

Technische Universität München

Department Chemie

Lehrstuhl für Biotechnologie

FOLDING AND ASSEMBLY OF ANTIBODIES

Matthias Johannes Feige

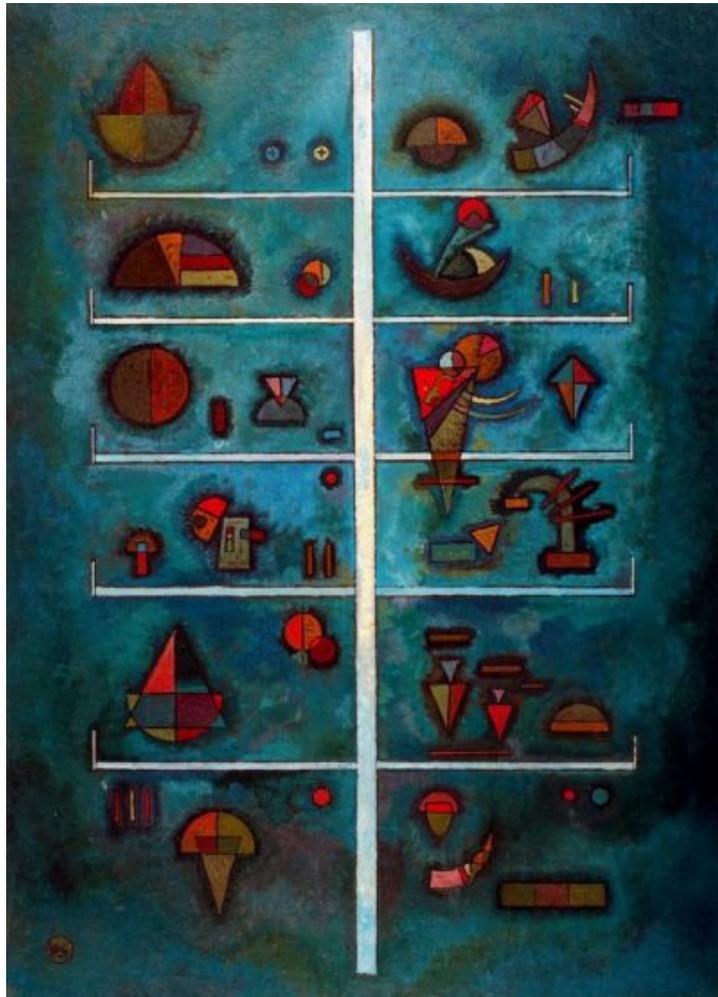
Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. M. Groll

Prüfer der Dissertation:

1. Univ.-Prof. Dr. J. Buchner
2. Univ.-Prof. Dr. T. Kiefhaber
3. Univ.-Prof. Dr. S. Weinkauf

Die Dissertation wurde am 28.09.2009 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 26.11.2009 angenommen.



Etagen, W. Kandinsky, 1929

MIT DEM BEOBACHTEN ERSCHAFFEN WIR.
(MJF)

Table of contents

PART A – INTRODUCTION AND SUMMARY

| | |
|--|-----------|
| I. A SHORT INTRODUCTION INTO PROTEIN FOLDING..... | 1 |
| I-1 GENERAL CONSIDERATIONS | 1 |
| I-2 A BRIEF HISTORY OF PROTEIN FOLDING | 1 |
| I-3 THE UNFOLDED STATE | 2 |
| I-3 MOVEMENTS ON THE FREE ENERGY HYPERSURFACE | 4 |
| I-4 THE TRANSITION STATE | 4 |
| I-5 PROTEIN FOLDING INTERMEDIATES | 5 |
| I-6 THE NATIVE STATE | 6 |
| I-7 MULTIDOMAIN AND MULTIMERIC PROTEINS | 7 |
| I-8 THE EVOLUTION OF FOLDABLE POLYPEPTIDES AND FOLDING MECHANISMS | 8 |
| I-9 A SYNOPSIS? | 9 |
| II. ANTIBODY FOLDING AND ASSEMBLY – CLASSICAL THEMES & NOVEL CONCEPTS..... | 11 |
| II-1 PROTEIN FOLDING IN THE ENDOPLASMIC RETICULUM | 11 |
| II-2 A SHORT OVERVIEW OVER ANTIBODY BIOLOGY | 12 |
| II-3 BIOSYNTHESIS OF ANTIBODIES IN THE CELL | 13 |
| II-4 ANTIBODY STRUCTURE AND THE EVOLUTION OF THE IMMUNOGLOBULIN FOLD | 14 |
| II-5 FROM THE FOLDING OF ANTIBODY DOMAINS TO COMPLETE MOLECULES | 17 |
| II- 6 QUALITY CONTROL OF ANTIBODY FOLDING IN VIVO | 21 |
| II- 7 A COMPREHENSIVE VIEW OF ANTIBODY FOLDING, ASSEMBLY AND QUALITY CONTROL – THE CURRENT STATUS AND CHALLENGES AHEAD | 24 |
| III. A SHORT SUMMARY OF THE WORK | 26 |
| IV. REFERENCES..... | 27 |

PART B – SCIENTIFIC PUBLICATIONS

| | |
|---|------------|
| V. PUBLISHED SCIENTIFIC PAPER..... | 38 |
| VI. PUBLISHED BOOK CHAPTERS..... | 106 |

I. A short introduction into protein folding

I-1 General considerations

Hardly any field of biological research has undergone such considerable changes and advancements in recent years as has the field of protein folding. Almost five decades ago the observation that protein folding in the test tube is reversible (Anfinsen and Haber, 1961) laid the basis of the field. Countless studies on protein folding have been published since then, and especially in the last few years, the advent of new techniques has pushed barriers and protein folding emerged as one of the first truly interdisciplinary fields in biological sciences. This work is supposed to contribute to this development, in particular by bridging the gap between *in silico*, *in vitro* and *in vivo* studies on protein folding for a well known model system: antibodies.

I-2 A brief history of protein folding

As soon as it had been appreciated that proteins are irregularly but well defined structured molecular objects, the question of how proteins could be able to self-organize began to bother scientists. The adverse effect of this motivation was that the structural perspective of the protein folding problem shaped its perception. In other words, early protein folding studies aimed at a well-defined description of the protein folding phenomenon with only few states involved, unfolded, native and intermediate, often applying the conceptual framework of organic chemistry developed for small molecules. The *new view of protein folding*, despite several shortcomings, has highlighted this conceptual dilemma of protein folding (Dill and Chan, 1997; Leopold et al., 1992; Onuchic et al., 1995). Some of its basic ideas had already been developed earlier and applied to the native state of proteins (Austin et al., 1973; Frauenfelder et al., 1979; Hartmann et al., 1982). Together they shifted the focus, and nowadays proteins are regarded as highly dynamical objects. This dynamical character should not only influence the native state of proteins but in particular be a dominant factor in folding (Hartmann et al., 1982; Leopold et al., 1992; Henzler-Wildman et al., 2007a; Henzler-Wildman et al., 2007b). Dynamics implies

heterogeneity and heterogeneity suggests parallel rather than strictly sequential pathways for protein folding. This was a main idea of the new view which arose from a merge of polymer physics and protein science (Dill and Chan, 1997; Leopold et al., 1992; Onuchic et al., 1995). It put forward that polypeptide chains navigate on a free energy hypersurface which is generally funnelled towards the native state (Figure I-1). Accordingly, already few interactions can tip the energetic balance of a polypeptide chain, predisposing it towards further movement down the funnel and finally the native state(s) (Dill and Chan, 1997; Onuchic et al., 1995). Even though the funnel concept alone does not necessarily explain different protein folding phenomena and is very often applied quite generously, it facilitated novel perspectives on protein folding as will be outlined in the following.

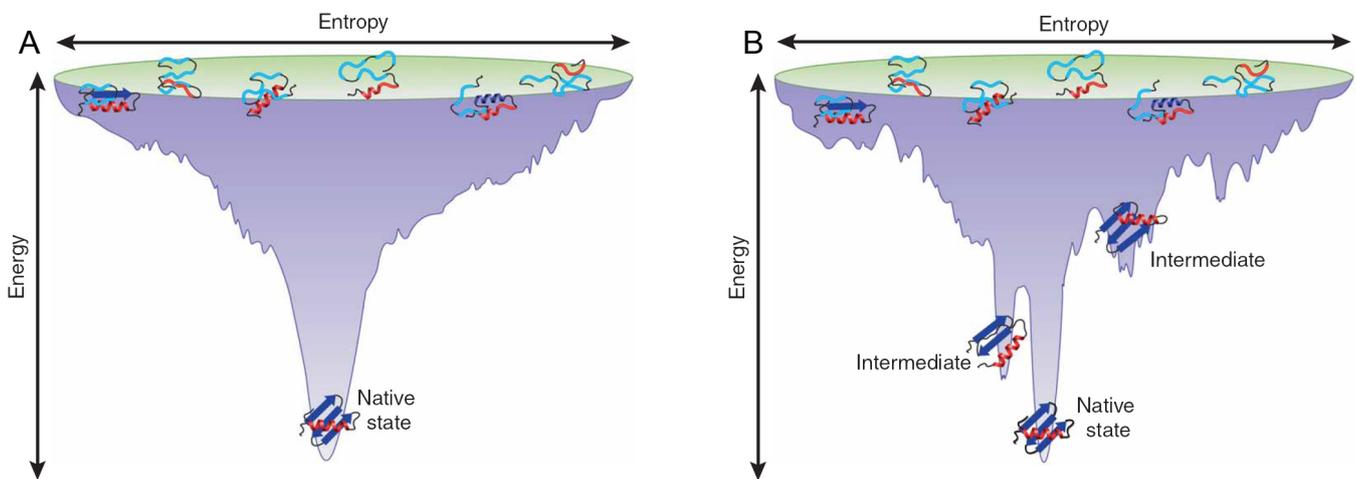


Figure I-1: A schematic protein folding funnel. Shown are possible folding funnels for a two-state (A) and a multi-state folder (B). The smoothness and the number and characteristics of the available energy minima on the free energy hypersurface will shape the folding mechanism. Figure adapted from (Bartlett and Radford, 2009).

I-3 The unfolded state

Originally, the unfolded state of proteins was termed *random coil*. This implies the absence of defined interactions or conformations of an unfolded protein. It is accordingly expected to behave like a random polymer in a good solvent. This view is out-dated. It was mainly founded on techniques with too low resolution or too low sensitivity which could hence only provide global properties for the investigated

proteins. In contrast, several recent studies on chemically denatured proteins, even under harsh conditions, have been able to detect significant residual structure, local and non-local, native and non-native (Crowhurst et al., 2002; Mok et al., 2007; Sanchez and Kiefhaber, 2003a; Le Duff et al., 2006; Smith et al., 1996). This residual structure can be expected to have a pronounced effect on the folding free energy landscape of a protein. If native, it can entropically destabilize the unfolded state thereby rendering folding more favourable. The classical examples are disulfide bridges (Pace et al., 1988; Clarke et al., 1995a; Clarke et al., 1995b; Abkevich and Shakhnovich, 2000; Eyles et al., 1994), which historically also played a significant role as probes to study protein folding (Creighton, 1974; Creighton, 1975). Non-covalent interactions, though, seem to be far from absent in unfolded states of proteins. If non-native, enthalpical stabilization of the unfolded state might be a consequence. It should be kept in mind, however, that most studies on the unfolded state of proteins dealt with rather harsh conditions of elevated temperature or high denaturant concentrations. Under physiological conditions, the amount of structure in unfolded states of proteins can be expected to be significantly higher as have revealed the few accessible cases (Religa et al., 2005; Marsh et al., 2007; Chugha et al., 2006) and a compaction of chemically unfolded proteins with lowering denaturant concentrations suggest (Sherman and Haran, 2006; Merchant et al., 2007).

Even more important, *in vivo*, proteins encounter quite a different scenario than dilution from a chemically denatured state. They begin their folding process often during translation or translocation; processes, which only make parts of a protein at a certain time amenable for folding, and in particular to form non-local interactions (Kosolapov and Deutsch, 2009; Zhang et al., 2009; Kimchi-Sarfaty et al., 2007; Tsai et al., 2008; Clarke and Clark, 2008; Hsu et al., 2007; Hsu et al., 2009). Furthermore, if folding occurs, it takes place in a confined space like the maximum 20 Å wide ribosome exit tunnel (Ban et al., 2000) which is chemically distinct from an aqueous solution (Lu et al., 2007).

I-3 Movements on the free energy hypersurface

Proteins are far more complex than the small molecules transition state theory has been developed for. Furthermore, in protein folding reactions a large number of weak non-covalent interactions are formed and broken and shaped by entropy, in contrast to the small molecules of organic chemistry. It is thus very unlikely that the dynamics of polypeptide chains can be adequately described by a system developed for a completely different class of molecules. Significant recent progress has been obtained in this field. Using fast spectroscopic techniques (Lapidus et al., 2000; Fierz et al., 2007), probes which are straightforward in the interpretation of the data (Krieger et al., 2003) and molecular dynamics simulations (Yeh and Hummer, 2002) has been a fruitful approach which provided significant new insights. Polypeptide chains are now believed to explore the free energy surface limited by local geometry around certain amino acids (Krieger et al., 2005) as well as solvent parameters and intrinsic hydrogen bonding (Moglich et al., 2006). Additionally, local chain motions have been found to be strongly coupled to motions of other chain segments (Fierz and Kiefhaber, 2007). Overall, polypeptides seem to move on a rough free energy surface within timescales of picoseconds to microseconds (Fierz et al., 2007) with some generalizable but also sequence-specific dynamic parameters (Krieger et al., 2003) not only in the unfolded state (Moglich et al., 2006), but also in partially folded or native states (Fierz et al., 2009).

I-4 The transition state

As the transition state of any chemical reaction represents a maximum on the free energy surface, it never becomes significantly populated in equilibrium. This makes indirect techniques necessary to explore its structural and dynamical features. Ablating specific interactions in a protein's native state and monitoring the impact on the folding kinetics, i.e. the stabilizing effect of this specific interaction on the transition state, has been coined *Phi-value* analysis for protein folding reactions (Matouschek et al., 1989). Being an energetics-based approach, it has been considerably debated in its structural interpretation, the adequate way of mutating a protein and the impact on other states than the native and the transition state

(Buchner and Kiefhaber, 1990; Sanchez and Kiefhaber, 2003a). Undoubtedly, though, it has provided insights into an otherwise elusive state of protein folding (Fersht et al., 1992; Matouschek et al., 1992a; Matouschek et al., 1992b; Matouschek et al., 1989; Serrano et al., 1992a; Serrano et al., 1992b; Serrano et al., 1992c; Cota et al., 2001; Hamill et al., 2000; Lappalainen et al., 2008). In particular, the combination with molecular dynamics simulations proved to be a rewarding approach as Phi-value analyses can only provide local information which, when coupled to molecular dynamics simulations, can be extended to provide global conformational features of protein folding transition states (Mayor et al., 2003; Geierhaas et al., 2004; Paci et al., 2003; Vendruscolo et al., 2001). Based on these analyses, protein folding transition states are now mostly believed to roughly resemble expanded counterparts of a proteins native state – with a native-like topology yet only very few tertiary interactions (Vendruscolo et al., 2001; Lindorff-Larsen et al., 2004).

I-5 Protein folding intermediates

Originally believed to be the key to circumvent the famous Levinthal protein folding paradoxon, intermediates in protein folding have lived through a quite diverse career. They are considered to be partially folded states of proteins, which can either accelerate, decelerate or even misguide protein folding (Wagner and Kiefhaber, 1999; Mucke and Schmid, 1994; Liu et al., 2000; Jahn et al., 2006; Jahn and Radford, 2008). As such, they were hard to position in classical protein folding descriptions. Yet, being considered as local minima on the protein folding energy hypersurface, their role and properties come quite natural (Brockwell and Radford, 2007). If separated by high free energy barriers, intermediates can be readily observable which directly implies that non-observability does not mean absence. Indeed, for a lot of proteins previously believed to be two-state folders partially folded states could be detected with recent techniques of higher sensitivity and time resolution (Sanchez and Kiefhaber, 2003b; Bai, 2006; Lipman et al., 2003; Korzhnev et al., 2004). Accordingly, the population of partially folded states is rather a gradual than an all-or-none process with the possible and highly debated extreme case of barrierless downhill folding (Onuchic and Wolynes, 2004; Garcia-Mira et al., 2002; Sadqi et al., 2006; Ma and Gruebele, 2005; Ferguson et al., 2007). Intermediates can be

stabilized by native or non-native secondary and tertiary structure interactions (Bai, 2006; Brockwell and Radford, 2007; Creighton et al., 1996; Feng et al., 2005; Kameda et al., 2005; Kiefhaber et al., 1995; Korzhnev et al., 2004; Liu et al., 2000; Religa et al., 2005; Udgaonkar and Baldwin, 1988). As the underlying interactions are *per se* unspecific, it is not surprising that a whole plethora of different intermediate characteristics is observed. First high resolution structural information on folding intermediates (Hughson et al., 1990; Kiefhaber et al., 1995; Udgaonkar and Baldwin, 1988; Kameda et al., 2005; Korzhnev et al., 2004; Mizuguchi et al., 2003; Religa et al., 2005) or complete folding pathways (Feng et al., 2005; Mayor et al., 2003) put them in a structural line with the transition states: well defined topology, sometimes ill-defined native interactions or non- native interactions underlining the hierarchical character of protein folding and the role of topology for protein folding mechanisms in general (Plaxco et al., 1998; Zarrine-Afsar et al., 2005). In contrast to the transition state, however, intermediates can be populated for significant amounts of time and therefore might be bifurcation points of a protein folding reaction towards the native or misfolded potentially harmful states (Liu et al., 2000; Kameda et al., 2005; Jahn et al., 2006; Jahn and Radford, 2008; Simpson et al., 2009).

I-6 The native state

Quite unexpectedly, the native state of proteins, formerly considered as the most “reliable” of all states, has undergone a drastic change in its perception, too. It has been considered dynamic for a long time (Austin et al., 1973; Frauenfelder et al., 1979; Hartmann et al., 1982) and recent technical advances have tightened the boundaries between dynamics and functions of proteins (Eisenmesser et al., 2005; Henzler-Wildman et al., 2007a; Henzler-Wildman et al., 2007b). In particular, studies on single protein molecules emphasized the ensemble character not only of unfolded, partially folded but also the native state of proteins (Lipman et al., 2003; Schuler et al., 2002). The most surprising discovery, however, was that a significant amount of naturally occurring proteins do not or hardly possess any defined native three-dimensional structure. This class of proteins has been named *intrinsically disordered* and is believed to make up almost 40% of the human proteome (Dunker et al., 2008; Uversky et al., 2005; Yang et al., 2005). Historically, this class of proteins has been

overlooked for a long time as most early biochemical studies focussed on enzymes where well defined structures are a prerequisite for well defined functions. Intrinsically disordered proteins are consequently underrepresented in this functional class. In contrast, a significant amount of proteins involved in signalling and regulation are thought to be unstructured (Dunker et al., 2008; Uversky et al., 2005; Yang et al., 2005). In general unstructured until assembled with their partners, which can either be small molecule ligands or, in most cases, other proteins (Dyson and Wright, 2002; Fink, 2005). Action through interaction thus in particular applies to these proteins which are able to couple high specificity with low affinity by enthalpy-entropy compensation. Sometimes promiscuous and often highly dynamic regulatory networks can thus be realized by intrinsically disordered proteins.

I-7 Multidomain and multimeric proteins

Being experimentally and theoretically much easier to address than oligomeric proteins, isolated one-domain proteins have dominated protein folding studies. Yet, in the organism, almost all proteins are found to be multidomain oligomers (Han et al., 2007) or at least involved in transient protein-protein interactions to fulfil their function. The evolutionary reasons to form oligomeric structures are diverse. Functionally, heterooligomers possess the advantage of being modular and therefore evolutionary more adaptable. Furthermore, oligomers or multidomain proteins can in principle work more efficiently on substrates by e.g. substrate channelling. In general, the stability of a protein is increasing with the hydrophobic surface buried upon folding (Miller et al., 1987). Therefore, based on simple surface to volume considerations, larger proteins have the possibility to reach higher stabilities than their smaller counterparts. This opens up the possibility for more non-ideal local geometries which might functionally be necessary. In particular, the symmetric or pseudo-symmetric association of smaller building blocks is energetically favourable (Ponstingl et al., 2005) and concomitantly reduces the problems of transcription/translation errors which are more likely to occur in monocistronically encoded large proteins. On the other hand, oligomeric proteins have to be kinetically stabilized to avoid dissociation in the crowded cellular milieu with often very low concentrations of the cognate association partners. Furthermore, in particular for multidomain proteins, individual

domains have to be evolved for robust folding in the local proximity of other domains being in the course of folding (Han et al., 2007).

Of course, the same principles deduced for the folding of small proteins still hold but often they have to be extended. One example is the stability of oligomeric structures. Up to 50% of the observed free energy change upon folding can be attributed to interactions of the association surfaces which often has a significant effect on the observed folding behavior (Jaenicke and Lilie, 2000). In many cases, partially structured intermediates or very unstable monomers form which only upon association are able to attain their native structure and/or stability. The association-coupled structure formation processes and the necessary recognition specificity of partially folded species, conceptually intimately linked to intrinsically disordered proteins, are still largely open questions.

1-8 The evolution of foldable polypeptides and folding mechanisms

Proteins are not primarily evolved for folding but for function. This becomes obvious for intrinsically disordered proteins. Though, as function is in most cases linked to structure, folding comes as a necessary additional selection criterion for proteins and dominant negative effects or protein misfolding have to be avoided. Indeed, folding mechanisms and protein structures seem to be more conserved than functions or sequences in protein families (Chothia and Lesk, 1986; Zarrine-Afsar et al., 2005). It should be kept in mind that at a first glance quite different folding reactions, e.g. two-state or multi-state, might be dictated by the same folding mechanisms (Friel et al., 2004; Spudich et al., 2004). The smoothness of the underlying free energy surface will govern the population of partially folded species; topology and few key interactions are believed to make up the general mechanism (Bartlett and Radford, 2009). On the one hand, this argues for a high robustness of folding pathways against mutational changes. On the other hand, it might imply that efficient folding to a defined structure is far from being trivial to achieve evolutionary and therefore conserved in the cause of evolution. Indeed, even though random polypeptide sequences can sometimes fold into a defined structure, they do so in an uncooperative manner (Davidson and Sauer, 1994) and structure-based design of proteins led to highly complex and uncooperative folding pathways (Dantas et al.,

2003; Scalley-Kim and Baker, 2004; Watters et al., 2007). This shows that simple and cooperative folding pathways are not physical prerequisites for a defined protein structure but evolutionary selected traits. Cooperativity is assumed to be achieved by avoidance of an over-stabilization of the native state, selective destabilization of partially folded states, favorization of non-local stabilizing interactions and overlapping folding nuclei – non-trivial features of a protein (Chan et al., 2004; Watters et al., 2007; Haglund et al., 2008). They are upheld even if in conflict with fast folding (Capraro et al., 2008; Jager et al., 2006) or functional needs (Jager et al., 2006; Nordlund et al., 2009; Friel et al., 2009) pointing towards their importance for the homeostasis of the organism.

