lejeune, A.; Stevens, F.; Hendershot, L. The in vivo association of BiP with newly synthesized proteins is dependent on the rate and stability of folding and not simply on the presence of sequences that can bind to BiP, *J. Cell Biol.* **1999**, *144*, 21 – 30.

[168] Herrmann, J.; Malkus, P.; Schekman, R. Out of the ER - outfitters, escorts and guides, *Trends Cell Biol.* **1999**, *9*, 5 – 7.

[169] Klausner, R.; Sitia, R. Protein degradation in the endoplasmic reticulum, *Cell* **1990**, *62*, 611 – 614.

[170] Brodsky, J.; McCracken, A. ER protein quality control and proteasome-mediated protein degradation, *Semin. Cell Dev. Biol.* **1999**, *10*, 507 – 513.

[171] Tsai, B.; Ye, Y.; Rapoport, T. Retro-translocation of proteins from the endoplasmic reticulum into the cytosol, *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 246 – 255.

[172] Jarosch, E.; Geiss-Friedlander, R.; Meusser, B.; Walter, J.; Sommer, T. Protein dislocation from the endoplasmic reticulum - pulling out the suspect, *Traffic* **2002**, *3*, 530 – 536.

[173] Turner, G.; Varshavsky, A. Detecting and measuring cotranslational protein degradation in vivo, *Science* **2000**, *289*, 2117 – 2120.

[174] Schubert, U.; Antón, L.; Gibbs, J.; Norbury, C.; Yewdell, J.; Bennink, J. Rapid degradation of a large fraction of newly synthesized proteins by proteasomes, *Nature* **2000**, *404*, 770 – 774.

[175] Yewdell, J.; Schubert, U.; Bennink, J. At the crossroads of cell biology and immunology: DRiPs and other sources of peptide ligands for MHC class I molecules, *J. Cell Sci.* **2001**, *114*, 845 – 851.

[176] Reits, E.; Vos, J.; Grommé, M.; Neefjes, J. The major substrates for TAP *in vivo* are derived from newly synthesized proteins, *Nature* **2000**, *404*, 774 – 778.

[177] Dobson, C. Protein-misfolding diseases: Getting out of shape, *Nature* **2002**, *418*, 729 – 730.

[178] Jiang, J.; Prasad, K.; Lafer, E.; Sousa, R. Structural basis of interdomain communication in the Hsc70 chaperone, *Mol. Cell* **2005**, *20*, 513 – 524.

[179] Zhu, X.; Zhao, X.; Burkholder, W.; Gragerov, A.; Ogata, C.; Gottesman, M.; Hendrickson, W. Structural analysis of substrate binding by the molecular chaperone DnaK, *Science* **1996**, *272*, 1606 – 1614.

[180] Kopito, R. Biosynthesis and degradation of CFTR, *Physiol. Rev.* **1999**, *79*, S167 – S173.

[181] Petäjä-Repo, U.; Hogue, M.; Laperrière, A.; Walker, P.; Bouvier, M. Export from the endoplasmic reticulum represents the limiting step in the maturation and cell surface expression of the human δ opioid receptor, *J. Biol. Chem.* **2000**, *275*, 13727 – 13736.

[182] Merlie, J.; Lindstrom, J. Assembly in vivo of mouse muscle acetylcholine receptor: identification of an α subunit species that may be an assembly intermediate, *Cell* **1983**, *34*, 747 – 757.

[183] Flynn, G.; Chappell, T.; Rothman, J. Peptide binding and release by proteins implicated as catalysts of protein assembly, *Science* **1989**, *245*, 385 – 390.

[184] Flynn, G.; Pohl, J.; Flocco, M.; Rothman, J. Peptide-binding specificity of the molecular chaperone BiP, *Nature* **1991**, *353*, 726 – 730.

[185] Blond-Elguindi, S.; Cwirla, S.; Dower, W.; Lipshutz, R.; Sprang, S.; Sambrook, J.; Gething, M. Affinity panning of a library of peptides displayed on bacteriophages reveals the binding specificity of BiP, *Cell* **1993**, *75*, 717 – 728.

[186] Knarr, G.; Kies, U.; Bell, S.; Mayer, M.; Buchner, J. Interaction of the chaperone BiP with an antibody domain: implications for the chaperone cycle, *J. Mol. Biol.* **2002**, *318*, 611 – 620.

[187] Hendershot, L.; Wei, J.; Gaut, J.; Melnick, J.; Aviel, S.; Argon, Y. Inhibition of immunoglobulin folding and secretion by dominant negative BiP ATPase mutants, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 5269 – 5274.

[188] Wei, J.; Hendershot, L. Characterization of the nucleotide binding properties and ATPase activity of recombinant hamster BiP purified from bacteria, *J. Biol. Chem.* **1995**, *270*, 26670 – 26676.

[189] Köhler, G.; Howe, S.; Milstein, C. Fusion between immunoglobulin-secreting and nonsecreting myeloma cell lines, *Eur. J. Immunol.* **1976**, *6*, 292 – 295.

[190] Anfinsen, C.; Haber, E.; Sela, M.; White, F. J. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain, *Proc. Natl. Acad. Sci. USA* **1961**, *47*, 1309 – 1314.

[191] Anfinsen, C. Principles that govern the folding of protein chains, *Science* **1973**, *181*, 223 – 230.

[192] Levinthal, C. Are there pathways for protein folding?, *J. Chim. Phys.* **1968**, *65*, 44 – 45.

[193] Zwanzig, R.; Szabo, A.; Bagchi, B. Levinthal's paradox, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 20 – 22.

[194] Onuchic, J.; Wolynes, P.; Luthey-Schulten, Z.; Socci, N. Toward an outline of the topography of a realistic protein-folding funnel, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 3626 – 3630.

[195] Wolynes, P.; Onuchic, J.; Thirumalai, D. Navigating the folding routes, *Science* **1995**, *267*, 1619 –1620.

[196] Bryngelson, J.; Onuchic, J.; Socci, N.; Wolynes, P. Funnels, pathways, and the energy landscape of protein folding: a synthesis, *Proteins: Struct. Funct. Genet.* **1995**, *21*, 167 – 195.

[197] Dinner, A.; Sali, A.; Smith, L.; Dobson, C.; Karplus, M. Understanding protein folding via free-energy surfaces from theory and experiment, *Trends Biochem. Sci.* **2000**, *25*, 331 – 339.

[198] Dobson, C.; Sali, A.; Karplus, M. Protein Folding: a perspective from theory and experiment, *Angew. Chem. Int. Ed.* **1998**, *37*, 868 –893.

[199] Fersht, A. Transition-state structure as a unifying basis in protein-folding mechanisms: contact order, chain topology, stability, and the extended nucleus mechanism, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 1525 – 1529.

[200] Onuchic, J.; Luthey-Schulten, Z.; Wolynes, P. Theory of protein folding: the energy landscape perspective, *Annu. Rev. Phys. Chem.* **1997**, *48*, 545 – 600.

[201] Gulukota, K.; Wolynes, P. Statistical mechanics of kinetic proofreading in protein folding in vivo, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 9292 – 9296.

[202] Ferreiro, D.; Hegler, J.; Komives, E.; Wolynes, P. Localizing frustration in native proteins and protein assemblies, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 19819 – 19824.

[203] Nymeyer, H.; García, A.; Onuchic, J. Folding funnels and frustration in off-lattice minimalist protein landscapes, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 5921 – 5928.

[204] Leopold, P.; Montal, M.; Onuchic, J. Protein folding funnels: a kinetic approach to the sequence-structure relationship, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 8721 – 8725.

[205] Dunker, A.; Lawson, J.; Brown, C.; Williams, R.; Romero, P.; Oh, J.; Oldfield, C.; Campen, A.; Ratliff, C.; Hipps, K.; Ausio, J.; Nissen, M.; Reeves, R.; Kang, C.; Kissinger, C.; Bailey, R.; Griswold, M.; Chiu, W.; Garner, E.; Obradovic, Z. Intrinsically disordered protein, *J. Mol. Graph. Model.* **2001**, *19*, 26 – 59.

[206] Uversky, V. Natively unfolded proteins: a point where biology waits for physics, *Protein Sci.* **2002**, *11*, 739 – 756.

[207] Wright, P.; Dyson, H. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm, *J. Mol. Biol.* **1999**, *293*, 321 – 331.

[208] Koshland, D. J. Application of a theory of enzyme specificity to protein synthesis, *Proc. Natl. Acad. Sci. USA* **1958**, *44*, 98 – 104.

[209] Dyson, H.; Wright, P. Coupling of folding and binding for unstructured proteins, *Curr. Opin. Struct. Biol.* **2002**, *12*, 54 – 60.

[210] Tompa, P. Intrinsically unstructured proteins, *Trends Biochem. Sci.* **2002**, *27*, 527 – 533.

[211] Papoian, G.; Wolynes, P. The physics and bioinformatics of binding and folding - an energy landscape perspective, *Biopolymers* **2003**, *68*, 333 – 349.

[212] Spolar, R.; Record, M. Coupling of local folding to site-specific binding of proteins to DNA, *Science* **1994**, *263*, 777 – 784.

[213] Kriwacki, R.; Hengst, L.; Tennant, L.; Reed, S.; Wright, P. Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 11504 – 11509.

[214] Gunasekaran, K.; Tsai, C.; Kumar, S.; Zanuy, D.; Nussinov, R. Extended disordered proteins: targeting function with less scaffold, *Trends Biochem. Sci.* **2003**, *28*, 81 – 85.

[215] Shoemaker, B.; Portman, J.; Wolynes, P. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 8868 – 8873.

[216] Goto, Y.; Hamaguchi, K. Unfolding and refolding of the constant fragment of the immunoglobulin light chain, *J. Mol. Biol.* **1982**, *156*, 891 – 910.

[217] Röthlisberger, D.; Honegger, A.; Plückthun, A. Domain interactions in the Fab fragment: a comparative evaluation of the single-chain Fv and Fab format engineered with variable domains of different stability, *J. Mol. Biol.* **2005**, *347*, 773 – 789.

[218] Yamazaki, T.; Lee, W.; Arrowsmith, C.; Muhandiram, D.; Kay, L. A suite of triple resonance NMR experiments for the backbone assignment of ^{15}N, ^{13}C, ^{2}H labeled proteins with high sensitivity, *J. Am. Chem. Soc.* **1994**, *116*, 11655 – 11666.