I-9 A synopsis?

Our understanding of protein folding has undergone considerable progress in recent years. Even structure-prediction and *de novo* protein design can now be carried out with a low but remarkable success rate (Dahiyat and Mayo, 1997; Kuhlman et al., 2003), often still relying on semi-empirical approaches, though.

A common understanding of protein folding, even though a lot of details are still highly debatable, entirely missing and most likely crucial, might be the following: Proteins start from a non-native state, which can already possess local and/or non-local fluctuating structural elements and thereby are limited in their conformational freedom. The polypeptide chain explores the ensemble of conformations accessible in a time scale of picoseconds for local conformational changes to nanoseconds or even microseconds for more global conformational changes. Once crucial residues have come into contact, folding can proceed via a transition state with a roughly native-like topology and different degrees of structural polarization. From this state of high free energy, folding can proceed to partially folded states which will be more native-like but might possess non-native elements and interactions. Dependent on the successive free energy barriers to complete folding, these states can be absent, populated very transiently or quite stably before complete folding to the native state occurs via much more native-like transition states. Dependent on the cooperativity of the process, the number of folding nuclei (“foldons” (Englander et al., 2007)), the nature and kind of rate limiting-steps etc. a variable number parallel pathways, mostly

differing in local properties, can be taken finally resulting in a dynamic native state (Figure I-2). Even though all these consideration should also hold *in vivo*, a considerable modulation of the outlined principles can be expected due to the nature of the environment and the ubiquitous presence of folding helpers and catalysts (Walter and Buchner, 2002; Hartl and Hayer-Hartl, 2009; Kramer et al., 2009).

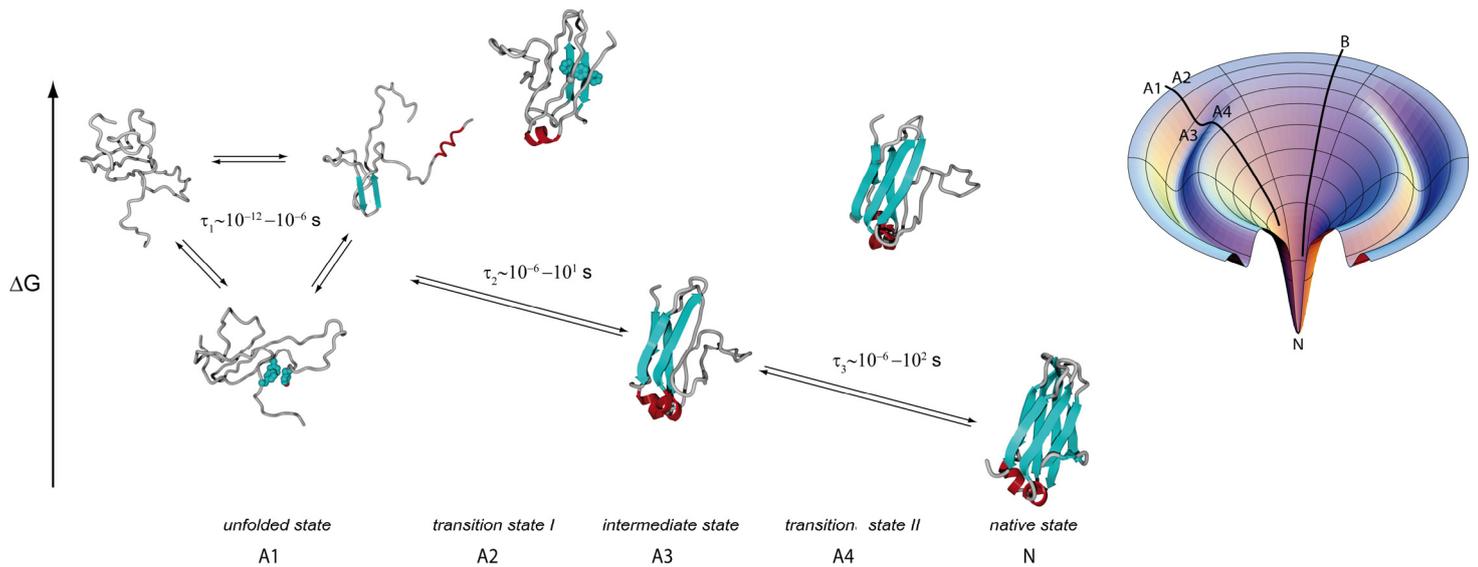


Figure I-2: A generalizing view on protein folding. From the unfolded ensemble, where the protein can sample multiple states within ps to μs, folding proceeds via a first transition state. This is assumed to be roughly native like in topology with only very few key interactions established. In a subsequent intermediate state, first structural elements rigidify before complete folding to the native state can occur. Typical timescales of the processes are indicated. As folding takes place on a funnel-shaped multidimensional energy landscape, the indicated pathway A is just one of the multiple pathways possible (e.g. an intermediate-free pathway B) and bifurcation might occur at different stages.

II. Antibody folding and assembly – classical themes & novel concepts

II-1 Protein folding in the endoplasmic reticulum

Naturally, proteins do not fold in the well controlled environment of the test tube but in the crowded cellular environment. Posttranslational modifications and chaperone interactions can and will occur in most cases – features which will heavily influence most folding reactions. Approximately one third of all human proteins is destined for secretion or transport to cellular membranes and therefore has to traverse the endoplasmic reticulum (ER). Polypeptides generally enter this compartment co-translationally. Accordingly, the ER represents the major folding compartment of the cell and sets the quality control standards all transported proteins have to pass. It is thus compulsory for the ER to be equipped with a large variety of folding helpers and catalysts (Meunier et al., 2002). The most prominent ER chaperones are Hsp70 family members like BiP (Haas and Wabl, 1983) and the accessory J-proteins, ERdj's (Dudek et al., 2009), the Hsp90 homologue Grp94 (Melnick et al., 1994; Ostrovsky et al., 2009) and the lectins calnexin and calreticulin (Helenius and Aebi, 2004; Trombetta and Helenius, 1998). Folding catalysts like protein disulfide isomerases (PDIs) (Freedman, 1989; Jessop et al., 2004; Ellgaard and Ruddock, 2005) and peptidyl-prolyl isomerases (PPIases) (Fischer et al., 1989; Lang et al., 1987) are also abundant in the ER.

Generally, folding and posttranslational modification of proteins in the endoplasmic reticulum occur with high efficiency and fidelity. Nevertheless, the ER folding machinery can become acutely or chronically overwhelmed. Therefore, the ER has evolved backup mechanisms to cope with ER-folding stress: the unfolded protein response (UPR) (Mori et al., 1992) and the ER-associated degradation (ERAD) (Lippincott-Schwartz et al., 1988; McCracken and Brodsky, 1996). ER-load is sensed by at least three sensors in the ER membrane of higher eukaryotes, IRE1, PERK and ATF6, which can, if necessary, induce an upregulation of the ER folding machinery via a signalling cascade to the nucleus (Ron and Walter, 2007) and additionally shut down the synthesis of new proteins via phosphorylation of translation initiation factors (Harding et al., 2000). If stress persists, the UPR can even lead to the induction of apoptosis (Lin et al., 2007). Terminally misfolded proteins can be retro-translocated

from the ER to the cytosol, ubiquitylated and degraded by the proteasome in the ERAD pathway (Vembar and Brodsky, 2008).

Key players of the folding machinery in the ER, as well as the major ER stress response mechanisms, UPR and ERAD, have been deciphered in recent years. What is still missing though is a detailed insight into the molecular mechanisms of folding and quality control in the ER. Only a combined biophysical and cell biological approach can address this complex issue and has been applied to the folding of antibodies within this work.

II-2 A short overview over antibody biology

For most genes, great care is taken to ensure that mutations are rapidly detected and repaired in order to maintain primary amino acid sequences that fold efficiently into biologically active proteins. In contrast, highly active mutational mechanisms are essential to generate antibody diversity and fast adaptability of the immune response, which clearly increases the likelihood of producing immunoglobulin (Ig) proteins that are incapable of folding, assembling or being transported to the cell surface. Failure to closely monitor and control maturation of antibodies would wreak havoc on the proper functioning of the immune system, which can only be avoided by the action of robust folding and quality control systems.

Ig proteins serve as cell surface antigen receptors on B cells, and upon antigen stimulation and plasma cell differentiation they are secreted as soluble effector molecules (antibodies) that provide protection against infections and foreign antigens. In their simplest form, the IgG antibodies, each molecule is composed of two identical heavy chains (HC) and two identical light chains (LC) that are linked by disulfide bonds. Both chains are composed of multiple domains; each composed of ~100 amino acids (Figure II-1A). The N-terminal domain of each chain is unique and is therefore called the variable domain (V_H respectively V_L). Together both variable domains form the antigen recognition region (Figure II-1A). Within the variable domains are stretches of particularly diverse amino acids (hypervariable regions) that provide the exquisite binding specificity of the antibody molecule. The remainder of the polypeptides is conserved within antibody classes (constant domains) and is

important for effector functions like complement activation or recruitment of macrophages and natural killer cells. Five different classes or isotypes of antibodies are made in most higher vertebrates: IgM, IgG, IgA, IgE, and IgD that differ in the HC constant regions used. Only two different types of LC (κ or λ) generally exist, which can assemble with all HC classes. However, in a given cell only one heavy chain and one light chain allele are expressed, so that antibodies with a single specificity are produced.

II-3 Biosynthesis of antibodies in the cell

The development of progenitor cells committed to the B cell lineage is characterized by the sequential expression of the HC and LC subunits. In preB cells, variable (V_H), diversity (D_H) and joining (J_H) gene segments are rearranged at the DNA level of a single allele to create a variable region (Maki et al., 1980; Early et al., 1980), which is then spliced to the constant region at the mRNA level (Calame et al., 1980). Once a functional HC is made, gene rearrangements begin for the LC, to form the light chain variable region (Bernard et al., 1978). The B cell is characterized by surface expression of Ig molecules, which are composed of two identical HC that possess a hydrophobic membrane anchor sequence and two identical LC.

DNA rearrangements that give rise to Ig variable regions are characterized by imprecise joining of the three gene segments, the addition of non-templated bases at the site of joining of these gene segments, and finally, during later stages of differentiation, the directed hypermutation of the variable region exons of HC and LC genes (Alt et al., 1987). These mechanisms are essential to generate antibody diversity and allow affinity maturation of the immune response, yet they clearly increase the likelihood of producing a protein that is incapable of folding and assembling properly, being transported to the cell surface or secreted, or engaging the appropriate signaling molecules thus compromising the functioning of the immune system. Accordingly, B lineage cells are particularly dependent on the ER quality control system to ensure that only correctly assembled Ig molecules are transported to the cell surface. In light of this, it is not surprising that many of the major components of the mammalian ER quality control machinery (e.g. BiP (Haas and

Wabl, 1983; Bole et al., 1986), calnexin (Hochstenbach et al., 1992), GRP170 (Lilie et al., 1993), ERdj3 (Meunier et al., 2002), and pERp1 (Meunier et al., 2002)) were first identified by virtue of their association with HC, and that Ig molecules were some of the earliest identified substrates of ER folding enzymes (e.g. protein disulfide isomerase (Roth and Pierce, 1987) and prolyl isomerases (Lang et al., 1987; Lilie et al., 1993).

Like all cell surface and secreted proteins, HC and LC are co-translationally translocated into the endoplasmic reticulum (ER) and folding begins even before the polypeptide chains are completely translated, starting with the V domain and proceeding through the C domains (Bergman and Kuehl, 1979). IgGs predominantly assemble first as HC dimers to which LC are added one at a time (Baumal et al., 1971). The C_H3 domain is important in this assembly, as IgG HC mutants with a deleted C_H3 domain do not form HC dimers readily and are often secreted as HCLC 'hemimers' (Zack et al., 1981). N-linked glycans are also added to nascent HC co-translationally (Bergman and Kuehl, 1978) and can restrict the potential folding pathways available to the nascent chain due to their highly polar nature. The heavily glycosylated heavy chains of IgM require the glycans for assembly and transport, while monoglycosylated heavy chains of IgG mature properly even without them (Hickman and Kornfeld, 1978). Although in vitro folding studies have not addressed the role of glycans in Ig domain folding due to the fact that the isolated domains in most cases have been produced in bacteria, they have provided a molecular explanation for most of the folding and assembly processes previously identified in cells.

II-4 Antibody structure and the evolution of the immunoglobulin fold

In the case of IgG, the complete molecule is composed of two four-domain heavy chains and two two-domain light chains (Figure II-1A), whose overall shape resembles a "Y". It should be noted, however, that the orientation of the two arms of the Y is flexible because of an unstructured hinge region between the first (C_H1) and the second (C_H2) constant domain of the heavy chain. The protease papain cleaves the IgG molecule in the hinge region and divides it into three functional segments, each of which is a dimer (Kalmanson and Bronfenbrenner, 1942; Porter, 1950). The

two N-terminal fragments are termed Fab fragments (for fragment antigen binding). The remaining fragment, the Fc fragment (for fragment crystallisable), is important for connecting antigen binding to antibody effector functions. Dimerization of the Fc fragment is largely mediated by interactions between the C_H3 domains and stabilized by disulfide bonds in the hinge region. The C_H2 domains only interact via the attached sugar residues (Figure II-1A). They determine the C_H2-C_H2 orientation and spacing and are therefore crucial for the binding of downstream effectors (Huber et al., 1976; Krapp et al., 2003; Feige et al., 2009b).

Each of the Ig domains forms a highly similar beta sandwich structure, known as the Ig fold. The Ig fold is characterized by a greek-key β -barrel topology in which the barrel is not continuously hydrogen-bonded but composed of two sheets forming a sandwich-like structure. The variable domains (Figure II-1B) are comprised of nine strands (abcc'c''defg) and the constant domains (Figure II-1C) of seven strands (abcdefg) (Bork et al., 1994). In most antibody domains, a buried disulfide bridge connects strands b and f, which spans ~60-70 residues (Huber et al., 1976; Bork et al., 1994). It is orientated roughly perpendicular to the individual sheets and significantly stabilizes the folded domain (Goto and Hamaguchi, 1979). Another characteristic feature antibody domains share is a conserved tryptophan residue that is located in proximity to the internal disulfide bridge. Since its fluorescence is specifically quenched only in the native state by the adjacent disulfide bond, it can be used as a reporter group for the conformational state of antibody domains (Goto and Hamaguchi, 1979). Additionally, (*cis*-) proline residues are unusually abundant in antibody domains, contributing up to 10% of the amino acids. Particularly characteristic is a conserved *cis*-proline residue in the loop connecting strand b and c of the constant domains which is often preceded by an aromatic amino acid (Figure II-1C).

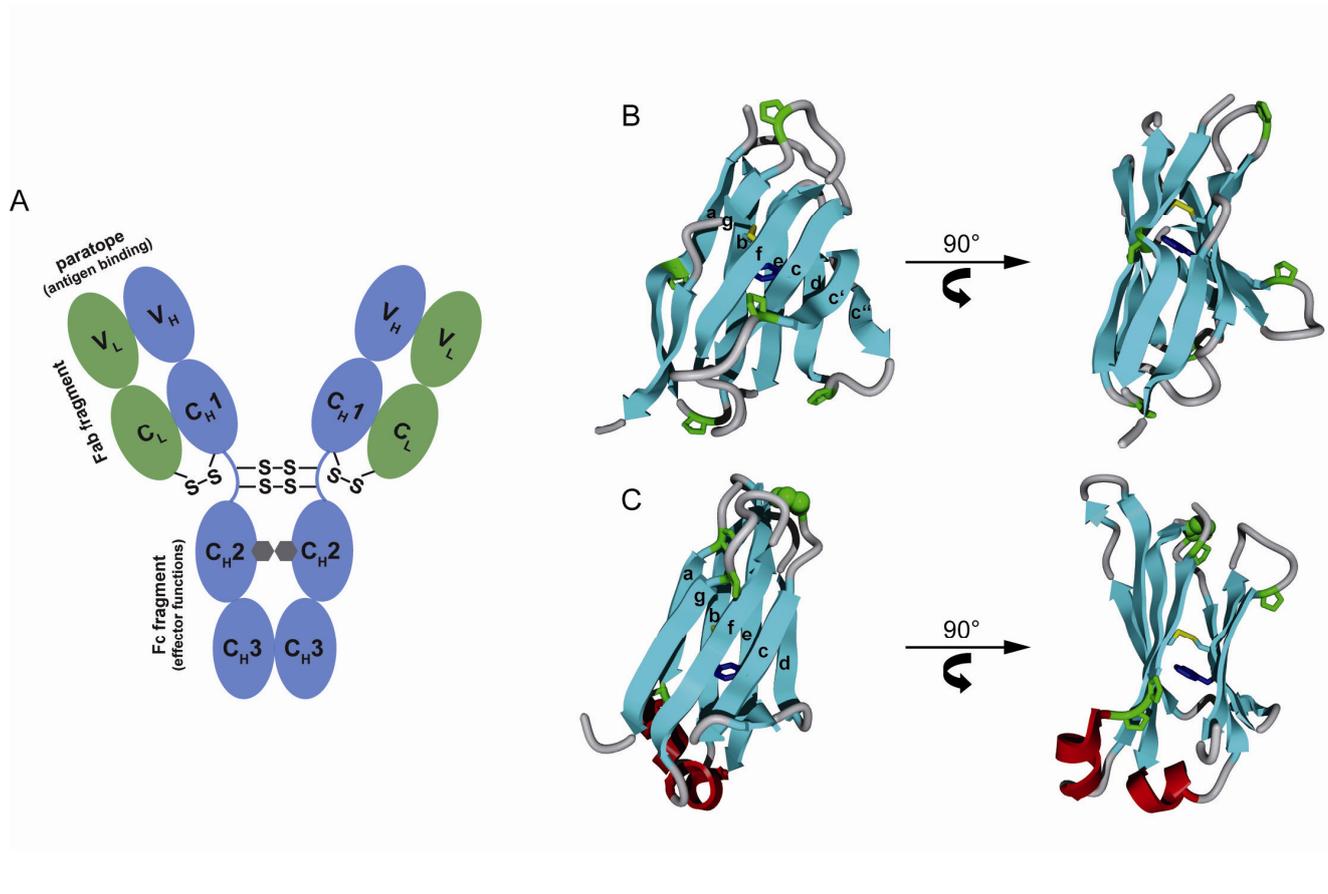


Figure II-1: Overall antibody structure and domain architecture. (A) Domain arrangement of an IgG antibody molecule. The light chains are shown in green, the heavy chains in blue. The oligosaccharide between the C_H2 domains is depicted as grey hexagon. Interchain disulfide bridges and important functional elements of the antibody (antigen binding paratope, Fab fragment, Fc fragment) are indicated. Domain architecture of the light chain variable domain (V_L) (B) and the light chain constant domain (C_L) (C). The strand nomenclature is indicated. The intrachain disulfide bridge (yellow) and the proximal conserved tryptophan residue (blue) are shown. The proline residues of the two domains are shown in green with the highly conserved *cis*-proline residue between strands b and c of C_L highlighted in a CPK representation. Small helices (red) connect strands a and b and strands e and f of the C_L domain.

Even though the Ig domain or “fold” is characteristic of antibodies, it is in fact one of the most widely used protein topologies in nature, giving rise to what is referred to as the Ig superfamily (IgSF). The origin of the Ig fold dates back to ~750 million years of evolutionary history, with the identification of IgSF members already in sponges (Du et al., 2004; Hsu et al., 2006). In contrast, the ability to produce antibodies is a more recent occurrence (~500 million years), dating back to cartilaginous fish, such as sharks, skates and rays (Hsu et al., 2006; Dooley and Flajnik, 2006). In vertebrates, the Ig fold is the major building block of extracellular recognition systems (Williams and Barclay, 1988; Barclay, 2003), while in invertebrates, IgSF members are more

limited to the neural system (Rougon and Hobert, 2003). The Ig fold has also been detected in prokaryotic and viral proteins, albeit it much less frequently, suggesting that it might have been acquired in these instances by horizontal gene transfer (Halaby and Mornon, 1998). The evolutionary success of the Ig superfamily can most likely be attributed to its robust fold, which provides stability against proteases and harsh environments, and the ability to build on this common core structure by incorporating highly diverse binding loops or edge strands.

II-5 From the folding of antibody domains to complete molecules

Antibodies are highly complex molecules. Dissection of the overall molecule into individual domains or fragments was necessary to determine folding differences between the structurally similar antibody domains. The pioneering studies on antibody folding were performed on secreted LCs that were denatured and allowed to refold in vitro (Goto and Hamaguchi, 1979). Further studies examined the constant domain of the light chain (C_L) (Goto and Hamaguchi, 1982a) and IgG heavy chain C_{H3} domain (Isenman et al., 1979) and established that individual Ig domains could fold autonomously. A key characteristic in the folding of antibody domains, is that proline-isomerization reactions, which are intrinsically slow due to their high activation energy, strongly influence antibody folding (Goto and Hamaguchi, 1982a; Ramm et al., 1999; Feige et al., 2004), often as the rate limiting step (Thies et al., 1999; Feige et al., 2008a). Recent in vitro studies have addressed the folding of Ig domains in more detail with a view to include all the IgG constant region domains (Freund et al., 1996; Ramm et al., 1999; Feige et al., 2004; Feige et al., 2007; Feige et al., 2008a; Feige et al., 2009a; Lilie et al., 1995; Simpson et al., 2009; Thies et al., 1999; Thies et al., 2002). These studies provided the first high resolution data on antibody folding and revealed that, although all Ig domains appear very similar in terms of the final structure, they can be grouped into three different folding categories (Figure II-2).