[219] Pachter, R.; Arrowsmith, C.; Jardetzky, O. The effect of selective deuteration on magnetization transfer in larger proteins, *J. Biomol. NMR* **1992**, *2*, 183 – 194.

[220] Mok, Y.; Kay, C.; Kay, L.; Forman-Kay, J. NOE data demonstrating a compact unfolded state for an SH3 domain under non-denaturing conditions, *J. Mol. Biol.* **1999**, *289*, 619 – 638.

[221] Jung, Y.; Zweckstetter, M. Mars - robust automatic backbone assignment of proteins, *J. Biomol. NMR* **2004**, *30*, 11 – 23.

[222] Jung, Y.; Zweckstetter, M. Backbone assignment of proteins with known structure using residual dipolar couplings, *J. Biomol. NMR* **2004**, *30*, 25 – 35.

[223] Rückert, M.; Otting, G. Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments, *J. Am. Chem. Soc.* **2000**, *122*, 7793 – 7797.

[224] Meunier, L.; Usherwood, Y.; Chung, K.; Hendershot, L. A subset of chaperones and folding enzymes form multiprotein complexes in endoplasmic reticulum to bind nascent proteins, *Mol. Biol. Cell* **2002**, *13*, 4456 – 4469.

[225] Vanhove, M.; Usherwood, Y.; Hendershot, L. Unassembled Ig heavy chains do not cycle from BiP in vivo but require light chains to trigger their release, *Immunity* **2001**, *15*, 105 – 114.

[226] Elkabetz, Y.; Argon, Y.; Bar-Nun, S. Cysteines in C_H1 underlie retention of unassembled Ig heavy chains, *J. Biol. Chem.* **2005**, *280*, 14402 – 14412.

[227] Knarr, G.; Gething, M.; Modrow, S.; Buchner, J. BiP binding sequences in antibodies, *J. Biol. Chem.* **1995**, *270*, 27589 – 27594.

[228] Burrows, P.; Lejeune, M.; Kearney, J. Evidence that murine pre-B cells synthesise μ heavy chains but no light chains, *Nature* **1979**, *280*, 838 – 841.

[229] Pillai, S.; Baltimore, D. Formation of disulphide-linked $\mu_2\omega_2$ tetramers in pre-B cells by the 18K ω-immunoglobulin light chain, *Nature* **1987**, *329*, 172 – 174.

[230] Cenci, S.; Sitia, R. Managing and exploiting stress in the antibody factory, *FEBS Lett.* **2007**, *581*, 3652 – 3657.

[231] Adetugbo, K. Spontaneous somatic mutations. Structural studies on mutant immunoglobulins, *J. Biol. Chem.* **1978**, *253*, 6076 – 6080.

[232] Wolfenstein-Todel, C.; Mihaesco, E.; Frangione, B. 'Alpha chain disease' protein Def: internal deletion of a human immunoglobulin A_1 heavy chain, *Proc. Natl. Acad. Sci. USA* **1974**, *71*, 974 – 978.

[233] Hatahet, F.; Ruddock, L. Substrate recognition by the protein disulfide isomerases, *FEBS J.* **2007**, *274*, 5223 – 5234.

[234] Lu, P.; Finn, G.; Lee, T.; Nicholson, L. Prolyl cis-trans isomerization as a molecular timer, *Nat. Chem. Biol.* **2007**, *3*, 619 – 629.

[235] Uversky, V.; Oldfield, C.; Dunker, A. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling, *J. Mol. Recognit.* **2005**, *18*, 343 – 384.

[236] Ward, J.; Sodhi, J.; McGuffin, L.; Buxton, B.; Jones, D. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life, *J. Mol. Biol.* **2004**, *337*, 635 – 645.

[237] Yang, Z.; Thomson, R.; McNeil, P.; Esnouf, R. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins, *Bioinformatics* **2005**, *21*, 3369 – 3376.

[238] Greenberg, A.; Avila, D.; Hughes, M.; Hughes, A.; McKinney, E.; Flajnik, M. A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks, *Nature* **1995**, *374*, 168 – 173.

[239] Okada, T.; Cyster, J. B cell migration and interactions in the early phase of antibody responses, *Curr. Opin. Immunol.* **2006**, *18*, 278 – 285.

[240] Cyster, J. Chemokines, sphingosine-1-phosphate, and cell migration in secondary lymphoid organs, *Annu. Rev. Immunol.* **2005**, *23*, 127 – 159.

[241] Okada, T.; Miller, M.; Parker, I.; Krummel, M.; Neighbors, M.; Hartley, S.; O'Garra, A.; Cahalan, M.; Cyster, J. Antigen-engaged B cells undergo chemotaxis toward the T zone and form motile conjugates with helper T cells, *PLoS Biol.* **2005**, *3*, 1047 – 1061.

[242] Critchley, D. Cytoskeletal proteins talin and vinculin in integrin-mediated adhesion, *Biochem. Soc. Trans.* **2004**, *32*, 831 – 836.

[243] Schmidt, J.; Zhang, J.; Lee, H.; Stromer, M.; Robson, R. Interaction of talin with actin: sensitive modulation of filament crosslinking activity, *Arch. Biochem. Biophys.* **1999**, *366*, 139 – 150.

[244] Goldmann, W.; Hess, D.; Isenberg, G. The effect of intact talin and talin tail fragment on actin filament dynamics and structure depends on pH and ionic strength, *Eur. J. Biochem.* **1999**, *260*, 439 – 445.

[245] Denker, S.; Barber, D. Cell migration requires both ion translocation and cytoskeletal anchoring by the Na-H exchanger NHE1, *J. Cell Biol.* **2002**, *159*, 1087 – 1096.

[246] Patel, H.; Barber, D. A developmentally regulated Na-H exchanger in Dictyostelium discoideum is necessary for cell polarity during chemotaxis, *J. Cell Biol.* **2005**, *169*, 321 – 329.

[247] Harguindey, S.; Orive, G.; Luis Pedraz, J.; Paradiso, A.; Reshkin, S. The role of pH dynamics and the Na^+/H^+ antiporter in the etiopathogenesis and treatment of cancer. Two faces of the same coin - one single nature, *Biochim. Biophys. Acta* **2005**, *1756*, 1 – 24.

[248] Hemmings, L.; Rees, D.; Ohanian, V.; Bolton, S.; Gilmore, A.; Patel, B.; Priddle, H.; Trevithick, J.; Hynes, R.; Critchley, D. Talin contains three actin-binding sites each of which is adjacent to a vinculin-binding site, *J. Cell Sci.* **1996**, *109*, 2715 – 2726.

[249] Ridley, A.; Schwartz, M.; Burridge, K.; Firtel, R.; Ginsberg, M.; Borisy, G.; Parsons, J.; Horwitz, A. Cell migration: integrating signals from front to back, *Science* **2003**, *302*, 1704 – 1709.

[250] Lauffenburger, D.; Horwitz, A. Cell migration: a physically integrated molecular process, *Cell* **1996**, *84*, 359 – 369.

[251] Devreotes, P.; Janetopoulos, C. Eukaryotic chemotaxis: distinctions between directional sensing and polarization, *J. Biol. Chem.* **2003**, *278*, 20445 – 20448.

[252] Merlot, S.; Firtel, R. Leading the way: directional sensing through phosphatidylinositol 3-kinase and other signaling pathways, *J. Cell Sci.* **2003**, *116*, 3471 – 3478.

[253] Schwartz, M.; Shattil, S. Signaling networks linking integrins and Rho family GTPases, *Trends Biochem. Sci.* **2000**, *25*, 388 – 391.

[254] Kiosses, W.; Shattil, S.; Pampori, N.; Schwartz, M. Rac recruits high-affinity integrin $\alpha_v\beta_3$ to lamellipodia in endothelial cell migration, *Nat. Cell Biol.* **2001**, *3*, 316 – 320.

[255] Rodriguez, O.; Schaefer, A.; Mandato, C.; Forscher, P.; Bement, W.; Waterman-Storer, C. Conserved microtubule-actin interactions in cell movement and morphogenesis, *Nat. Cell Biol.* **2003**, *5*, 599 – 609.

[256] Evers, E.; Zondag, G.; Malliri, A.; Price, L.; ten Klooster, J.; van der Kammen, R.; Collard, J. Rho family proteins in cell adhesion and cell migration, *Eur. J. Cancer* **2000**, *36*, 1269 – 1274.

[257] Worthylake, R.; Burridge, K. RhoA and ROCK promote migration by limiting membrane protrusions, *J. Biol. Chem.* **2003**, *278*, 13578 – 13584.

[258] Xu, J.; Wang, F.; Van Keymeulen, A.; Herzmark, P.; Straight, A.; Kelly, K.; Takuwa, Y.; Sugimoto, N.; Mitchison, T.; Bourne, H. Divergent signals and cytoskeletal assemblies regulate self-organizing polarity in neutrophils, *Cell* **2003**, *114*, 201 – 214.

[259] Pollard, T.; Borisy, G. Cellular motility driven by assembly and disassembly of actin filaments, *Cell* **2003**, *112*, 453 – 465.

[260] Dos Remedios, C.; Chhabra, D.; Kekic, M.; Dedova, I.; Tsubakihara, M.; Berry, D.; Nosworthy, N. Actin binding proteins: regulation of cytoskeletal microfilaments, *Physiol. Rev.* **2003**, *83*, 433 – 473.

[261] Kim, M.; Carman, C.; Springer, T. Bidirectional transmembrane signaling by cytoplasmic domain separation in integrins, *Science* **2003**, *301*, 1720 – 1725.

[262] Tadokoro, S.; Shattil, S.; Eto, K.; Tai, V.; Liddington, R.; de Pereda, J.; Ginsberg, M.; Calderwood, D. Talin binding to integrin β tails: a final common step in integrin activation, *Science* **2003**, *302*, 103 – 106.

[263] Beningo, K.; Dembo, M.; Kaverina, I.; Small, J.; Wang, Y. Nascent focal adhesions are responsible for the generation of strong propulsive forces in migrating fibroblasts, *J. Cell Biol.* **2001**, *153*, 881 – 888.