In the first category, exemplified by the well-studied C_L protein, domains are able to fold autonomously to a monomeric state. In the denatured protein, no residual structure was detected, regardless of whether or not its internal disulfide bridge was present (Feige et al., 2007; Feige et al., 2008a). However, once folding was initiated, the presence of an internal disulfide bridge exerted an important guiding impact on

the folding pathway (Goto and Hamaguchi, 1982b; Feige et al., 2007; Ramm et al., 1999). The reason for this became clear when studies revealed that the folding nucleus for Ig domains involves the clustering of hydrophobic residues in strands b, c, e and f (Freund et al., 1996; Geierhaas et al., 2004; Lappalainen et al., 2008). The term folding nucleus implies that these residues cluster together in an early reaction when the overall topology of the polypeptide chain starts to arrange. The covalent linkage of residues in strands b and f via a disulfide bond facilitates the establishment of this folding nucleus, creating a structured, on-pathway intermediate and preventing unproductive off-pathway interactions. Indeed, a population of misfolded, aggregation-prone off-pathway folding intermediates for C_L was detected in the absence of the internal disulfide bridge (Feige et al., 2007).

While folding intermediates are usually quite transient and elusive, they can be populated for longer times if a slow reaction limits the subsequent folding step. This is the case for C_L, where the major on-pathway folding intermediate is relatively long-lived due to the non-native *trans* isomerization state of a proline residue between strands b and c, which must isomerize to its *cis* state before folding can proceed (Figure II-2) (Feige et al., 2008a). Using NMR combined with molecular dynamics simulations, it was possible to follow the changes in the chemical environment of most amino acids and to obtain an atomic resolution view of the intermediate structure. It revealed that the core β -strands of the Ig topology were almost completely formed, while higher flexibility was observed for flanking strands, particularly strand d (Figure II-2) (Feige et al., 2008a), in keeping with data on other IgSF members (Mizuguchi et al., 2003; Kameda et al., 2005; Jahn et al., 2006). Importantly, two small helices linking strands a and b and strands e and f (Figure II-1C, Figure II-2) are natively structured very early (Feige et al., 2008a). They can be assumed to stabilize the orientation and spacing of the beta-strands of the Ig fold and furthermore correctly position a bulky hydrophobic residue in the core of the protein (Feige et al., 2008a). Thus, if present, these helices should render the folding of Ig domains more robust as they stabilize a highly-structured on-pathway intermediate in a conformation which is poised for subsequent productive folding. These small helices are highly conserved in most constant region domains and other members of the Ig superfamily (Bork et al., 1994; Barclay, 2003; Williams and Barclay, 1988) but are absent in several members of the IgSF, in particular those few, where high resolution data is available on folding intermediates (Mizuguchi et al., 2003; Kameda

et al., 2005; Jahn et al., 2006). Importantly, no such helices are found in variable antibody domains (Figure II-1B) or in several IgSF members that are prone to misfolding and amyloid formation (Kameda et al., 2005; Jahn et al., 2006; Qin et al., 2007), suggesting that differences between partially structured folding intermediates, e.g. the presence or absence of small, highly structured helices, might hold the clue for the tendency of IgSF members to either reliably fold or tend to misfold as could be shown for β_2 -microglobulin (Feige et al., 2008a). The structural insights gained for the major C_L folding intermediate seem to be transferable to other HC constant domains, in particular C_{H2} .

The second category of antibody domain folding pathways (Figure II-2) is represented by the C_{H3} domain (Isenman et al., 1979; Thies et al., 1999). In addition to folding slower than the domains in the first category, this domain forms an obligatory homodimer. As with the C_L domain, a partially folded species was observed, which was trapped by a non-native prolyl isomerization state (Thies et al., 1999). Interestingly, the formation of the correct prolyl isomers was required for dimerization to occur, whereas the internal disulfide bridge was dispensable for folding and self-association (Thies et al., 2002). Thus, proline switches can influence not only folding but also dimerization.

The third, and most unexpected, category of antibody domain folding is the recently discovered, template-assisted folding of the C_{H1} domain, which interacts with the C_L domain in the intact antibody (Figure II-1A). The isolated C_{H1} domain is intrinsically disordered as determined by various spectroscopic techniques (Feige et al., 2009a). This is in marked contrast to all antibody domains previously studied and unexpected from the crystal structure of antibodies (Huber et al., 1976). To form a hydrophobic folding nucleus in the C_{H1} domain, the interaction with a few key residues in the dimerization interface of the folded C_L domain is required. Unlike other domains where disulfide bond formation only enhances the folding process, it was absolutely required for C_{H1} . Another unique twist is that the rate-limiting proline isomerization between strand b and c observed in the folding of other Ig domains can only occur after association with the C_L domain. These observations are in agreement with previous studies on the complete antibody Fab fragment where C_{H1} folding was proposed to be the slowest step occurring after association of the HC and the LC (Lilie et al., 1995), and underline the special role of the C_{H1} domain in B cell biology.

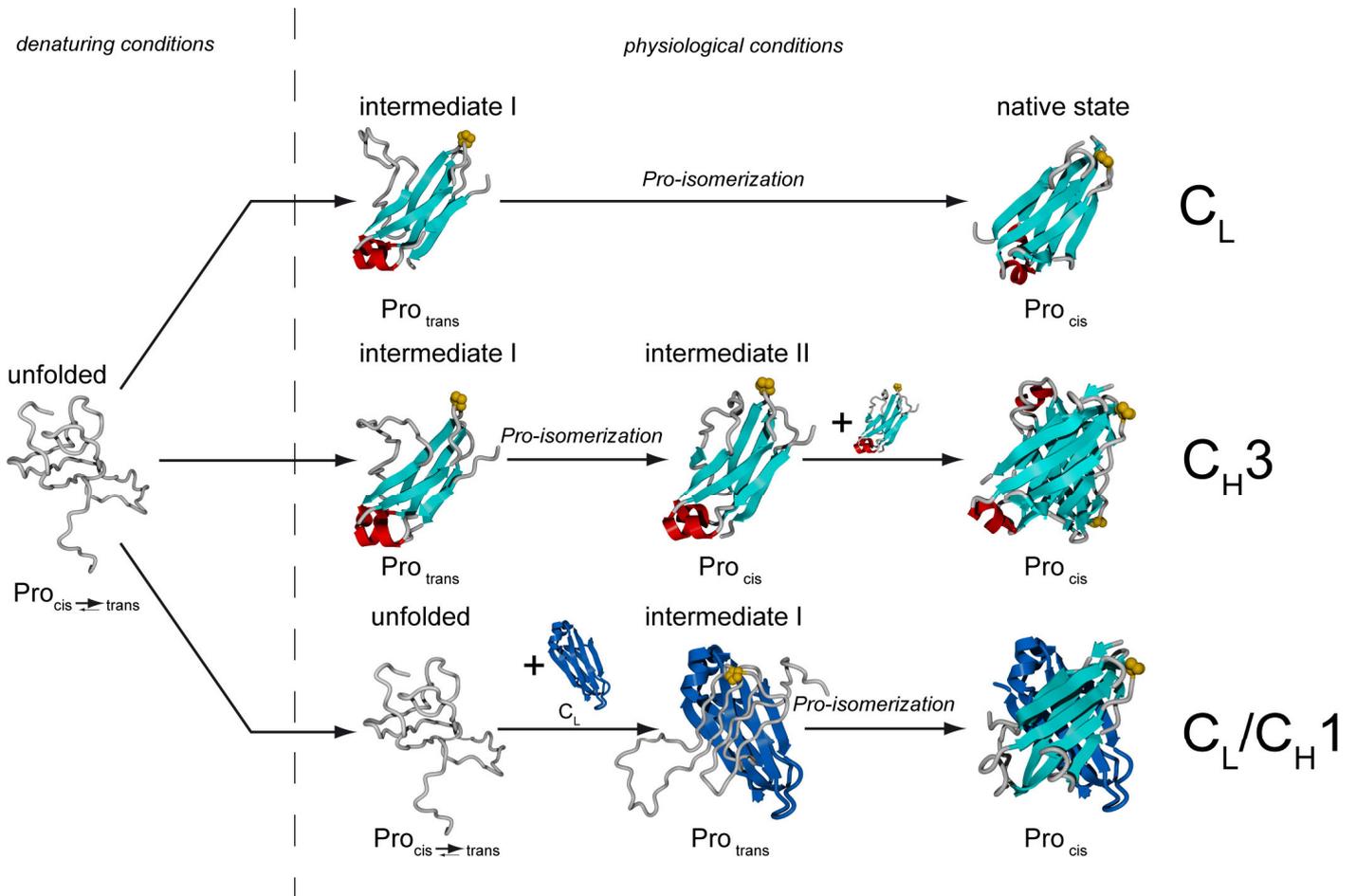


Figure II-2: Three pathways of antibody domain folding. C_L folds via a highly structured on-pathway intermediate that is trapped by the *trans* state of a proline residue in the loop connecting strands b and c (highlighted in yellow). In the intermediate, the core β -sheet structure and the two short helices connecting strands a and b and strands e and f are fully formed (shown in red). The obligate dimer C_H3 folds via two intermediates, both most likely similar in structure to the one of the C_L domain. In a first, rapidly formed intermediate, a critical proline residue (highlighted in yellow) must isomerize to its native *cis* state, leading to a dimerization competent second intermediate. C_H1 is intrinsically disordered in isolation. Upon association with C_L it forms a loosely folded intermediate. In this intermediate, isomerization of the conserved proline residue between strands b and c (highlighted in yellow) limits the complete folding to the native state and formation of the interchain disulfide bridge between C_H1 and C_L . Accordingly, for most constant antibody domains, isomerization of a proline residue between strands b and c, which is usually preceded by an aromatic amino acid, is the rate limiting step for the folding of the major part of the molecule.

Together, the data reveal that the same final fold is reached for all antibody domains by overall conserved mechanisms, yet with strikingly different details. The biological significance of these differences becomes clear when one considers the biosynthesis of antibodies, where the C_H1 domain is indeed a built-in assembly sensor.

II- 6 Quality control of antibody folding in vivo

The mechanisms for generating between 10^7 and 10^9 different antibody specificities from a limited amount of genetic information are equivalent to extensive mutagenesis. Therefore, it is not surprising that a number of quality control checkpoints are required during B cell development and differentiation that monitor the integrity of the antibody (Figure II-3). The newly created HC produced by a preB cell is retained in the ER unless it associates with a “surrogate LC”. This LC-mimetic is assembled from the V_{preB} protein (contributing the variable domain) (Kudo and Melchers, 1987) and the λ_5 protein (supplying the constant domain) (Sakaguchi and Melchers, 1986). The preB-specific surrogate LC “tests” the ability of HC to assemble properly with a LC-like protein and to pass ER quality control measures. If properly associated, the HC will be transported to the cell surface and engage the B cell receptor signaling components (Karasuyama et al., 1990). The surrogate LC focuses in large part on the C_{H1} domain, as unassembled C_{H1} -deleted HC can traffic to the cell surface of the preB cell and transmit a “functional” signal (Shaffer and Schlissel, 1997). Once the heavy chain is judged functional by these criteria, conventional light chain gene rearrangements commence. The C_{H1} domain remains a critical focus of Ig quality control efforts throughout B cell development and plasma cell differentiation. Unlike LC, which can be secreted alone, HCs are retained in the ER and eventually degraded unless they assemble with LC (Figure II-3) (Mains and Sibley, 1983). LC loss variants of plasmacytomas were very rarely observed, whereas HC loss variants occurred much more frequently (Coffino and Scharff, 1971). This was argued to be due to the “toxicity” of free HC, which were neutralized by LC (Kohler, 1980). Exceptions to this rule occur in the rare B cell lymphoproliferative disorder known as Heavy Chain Disease, where truncated Ig HCs are secreted from cells without LCs. Notably, although these short HC have been identified for a number of different isotypes (i.e., IgA, IgG, and IgM), the deletions nearly always involve portions of the V_H and C_{H1} domains (Seligmann et al., 1979). Similarly, mouse plasmacytoma lines expressing HC with deletions of the C_{H1} domain can secrete free HC (Morrison, 1978), whereas deletion of any of the other constant region domains does not permit this. Finally, the serum of Camelidae (i.e., single humped dromedary and llamas) contains a significant fraction of antibodies that are naturally devoid of LC. These HC only antibodies do not possess a C_{H1} domain (Hamers-Casterman et al., 1993; Vu et

al., 1997), further underscoring the evolutionary conserved importance of this domain in regulating Ig transport and quality control.

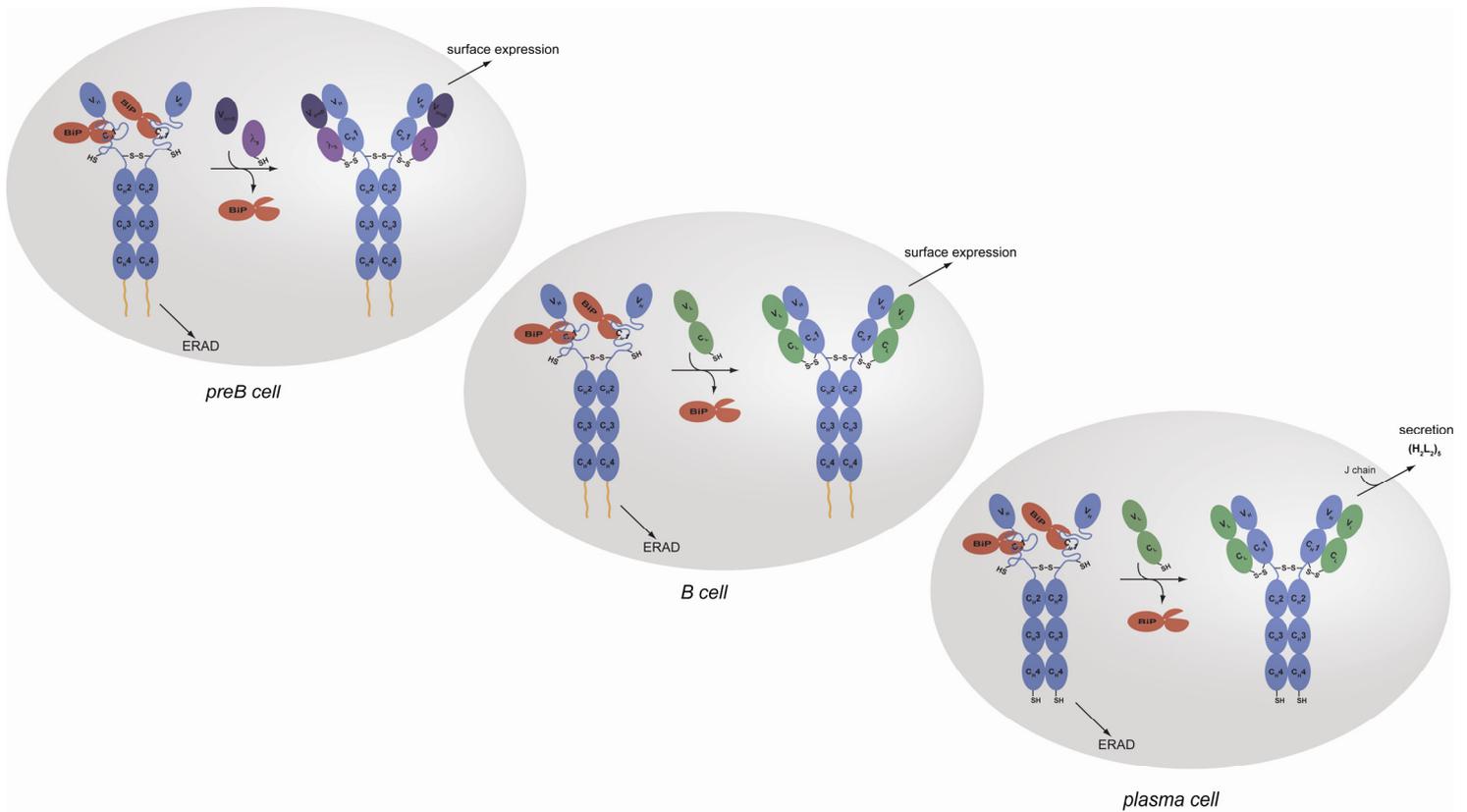


Figure II-3: Immunoglobulin quality control checkpoints at various stages in B cell development. After HC gene rearrangements, preB cells produce BiP-bound μ HC. If their association with the surrogate LC, which is assembled from the V_{preB} and λ_5 proteins, induces BiP release and folding of the C_{H1} domain and if the rest of the Ig domains fold properly, the HC can traffic to the plasma membrane and engage BCR signalling molecules. If there is a failure in any of these steps, the μ HCs are retrotranslocated to the cytosol for degradation by the 26S proteasome. Once conventional LC are produced in the B cell, they similarly assemble with μ HC, displace BiP from the C_{H1} domain and induce its folding. As the rest of the constant region domains were tested at the preB cell stage, only the pairing and folding of the V domains must pass the quality control scrutiny. Plasma cell differentiation leads to the synthesis of extremely high levels of antibodies. The ability of the specific HC and LC combination to assemble and fold properly was verified at earlier stages of development, thus quality control at this point involves monitoring the completeness of Ig assembly, focusing on the LC-induced release of BiP from the C_{H1} domain and its concomitant folding. In addition, there is a shift to the secretory form of μ HC in plasma cells, which possess a terminal cysteine that is involved in assembly with J chain and pentamer formation (IgM). Thiol-mediated retention mechanisms monitor this assembly and prevent IgM monomers from being secreted (Anelli et al., 2003).

The term “ER quality control” refers to the process of monitoring the maturation of nascent secretory proteins and allowing only properly folded and assembled proteins to transit further along the secretory pathway. Proteins that fail to mature properly are retained and eventually retrotranslocated to the cytosol for degradation by the 26S proteasome in a process known as ER associated degradation (ERAD). Immunoglobulin heavy chain binding protein (BiP), the first component of the eukaryotic ER quality control apparatus to be identified, was found by virtue of its association with the unassembled, non-transported HCs produced in pre-B cell lines (Haas and Wabl, 1983). It interacts transiently with Ig assembly intermediates but not with completely assembled H₂L₂ molecules in plasmacytomas (Bole et al., 1986). BiP is the ER orthologue of the Hsp70 family of chaperones (Haas and Meo, 1988) and is retained in the ER, along with any protein associated with it, by virtue of a C-terminal KDEL tetrapeptide (Munro and Pelham, 1987). Similar to the differences detected by in vitro folding experiments, the folding requirements and dependence on BiP for the various Ig domains are quite different in cells. BiP binds transiently to some Ig domains (i.e., V_L, V_H, and some C_H domains) (Kaloff and Haas, 1995), while other domains appear to fold rapidly and stably without ever interacting with BiP (i.e., C_L) (Hellman et al., 1999; Lee et al., 1999). Only the C_H1 domain interacts stably with BiP in the absence of LC (Hendershot et al., 1987) (Figure II-3). The association of this domain with BiP is crucial for controlling Ig assembly and transport, because its deletion, and the resulting ablation of BiP binding, leads to the secretion of incompletely assembled Ig intermediates. Studies to determine the folding status of proteins or domains in the ER of cells rely on the analysis of the oxidation status of cysteine residues, which often form disulfide bonds upon folding making the protein more compact and faster migrating on non-reducing SDS polyacrylamide gels (Braakman et al., 1992). Surprisingly, unlike other Ig domains, the C_H1 domain remains reduced in the absence of LC (Lee et al., 1999). In contrast to the in vitro studies described above (Feige et al., 2009a), there was no evidence of BiP associating with an oxidized C_H1 domain. This may be due to the fact that BiP does not appear to continuously cycle from HC in the absence of LC in cells, which is required for oxidation (Vanhove et al., 2001), or that association with light chain, oxidation, and folding of C_H1 are more tightly coupled in vivo. In keeping with in vitro studies, although the ATP-mediated release of BiP from isolated HC resulted in C_H1 domain oxidation, this was not sufficient for proper folding (Vanhove et al., 2001).

This argues that LC association is also required for folding of this domain in vivo. Only LC in which both domains are folded led to C_H1 domain oxidation and secretion from cells. In vivo experiments also confirmed the requirement for proline isomerization in promoting C_H1 oxidation, assembly with LC and secretion (Feige et al., 2009a). Thus, the evolution of a unique C_H1 domain that absolutely requires assembly with a C_L domain for folding, allows the cell to ensure that newly produced HCs in B cells will be retained unless they are able to pass that first important test: their ability to combine with a LC and be transported only after assembly. It also ensures that plasma cells, which have been estimated to synthesize up to 10³ antibody molecules a second, are not releasing partially assembled subunits that cannot properly bind to antigen or perform effector functions.

Many LC fold readily and can be secreted by themselves as either monomers (Dul et al., 1996) or dimers (Leitzgen et al., 1997), suggesting that the V_L domains of these LCs are likely to belong to categories 1 and 2 described above. However, LC exist that cannot be secreted alone due to a failure of the V_L domain to fold properly by itself (Skowronek et al., 1998). These LC fold perfectly well and are secreted when they assemble with a HC (Horibata and Harris, 1970), suggesting that their V_L domains may belong to category 3. This requirement for assembly assisted folding of some V_L, and presumably V_H domains, could explain the restriction of possible HC and LC pairings (Wiens et al., 1998) and might play a role in antibody misfolding diseases.

II- 7 A comprehensive view of antibody folding, assembly and quality control – the current status and challenges ahead

In summary, detailed biophysical and in vivo studies on individual Ig domains, antibody fragments and complete antibodies have identified common themes that provide a detailed picture of IgG folding (Figure II-4). Once the internal disulfide bridge is formed, most domains will autonomously fold in a three state reaction via on-pathway folding intermediates whose lifetime is increased by incorrect peptidyl-prolyl isomerization states. Peptidyl-prolyl isomerization reactions control folding, assembly and formation of disulfide bridges and might also play a role in inhibiting

aggregation. The very first steps in domain folding involve hydrophobic interactions in the inner face of strands b, c, e and f. Once folded, C_{H3} induces Fc dimerization and the covalent linkage of the heavy chains in the hinge region (Figure II-4). At this stage, all the constant region domains except C_{H1} are folded in a heavy chain dimer. Association of a folded light chain will induce folding of the C_{H1} domain, and the assembly of heavy and light chains will be stabilized by an interchain disulfide bond (Figure II-4). This picture shows nicely how the combination of biophysical properties of the individual domains revealed through in vitro studies and the delineation of cellular components controlling the synthesis of antibodies provide a comprehensive view of the basic mechanisms that govern antibody synthesis. Future work must focus on further integrating the role of the complex ER folding machinery in modulating, synchronizing and controlling assembly and folding of antibodies and other IgSF members.