[264] Galbraith, C.; Yamada, K.; Sheetz, M. The relationship between force and focal complex development, *J. Cell Biol.* **2002**, *159*, 695 – 705.

[265] Lee, J.; Ishihara, A.; Oxford, G.; Johnson, B.; Jacobson, K. Regulation of cell movement is mediated by stretch-activated calcium channels, *Nature* **1999**, *400*, 382 – 386.

[266] Hendey, B.; Klee, C.; Maxfield, F. Inhibition of neutrophil chemokinesis on vitronectin by inhibitors of calcineurin, *Science* **1992**, *258*, 296 – 299.

[267] Glading, A.; Lauffenburger, D.; Wells, A. Cutting to the chase: calpain proteases in cell motility, *Trends Cell Biol.* **2002**, *12*, 46 –54.

[268] Vinogradova, O.; Velyvis, A.; Velyviene, A.; Hu, B.; Haas, T.; Plow, E.; Qin, J. A structural mechanism of integrin $\alpha_{\mathrm{IIb}}\beta_3$ 'inside-out' activation as regulated by its cytoplasmic face, *Cell* **2002**, *110*, 587 – 597.

[269] García-Alvarez, B.; de Pereda, J.; Calderwood, D.; Ulmer, T.; Critchley, D.; Campbell, I.; Ginsberg, M.; Liddington, R. Structural determinants of integrin recognition by talin, *Mol. Cell* **2003**, *11*, 49 – 58.

[270] Yamada, K.; Geiger, B. Molecular interactions in cell adhesion complexes, *Curr. Opin. Cell Biol.* **1997**, *9*, 76 – 85.

[271] Zamir, E.; Geiger, B. Molecular complexity and dynamics of cell-matrix adhesions, *J. Cell Sci.* **2001**, *114*, 3583 – 3590.

[272] Webb, D.; Parsons, J.; Horwitz, A. Adhesion assembly, disassembly and turnover in migrating cells - over and over and over again, *Nat. Cell Biol.* **2002**, *4*, E97 – E100.

[273] Zaidel-Bar, R.; Ballestrem, C.; Kam, Z.; Geiger, B. Early molecular events in the assembly of matrix adhesions at the leading edge of migrating cells, *J. Cell Sci.* **2003**, *116*, 4605 – 4613.

[274] Zaidel-Bar, R.; Cohen, M.; Addadi, L.; Geiger, B. Hierarchical assembly of cell-matrix adhesion complexes, *Biochem. Soc. Trans.* **2004**, *32*, 416 – 420.

[275] Zamir, E.; Katz, M.; Posen, Y.; Erez, N.; Yamada, K.; Katz, B.; Lin, S.; Lin, D.; Bershadsky, A.; Kam, Z.; Geiger, B. Dynamics and segregation of cell-matrix adhesions in cultured fibroblasts, *Nat. Cell Biol.* **2000**, *2*, 191 – 196.

[276] Horwitz, A.; Duggan, K.; Buck, C.; Beckerle, M.; Burridge, K. Interaction of plasma membrane fibronectin receptor with talin - a transmembrane linkage, *Nature* **1986**, *320*, 531 – 533.

[277] Critchley, D. Focal adhesions - the cytoskeletal connection, *Curr. Opin. Cell Biol.* **2000**, *12*, 133 – 139.

[278] Xing, B.; Jedsadayanmata, A.; Lam, S. Localization of an integrin binding site to the C terminus of talin, *J. Biol. Chem.* **2001**, *276*, 44373 – 44378.

[279] Goldmann, W.; Bremer, A.; Häner, M.; Aebi, U.; Isenberg, G. Native talin is a dumbbell-shaped homodimer when it interacts with actin, *J. Struct. Biol.* **1994**, *112*, 3 – 10.

[280] Calderwood, D.; Zent, R.; Grant, R.; Rees, D.; Hynes, R.; Ginsberg, M. The talin head domain binds to integrin β subunit cytoplasmic tails and regulates integrin activation, *J. Biol. Chem.* **1999**, *274*, 28071 – 28074.

[281] Calderwood, D.; Yan, B.; de Pereda, J.; García Alvarez, B.; Fujioka, Y.; Liddington, R.; Ginsberg, M. The phosphotyrosine binding-like domain of talin activates integrins, *J. Biol. Chem.* **2002**, *277*, 21749 – 21758.

[282] Chen, H.; Appeddu, P.; Parsons, J.; Hildebrand, J.; Schaller, M.; Guan, J. Interaction of focal adhesion kinase with cytoskeletal protein talin, *J. Biol.Chem.* **1995**, *270*, 16995 – 16999.

[283] Martel, V.; Racaud-Sultan, C.; Dupe, S.; Marie, C.; Paulhe, F.; Galmiche, A.; Block, M.; Albiges-Rizo, C. Conformation, localization, and integrin binding of talin depend on its interaction with phosphoinositides, *J. Biol. Chem.* **2001**, *276*, 21217 – 21227.

[284] Bass, M.; Patel, B.; Barsukov, I.; Fillingham, I.; Mason, R.; Smith, B.; Bagshaw, C.; Critchley, D. Further characterization of the interaction between the cytoskeletal proteins talin and vinculin, *Biochem. J.* **2002**, *362*, 761 – 768.

[285] Gottschalk, K.; Adams, P.; Brunger, A.; Kessler, H. Transmembrane signal transduction of the $\alpha_{\mathrm{IIb}}\beta_3$ integrin, *Protein Sci.* **2002**, *11*, 1800 – 1812.

[286] Gottschalk, K.; Kessler, H. Evidence for hetero-association of transmembrane helices of integrins, *FEBS Lett.* **2004**, *557*, 253 – 258.

[287] Gottschalk, K.; Kessler, H. A computational model of transmembrane integrin clustering, *Structure* **2004**, *12*, 1109 – 1116.

[288] Brown, N.; Gregory, S.; Rickoll, W.; Fessler, L.; Prout, M.; White, R.; Fristrom, J. Talin is essential for integrin function in Drosophila, *Dev. Cell* **2002**, *3*, 569 – 579.

[289] Perrin, B.; Huttenlocher, A. Calpain, *Int. J. Biochem. Cell Biol.* **2002**, *34*, 722 – 725.

[290] Guroff, G. A neutral, calcium-activated proteinase from the soluble fraction of rat brain, *J. Biol. Chem.* **1964**, *239*, 149 – 155.

[291] Franco, S.; Huttenlocher, A. Regulating cell migration: calpains make the cut, *J. Cell Sci.* **2005**, *118*, 3829 – 3838.

[292] Saido, T.; Shibata, M.; Takenawa, T.; Murofushi, H.; Suzuki, K. Positive regulation of μ-calpain action by polyphosphoinositides, *J. Biol. Chem.* **1992**, *267*, 24585 – 24590.

[293] Salamino, F.; De Tullio, R.; Mengotti, P.; Viotti, P.; Melloni, E.; Pontremoli, S. Site-directed activation of calpain is promoted by a membrane-associated natural activator protein, *Biochem. J.* **1993**, *290*, 191 – 197.

[294] Friedrich, P. The intriguing Ca^{2+} requirement of calpain activation, *Biochem. Biophys. Res. Commun.* **2004**, *323*, 1131 – 1133.

[295] Glading, A.; Bodnar, R.; Reynolds, I.; Shiraha, H.; Satish, L.; Potter, D.; Blair, H.; Wells, A. Epidermal growth factor activates m-calpain (calpain II), at least in part, by extracellular signal-regulated kinase-mediated phosphorylation, *Mol. Cell Biol.* **2004**, *24*, 2499 – 2512.

[296] Franco, S.; Rodgers, M.; Perrin, B.; Han, J.; Bennin, D.; Critchley, D.; Huttenlocher, A. Calpain-mediated proteolysis of talin regulates adhesion dynamics, *Nat. Cell Biol.* **2004**, *6*, 977 – 983.

[297] Jalali, S.; del Pozo, M.; Chen, K.; Miao, H.; Li, Y.; Schwartz, M.; Shyy, J.; Chien, S. Integrin-mediated mechanotransduction requires its dynamic interaction with specific extracellular matrix (ECM) ligands, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 1042 – 1046.

[298] Mitchison, T.; Cramer, L. Actin-based cell motility and cell locomotion, *Cell* **1996**, *84*, 371 – 379.

[299] Diez, S.; Gerisch, G.; Anderson, K.; Müller-Taubenberger, A.; Bretschneider, T. Subsecond reorganization of the actin network in cell motility and chemotaxis, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 7601 – 7606.

[300] Schwab, A. Function and spatial distribution of ion channels and transporters in cell migration, *Am. J. Physiol. Renal Physiol.* **2001**, *280*, F739 – F747.

[301] Denker, S.; Barber, D. Ion transport proteins anchor and regulate the cytoskeleton, *Curr. Opin. Cell Biol.* **2002**, *14*, 214 – 220.

[302] Bretscher, M.; Aguado-Velasco, C. EGF induces recycling membrane to form ruffles, *Curr. Biol.* **1998**, *8*, 721 – 728.

[303] Bretscher, M.; Aguado-Velasco, C. Membrane traffic during cell locomotion, *Curr. Opin. Cell Biol.* **1998**, *10*, 537 – 541.

[304] Stock, C.; Schwab, A. Role of the Na^+/H^+ exchanger NHE1 in cell migration, *Acta Physiol.* **2006**, *187*, 149 – 157.

[305] Pouysségur, J.; Sardet, C.; Franchi, A.; L'Allemain, G.; Paris, S. A specific mutation abolishing Na^+/H^+ antiport activity in hamster fibroblasts precludes growth at neutral and acidic pH, *Proc. Natl. Acad. Sci. USA* **1984**, *81*, 4833 – 4837.

[306] Schwartz, M.; Lechene, C.; Ingber, D. Insoluble fibronectin activates the Na/H antiporter by clustering and immobilizing integrin $\alpha_5\beta_1$, independent of cell shape, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 7849 – 7853.

[307] Tominaga, T.; Barber, D. Na-H exchange acts downstream of RhoA to regulate integrin-induced cell adhesion and spreading, *Mol. Biol. Cell* **1998**, *9*, 2287 – 2303.