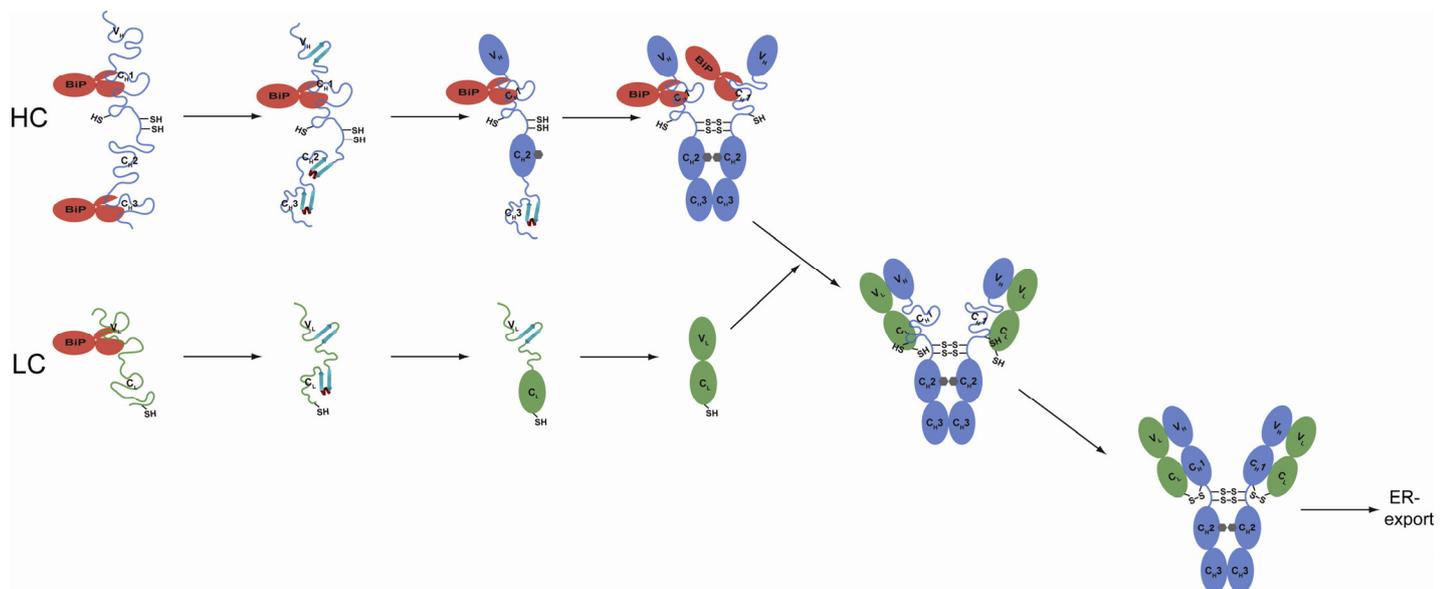


Figure II-4: An overall view over antibody folding and assembly. Folding, formation of disulfide bridges and glycosylation of the heavy chain (blue) and light chain (green) begins co-translationally in the ER. The molecular chaperone BiP (red) interacts with most of the domains before folding is completed. All constant domains and most variable domains fold autonomously, populating an on-pathway intermediate on the way to the native state. C_L is known to fold particularly fast in the cell. Once C_{H3} is folded it induces HC dimerization which will be solidified by disulfide bridges in the hinge region. C_{H1} remains unfolded, unoxidized and stably bound to BiP until the LC displaces BiP and C_L induces folding of the C_{H1} domain. Once all C_{H1} prolines are in the correct isomerization state and C_{H1} is folded, a disulfide bridge between the LC and the HC forms. Most of the steps are likely to hold for other Ig classes. Chaperones and folding catalysts, like Grp94, protein disulfide isomerase PDI and the peptidyl-prolyl isomerase CyclophilinB are likely to contribute to the individual steps in immunoglobulin biogenesis.

III. A short summary of the work

Within this work, novel insights into the dynamics, folding and association of polypeptide chains *in silico*, *in vitro* and *in vivo* could be obtained. For simple oligopeptides, we were able to show that rugged energy landscapes govern their behaviour (Feige et al., 2008b) and laid the basis for future studies on biologically relevant molecules and molecular processes. Our structural work on the Fc fragment helped to answer a long standing functional question about IgG: why deglycosylation goes hand in hand with the complete loss of effector functions (Feige et al., 2009b). The major efforts of our *in vitro* folding studies focussed on the constant domain of the antibody light chain, C_L. There, we were able to show that the evolutionary conserved intrinsic disulfide bridge in antibody domains not only confers stability, but, as it is located in the folding nucleus, also guides the folding pathway and inhibits misfolding (Feige et al., 2007). Follow-up studies on the C_L domain focussed on its major folding intermediate which we could trap in equilibrium and characterize at atomic resolution by a combined experimental and molecular dynamics approach. Two small helices, previously unrecognized as important in immunoglobulin folding, were found to be completely structured in the intermediate and to decisively influence immunoglobulin folding for another member of the superfamily (Feige et al., 2008a). Finally, we were able by a combined *in vitro* and *in vivo* study to provide a molecular description of the quality control mechanism antibodies have to pass prior to secretion from the endoplasmic reticulum. The key is an unfolded domain within the heavy chain, which strongly binds to the molecular chaperone BiP. Thereby it retains the heavy chain in the ER until paired with the light chain which induces folding (Feige et al., 2009a). Taken together, our work has provided significant new insights into the folding and assembly of antibodies from the test tube to the cell and paved the path for rational optimization of antibodies as well as further detailed studies on their folding, assembly and evolution.

IV. References

- Abkevich, V.I. and Shakhnovich, E.I. (2000). What can disulfide bonds tell us about protein energetics, function and folding: Simulations and bioinformatics analysis. *Journal of Molecular Biology* 300, 975-985.
- Alt, F.W., Blackwell, T.K., and Yancopoulos, G.D. (1987). Development of the primary antibody repertoire. *Science* 238, 1079-1087.
- Anelli, T., Alessio, M., Bachi, A., Bergamelli, L., Bertoli, G., Camerini, S., Mezghrani, A., Ruffato, E., Simmen, T., and Sitia, R. (2003). Thiol-mediated protein retention in the endoplasmic reticulum: the role of ERp44. *EMBO J.* 22, 5015-5022.
- Anfinsen, C.B. and Haber, E. (1961). Studies on the reduction and re-formation of protein disulfide bonds. *J. Biol. Chem.* 236, 1361-1363.
- Austin, R.H., Beeson, K., Eisenstein, L., Frauenfelder, H., Gunsalus, I.C., and Marshall, V.P. (1973). Dynamics of carbon monoxide binding by heme proteins. *Science* 181, 541-543.
- Bai, Y. (2006). Energy barriers, cooperativity, and hidden intermediates in the folding of small proteins. *Biochemical and Biophysical Research Communications* 340, 976-983.
- Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* 289, 905-920.
- Barclay, A.N. (2003). Membrane proteins with immunoglobulin-like domains--a master superfamily of interaction molecules. *Semin. Immunol.* 15, 215-223.
- Bartlett, A.I. and Radford, S.E. (2009). An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms. *Nat. Struct. Mol. Biol.* 16, 582-588.
- Baumal, R., Potter, M., and Scharff, M.D. (1971). Synthesis, assembly, and secretion of gamma globulin by mouse myeloma cells. 3. Assembly of the three subclasses of IgG. *J. Exp. Med.* 134, 1316-1334.
- Bergman, L.W. and Kuehl, W.M. (1978). Temporal relationship of translation and glycosylation of immunoglobulin heavy and light chains. *Biochemistry* 17, 5174-5180.
- Bergman, L.W. and Kuehl, W.M. (1979). Formation of an intrachain disulfide bond on nascent immunoglobulin light chains. *J. Biol. Chem.* 254, 8869-8876.
- Bernard, O., Hozumi, N., and Tonegawa, S. (1978). Sequences of mouse immunoglobulin light chain genes before and after somatic changes. *Cell* 15, 1133-1144.
- Bole, D.G., Hendershot, L.M., and Kearney, J.F. (1986). Posttranslational Association of Immunoglobulin Heavy-Chain Binding-Protein with Nascent Heavy-Chains in Nonsecreting and Secreting Hybridomas. *Journal of Cell Biology* 102, 1558-1566.
- Bork, P., Holm, L., and Sander, C. (1994). The Immunoglobulin Fold - Structural Classification, Sequence Patterns and Common Core. *Journal of Molecular Biology* 242, 309-320.
- Braakman, I., Helenius, J., and Helenius, A. (1992). Manipulating disulfide bond formation and protein folding in the endoplasmic reticulum. *EMBO J.* 11, 1717-1722.
- Brockwell, D.J. and Radford, S.E. (2007). Intermediates: ubiquitous species on folding energy landscapes? *Curr. Opin. Struct. Biol.* 17, 30-37.
- Buchner, J. and Kiefhaber, T. (1990). Folding pathway enigma. *Nature* 343, 601-602.

- Calame,K., Rogers,J., Early,P., Davis,M., Livant,D., Wall,R., and Hood,L. (1980). Mouse Cmu heavy chain immunoglobulin gene segment contains three intervening sequences separating domains. *Nature* 284, 452-455.
- Capraro,D.T., Roy,M., Onuchic,J.N., and Jennings,P.A. (2008). Backtracking on the folding landscape of the beta-trefoil protein interleukin-1beta? *Proc. Natl. Acad. Sci. U. S. A* 105, 14844-14848.
- Chan,H.S., Shimizu,S., and Kaya,H. (2004). Cooperativity principles in protein folding. *Methods Enzymol.* 380, 350-379.
- Chothia,C. and Lesk,A.M. (1986). The relation between the divergence of sequence and structure in proteins. *EMBO J.* 5, 823-826.
- Chugha,P., Sage,H.J., and Oas,T.G. (2006). Methionine oxidation of monomeric lambda repressor: the denatured state ensemble under nondenaturing conditions. *Protein Sci.* 15, 533-542.
- Clarke,J., Henrick,K., and Fersht,A.R. (1995a). Disulfide Mutants of Barnase .1. Changes in Stability and Structure Assessed by Biophysical Methods and X-Ray Crystallography. *Journal of Molecular Biology* 253, 493-504.
- Clarke,J., Hounslow,A.M., and Fersht,A.R. (1995b). Disulfide Mutants of Barnase .2. Changes in Structure and Local Stability Identified by Hydrogen-Exchange. *Journal of Molecular Biology* 253, 505-513.
- Clarke,T.F. and Clark,P.L. (2008). Rare codons cluster. *PLoS. One.* 3, e3412.
- Coffino,P. and Scharff,M.D. (1971). Rate of somatic mutation in immunoglobulin production by mouse myeloma cells. *Proc. Natl. Acad. Sci. U. S. A* 68, 219-223.
- Cota,E., Steward,A., Fowler,S.B., and Clarke,J. (2001). The folding nucleus of a fibronectin type III domain is composed of core residues of the immunoglobulin-like fold. *Journal of Molecular Biology* 305, 1185-1194.
- Creighton,T.E. (1974). The single-disulphide intermediates in the refolding of reduced pancreatic trypsin inhibitor. *J. Mol. Biol.* 87, 603-624.
- Creighton,T.E. (1975). The two-disulphide intermediates and the folding pathway of reduced pancreatic trypsin inhibitor. *J. Mol. Biol.* 95, 167-199.
- Creighton,T.E., Darby,N.J., and Kemmink,J. (1996). The roles of partly folded intermediates in protein folding. *Faseb Journal* 10, 110-118.
- Crowhurst,K.A., Tollinger,M., and Forman-Kay,J.D. (2002). Cooperative interactions and a non-native buried Trp in the unfolded state of an SH3 domain. *J. Mol. Biol.* 322, 163-178.
- Dahiyat,B.I. and Mayo,S.L. (1997). De novo protein design: fully automated sequence selection. *Science* 278, 82-87.
- Dantas,G., Kuhlman,B., Callender,D., Wong,M., and Baker,D. (2003). A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins. *J. Mol. Biol.* 332, 449-460.
- Davidson,A.R. and Sauer,R.T. (1994). Folded proteins occur frequently in libraries of random amino acid sequences. *Proc. Natl. Acad. Sci. U. S. A* 91, 2146-2150.
- Dill,K.A. and Chan,H.S. (1997). From Levinthal to pathways to funnels. *Nature Structural Biology* 4, 10-19.
- Dooley,H. and Flajnik,M.F. (2006). Antibody repertoire development in cartilaginous fish. *Dev. Comp Immunol.* 30, 43-56.

- Du, P.L., Zucchetti, I., and De Santis, R. (2004). Immunoglobulin superfamily receptors in protochordates: before RAG time. *Immunol. Rev.* *198*, 233-248.
- Dudek, J., Benedix, J., Cappel, S., Greiner, M., Jalal, C., Muller, L., and Zimmermann, R. (2009). Functions and pathologies of BiP and its interaction partners. *Cell Mol. Life Sci.* *66*, 1556-1569.
- Dul, J.L., Aviel, S., Melnick, J., and Argon, Y. (1996). Ig light chains are secreted predominantly as monomers. *J. Immunol.* *157*, 2969-2975.
- Dunker, A.K., Silman, I., Uversky, V.N., and Sussman, J.L. (2008). Function and structure of inherently disordered proteins. *Curr. Opin. Struct. Biol.* *18*, 756-764.
- Dyson, H.J. and Wright, P.E. (2002). Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* *12*, 54-60.
- Early, P., Huang, H., Davis, M., Calame, K., and Hood, L. (1980). An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: VH, D and JH. *Cell* *19*, 981-992.
- Eisenmesser, E.Z., Millet, O., Labeikovsky, W., Korzhnev, D.M., Wolf-Watz, M., Bosco, D.A., Skalicky, J.J., Kay, L.E., and Kern, D. (2005). Intrinsic dynamics of an enzyme underlies catalysis. *Nature* *438*, 117-121.
- Ellgaard, L. and Ruddock, L.W. (2005). The human protein disulphide isomerase family: substrate interactions and functional properties. *EMBO Rep.* *6*, 28-32.
- Englander, S.W., Mayne, L., and Krishna, M.M. (2007). Protein folding and misfolding: mechanism and principles. *Q. Rev. Biophys.* *40*, 287-326.
- Eyles, S.J., Radford, S.E., Robinson, C.V., and Dobson, C.M. (1994). Kinetic Consequences of the Removal of A Disulfide Bridge on the Folding of Hen Lysozyme. *Biochemistry* *33*, 13038-13048.
- Feige, M.J., Groscurth, S., Marcinowski, M., Shimizu, Y., Kessler, H., Hendershot, L.M., and Buchner, J. (2009a). An unfolded CH1 domain controls the assembly and secretion of IgG antibodies. *Mol. Cell* *34*, 569-579.
- Feige, M.J., Nath, S., Catharino, S.R., Weinfurter, D., Steinbacher, S., and Buchner, J. (2009b). Structure of the murine unglycosylated IgG1 Fc fragment. *J. Mol. Biol.* *391*, 599-608.
- Feige, M.J., Groscurth, S., Marcinowski, M., Yew, Z.T., Truffault, V., Paci, E., Kessler, H., and Buchner, J. (2008a). The structure of a folding intermediate provides insight into differences in immunoglobulin amyloidogenicity. *Proc. Natl. Acad. Sci. U. S. A* *105*, 13373-13378.
- Feige, M.J. and Paci, E. (2008b). Rate of loop formation in peptides: a simulation study. *J. Mol. Biol.* *382*, 556-565.
- Feige, M.J., Hagn, F., Esser, J., Kessler, H., and Buchner, J. (2007). Influence of the internal disulfide bridge on the folding pathway of the C-L antibody domain. *Journal of Molecular Biology* *365*, 1232-1244.
- Feige, M.J., Walter, S., and Buchner, J. (2004). Folding mechanism of the C(H)2 antibody domain. *Journal of Molecular Biology* *344*, 107-118.
- Feng, H.Q., Zhou, Z., and Bai, Y.W. (2005). A protein folding pathway with multiple folding intermediates at atomic resolution. *Proceedings of the National Academy of Sciences of the United States of America* *102*, 5026-5031.
- Ferguson, N., Sharpe, T.D., Johnson, C.M., Schartau, P.J., and Fersht, A.R. (2007). Structural biology: analysis of 'downhill' protein folding. *Nature* *445*, E14-E15.
- Fersht, A.R., Matouschek, A., and Serrano, L. (1992). The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* *224*, 771-782.

- Fierz,B. and Kiefhaber,T. (2007). End-to-end vs interior loop formation kinetics in unfolded polypeptide chains. *J. Am. Chem. Soc.* *129*, 672-679.
- Fierz,B., Reiner,A., and Kiefhaber,T. (2009). Local conformational dynamics in alpha-helices measured by fast triplet transfer. *Proc. Natl. Acad. Sci. U. S. A* *106*, 1057-1062.
- Fierz,B., Satzger,H., Root,C., Gilch,P., Zinth,W., and Kiefhaber,T. (2007). Loop formation in unfolded polypeptide chains on the picoseconds to microseconds time scale. *Proc. Natl. Acad. Sci. U. S. A* *104*, 2163-2168.
- Fink,A.L. (2005). Natively unfolded proteins. *Curr. Opin. Struct. Biol.* *15*, 35-41.
- Fischer,G., Wittmann-Liebold,B., Lang,K., Kiefhaber,T., and Schmid,F.X. (1989). Cyclophilin and peptidyl-prolyl cis-trans isomerase are probably identical proteins. *Nature* *337*, 476-478.
- Frauenfelder,H., Petsko,G.A., and Tsernoglou,D. (1979). Temperature-dependent X-ray diffraction as a probe of protein structural dynamics. *Nature* *280*, 558-563.
- Freedman,R.B. (1989). Protein disulfide isomerase: multiple roles in the modification of nascent secretory proteins. *Cell* *57*, 1069-1072.
- Freund,C., Honegger,A., Hunziker,P., Holak,T.A., and Pluckthun,A. (1996). Folding nuclei of the scFv fragment of an antibody. *Biochemistry* *35*, 8457-8464.
- Friel,C.T., Beddard,G.S., and Radford,S.E. (2004). Switching two-state to three-state kinetics in the helical protein Im9 via the optimisation of stabilising non-native interactions by design. *J. Mol. Biol.* *342*, 261-273.
- Friel,C.T., Smith,D.A., Vendruscolo,M., Gsponer,J., and Radford,S.E. (2009). The mechanism of folding of Im7 reveals competition between functional and kinetic evolutionary constraints. *Nat. Struct. Mol. Biol.* *16*, 318-324.
- Garcia-Mira,M.M., Sadqi,M., Fischer,N., Sanchez-Ruiz,J.M., and Munoz,V. (2002). Experimental identification of downhill protein folding. *Science* *298*, 2191-2195.
- Geierhaas,C.D., Paci,E., Vendruscolo,M., and Clarke,J. (2004). Comparison of the transition states for folding of two Ig-like proteins from different superfamilies. *J. Mol. Biol.* *343*, 1111-1123.
- Goto,Y. and Hamaguchi,K. (1979). Role of the Intrachain Disulfide Bond in the Conformation and Stability of the Constant Fragment of the Immunoglobulin Light Chain. *Journal of Biochemistry* *86*, 1433-1441.
- Goto,Y. and Hamaguchi,K. (1982a). Unfolding and Refolding of the Constant Fragment of the Immunoglobulin Light Chain. *Journal of Molecular Biology* *156*, 891-910.
- Goto,Y. and Hamaguchi,K. (1982b). Unfolding and Refolding of the Reduced Constant Fragment of the Immunoglobulin Light Chain - Kinetic Role of the Intrachain Disulfide Bond. *Journal of Molecular Biology* *156*, 911-926.
- Haas,I.G. and Meo,T. (1988). cDNA cloning of the immunoglobulin heavy chain binding protein. *Proc. Natl. Acad. Sci. U. S. A* *85*, 2250-2254.
- Haas,I.G. and Wabl,M. (1983). Immunoglobulin heavy chain binding protein. *Nature* *306*, 387-389.
- Haglund,E., Lindberg,M.O., and Oliveberg,M. (2008). Changes of protein folding pathways by circular permutation. Overlapping nuclei promote global cooperativity. *J. Biol. Chem.* *283*, 27904-27915.
- Halaby,D.M. and Mornon,J.P. (1998). The immunoglobulin superfamily: an insight on its tissular, species, and functional diversity. *J. Mol. Evol.* *46*, 389-400.

- Hamers-Casterman,C., Atarhouch,T., Muyldermans,S., Robinson,G., Hamers,C., Songa,E.B., Bendahman,N., and Hamers,R. (1993). Naturally-Occurring Antibodies Devoid of Light-Chains. *Nature* 363, 446-448.
- Hamill,S.J., Steward,A., and Clarke,J. (2000). The folding of an immunoglobulin-like Greek key protein is defined by a common-core nucleus and regions constrained by topology. *Journal of Molecular Biology* 297, 165-178.
- Han,J.H., Batey,S., Nickson,A.A., Teichmann,S.A., and Clarke,J. (2007). The folding and evolution of multidomain proteins. *Nat. Rev. Mol. Cell Biol.* 8, 319-330.
- Harding,H.P., Zhang,Y., Bertolotti,A., Zeng,H., and Ron,D. (2000). Perk is essential for translational regulation and cell survival during the unfolded protein response. *Mol. Cell* 5, 897-904.
- Hartl,F.U. and Hayer-Hartl,M. (2009). Converging concepts of protein folding in vitro and in vivo. *Nat. Struct. Mol. Biol.* 16, 574-581.
- Hartmann,H., Parak,F., Steigemann,W., Petsko,G.A., Ponzi,D.R., and Frauenfelder,H. (1982). Conformational substates in a protein: structure and dynamics of metmyoglobin at 80 K. *Proc. Natl. Acad. Sci. U. S. A* 79, 4967-4971.
- Helenius,A. and Aebi,M. (2004). Roles of N-linked glycans in the endoplasmic reticulum. *Annu. Rev. Biochem.* 73, 1019-1049.
- Hellman,R., Vanhove,M., Lejeune,A., Stevens,F.J., and Hendershot,L.M. (1999). 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.* 144, 21-30.
- Hendershot,L., Bole,D., Kohler,G., and Kearney,J.F. (1987). Assembly and Secretion of Heavy-Chains That do Not Associate Posttranslationally with Immunoglobulin Heavy-Chain Binding-Protein. *Journal of Cell Biology* 104, 761-767.
- Henzler-Wildman,K.A., Lei,M., Thai,V., Kerns,S.J., Karplus,M., and Kern,D. (2007a). A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* 450, 913-916.
- Henzler-Wildman,K.A., Thai,V., Lei,M., Ott,M., Wolf-Watz,M., Fenn,T., Pozharski,E., Wilson,M.A., Petsko,G.A., Karplus,M., Hubner,C.G., and Kern,D. (2007b). Intrinsic motions along an enzymatic reaction trajectory. *Nature* 450, 838-844.
- Hickman,S. and Kornfeld,S. (1978). Effect of tunicamycin on IgM, IgA, and IgG secretion by mouse plasmacytoma cells. *J. Immunol.* 121, 990-996.
- Hochstenbach,F., David,V., Watkins,S., and Brenner,M.B. (1992). Endoplasmic reticulum resident protein of 90 kilodaltons associates with the T- and B-cell antigen receptors and major histocompatibility complex antigens during their assembly. *Proc. Natl. Acad. Sci. U. S. A* 89, 4734-4738.
- Horibata,K. and Harris,A.W. (1970). Mouse myelomas and lymphomas in culture. *Exp. Cell Res.* 60, 61-77.
- Hsu,E., Pulham,N., Ruffelt,L.L., and Flajnik,M.F. (2006). The plasticity of immunoglobulin gene systems in evolution. *Immunol. Rev.* 210, 8-26.
- Hsu,S.T., Cabrita,L.D., Fucini,P., Christodoulou,J., and Dobson,C.M. (2009). Probing side-chain dynamics of a ribosome-bound nascent chain using methyl NMR spectroscopy. *J. Am. Chem. Soc.* 131, 8366-8367.
- Hsu,S.T., Fucini,P., Cabrita,L.D., Launay,H., Dobson,C.M., and Christodoulou,J. (2007). Structure and dynamics of a ribosome-bound nascent chain by NMR spectroscopy. *Proc. Natl. Acad. Sci. U. S. A* 104, 16516-16521.

- Huber,R., Deisenhofer,J., Colman,P.M., Matsushima,M., and Palm,W. (1976). Crystallographic structure studies of an IgG molecule and an Fc fragment. *Nature* 264, 415-420.
- Hughson,F.M., Wright,P.E., and Baldwin,R.L. (1990). Structural characterization of a partly folded apomyoglobin intermediate. *Science* 249, 1544-1548.
- Isenman,D., Lancet,D., and Pecht,I. (1979). Folding Pathways of Immunoglobulin Domains. The Folding Kinetics of the C3 Domain of Human IgG. *Biochemistry* 18, 3327-3336.
- Jaenicke,R. and Lilie,H. (2000). Folding and association of oligomeric and multimeric proteins. *Adv. Protein Chem.* 53, 329-401.
- Jager,M., Zhang,Y., Bieschke,J., Nguyen,H., Dendle,M., Bowman,M.E., Noel,J.P., Gruebele,M., and Kelly,J.W. (2006). Structure-function-folding relationship in a WW domain. *Proc. Natl. Acad. Sci. U. S. A* 103, 10648-10653.
- Jahn,T.R., Parker,M.J., Homans,S.W., and Radford,S.E. (2006). Amyloid formation under physiological conditions proceeds via a native-like folding intermediate. *Nature Structural & Molecular Biology* 13, 195-201.
- Jahn,T.R. and Radford,S.E. (2008). Folding versus aggregation: polypeptide conformations on competing pathways. *Arch. Biochem. Biophys.* 469, 100-117.
- Jessop,C.E., Chakravarthi,S., Watkins,R.H., and Bulleid,N.J. (2004). Oxidative protein folding in the mammalian endoplasmic reticulum. *Biochem. Soc. Trans.* 32, 655-658.
- Kalmanson,G.M. and Bronfenbrenner,J. (1942). The reversal of pneumococcus quellung by digestion of the antibody with papain. *Science* 96, 21-22.
- Kaloff,C.R. and Haas,I.G. (1995). Coordination of Immunoglobulin Chain Folding and Immunoglobulin Chain Assembly Is Essential for the Formation of Functional Igg. *Immunity* 2, 629-637.
- Kameda,A., Hoshino,M., Higurashi,T., Takahashi,S., Naiki,H., and Goto,Y. (2005). Nuclear magnetic resonance characterization of the refolding intermediate of beta(2)-microglobulin trapped by non-native prolyl peptide bond. *Journal of Molecular Biology* 348, 383-397.
- Karasuyama,H., Kudo,A., and Melchers,F. (1990). The proteins encoded by the VpreB and lambda 5 pre-B cell-specific genes can associate with each other and with mu heavy chain. *J. Exp. Med.* 172, 969-972.
- Kimchi-Sarfaty,C., Oh,J.M., Kim,I.W., Sauna,Z.E., Calcagno,A.M., Ambudkar,S.V., and Gottesman,M.M. (2007). A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science* 315, 525-528.
- Kiefhaber,T., Labhardt,A.M., and Baldwin,R.L. (1995). Direct NMR evidence for an intermediate preceding the rate-limiting step in the unfolding of ribonuclease A. *Nature* 375, 513-515.
- Kohler,G. (1980). Immunoglobulin chain loss in hybridoma lines. *Proc. Natl. Acad. Sci. U. S. A* 77, 2197-2199.
- Korzhev,D.M., Salvatella,X., Vendruscolo,M., Di Nardo,A.A., Davidson,A.R., Dobson,C.M., and Kay,L.E. (2004). Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* 430, 586-590.
- Kosolapov,A. and Deutsch,C. (2009). Tertiary interactions within the ribosomal exit tunnel. *Nat. Struct. Mol. Biol.* 16, 405-411.
- Kramer,G., Boehringer,D., Ban,N., and Bukau,B. (2009). The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nat. Struct. Mol. Biol.* 16, 589-597.

- Krapp,S., Mimura,Y., Jefferis,R., Huber,R., and Sondermann,P. (2003). Structural analysis of human IgG-Fc glycoforms reveals a correlation between glycosylation and structural integrity. *Journal of Molecular Biology* 325, 979-989.
- Krieger,F., Fierz,B., Bieri,O., Drewello,M., and Kiefhaber,T. (2003). Dynamics of unfolded polypeptide chains as model for the earliest steps in protein folding. *J. Mol. Biol.* 332, 265-274.
- Krieger,F., Moglich,A., and Kiefhaber,T. (2005). Effect of proline and glycine residues on dynamics and barriers of loop formation in polypeptide chains. *J. Am. Chem. Soc.* 127, 3346-3352.
- Kudo,A. and Melchers,F. (1987). A second gene, VpreB in the lambda 5 locus of the mouse, which appears to be selectively expressed in pre-B lymphocytes. *EMBO J.* 6, 2267-2272.
- Kuhlman,B., Dantas,G., Ireton,G.C., Varani,G., Stoddard,B.L., and Baker,D. (2003). Design of a novel globular protein fold with atomic-level accuracy. *Science* 302, 1364-1368.
- Lang,K., Schmid,F.X., and Fischer,G. (1987). Catalysis of protein folding by prolyl isomerase. *Nature* 329, 268-270.
- Lapidus,L.J., Eaton,W.A., and Hofrichter,J. (2000). Measuring the rate of intramolecular contact formation in polypeptides. *Proc. Natl. Acad. Sci. U. S. A* 97, 7220-7225.
- Lappalainen,I., Hurley,M.G., and Clarke,J. (2008). Plasticity within the obligatory folding nucleus of an immunoglobulin-like domain. *J. Mol. Biol.* 375, 547-559.
- Le Duff,C.S., Whittaker,S.B., Radford,S.E., and Moore,G.R. (2006). Characterisation of the conformational properties of urea-unfolded Im7: implications for the early stages of protein folding. *J. Mol. Biol.* 364, 824-835.
- Lee,Y.K., Brewer,J.W., Hellman,R., and Hendershot,L.M. (1999). BiP and immunoglobulin light chain cooperate to control the folding of heavy chain and ensure the fidelity of immunoglobulin assembly. *Molecular Biology of the Cell* 10, 2209-2219.
- Leitzgen,K., Knittler,M.R., and Haas,I.G. (1997). Assembly of immunoglobulin light chains as a prerequisite for secretion. A model for oligomerization-dependent subunit folding. *J. Biol. Chem.* 272, 3117-3123.
- Leopold,P.E., Montal,M., and Onuchic,J.N. (1992). Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc. Natl. Acad. Sci. U. S. A* 89, 8721-8725.
- Lilie,H., Lang,K., Rudolph,R., and Buchner,J. (1993). Prolyl isomerases catalyze antibody folding in vitro. *Protein Sci* 2, 1490-1496.
- Lilie,H., Rudolph,R., and Buchner,J. (1995). Association of antibody chains at different stages of folding: prolyl isomerization occurs after formation of quaternary structure. *J. Mol. Biol.* 248, 190-201.
- Lin,J.H., Li,H., Yasumura,D., Cohen,H.R., Zhang,C., Panning,B., Shokat,K.M., Lavail,M.M., and Walter,P. (2007). IRE1 signaling affects cell fate during the unfolded protein response. *Science* 318, 944-949.
- Lindorff-Larsen,K., Vendruscolo,M., Paci,E., and Dobson,C.M. (2004). Transition states for protein folding have native topologies despite high structural variability. *Nat. Struct. Mol. Biol.* 11, 443-449.
- Lipman,E.A., Schuler,B., Bakajin,O., and Eaton,W.A. (2003). Single-molecule measurement of protein folding kinetics. *Science* 301, 1233-1235.
- Lippincott-Schwartz,J., Bonifacino,J.S., Yuan,L.C., and Klausner,R.D. (1988). Degradation from the endoplasmic reticulum: disposing of newly synthesized proteins. *Cell* 54, 209-220.
- Liu,K., Cho,H.S., Lashuel,H.A., Kelly,J.W., and Wemmer,D.E. (2000). A glimpse of a possible amyloidogenic intermediate of transthyretin. *Nature Structural Biology* 7, 754-757.

- Lu, J., Kobertz, W.R., and Deutsch, C. (2007). Mapping the electrostatic potential within the ribosomal exit tunnel. *J. Mol. Biol.* *371*, 1378-1391.
- Ma, H. and Gruebele, M. (2005). Kinetics are probe-dependent during downhill folding of an engineered lambda6-85 protein. *Proc. Natl. Acad. Sci. U. S. A* *102*, 2283-2287.
- Mains, P.E. and Sibley, C.H. (1983). The requirement of light chain for the surface deposition of the heavy chain of immunoglobulin M. *J. Biol. Chem.* *258*, 5027-5033.
- Maki, R., Traunecker, A., Sakano, H., Roeder, W., and Tonegawa, S. (1980). Exon shuffling generates an immunoglobulin heavy chain gene. *Proc. Natl. Acad. Sci. U. S. A* *77*, 2138-2142.
- Marsh, J.A., Neale, C., Jack, F.E., Choy, W.Y., Lee, A.Y., Crowhurst, K.A., and Forman-Kay, J.D. (2007). Improved structural characterizations of the drkN SH3 domain unfolded state suggest a compact ensemble with native-like and non-native structure. *J. Mol. Biol.* *367*, 1494-1510.
- Matouschek, A., Kellis, J.T., Jr., Serrano, L., and Fersht, A.R. (1989). Mapping the transition state and pathway of protein folding by protein engineering. *Nature* *340*, 122-126.
- Matouschek, A., Serrano, L., and Fersht, A.R. (1992a). The folding of an enzyme. IV. Structure of an intermediate in the refolding of barnase analysed by a protein engineering procedure. *J. Mol. Biol.* *224*, 819-835.
- Matouschek, A., Serrano, L., Meiering, E.M., Bycroft, M., and Fersht, A.R. (1992b). The folding of an enzyme. V. H/2H exchange-nuclear magnetic resonance studies on the folding pathway of barnase: complementarity to and agreement with protein engineering studies. *J. Mol. Biol.* *224*, 837-845.
- Mayor, U., Guydosh, N.R., Johnson, C.M., Grossmann, J.G., Sato, S., Jas, G.S., Freund, S.M.V., Alonso, D.O.V., Daggett, V., and Fersht, A.R. (2003). The complete folding pathway of a protein from nanoseconds to microseconds. *Nature* *421*, 863-867.
- McCracken, A.A. and Brodsky, J.L. (1996). Assembly of ER-associated protein degradation in vitro: dependence on cytosol, calnexin, and ATP. *J. Cell Biol.* *132*, 291-298.
- Melnick, J., Dul, J.L., and Argon, Y. (1994). Sequential interaction of the chaperones BiP and GRP94 with immunoglobulin chains in the endoplasmic reticulum. *Nature* *370*, 373-375.
- Merchant, K.A., Best, R.B., Louis, J.M., Gopich, I.V., and Eaton, W.A. (2007). Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. *Proc. Natl. Acad. Sci. U. S. A* *104*, 1528-1533.
- Meunier, L., Usherwood, Y.K., Chung, K.T., and Hendershot, L. (2002). A sub-set of chaperones and folding enzymes form multi-protein complexes in the ER to bind nascent proteins. *Molecular Biology of the Cell* *12*, 488A.
- Miller, S., Lesk, A.M., Janin, J., and Chothia, C. (1987). The accessible surface area and stability of oligomeric proteins. *Nature* *328*, 834-836.
- Mizuguchi, M., Kroon, G.J., Wright, P.E., and Dyson, H.J. (2003). Folding of a beta-sheet protein monitored by real-time NMR spectroscopy. *Journal of Molecular Biology* *328*, 1161-1171.
- Moglich, A., Joder, K., and Kiefhaber, T. (2006). End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation. *Proc. Natl. Acad. Sci. U. S. A* *103*, 12394-12399.
- Mok, K.H., Kuhn, L.T., Goetz, M., Day, I.J., Lin, J.C., Andersen, N.H., and Hore, P.J. (2007). A pre-existing hydrophobic collapse in the unfolded state of an ultrafast folding protein. *Nature* *447*, 106-109.
- Mori, K., Sant, A., Kohno, K., Normington, K., Gething, M.J., and Sambrook, J.F. (1992). A 22 bp cis-acting element is necessary and sufficient for the induction of the yeast KAR2 (BiP) gene by unfolded proteins. *EMBO J.* *11*, 2583-2593.

- Morrison, S.L. (1978). Murine heavy chain disease. *Eur. J. Immunol.* **8**, 194-199.
- Mucke, M. and Schmid, F.X. (1994). Intact Disulfide Bonds Decelerate the Folding of Ribonuclease-T1. *Journal of Molecular Biology* **239**, 713-725.
- Munro, S. and Pelham, H.R. (1987). A C-terminal signal prevents secretion of luminal ER proteins. *Cell* **48**, 899-907.
- Nordlund, A., Leinartaitė, L., Saraboji, K., Aisenbrey, C., Grobner, G., Zetterstrom, P., Danielsson, J., Logan, D.T., and Oliveberg, M. (2009). Functional features cause misfolding of the ALS-provoking enzyme SOD1. *Proc. Natl. Acad. Sci. U. S. A* **106**, 9667-9672.
- Onuchic, J.N. and Wolynes, P.G. (2004). Theory of protein folding. *Curr. Opin. Struct. Biol.* **14**, 70-75.
- Onuchic, J.N., Wolynes, P.G., Luthey-Schulten, Z., and Socci, N.D. (1995). Toward an outline of the topography of a realistic protein-folding funnel. *Proc. Natl. Acad. Sci. U. S. A* **92**, 3626-3630.
- Ostrovsky, O., Makarewich, C.A., Snapp, E.L., and Argon, Y. (2009). An essential role for ATP binding and hydrolysis in the chaperone activity of GRP94 in cells. *Proc. Natl. Acad. Sci. U. S. A.*
- Pace, C.N., Grimsley, G.R., Thomson, J.A., and Barnett, B.J. (1988). Conformational Stability and Activity of Ribonuclease-T1 with Zero, One, and 2 Intact Disulfide Bonds. *Journal of Biological Chemistry* **263**, 11820-11825.
- Paci, E., Clarke, J., Steward, A., Vendruscolo, M., and Karplus, M. (2003). Self-consistent determination of the transition state for protein folding: Application to a fibronectin type III domain. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 394-399.
- Plaxco, K.W., Simons, K.T., and Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277**, 985-994.
- Ponstingl, H., Kabir, T., Gorse, D., and Thornton, J.M. (2005). Morphological aspects of oligomeric protein structures. *Prog. Biophys. Mol. Biol.* **89**, 9-35.
- Porter, R.R. (1950). The formation of a specific inhibitor by hydrolysis of rabbit antioalbumin. *Biochem. J.* **46**, 479-484.
- Qin, Z.J., Hu, D.M., Zhu, M., and Fink, A.L. (2007). Structural characterization of the partially folded intermediates of an immunoglobulin light chain leading to amyloid fibrillation and amorphous aggregation. *Biochemistry* **46**, 3521-3531.
- Ramm, K., Gehrig, P., and Pluckthun, A. (1999). Removal of the conserved disulfide bridges from the scFv fragment of an antibody: effects on folding kinetics and aggregation. *J. Mol. Biol.* **290**, 535-546.
- Religa, T.L., Markson, J.S., Mayor, U., Freund, S.M.V., and Fersht, A.R. (2005). Solution structure of a protein denatured state and folding intermediate. *Nature* **437**, 1053-1056.
- Ron, D. and Walter, P. (2007). Signal integration in the endoplasmic reticulum unfolded protein response. *Nat. Rev. Mol. Cell Biol.* **8**, 519-529.
- Roth, R.A. and Pierce, S.B. (1987). In vivo cross-linking of protein disulfide isomerase to immunoglobulins. *Biochemistry* **26**, 4179-4182.
- Rougon, G. and Hobert, O. (2003). New insights into the diversity and function of neuronal immunoglobulin superfamily molecules. *Annu. Rev. Neurosci.* **26**, 207-238.
- Sadqi, M., Fushman, D., and Munoz, V. (2006). Atom-by-atom analysis of global downhill protein folding. *Nature* **442**, 317-321.
- Sakaguchi, N. and Melchers, F. (1986). Lambda 5, a new light-chain-related locus selectively expressed in pre-B lymphocytes. *Nature* **324**, 579-582.

- Sanchez,I.E. and Kiefhaber,T. (2003a). Hammond behavior versus ground state effects in protein folding: evidence for narrow free energy barriers and residual structure in unfolded states. *J. Mol. Biol.* **327**, 867-884.
- Sanchez,I.E. and Kiefhaber,T. (2003b). Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *Journal of Molecular Biology* **325**, 367-376.
- Scalley-Kim,M. and Baker,D. (2004). Characterization of the folding energy landscapes of computer generated proteins suggests high folding free energy barriers and cooperativity may be consequences of natural selection. *J. Mol. Biol.* **338**, 573-583.
- Schuler,B., Lipman,E.A., and Eaton,W.A. (2002). Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* **419**, 743-747.
- Seligmann,M., Mihaesco,E., Preud'homme,J.L., Danon,F., and Brouet,J.C. (1979). Heavy chain diseases: current findings and concepts. *Immunol. Rev.* **48**, 145-167.
- Serrano,L., Kellis,J.T., Jr., Cann,P., Matouschek,A., and Fersht,A.R. (1992a). The folding of an enzyme. II. Substructure of barnase and the contribution of different interactions to protein stability. *J. Mol. Biol.* **224**, 783-804.
- Serrano,L., Matouschek,A., and Fersht,A.R. (1992b). The folding of an enzyme. III. Structure of the transition state for unfolding of barnase analysed by a protein engineering procedure. *J. Mol. Biol.* **224**, 805-818.
- Serrano,L., Matouschek,A., and Fersht,A.R. (1992c). The folding of an enzyme. VI. The folding pathway of barnase: comparison with theoretical models. *J. Mol. Biol.* **224**, 847-859.
- Shaffer,A.L. and Schliessel,M.S. (1997). A truncated heavy chain protein relieves the requirement for surrogate light chains in early B cell development. *J. Immunol.* **159**, 1265-1275.
- Sherman,E. and Haran,G. (2006). Coil-globule transition in the denatured state of a small protein. *Proc. Natl. Acad. Sci. U. S. A* **103**, 11539-11543.
- Simpson,E.R., Herold,E.M., and Buchner,J. (2009). The Folding Pathway of the Antibody V(L) Domain. *J. Mol. Biol.*
- Skowronek,M.H., Hendershot,L.M., and Haas,I.G. (1998). The variable domain of nonassembled Ig light chains determines both their half-life and binding to the chaperone BiP. *Proc. Natl. Acad. Sci U. S. A* **95**, 1574-1578.
- Smith,L.J., Fiebig,K.M., Schwalbe,H., and Dobson,C.M. (1996). The concept of a random coil. Residual structure in peptides and denatured proteins. *Fold. Des* **1**, R95-106.
- Spudich,G.M., Miller,E.J., and Marqusee,S. (2004). Destabilization of the Escherichia coli RNase H kinetic intermediate: switching between a two-state and three-state folding mechanism. *J. Mol. Biol.* **335**, 609-618.
- Thies,M.J., Talamo,F., Mayer,M., Bell,S., Ruoppolo,M., Marino,G., and Buchner,J. (2002). Folding and oxidation of the antibody domain C(H)3. *J. Mol. Biol.* **319**, 1267-1277.
- Thies,M.J.W., Mayer,J., Augustine,J.G., Frederick,C.A., Lillie,H., and Buchner,J. (1999). Folding and association of the antibody domain C(H)3: Prolyl isomerization precedes dimerization. *Journal of Molecular Biology* **293**, 67-79.
- Trombetta,E.S. and Helenius,A. (1998). Lectins as chaperones in glycoprotein folding. *Curr. Opin. Struct. Biol.* **8**, 587-592.
- Tsai,C.J., Sauna,Z.E., Kimchi-Sarfaty,C., Ambudkar,S.V., Gottesman,M.M., and Nussinov,R. (2008). Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima. *J. Mol. Biol.* **383**, 281-291.