[308] Bernstein, B.; Painter, W.; Chen, H.; Minamide, L.; Abe, H.; Bamburg, J. Intracellular pH modulation of ADF/cofilin proteins, *Cell Motil. Cytoskeleton* **2000**, *47*, 319 – 336.

[309] Bowman, G.; Nodelman, I.; Hong, Y.; Chua, N.; Lindberg, U.; Schutt, C. A comparative structural analysis of the ADF/cofilin family, *Proteins* **2000**, *41*, 374 – 384.

[310] Pope, B.; Zierler-Gould, K.; Kühne, R.; Weeds, A.; Ball, L. Solution structure of human cofilin: actin binding, pH sensitivity, and relationship to actin-depolymerizing factor, *J. Biol. Chem.* **2004**, *279*, 4840 – 4848.

[311] Denker, S.; Huang, D.; Orlowski, J.; Furthmayr, H.; Barber, D. Direct binding of the Na-H exchanger NHE1 to ERM proteins regulates the cortical cytoskeleton and cell shape independently of H^+ translocation, *Mol. Cell* **2000**, *6*, 1425 – 1436.

[312] Webb, D.; Donais, K.; Whitmore, L.; Thomas, S.; Turner, C.; Parsons, J.; Horwitz, A. FAK-Src signalling through paxillin, ERK and MLCK regulates adhesion disassembly, *Nat. Cell Biol.* **2004**, *6*, 154 – 161.

[313] Nayal, A.; Webb, D.; Horwitz, A. Talin: an emerging focal point of adhesion dynamics, *Curr. Opin. Cell Biol.* **2004**, *16*, 94 – 98.

[314] Brown, C.; Hebert, B.; Kolin, D.; Zareno, J.; Whitmore, L.; Horwitz, A.; Wiseman, P. Probing the integrin-actin linkage using high-resolution protein velocity mapping, *J. Cell Sci.* **2006**, *119*, 5204 – 5214.

[315] Hu, K.; Ji, L.; Applegate, K.; Danuser, G.; Waterman-Storer, C. Differential transmission of actin motion within focal adhesions, *Science* **2007**, *315*, 111 – 115.

[316] Stock, C.; Gassner, B.; Hauck, C.; Arnold, H.; Mally, S.; Eble, J.; Dieterich, P.; Schwab, A. Migration of human melanoma cells depends on extracellular pH and Na^+/H^+ exchange, *J. Physiol.* **2005**, *567*, 225 – 238.

[317] Gingras, A.; Bate, N.; Goult, B.; Hazelwood, L.; Canestrelli, I.; Grossmann, J.; Liu, H.; Putz, N.; Roberts, G.; Volkmann, N.; Hanein, D.; Barsukov, I.; Critchley, D. The structure of the C-terminal actin-binding domain of talin, *EMBO J.* **2008**, *27*, 458 – 469.

[318] Brett, T.; Legendre-Guillemin, V.; McPherson, P.; Fremont, D. Structural definition of the F-actin-binding THATCH domain from HIP1R, *Nat. Struct. Mol. Biol.* **2006**, *13*, 121 – 130.

[319] Tanokura, M. ^{1}H-NMR study on the tautomerism of the imidazole ring of histidine residues: I. Microscopic pK values and molar ratios of tautomers in histidine-containing peptides, *Biochim. Biophys. Acta* **1983**, *742*, 576 – 585.

[320] Smith, S.; McCann, R. A C-terminal dimerization motif is required for focal adhesion targeting of talin1 and the interaction of the talin1 I/LWEQ module with F-actin, *Biochemistry* **2007**, *46*, 10886 – 10898.

[321] Tousignant, A.; Pelletier, J. Protein motions promote catalysis, *Chem. Biol.* **2004**, *11*, 1037 – 1042.

[322] Rousseau, F.; Schymkowitz, J. A systems biology perspective on protein structural dynamics and signal transduction, *Curr. Opin. Struct. Biol.* **2005**, *15*, 23 – 30.

[323] Kern, D.; Zuiderweg, E. The role of dynamics in allosteric regulation, *Curr. Opin. Struct. Biol.* **2003**, *13*, 748 – 757.

[324] Prestegard, J.; Al-Hashimi, H.; Tolman, J. NMR structures of biomolecules using field oriented media and residual dipolar couplings, *Q. Rev. Biophys.* **2000**, *33*, 371 – 424.

[325] Bouvignies, G.; Bernadó, P.; Meier, S.; Cho, K.; Grzesiek, S.; Brüschweiler, R.; Blackledge, M. Identification of slow correlated motions in proteins using residual dipolar and hydrogen-bond scalar couplings, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 13885 – 13890.

[326] Zweckstetter, M.; Bax, A. Prediction of sterically induced alignment in a dilute liquid crystalline phase: aid to protein structure determination by NMR, *J. Am. Chem. Soc.* **2000**, *122*, 3791 – 3792.

[327] Das, R.; Abu-Abed, M.; Melacini, G. Mapping allostery through equilibrium perturbation NMR spectroscopy, *J. Am. Chem. Soc.* **2006**, *128*, 8406 – 8407.

[328] Frantz, C.; Karydis, A.; Nalbant, P.; Hahn, K.; Barber, D. Positive feedback between Cdc42 activity and H^+ efflux by the Na-H exchanger NHE1 for polarity of migrating cells, *J. Cell Biol.* **2007**, *179*, 403 – 410.