- Udgaonkar, J.B. and Baldwin, R.L. (1988). NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. *Nature* **335**, 694-699.
- Uversky, V.N., Oldfield, C.J., and Dunker, A.K. (2005). Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J. Mol. Recognit.* **18**, 343-384.
- Vanhove, M., Usherwood, Y.K., and Hendershot, L.M. (2001). Unassembled Ig heavy chains do not cycle from BiP in vivo but require light chains to trigger their release. *Immunity* **15**, 105-114.
- Vembar, S.S. and Brodsky, J.L. (2008). One step at a time: endoplasmic reticulum-associated degradation. *Nat. Rev. Mol. Cell Biol.* **9**, 944-957.
- Vendruscolo, M., Paci, E., Dobson, C.M., and Karplus, M. (2001). Three key residues form a critical contact network in a protein folding transition state. *Nature* **409**, 641-645.
- Vu, K.B., Ghahroudi, M.A., Wyns, L., and Muyldermans, S. (1997). Comparison of llama VH sequences from conventional and heavy chain antibodies. *Mol. Immunol.* **34**, 1121-1131.
- Wagner, C. and Kiefhaber, T. (1999). Intermediates can accelerate protein folding. *Proc. Natl. Acad. Sci. U. S. A* **96**, 6716-6721.
- Walter, S. and Buchner, J. (2002). Molecular chaperones--cellular machines for protein folding. *Angew. Chem. Int. Ed Engl.* **41**, 1098-1113.
- Watters, A.L., Deka, P., Corrent, C., Callender, D., Varani, G., Sosnick, T., and Baker, D. (2007). The highly cooperative folding of small naturally occurring proteins is likely the result of natural selection. *Cell* **128**, 613-624.
- Wiens, G.D., Roberts, V.A., Whitcomb, E.A., O'Hare, T., Stenzel-Poore, M.P., and Rittenberg, M.B. (1998). Harmful somatic mutations: lessons from the dark side. *Immunol. Rev.* **162**, 197-209.
- Williams, A.F. and Barclay, A.N. (1988). The immunoglobulin superfamily--domains for cell surface recognition. *Annu. Rev. Immunol.* **6**, 381-405.
- Yang, Z.R., Thomson, R., McNeil, P., and Esnouf, R.M. (2005). RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics.* **21**, 3369-3376.
- Yeh, I.C. and Hummer, G. (2002). Peptide loop-closure kinetics from microsecond molecular dynamics simulations in explicit solvent. *J. Am. Chem. Soc.* **124**, 6563-6568.
- Zack, D.J., Morrison, S.L., Cook, W.D., Dackowski, W., and Scharff, M.D. (1981). Somatic generated mouse myeloma variants synthesizing IgA half-molecules. *J. Exp. Med.* **154**, 1554-1569.
- Zarrine-Afsar, A., Larson, S.M., and Davidson, A.R. (2005). The family feud: do proteins with similar structures fold via the same pathway? *Curr. Opin. Struct. Biol.* **15**, 42-49.
- Zhang, G., Hubalewska, M., and Ignatova, Z. (2009). Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat. Struct. Mol. Biol.* **16**, 274-280.

V. Published scientific papers

Rate of Loop Formation in Peptides: A Simulation Study

Matthias J. Feige¹ and Emanuele Paci^{2*}

¹Department Chemie,
Technische Universität
München, 85747 Garching,
Germany

²Physics and Astronomy,
Institute of Molecular and
Cellular Biology and Astbury
Centre for Structural Molecular
Biology, University of Leeds,
Leeds LS2 9JT, UK

Received 18 April 2008;
received in revised form
2 July 2008;
accepted 3 July 2008
Available online
11 July 2008

Experimental techniques with high temporal and spatial resolution extend our knowledge of how biological macromolecules self-organise and function. Here, we provide an illustration of the convergence between simulation and experiment made possible by techniques such as triplet–triplet energy transfer and fluorescence quenching with long-lifetime and fast-quenching fluorescent probes. These techniques have recently been used to determine the average time needed for two residues in a peptide or protein segment to form a contact. The timescale of this process is accessible to computer simulation, providing a microscopic interpretation of the data and yielding new insight into the disordered state of proteins. Conversely, such experimental data also provide a test of the validity of alternative choices for the molecular models used in simulations, indicating their possible deficiencies. We carried out simulations of peptides of various composition and length using several models. End-to-end contact formation rates and their dependence on peptide length agree with experimental estimates for some sequences and some force fields but not for others. The deviations are due to artefactual structuring of some peptides, which is not observed when an atomistic model for the solvation water is used. Simulations show that the observed experimental rates are compatible with considerably different distributions of the end-to-end distance; for realistic models, these are never Gaussian but indicative of a rugged energy landscape.

© 2008 Elsevier Ltd. All rights reserved.

Edited by D. Case

Keywords: loop formation; protein folding; peptide dynamics; molecular dynamics simulation

Introduction

Molecular dynamics (MD) simulations in principle provide information on proteins and other biological macromolecules that is unrivalled in spatial and temporal resolution. However, only short timescales are accessible due to limitations on the speed of computers. This prevents, for example, the direct simulation of given amino acid sequences folding to their stable biologically active conformation and also limits the extent to which the different force fields generally used in simulations can be evaluated by direct comparison to experi-

ments. Proteins usually adopt their native state within milliseconds to seconds under ambient conditions; this is orders of magnitude away from the nanosecond regime accessible to MD simulations of “large” systems such as even a single protein in its aqueous environment.¹ An essential step in a protein folding reaction is the formation of a contact between two residues in the folding chain, which determines the closure of loops or formation of turns; these events occur on a submicrosecond timescale.^{2,3} The formation of contacts, whether native or not, dramatically reduces the fraction of conformations that have to be explored by a folding polypeptidic chain and is of fundamental importance in shaping the free-energy landscape of a foldable polypeptide sequence.^{4–6} The formation of contacts in an unfolded chain (depending on the length of the chain and the separation in sequence between the pair of residues) takes place on a timescale directly accessible to MD simulations.

A variety of experimental approaches have been developed to characterise these basic folding reactions.³ A particularly promising one is based

*Corresponding author. E-mail address: e.paci@leeds.ac.uk.

Abbreviations used: MD, molecular dynamics; TTET, triplet–triplet energy transfer; SASA, solvent-accessible surface area; ORI, only repulsive interactions; DBO, 2,3-diazabicyclo[2.2.2]oct-2-ene; EEF1, effective energy function 1; ACE, analytical continuum electrostatics; GBSW, generalised Born with simple switching.

on triplet–triplet energy transfer (TTET).² In TTET experiments, the donor is excited by a laser pulse and is able to transfer its triplet state to a suitable acceptor within a few picoseconds, but only when the donor and acceptor are in van der Waals contact,² so that donor quenching is limited by contact formation rather than by reaction time. TTET has been successfully applied to end-to-end contact formation in unstructured peptides from short to intermediate length and from picoseconds to nanoseconds under a variety of conditions.^{2,7–10} Another approach to measuring contact formation close to the diffusion limit has recently been proposed;^{11,12} it makes use of the rapid quenching of a long-lifetime fluorophore [2,3-diazabicyclo[2.2.2]oct-2-ene (DBO)] by tryptophan. A similar technique, in which a tryptophan at one end of the peptide is optically excited into a triplet state that is quenched upon contact formation with a cysteine at the other end, has also been applied for the same purpose,¹³ however, as molecular simulation clearly demonstrated,¹⁴ triplet quenching by cysteine is mainly reaction-limited and, therefore, does not allow direct measurement of contact formation kinetics.

Such experimental data can provide direct insight into the dynamics of an unfolded polypeptide chain. The conformational heterogeneity and the dynamics by which the conformation space can be explored play an essential role in the folding of a polypeptide chain. To date, most MD studies of peptides have focused on structured peptides, aiming at the reproduction of the experimentally observed structure (see, e.g., Ref. ¹⁵ for a review) and neglecting to a certain extent the role the unfolded state plays in a protein folding reaction. However, neglect of the denatured state limits the utility of such simulations for predicting the stability of the native state and the rates of folding and unfolding, because the thermodynamics of the system as a whole depends fundamentally on the properties of both the native and the denatured states. Therefore, it remains essential to learn how best to simulate the dynamics of unfolded peptides and to validate these simulations against appropriate experimental data.

We carried out equilibrium simulations of peptides of various composition and length using several force fields, including implicit (i.e., a continuum model) and explicit (i.e., an atomistic model) solvation; beside such “realistic” models, we also used a model in which only the covalent bonds and the repulsive part of the van der Waals interactions have been retained. Polyserine, S_x , and poly(glycine–serine), $(GS)_x$, which have been systematically characterised with TTET, have been chosen as models in our studies. Furthermore, we studied $Y(GS)_x$ FSG peptides of different lengths in which the aromatic residues have been incorporated to mimic the donor–acceptor pair used in the TTET experiments. Our simulations provide a microscopic interpretation of the experimental data, yielding new insight into the disordered state of proteins. Conversely, the experimental data provide a new test of the validity and performance of the alter-

native molecular models used in the simulations, indicating deficiencies and strengths of each of these models.

Theory and Models

Implicit solvent simulations

Implicit solvent simulations have been carried out using different force fields. In one [SASA (solvent-accessible surface area)],¹⁶ the polar and nonpolar contributions of each atom to the free energy of solvation are assumed to be proportional to their solvent-accessible surface areas approximated with a simple analytical function; the electrostatic screening between solute charges is accounted for by a dielectric constant $\epsilon = 2r$, where r is the distance between charges, and by neutralising the formal charges (Asp, Glu, Arg, Lys, and the termini) as in the EEF1 (effective energy function 1) model.¹⁷ The EEF1 model¹⁷ is similar to SASA but uses excluded volumes instead of surfaces and a dielectric constant $\epsilon = r$ (i.e., charge screening is weaker than that in SASA). In ACE (analytical continuum electrostatics),¹⁸ the electrostatic solvation free-energy and the nonpolar (i.e., nonelectrostatic) solvation free-energy contributions are calculated separately. The electrostatic contribution is an analytical approximation to the solution of the Poisson equation; the nonpolar solvation free energy is approximated by a pairwise potential that yields results similar to the surface area approximation solvation energy; the version used here (ACE2) takes account of the overestimation of the desolvation of charges by the pairwise descreening potential. GBSW (generalised Born with simple switching)¹⁹ uses a generalised Born method for the electrostatic part and the solvent-exposed surface area for the nonpolar part with a phenomenological surface tension coefficient; the self-electrostatic solvation energy is calculated using a simple smoothing function at the dielectric boundary.

While SASA, EEF1, and ACE use a protein model in which nonpolar hydrogens are not explicitly included, all hydrogens are present with GBSW. Additional simulations have been performed with an all-atom model in which all the attractive interactions have been removed [ORI (only repulsive interactions)]. Langevin dynamics simulations were performed with the programs CHARMM²⁰ at 300 K, using the Brooks–Brunger–Karplus algorithm.²¹ Simulations with an alternative integrator²² have also been performed in a few cases and gave identical results. Simulations have been performed with a friction coefficient of 1 ps^{-1} . This is lower than the friction coefficient corresponding to the viscosity of water by a factor of ~ 60 at 300 K†. Since the Kramers theory^{23,24} predicts linear dependence of rates on

† www.iapws.org/relguide/visc.pdf

friction coefficient in the limit of large friction, meaningful rates could be determined in this way while economising a factor of ~ 60 of the simulation time. For a few peptides, a broad range of frictions was explored to precisely extract the rate of contact formation in a medium of viscosity comparable to that of water and, thus, test the validity of Kramers' prediction in our system. Simulation time ranged from 0.25 to 2 μs , depending on the peptide length. Full convergence of the average contact formation time and the distribution of the end-to-end distance was assessed by comparing results from the first and second half of each trajectory.

Explicit solvent simulations

Explicit solvent simulations were performed using GROMACS²⁵ and the OPLS-AA force field²⁶ in TIP4P water²⁷ at 300 K. A cubic water box with a minimum distance of 13 Å to the ends of the extended initial peptide conformation was used.²⁸ Electrostatic interactions were calculated using particle mesh Ewald. Bond lengths were constrained using the LINCS algorithm.²⁹ Both the temperature and pressure were maintained close to 300 K and 1 atm, respectively, using the weak coupling algorithm³⁰ ($\tau_T=0.1$ ps and $\tau_P=1$ ps). Explicit solvent simulations converge much slower than implicit solvent ones. All explicit solvent simulations were sufficiently long to observe relaxation from the initial extended conformation and continued for a time ranging between 160 ns for S4, GS₂, and GS₄ and 2.5 ns for GS₁₆. Contact formation times have only been estimated for peptides up to 12 residues long, where simulations at least 10 times longer than the average contact time could be performed. To compare contact formation rates with experimental ones, we estimated the diffusion coefficient of the TIP4P water at 300 K temperature and 1 atm pressure by simulating 1000 water molecules for 20 ns; this turns out to be $D_{\text{TIP4P}}=(3.93\pm 0.04) 10^{-9} \text{ m}^2/\text{s}$, as compared to $D_{\text{exp}}=(2.3\pm 0.04) 10^{-9} \text{ m}^2/\text{s}$;³¹ the reciprocal of the ratio (~ 1.7) provides an estimate for the corresponding ratio of viscosities (based on the Stokes–Einstein relation).

End-to-end distance and contact formation rates

In the TTET experiment,⁷ xanthone was used as donor and 1-naphthyl alanine was used as acceptor. In the simulations, contact formation rates were determined between the C- and N-terminal amino acid for the (GS)_x and S_x peptides, whereas in the case of the Y(GS)_xFSG peptide, the Tyr and the Phe moieties were defined as donor and acceptor, respectively. The contact distance was calculated as the minimum distance between any non-hydrogen side-chain atom of the defined acceptor and donor. In the case of glycine residues, the C^α atom was used instead of the side-chain atoms. Contact formation was assumed when the distance between donor and acceptor is less than 4 Å. This takes the side-chain hydrogens and the necessary van der Waals contact

for TTET into account. Contact formation rates were calculated as

$$k^{-1} = \int_0^{\infty} S(t) dt \quad (1)$$

where $S(t)$ is the survival probability of the triplet state (i.e., the probability that the two ends are not in contact). If the decay is single exponential, $S(t)=e^{-t/\tau}$, $k^{-1}=\tau$. The original TTET experiments⁷ used a nanosecond laserflash to excite the triplet state; the relaxation appeared to be single exponential for any time above the dead time of the experiment. A more recent femtosecond laser experiment¹⁰ showed that, for short times, the relaxation is not single exponential, as predicted by a theory for a simple polymer model.³² The distribution of survival times of the triplet state displays a long time exponential decay and a nonexponential short-time decay, which fits well to a power law as the theory predicts. Despite a nonexponential distribution, the average contact formation time is still very close to the time constant extracted from the exponential long-time part of the distribution.

Results

Most of the simulations in this work have been performed with a simple, computationally efficient implicit solvent model where the solvation effect is based on the solvent-accessible surface (SASA, see Theory and Models). This model allowed a broad and rigorous exploration of different peptide lengths [from 4 to 52 amino acids for each of the (GS)_x, S_x, and Y(GS)_xFSG sequences]. We used uncapped peptides; simulations with acetylated N-terminus showed that this does not significantly influence the rates or the distributions. For some peptides only, simulations were carried out with three alternative implicit solvents (GBSW, ACE, and EEF1, see Theory and Models) as well as an explicit solvent representation (OPLS/AA+TIP4P).

The dynamics of polypeptides

Figure 1 shows the dependence of the average time (or inverse rate) of contact formation as a function of the number of peptide bonds for the three different sequences. For both an implicit (SASA) and explicit solvent, qualitative agreement between experimentally determined and computed values is observed for (GS)_x and Y(GS)_xFSG. Since implicit solvent simulations were carried out at a low friction compared to water, absolute rates differ. At the friction of water ($\sim 60 \text{ ps}^{-1}$, empty black circles in Fig. 1), the agreement with the experiment is excellent (the whole friction dependence of the rates for some peptides is shown in Fig. 2). Experimental TTET⁷ contact times fit the equation $\tau=\tau_0+\tau_1 N^m$ (solid magenta lines in Fig. 1), with $m=1.72$ for (GS)_x. The exponent m characterises the asymptotic behaviour for long chains and is 1.5 for

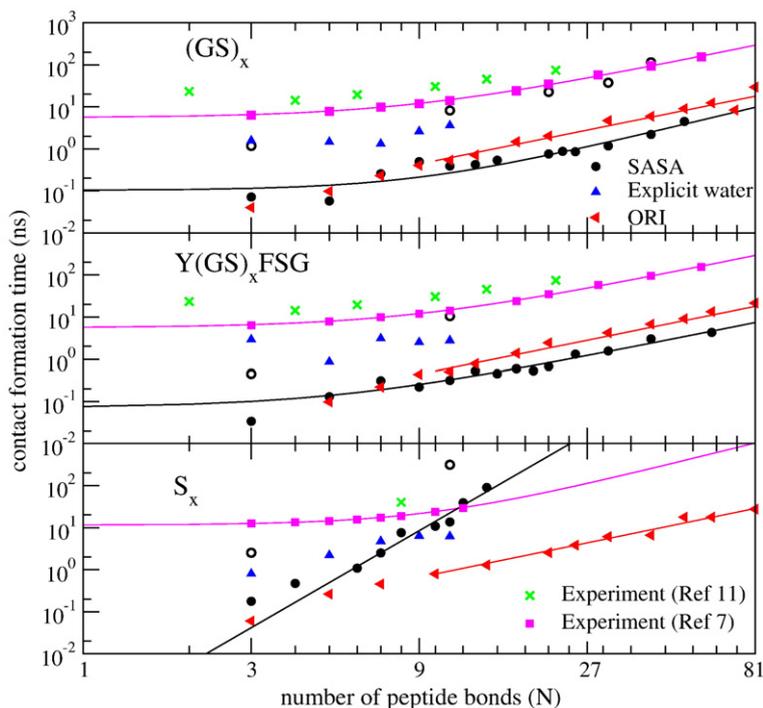


Fig. 1. End-to-end contact formation times (inverse rates) in implicit solvent (SASA) at a friction coefficient $\gamma=1 \text{ ps}^{-1}$ (black circles) and extrapolated at water-like friction (black empty circles). Blue triangles represent the result for the explicit solvent simulations. Red (left-pointing) triangles correspond to the result of the ORI simulations. TET experimental data⁷ are shown as magenta squares; the fluorescence lifetimes for DBO/Trp-containing polypeptides^{11,12} are shown as green crosses. Continuous lines are the fits to $\tau = \tau_0 + \tau_1 N^m$.

an ideal freely jointed chain,³³ larger values are expected when excluded volume effects are present.³³ From SASA simulations, we obtain $m=2.0$ for $(GS)_x$ and $m=1.7$ for $Y(GS)_x \text{FSG}$; these values are close to the experimental estimation although both are potentially subject to error because much longer peptides cannot be studied either *in vitro* due to limitations in peptide synthesis⁷ or *in silico* due to the required computation time, which is exponential in the peptide length.

The value τ_0 represents the limiting contact formation time for short peptides. For $(GS)_x$ and $Y(GS)_x \text{FSG}$, the fits give 0.07 and 0.1 ns, respectively;

corrected by a factor of 60, this gives a value of $\tau_0 \sim 4.2\text{--}6$ ns as compared with the experimental value of $\tau_0 = 5.6$ ns. For a few peptides, we performed simulations at water-like friction and direct agreement for τ_0 is found between simulation and experiment; for $Y(GS)_5 \text{FSG}$, we obtain a contact formation time of 7 ± 1 ns (see Fig. 2) in agreement with the experiment for $(GS)_5$. Overall, for peptides containing the GS repeat, agreement between contact formation times estimated by implicit solvent simulation and experiment is excellent. In explicit solvent, while fewer peptides were studied, the agreement with the experiment is good when

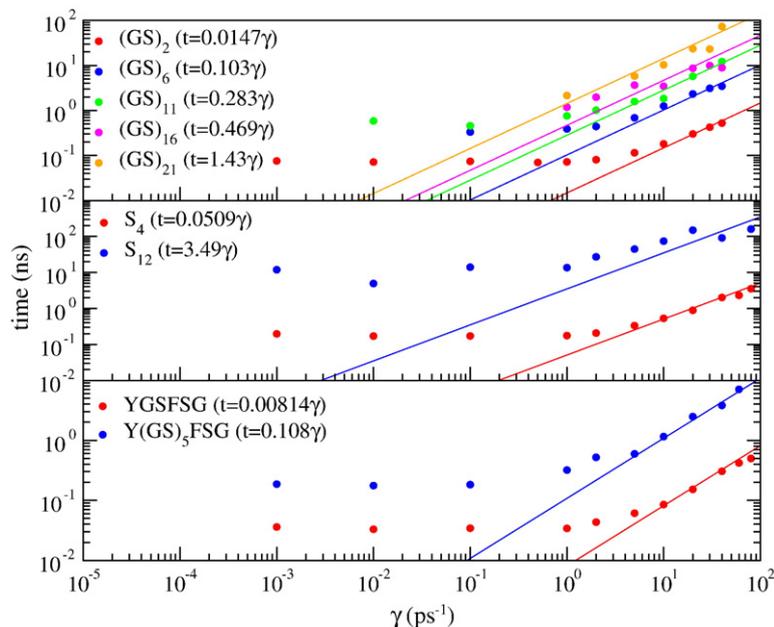


Fig. 2. End-to-end contact formation times in implicit solvent (SASA) as a function of the friction coefficient γ . This dependence is linear (continuous lines) for large friction ($\gamma \geq 10$), as expected from the Kramers theory.^{23,24}

corrected for the lower viscosity of the TIP4P water model relative to real water (by a factor of ~ 1.7 , see Theory and Models).

The case of S_x is different. From TTET experiments, a very similar behaviour was found to that of the $(GS)_x$ series, that is, a similar functional dependence with $m=2.1$ and $\tau_0=12$ ns, reflecting a slight increase in rigidity. However, in simulations of S_x in implicit solvent (at least with the SASA model), we obtain $m\approx 4.8$, which reveals an unusually high rigidity, due to artefactual formation of persistent structure in the simulations (discussed further below). This behaviour is not observed for S_x in explicit solvent where the computed values are always close to the experimental ones, at least for peptides of lengths up to 12 residues (the maximum we simulated with explicit solvent).

Figure 1 also shows the result for a simpler model of the peptides in which only covalent bonds and steric interactions are retained (ORI). For large N , we observe the expected power-law behaviour with exponents $m=1.7$ for $(GS)_x$ and $Y(GS)_x$ FSG and $m=1.9$ for S_x ; for small N , no plateau is evident for short peptides, suggesting that the rigidity of short peptides depends, in part, on attractive interactions; this is in line with the experimental observation⁷ that in 8 M GdmCl, the onset of length-independent rate occurs at shorter distances, indicating decreased chain stiffness in 8 M GdmCl compared to water.

Thus, both the overall experimental kinetics of contact formation and the small increased rigidity of S_x relative to $(GS)_x$ can be reproduced well by a simple steric model; indeed, the latter, for S_x , performs much better than an implicit solvent model such as SASA.

Figure 1 also shows, as green crosses, fluorescence lifetimes for DBO/Trp-containing peptides determined by Nau *et al.*,^{11,12} these are about 2.5 times longer than both those measured by TTET⁷ and calculated (at water-like friction) in the present simulations. Interestingly, Nau *et al.* observed an increase in contact formation time for the shortest peptide studied,^{11,12} as predicted by the theory.³⁴ Ignoring the shortest peptide, the looping time scales as $\tau=\tau_0+\tau_1N^m$, where $m\approx 1$; such value may reflect an important role of hydrodynamic interactions, which are neglected in our implicit solvent simulations and which have been shown by theory and simulation to reduce the exponent relative to the ideal case.³⁵ However, given the similarity of such times with those determined by TTET and simulation (see Fig. 1), it seems more likely that the problem is that small errors in the times for short peptides affect enormously the asymptotic behaviour for long ones.

End-to-end distance distributions

Unlike in experiments, end-to-end distance distributions are readily available from the simulations. This provides information on the conformational properties of the different peptides under simulation, including the differences between force fields,

and helps to clarify the discrepancy between the experimentally derived and computed contact formation rates for S_x observed with an implicit solvent model, as mentioned above. The distributions of the end-to-end distances for $(GS)_x$, $Y(GS)_x$ FSG, and S_x for various different lengths are shown in Fig. 3 for three different implicit solvents, explicit solvent, and ORI. Some general tendencies emerge. For the pure steric model (ORI), a skewed Gaussian distribution, $p(r)=cr^2\exp(-b(r-a)^2)$ with $a=AN^\alpha$ where $A\approx 5.3$ and $\alpha\approx 0.44$ and with $b\approx BN^{-\beta}$ where $B\approx 0.41$ and $\beta\approx 1.5$ [N is the number of peptide bonds and $c=c(a, b)$ a normalisation constant], is observed for all peptide sequences and length. This type of end-to-end distribution has often been assumed to underlie experimental data.^{36–38} However, a clear deviation from the skewed Gaussian distribution is observed for all more realistic force fields. In shorter peptides, the distributions show several peaks and are rather similar for different models due to the intrinsic stiffness of the peptidic chain as the determining factor. For longer peptides, all implicit solvent representations favour more compact conformations than the explicit solvent model. A significant peak at the donor–acceptor contact distance is found for the $Y(GS)_x$ FSG peptide with all force fields. This likely reflects hydrophobic interactions of the aromatic donor–acceptor pair. The area under the first peak, that is, the fraction of peptides with donor and acceptor in contact, decreases with the peptide length, and for $(GS)_{16}$, it is about 15%, in agreement with that found experimentally, $\sim 10\%$.⁷

In some cases, we observed a pronounced difference between random-coil behaviour and computed distance distributions. In the case of the GBSW force field and for all sequences, the difference can likely be attributed to an overestimation of electrostatic interactions; for example, for S_8 , the peak at ~ 5.4 Å is due to structures where N-terminal NH_3^+ and the C-terminal COO^- groups are in contact; the peak at ~ 10.4 Å is due to structures where COO^- is in contact to backbone NH amide groups or NH_3^+ in contact with CO groups or Ser oxygen atoms. The peculiar behaviour observed for S_x with the SASA implicit solvent is due to a strong helical propensity, as revealed by a peak at end-to-end distances $\sim x\times 1.4$ Å.

Structure formation

To rationalise the differences in the distance distribution for different force fields, solvent representation, and different peptides on a structural basis, we also analysed hydrogen bonds and secondary-structure formation.

In all cases, the average number of intramolecular hydrogen bonds present depends linearly on the peptide length (Fig. 4). Nevertheless, the slope differs for different models. For all implicit solvation models (SASA, GBSW, ACE, and EEF1), the number of hydrogen bonds increases by ~ 0.35 with each residue. In Fig. 4, the result is shown for $(GS)_x$. For S_x (not shown), ACE and GBSW display the

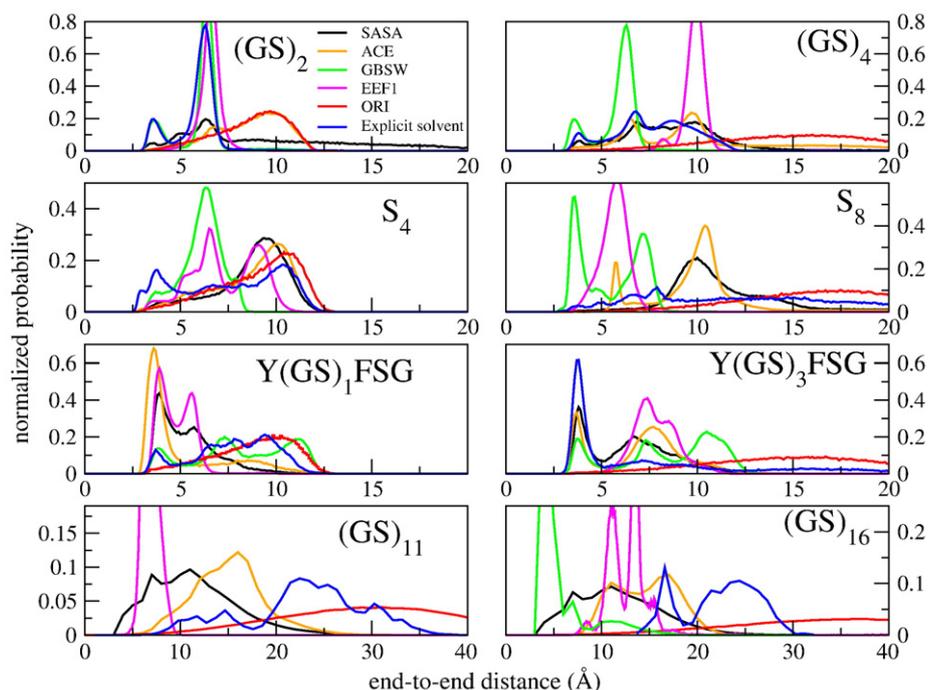


Fig. 3. The donor–acceptor distance distributions for the three peptides (see Theory and Models for the definition of such distances). Results for four different implicit solvent models (SASA, ACE, EEf1, and GBSW), explicit solvent, and a model where only repulsive interactions are retained (ORI).

same dependence as for $(GS)_x$, while with SASA, the proportionality factor turns out to be ~ 0.8 , consistent with the overestimation of the helical propensity mentioned above. In the case of explicit solvent, the number of intrapeptide hydrogen bonds increases much more weakly, at a rate of 0.18 hydrogen bonds per residue. In summary, all implicit solvent models used here seem to overestimate the intramolecular hydrogen-bonding propensity relative to explicit solvent simulations.

The secondary-structure analysis provides additional insight into the differences between force fields. With the explicit solvent model, the residual

secondary structure is low but not absent [e.g., for $(GS)_{11}$, about 4% of the residues are in either a helical or an extended conformation, according to DSSP³⁹]. Implicit solvent models suggest instead that there is more secondary structure: GBSW and ACE give predictions only slightly larger than the explicit solvent; SASA gives, also taking $(GS)_{11}$ as example, a helical and extended propensity of 17% and 15%, respectively; while the helical propensity decreases with the length, that of extended structure increases. For S_x the implicit solvents GBSW and ACE and the explicit solvent provide secondary structural propensity very similar to $(GS)_x$, while SASA predicts a

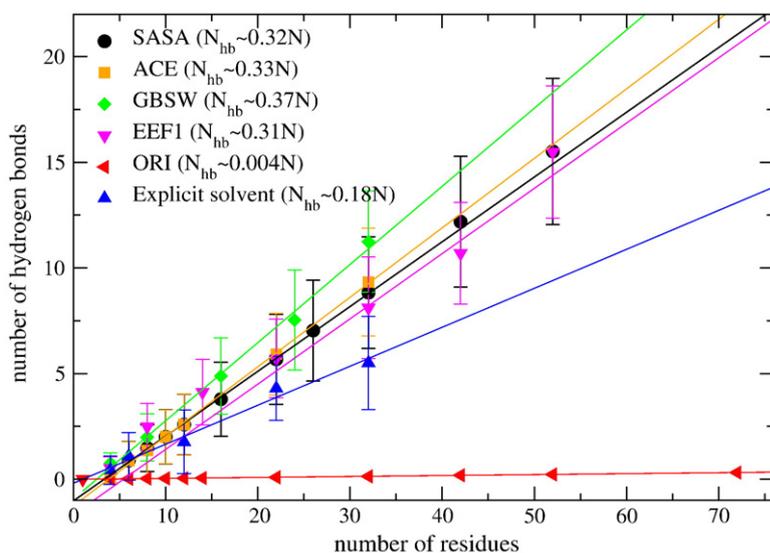


Fig. 4. The average number of intrapeptide hydrogen bonds increases linearly for the $(GS)_x$ peptides as a function of the peptide length for all force fields.

helical propensity of 87% for S_{13} , increasing for longer peptides and again indicating that artefactual helical structure is the source of the poor agreement with experimental data for this force field and peptide sequence.

Radius of gyration

The most unambiguous measurable characteristic of the state of a disordered polypeptide chain is its radius of gyration. This is shown for $(GS)_x$ peptides with different force fields in Fig. 5. The radius of gyration has a power-law dependence on the number of residues for all models, while the exponent depends on the molecular model.

This situation mirrors the experimental estimation of compactness of disordered peptides and proteins. Kohn *et al.*⁴⁰ recently analysed small-angle x-ray scattering data for 19 proteins spanning a broad range of size and found that under strongly denaturing conditions, the radius of gyration scales as $R_g = R_g^0 N^\nu$, where $\nu = 0.598$, in good agreement with the analytical result for a random coil with excluded volume effect ($\nu = 0.588$).⁴¹ In our simulations, the purely steric (ORI) model leads to $\nu = 0.68$, slightly larger than the recent experimental estimations but identical with the value reported by Tanford *et al.* using intrinsic viscosity measurement for 12 proteins.⁴² All implicit solvents give a considerably lower exponent ($\nu \sim 0.31\text{--}0.37$). This value is not far from the experimental value found for native proteins⁴³ and highlights overcompaction of unstructured peptides in implicit solvent. For explicit solvent simulations, we found an exponent $\nu \sim 0.5$, which is in agreement with that experimentally observed for natively unfolded proteins under native conditions.⁴³

While small-angle x-ray scattering data on the peptides studied here are not available, the root-mean-square end-to-end distance for $(GS)_{16}$ has been recently determined by FRET³⁸ to be 18.9 Å in water and 39.2 Å under highly denaturing con-

ditions (8 M GdmCl). Simulations with all implicit solvents provide values that are lower than the experimental value in native conditions (4.2 Å for GBSW, 12.1 Å for SASA, 14.5 Å for ACE, and 12.6 Å for EOF1). Simulation with a purely steric potential (ORI) gives a value (39.1 Å) that is close to the high denaturant one. In explicit solvent, we found 21.9 ± 3.4 Å (averaged over the last 2.5 ns of a 6.5-ns simulation). Thus, all implicit solvents estimate the two ends of the peptide to be closer on average than for the peptide in pure water. Conversely, a pure steric model gives a value close to that observed at high denaturant concentration. Explicit solvent gives a value intermediate between the two.

Discussion

The unprecedented resolution of recent measurements of the rate of loop formation in peptides has added important elements to our knowledge of the dynamical properties of disordered polypeptide systems. Here, we have shown how simulation is instrumental in interpreting and provoking such experiments. Implicit solvent simulations reproduce well the experimental rates of contact formation for some sequences. Nevertheless, comparison with explicit solvent simulations shows that a general trend of implicit solvent models is to overestimate the compactness, the degree of secondary structure, and the number of hydrogen bonds present. For some peptides, despite these artefacts, we were still able to obtain good estimates of the rates of contact formation with implicit solvents, explicit solvent, and even with the purely steric model. In a pioneering simulation study, Yeh and Hummer also observed that differences in distribution of the end-to-end distance for the pentapeptide CAGQW given by two different explicit solvents (AMBER and CHARMM) were also compatible with identical rates of contact formation.¹⁴ One possible reason is that contact formation is an activated process and

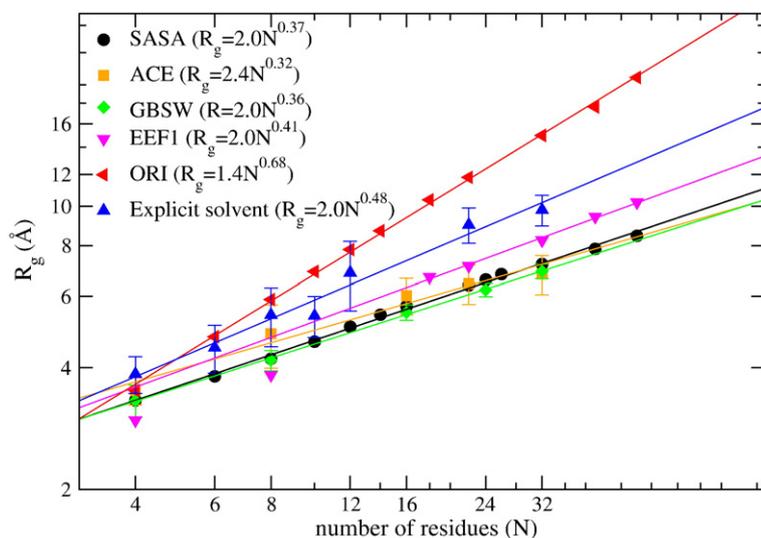


Fig. 5. The radius of gyration as a function of the number of residues follows a power-law $R_g = R_g^0 N^\nu$ for the $(GS)_x$ peptides for all force fields.

that the height of the main free-energy barrier is similar for the different force fields. However, this is certainly not the case for the steric model whose end-to-end distribution, a skewed Gaussian, shows no barriers between contact and noncontact and which itself provides a surprisingly accurate prediction of experimental rates of contact formation. This suggests that excluded volume and chain flexibility may be the main determinant of the kinetics of contact formation. Attractive or long-range interactions only have an effect on the rates of contact formation when they induce a state (e.g., the helical state for S_x seen above) segregated by a sizable free-energy barrier from all the other conformations.

Similarly, the effect of the hydrophobic residues Tyr and Phe (included here to model the triplet donor and acceptor) is negligible on the rates of contact formation but considerable on the distribution of end-to-end distance independent of the underlying force field. While a peak corresponding to conformations where the two ends are in contact is always observed, the peak is considerably larger when the hydrophobic groups are present.

The power-law relationship between the radius of gyration and the peptide length, with an exponent close to that of a random coil with excluded volume and close to that determined experimentally for proteins in strong denaturing conditions,⁴⁰ is reproduced only by a purely steric model. This suggests that ignoring all interactions except steric ones may represent proteins at highly denaturing conditions better than more sophisticated models of unstructured peptides.

Polypeptides that do not fold in native conditions are particularly interesting because they share properties with natively unfolded proteins. Natively unfolded proteins have an important biological role in phenomena such as transcription, signalling, and regulation;^{44,45} they constitute an important class of proteins, whose conformational properties can help elucidate much of what has not yet been understood on how structured proteins fold. The analysis of the hydrodynamic radius of natively unfolded proteins⁴³ showed that the radius of gyration has a power-law dependence on the number of residues; interestingly, proteins divide into two groups: one, identified as "coil-like," is characterised by an exponent ~ 0.5 , identical with our findings for explicit water. Intriguingly, for a second group, identified as "pre-molten-globule-like,"⁴³ the exponent is 0.41, closer to our results for all implicit solvents. A difference in hydrogen bond formation properties observed in the simulations with either implicit or explicit solvents may explain the origin of different groups of natively disordered proteins.

In conclusion, we found that the properties of naturally unstructured peptides calculated in computer simulations are sensitive to variations in the chosen force field; with implicit solvation models, the peptides are very compact, contain a considerable number of (generally short lived) hydrogen bonds, and are mostly unstructured. Interestingly, for a number of implicit solvents, such "odd" features do

not seem to be an obstacle to the simulation of the folding of peptides or miniproteins to their experimental structure.^{15,46,47} The scaling behaviour of the radius of gyration observed experimentally for structured proteins at high denaturant concentrations is reproduced only by a model in which all attractive interactions are switched off, while all more realistic models are far off. On the other hand, while scaling of the radius of gyration with implicit and explicit solvents is quite different, both behaviours have been observed for natively unfolded proteins in native conditions. Thus, while it is difficult to decide which model is most appropriate, it is certainly interesting to observe that all the models provide results that are within the spectrum of the experimental observations.

Apart from one exception, we observe highly disordered conformations with little or no persistent structure; however, this is not necessarily due to a Gaussian end-to-end distance distribution or to an absence of secondary structure and hydrogen bonds.⁴⁸

We note that all force fields utilised here have been optimised to accurately predict experimental structures for naturally folded proteins, yet they give markedly divergent results for unfolded peptides except in the case of end-to-end contact formation rate. This suggests that the ability of peptides under simulation to explore the accessible conformation space and fold to a specific structure may be relatively independent of their global behaviour in the denatured state (which we have shown, in this study, to be model dependent in some important aspects) but strongly dependent on rates of contact formation, in keeping with a model of the folding process dominated by diffusion and bifurcations in the folding pathway induced by closure of loops with resulting constraints on accessible conformation space. Finally, we observe that the different representation of the denatured state between models may correspond to different conditions in which proteins happen to fold, either *in vivo* or *in vitro*.

Acknowledgements

The authors thank M. Schaefer, R.L. Hayward, and L.R. Allen for helpful comments on the manuscript. M.J.F. acknowledges the Studienstiftung des Deutschen Volkes for a PhD studentship and E.P. acknowledges INTAS for financial support.

References

1. Karplus, M. & McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **9**, 646–652.
2. Bieri, O., Wirz, J., Hellrung, B., Schutkowski, M., Drewello, M. & Kiefhaber, T. (1999). The speed limit

- for protein folding measured by triplet-triplet energy transfer. *Proc. Natl Acad. Sci. USA*, **96**, 9597–9601.
3. Eaton, W. A., Muñoz, V., Hagen, S. J., Jas, G. S., Lapidus, L. J., Henry, E. R. & Hofrichter, J. (2000). Fast kinetics and mechanisms in protein folding. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 327–359.
 4. Karplus, M. & Weaver, D. L. (1994). Protein folding dynamics: the diffusion–collision model and experimental data. *Protein Sci.* **3**, 650–668.
 5. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. (1995). Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins: Struct. Funct. Genet.* **21**, 167–195.
 6. McLeish, T. C. (2005). Protein folding in high-dimensional spaces: hypergutters and the role of non-native interactions. *Biophys. J.* **88**, 172–183.
 7. Krieger, F., Fierz, B., Bieri, O., Drewello, M. & Kiefhaber, T. (2003). Dynamics of unfolded polypeptide chains as model for the earliest steps in protein folding. *J. Mol. Biol.* **332**, 265–274.
 8. Krieger, F., Möglich, A. & Kiefhaber, T. (2005). Effect of proline and glycine residues on dynamics and barriers of loop formation in polypeptide chains. *J. Am. Chem. Soc.* **127**, 3346–3352.
 9. Möglich, A., Krieger, F. & Kiefhaber, T. (2005). Molecular basis for the effect of urea and guanidinium chloride on the dynamics of unfolded polypeptide chains. *J. Mol. Biol.* **345**, 153–162.
 10. Fierz, B., Satzger, H., Root, C., Gilch, P., Zinth, W. & Kiefhaber, T. (2007). Loop formation in unfolded polypeptide chains on the picoseconds to microseconds time scale. *Proc. Natl Acad. Sci. USA*, **104**, 2163–2168.
 11. Hudgins, R. R., Huang, F., Gramlich, G. & Nau, W. M. (2002). A fluorescence-based method for direct measurement of submicrosecond intramolecular contact formation in biopolymers: an exploratory study with polypeptides. *J. Am. Chem. Soc.* **124**, 556–564.
 12. Huang, F. & Nau, W. M. (2003). A conformational flexibility scale for amino acids in peptides. *Angew. Chem., Int. Ed.* **42**, 2269–2272.
 13. Lapidus, L. J., Eaton, W. A. & Hofrichter, J. (2000). Measuring the rate of intramolecular contact formation in polypeptides. *Proc. Natl Acad. Sci. USA*, **97**, 7220–7225.
 14. Yeh, I. C. & Hummer, G. (2002). Peptide loop-closure kinetics from microsecond molecular dynamics simulations in explicit solvent. *J. Am. Chem. Soc.* **124**, 6563–6568.
 15. Gnanakaran, S., Nymeyer, H., Portman, J. J., Sanbonmatsu, K. Y. & Garcia, A. E. (2003). Peptide folding simulations. *Curr. Opin. Struct. Biol.* **13**, 168–174.
 16. Ferrara, P., Apostolakis, J. & Caflisch, A. (2002). Evaluation of a fast implicit solvent model for molecular dynamics simulations. *Proteins: Struct. Funct. Genet.* **46**, 24–33.
 17. Lazaridis, T. & Karplus, M. (1999). Effective energy function for protein dynamics and thermodynamics. *Proteins: Struct. Funct. Genet.* **35**, 133–152.
 18. Schaefer, M. & Karplus, M. (1996). A comprehensive treatment of continuum electrostatic. *J. Phys. Chem.* **100**, 1578–1599.
 19. Im, W., Feig, M. & Brooks, C. L. I. (2003). An implicit membrane generalized Born theory for the study of structure, stability, and interactions of membrane proteins. *Biophys. J.* **85**, 2900–2918.
 20. Brooks, B. R., Brucoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization and dynamics calculations. *J. Comp. Chem.* **4**, 187–217.
 21. Brunger, A., Brooks, C. L. & Karplus, M. (1984). Stochastic boundary conditions for molecular dynamics simulations of ST2 water. *Chem. Phys. Lett.* **105**, 495–500.
 22. Skeel, R. D. & Izaguirre, J. (2002). An impulse integrator for Langevin dynamics. *Mol. Phys.* **100**, 3885–3891.
 23. Kramers, H. A. (1940). Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica (Amsterdam)*, **7**, 284.
 24. Hänggi, P., Talkner, P. & Borkovec, M. (1990). Reaction-rate theory: fifty years after Kramers. *Rev. Mod. Phys.* **62**, 251–341.
 25. van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E. & Berendsen, H. J. C. (2005). GROMACS: fast, flexible and free. *J. Comp. Chem.* **26**, 1701–1718.
 26. Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **45**, 11225–11236.
 27. Jorgensen, W. J., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935.
 28. Darden, T. A., York, D. M. & Pedersen, L. (1993). Particle mesh Ewald: an $N \log(N)$ method for computing Ewald sums. *J. Chem. Phys.* **98**, 10089–10092.
 29. Hess, B., Bekker, H. & Berendsen, H. J. C. (1997). LINCS: a linear constraint solver for molecular simulations. *J. Comp. Chem.* **18**, 1463–1472.
 30. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690.
 31. Eisenberg, D. & Kauzmann, W. (1969). *The structure and properties of water*. Oxford University Press, New York.
 32. Likhtman, A. E. & Marques, C. M. (2006). First-passage problem for the Rouse polymer chain: an exact solution. *Europhys. Lett.* **75**, 971–977.
 33. Doi, M. & Edwards, S. F. (1986). *The theory of polymer dynamics*. Oxford Science Publications, Oxford, UK.
 34. Jun, S., Bechhoefer, J. & Ha, B. Y. (2003). Diffusion-limited loop formation of semiflexible polymers: Kramers theory and the intertwined time scales of chain relaxation and closing. *Europhys. Lett.* **64**, 420–426.
 35. Friedman, B. A. & Yeung, C. (2006). Renormalization group analysis of polymer cyclization with non-equilibrium initial conditions. *Eur. Phys. J., E Soft Matter*, **21**, 25–31.
 36. Haas, E., Wilchek, M., Katchalski-Katzir, E. & Steinberg, I. Z. (1975). Distribution of end-to-end distances of oligopeptides in solution as estimated by energy transfer. *Proc. Natl. Acad. Sci. USA*, **72**, 1807–1811.
 37. Haas, E., Katchalski-Katzir, E. & Steinberg, I. Z. (1978). Brownian motion of the ends of oligopeptide chains in solution as estimated by energy transfer between chain ends. *Biopolymers*, **17**, 11–31.
 38. Möglich, A., Joder, K. & Kiefhaber, T. (2006). End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation. *Proc. Natl. Acad. Sci. USA*, **103**, 12394–12399.
 39. Kabsch, W. & Sander, C. (1983). Dictionary of protein

- secondary structure: pattern recognition of hydrogen-bonded and geometric features. *Biopolymers*, **22**, 2577–2637.
40. Kohn, J. E., Millett, I. S., Jacob, J., Zagrovic, B., Dillon, T. M., Cingel, N. *et al.* (2004). Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl. Acad. Sci. USA*, **101**, 12491–12496.
 41. Guillou, J. C. L. & Zinn-Justin, J. (1977). Critical exponents for the n -vector model in three dimensions from field theory. *Phys. Rev. Lett.* **39**, 95–98.
 42. Tanford, C., Kawahara, K. & Lapanje, S. (1966). Proteins in 6-M guanidine hydrochloride. Demonstration of random coil behavior. *J. Biol. Chem.* **241**, 1921–1923.
 43. Uversky, V. N. (2002). Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* **11**, 739–756.
 44. Dyson, H. J. & Wright, P. E. (2004). Unfolded proteins and protein folding studied by NMR. *Chem. Rev.* **104**, 3607–3622.
 45. Dyson, H. J. & Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nat. Rev., Mol. Cell Biol.* **6**, 197–208.
 46. Caflisch, A. & Paci, E. (2004). *Molecular dynamics simulations to study protein folding and unfolding* (chapter 32). Wiley-VCH, Weinheim.
 47. Chen, J., Brooks, C. L. & Khandogin, J. (2008). Recent advances in implicit solvent-based methods for biomolecular simulations. *Curr. Opin. Struct. Biol.* **18**, 140–148.
 48. Fitzkee, N. C. & Rose, G. D. (2004). Reassessing random-coil statistics in unfolded proteins. *Proc. Natl Acad. Sci. USA*, **101**, 12497–12502.