[329] Jiang, G.; Giannone, G.; Critchley, D.; Fukumoto, E.; Sheetz, M. Two-piconewton slip bond between fibronectin and the cytoskeleton depends on talin, *Nature* **2003**, *424*, 334 – 337.

[330] Franco, S.; Senetar, M.; Simonson, W.; Huttenlocher, A.; McCann, R. The conserved C-terminal I/LWEQ module targets talin1 to focal adhesions, *Cell Motil. Cytoskeleton* **2006**, *63*, 563 – 581.

[331] Schwartz, M.; Ingber, D.; Lawrence, M.; Springer, T.; Lechene, C. Multiple integrins share the ability to induce elevation of intracellular pH, *Exp. Cell Res.* **1991**, *195*, 533 – 535.

[332] Yan, W.; Nehrke, K.; Choi, J.; Barber, D. The Nck-interacting kinase (NIK) phosphorylates the Na^+-H^+ exchanger NHE1 and regulates NHE1 activation by platelet-derived growth factor, *J. Biol. Chem.* **2001**, *276*, 31349 – 31356.

[333] Putney, L.; Barber, D. Na-H exchange-dependent increase in intracellular pH times G_2/M entry and transition, *J. Biol. Chem.* **2003**, *278*, 44645 – 44649.

[334] Wang, H.; Singh, D.; Fliegel, L. The Na^+/H^+ antiporter potentiates growth and retinoic acid-induced differentiation of P19 embryonal carcinoma cells, *J. Biol. Chem.* **1997**, *272*, 26545 – 26549.

[335] Ritter, M.; Schratzberger, P.; Rossmann, H.; Wöll, E.; Seiler, K.; Seidler, U.; Reinisch, N.; Kähler, C.; Zwierzina, H.; Lang, H.; Lang, F.; Paulmichl, M.; Wiedermann, C. Effect of inhibitors of Na^+/H^+-exchange and gastric H^+/K^+ ATPase on cell volume, intracellular pH and migration of human polymorphonuclear leucocytes, *Br. J. Pharmacol.* **1998**, *124*, 627 – 638.

[336] Reshkin, S.; Bellizzi, A.; Caldeira, S.; Albarani, V.; Malanchi, I.; Poignee, M.; Alunni-Fabbroni, M.; Casavola, V.; Tommasino, M. Na^+/H^+ exchanger-dependent intracellular alkalinization is an early event in malignant transformation and plays an essential role in the development of subsequent transformation-associated phenotypes, *FASEB J.* **2000**, *14*, 2185 – 2197.

[337] Boron, W. Regulation of intracellular pH, *Advan. Physiol. Edu.* **2004**, *28*, 160 – 179.

[338] King, K.; Essig, J.; Roberts, T.; Moerland, T. Regulation of the Ascaris major sperm protein (MSP) cytoskeleton by intracellular pH, *Cell Motil. Cytoskeleton* **1994**, *27*, 193 – 205.

[339] Senetar, M.; Foster, S.; McCann, R. Intrasteric inhibition mediates the interaction of the I/LWEQ module proteins talin1, talin2, Hip1, and Hip12 with actin, *Biochemistry* **2004**, *43*, 15418 – 15428.

[340] Grey, M.; Tang, Y.; Alexov, E.; McKnight, C.; Raleigh, D.; Palmer, A. Characterizing a partially folded intermediate of the villin headpiece domain under non-denaturing conditions: contribution of His41 to the pH-dependent stability of the N-terminal subdomain, *J. Mol. Biol.* **2006**, *355*, 1078 – 1094.

[341] DiGiammarino, E.; Lee, A.; Cadwell, C.; Zhang, W.; Bothner, B.; Ribeiro, R.; Zambetti, G.; Kriwacki, R. A novel mechanism of tumorigenesis involving pH-dependent destabilization of a mutant p53 tetramer, *Nat. Struct. Biol.* **2002**, *9*, 12 – 16.

[342] Orive, G.; Reshkin, S.; Harguindey, S.; Pedraz, J. Hydrogen ion dynamics and the Na^+/H^+ exchanger in cancer angiogenesis and antiangiogenesis, *Br. J. Cancer* **2003**, *89*, 1395 – 1399.

[343] Putney, L.; Denker, S.; Barber, D. The changing face of the Na^+/H^+ exchanger, NHE1: structure, regulation, and cellular actions, *Annu. Rev. Pharmacol. Toxicol.* **2002**, *42*, 527 – 552.

[344] Iwasaki, T.; Nakata, A.; Mukai, M.; Shinkai, K.; Yano, H.; Sabe, H.; Schaefer, E.; Tatsuta, M.; Tsujimura, T.; Terada, N.; Kakishita, E.; Akedo, H. Involvement of phosphorylation of Tyr-31 and Tyr-118 of paxillin in MM1 cancer cell migration, *Int. J. Cancer* **2002**, *97*, 330 – 335.

[345] Thies, M.; Pirkl, F. Chromatographic purification of the C_H2 domain of the monoclonal antibody MAK33, *J. Chromatog. sect. B* **2000**, *737*, 63 – 69.