Influence of the Internal Disulfide Bridge on the Folding Pathway of the C_L Antibody Domain

Matthias J. Feige, Franz Hagn, Julia Esser, Horst Kessler and Johannes Buchner*

*Institut für Organische Chemie und Biochemie, Technische Universität München
Lichtenbergstr. 4, 85747 Garching, Germany*

Disulfide bridges are one of the most important factors stabilizing the native structure of a protein. Whereas the basis for their stabilizing effect is well understood, their role in a protein folding reaction still seems to require further attention. We used the constant domain of the antibody light chain (C_L), a representative of the ubiquitous immunoglobulin (Ig)-superfamily, to delineate the kinetic role of its single buried disulfide bridge. Independent of its redox state, the monomeric C_L domain adopts a typical Ig-fold under native conditions and does not retain significant structural elements when unfolded. Interestingly, its folding pathway is strongly influenced by the disulfide bridge. The more stable oxidized protein folds *via* a highly structured on-pathway intermediate, whereas the destabilized reduced protein populates a misfolded off-pathway species on its way to the native state. In both cases, the formation of the intermediate species is shown to be independent of the isomerization state of the Tyr₁₄₁-Pro₁₄₂ bond. Our results demonstrate that the internal disulfide bridge in an antibody domain restricts the folding pathway by bringing residues of the folding nucleus into proximity thus facilitating the way to the native state.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: antibody folding; disulfide bridges; folding intermediates; C_L; NMR

*Corresponding author

Introduction

To maintain the integrity of secreted proteins, internal stabilizing factors are of outstanding importance. The most prominent example is the disulfide bridge. Disulfide bonds are able to covalently link amino acids far apart in the primary sequence. The resulting conformational restriction of the unfolded state destabilizes it and thus increases the free energy difference to the native state.^{1,2} In practice, however, this desirable effect is sometimes foiled by strain or reduced dynamics imposed on the native state, especially in the case of artificially introduced disulfide bonds.^{3–5}

Abbreviations used: ANS, 8-anilino-1-naphthalene sulfonic acid; C_L, constant domain of the antibody light chain; CD, circular dichroism; DJ, double jump; EDTA, ethylene diamine tetra-acetic acid; GdmCl, guanidinium chloride; Ig, immunoglobulin; MAK33, monoclonal antibody from mouse of the subtype κ/IgG1; NOESY, nuclear Overhauser enhancement spectroscopy.

E-mail address of the corresponding author:
johannes.buchner@ch.tum.de

Even though disulfide bridges are one of the best understood factors stabilizing the native structure of a protein,^{6–8} their kinetic role in protein folding is not comprehensively understood. Disulfide bonds can lead to the persistence of partially structured or at least conformationally restricted regions in otherwise unfolded polypeptides.^{9,10} These preformed tertiary interactions may have a profound influence on the kinetics of a protein folding reaction.^{11–14} Accordingly, genetically engineered disulfide bonds have been used successfully as a tool to probe nucleation sites of the protein folding process.^{5,15–17} Additionally, a number of studies dealt with the formation of disulfide bridges within the folding process.^{18–22} However, the influence of naturally occurring disulfide bridges on the mechanism of a protein folding reaction is still not completely understood even though theoretical considerations suggest it to be of pronounced importance.²³

To address this issue experimentally, the constant domain of the antibody light chain (C_L) of the murine monoclonal antibody MAK33 of the subtype κ/IgG1 was chosen.²⁴ Antibody domains represent a well suited model system, since their Greek-key

β -barrel topology is stabilized by a single internal disulfide bond located in the hydrophobic core bridging the two β -sheets.²⁵ The folding of this domain structure has already been investigated, in most cases with intact disulfide bridges^{26–30} or in naturally disulfide-deficient model systems.³¹ A unifying feature for all the different immunoglobulin (Ig) proteins analyzed in this respect so far appears to be the formation of a folding nucleus which is mainly established by hydrophobic residues located in the inner part of the β -sheets.^{32–34} These residues are part of the hydrophobic core in the native state. Details of the folding process vary for the different Ig proteins studied. In some cases, two-state folding behavior has been reported,³¹ in others, folding *via* a highly structured intermediate^{27,29,30} or a misfolded species was assumed.³⁵ These differences do of course not rule out a conserved folding mechanism. They might rather represent variations of a motif due to sequence differences which influence the stabilities of different populated species.^{36–38} For the C_L domain, it was shown that folding in the presence and in the absence of its disulfide bridge leads to an identical structure.^{27,29,40} Thus, the C_L domain is a well suited model system to delineate the kinetic role of the disulfide bridge in the structure formation process of the Ig-topology in antibodies.

Results

Structural characterization of the oxidized and reduced C_L domain in its native and unfolded state

All members of the immunoglobulin superfamily are characterized by the Greek-key β -barrel topology.⁴¹ The recombinant C_L domain shows the typical far-UV circular dichroism (CD) spectrum of an all- β protein characterized by a minimum at 218 nm and a low overall intensity (Figure 1(a)). Importantly, the far-UV CD spectrum, and thus the secondary structure of the C_L domain, does not change significantly upon reduction of its internal disulfide bridge (Figure 1(a)). Furthermore, near-UV CD and 1D ¹H NMR/2D ¹H-¹H-NOESY spectra of the reduced and oxidized C_L domain show that the structure of the domain remains essentially unchanged upon reduction of its disulfide bridge (Figure 1(b) and (c) and data not shown).

The presence of a covalent cross-link bridging 60 residues influences the conformational freedom of the unfolded chain. To investigate whether significant structural elements persist in the unfolded state

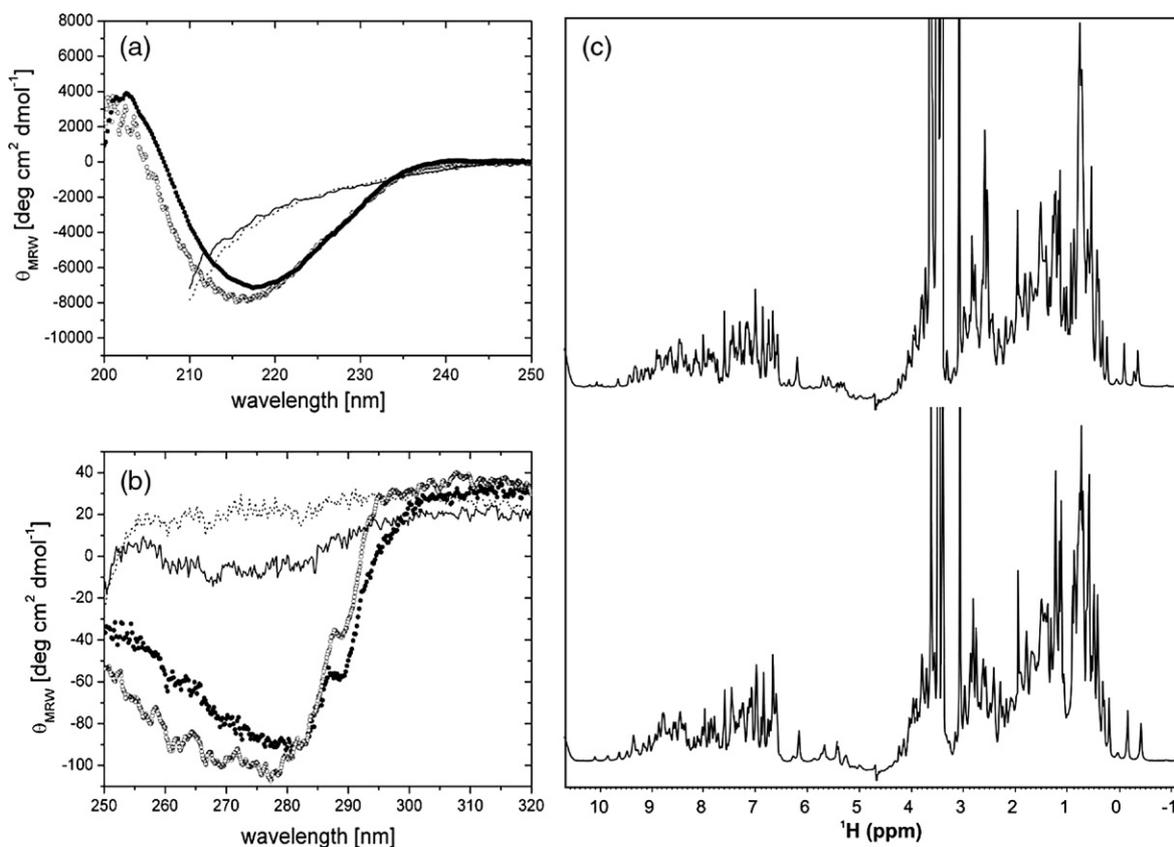


Figure 1. Spectroscopic characterization of the C_L domain. Far-UV CD, near-UV CD and 1D ¹H NMR spectra of the reduced and oxidized C_L domain were measured. In the CD experiments (a, b), filled circles denote native oxidized protein, open circles the reduced domain under native conditions (standard buffer). Spectra of the unfolded domain were measured in the presence of 2 M GdmCl. CD data of the unfolded and oxidized protein are shown as straight lines, its reduced counterpart is depicted as dashed lines (a, b). The 1D ¹H NMR spectrum of the reduced C_L domain is shown in the upper panel, the spectrum of the oxidized domain in the lower panel (c).

of the oxidized C_L domain, compared to its reduced counterpart, the spectroscopic techniques used to characterize the native state were also applied to the denatured state of the oxidized and reduced protein. The far-UV CD spectra of both forms are very similar and typical for random coil polypeptides (Figure 1(a)). Near-UV CD spectra clearly show the absence of significant asymmetric environment around the aromatic residues for the reduced and the oxidized C_L domain (Figure 1(b)). The small and broad signal in the latter case can likely be attributed to the intrinsically asymmetric disulfide bridge. NMR spectra also argue against significant differences between the unfolded states. In the presence of 2 M guanidinium chloride (GdmCl), both forms show no significant signal dispersion in the 1D ¹H NMR spectrum (data not shown). Additionally, iodide fluorescence quenching experiments were carried out to investigate the solvent accessibility of the two tryptophanes in the unfolded states of the oxidized and reduced C_L domain. Solvent accessibility of tryptophane residues was observed to be slightly higher in the reduced protein (data not shown). Upon reduction, the Stern-Volmer constant K_{SV} increases from 4.6 M⁻¹ to 4.9 M⁻¹. In the native state, no quenching difference for the oxidized and the reduced C_L domain is observed (data not shown).

The hydrodynamic properties of the C_L domain are influenced by the disulfide bridge as expected. Pulsed field gradient diffusion NMR experiments of the native proteins show a decrease in the diffusion coefficient of ~10% when the disulfide bridge is reduced, i.e. the reduced form is more extended. For the native state, the calculated diffusion constant using the pdb structure file (pdb code 1FH5) containing oxidized C_L is $1.083 \times 10^{-10} \text{ m}^2 \text{ s}^{-1}$ which is in good agreement with the measured value of $1.255 \times 10^{-10} \text{ m}^2 \text{ s}^{-1}$. In the presence of 2 M GdmCl, the diffusion constant decreases ~tenfold as compared to the native state (data not shown). Using the Debye-Einstein equation, the hydrodynamic radii of the different C_L species can be calculated assuming a spherical shape for all proteins.^{42,43} Upon reduction, the hydrodynamic radius increases from 1.46 nm to 1.60 nm. These data emphasize the role of the disulfide bridge for the compact structure of the folded C_L domain. In the presence of 2 M GdmCl, the increase in the hydrodynamic radius is around 40% for both the oxidized and the reduced form showing that the presence of the disulfide bond does not significantly restrict the unfolded state of the C_L domain.

Stabilization of the C_L domain by its internal disulfide bridge

Equilibrium unfolding experiments were carried out to determine how the internal disulfide bridge contributes to the stability of the C_L domain. Unfolding was monitored either by changes in the far-UV CD signal at 218 nm or by changes in the intrinsic tryptophane fluorescence. Two tryptophanes are

found in the C_L domain, one is located in strand C and is part of the hydrophobic core of the Ig-fold.⁴¹ The other one is located in strand D and is rather solvent-exposed.²⁴ As can be seen in Figure 2(a), the disulfide bridge is a major stabilizing factor of the C_L domain. In its absence, the midpoint of the GdmCl transition is shifted from 0.8 M GdmCl to 0.3 M GdmCl. This corresponds to a reduction in stability of 8.2 kJ mol⁻¹, from 14.6 kJ mol⁻¹ in the oxidized form to 6.4 kJ mol⁻¹ in the reduced form. The cooperativity parameter m of the transition is found to be 17.4 kJ M⁻¹ mol⁻¹ in the oxidized and 20.9 kJ M⁻¹ mol⁻¹ in the reduced state. Since the respective far-UV CD and fluorescence transitions of the reduced and oxidized domain coincide, significantly populated equilibrium unfolding intermediates are very unlikely. Independent of the redox state, the temperature-induced unfolding of the C_L domain was not completely reversible. Nevertheless, the major stabilizing effect of the disulfide bond is obvious from the decrease in the melting point by 20 °C, from 51 °C to 31 °C upon reduction (Figure 2(b)).

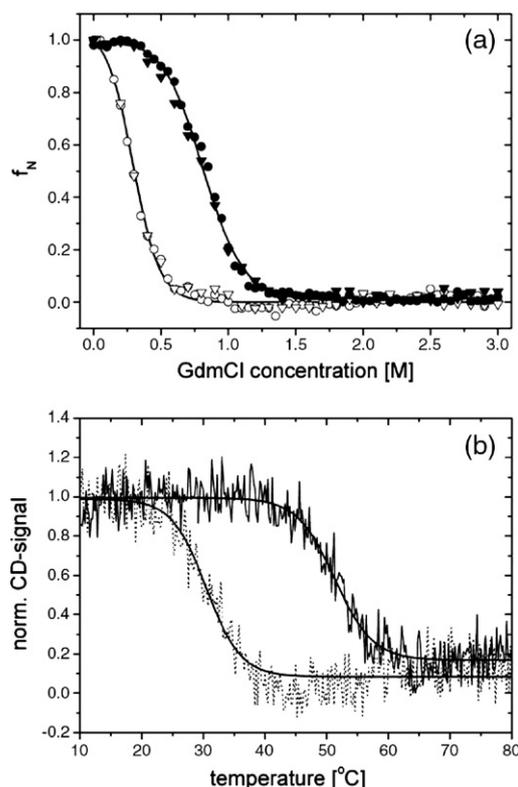


Figure 2. Influence of the disulfide bridge on the stability of the C_L domain. GdmCl-induced unfolding of the C_L domain (a) was monitored by the change in its intrinsic fluorescence (circles) or far-UV CD spectroscopy (down triangles). Filled symbols stand for the oxidized protein, open symbols denote the reduced protein. The temperature-induced unfolding of the C_L domain was monitored by far-UV CD spectroscopy (b). Data for the oxidized domain are shown as a continuous line, the unfolding of the reduced protein as a dashed line. GdmCl transitions were fit to a two-state model, and the temperature melts to a Boltzmann function.

Folding mechanism of the C_L domain

To investigate the role of the disulfide bridge in the folding process of the C_L domain, the kinetics of folding and unfolding of the reduced and the oxidized protein were monitored at different GdmCl concentrations by fluorescence spectroscopy. The experiments were performed in standard buffer at 20 °C. In all cases, a small signal loss was observed during the dead time of the stopped-flow experiments which can likely be attributed to the rearrangement of the polypeptide chain upon change of the solvent conditions. From the chevron plots of the respective rate constants for the oxidized and the reduced protein, it is evident that the folding reaction consists of three observable, kinetically well separated phases (Figure 3 and Table 1). The disulfide bridge has a pronounced effect on the

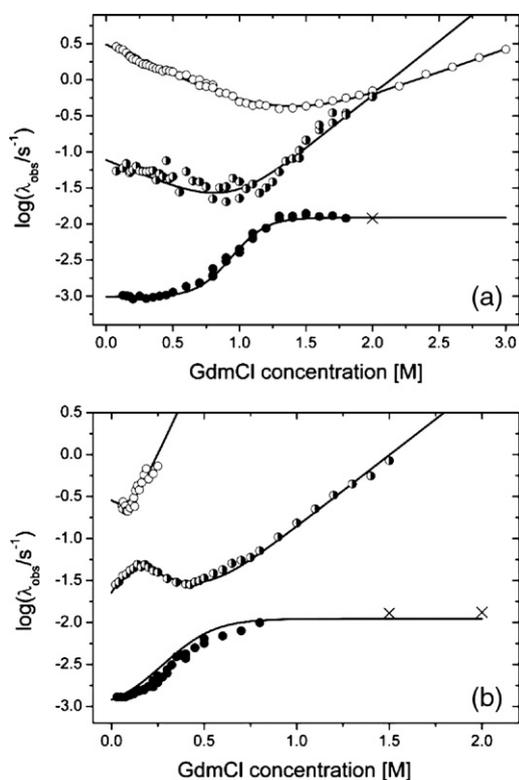


Figure 3. Chevron plots of the oxidized and reduced C_L domain. Unfolding and refolding of the oxidized (a) or reduced (b) C_L domain at different GdmCl concentrations were measured. The fast phase (open circles) was monitored by stopped-flow fluorescence spectroscopy and the slowest phase (filled circles) by manual mixing fluorescence spectroscopy. The intermediate phase (half-filled circles) was monitored by both techniques. Additionally, rate constants obtained from DJ experiments are shown (crosses). The data obtained for the oxidized C_L domain were fit to a three-state on-pathway model, whereas the data for the reduced protein were best described by a three-state off-pathway model (continuous lines). The slowest phase was fit to a Pro-isomerization reaction in both cases. The rate constants and kinetic *m*-values for the formation of the intermediate species were assumed to be independent of the isomerization state of the Tyr₁₄₁-Pro₁₄₂ bond.

Table 1. Microscopic rate constants and kinetic *m*-values derived from the chevron plots for the oxidized and the reduced C_L domain

| Parameter | Value for oxidized C _L | Value for reduced C _L |
|---|-----------------------------------|----------------------------------|
| $k_{U}^{\text{H}_2\text{O}} \text{ cis/trans} \rightarrow \text{I cis/trans} [\text{s}^{-1}]$ ($m_{U \text{ cis/trans} \rightarrow \text{I cis/trans}} [\text{M}^{-1}]$) | 3.01 (-0.87) | 0.15 (-3.0) |
| $k_{I}^{\text{H}_2\text{O}} \text{ cis/trans} \rightarrow \text{U cis/trans} [\text{s}^{-1}]$ ($m_{I \text{ cis/trans} \rightarrow \text{U cis/trans}} [\text{M}^{-1}]$) | 0.032 (0.65) | 0.065 (4.75) |
| $k_{I}^{\text{H}_2\text{O}} \text{ cis} \rightarrow \text{N} [\text{s}^{-1}]$ ($m_{I \text{ cis} \rightarrow \text{N}} [\text{M}^{-1}]$) | 0.077 (-0.75) | |
| $k_{N}^{\text{H}_2\text{O}} \text{ cis} [\text{s}^{-1}]$ ($m_{N \rightarrow \text{I cis}} [\text{M}^{-1}]$) | 0.00068 (1.48) | |
| $k_{I}^{\text{H}_2\text{O}} \text{ trans} \rightarrow \text{I cis} [\text{s}^{-1}]$ ($m_{I \text{ trans} \rightarrow \text{I cis}} [\text{M}^{-1}]$) | 0.00098 (0.0) | |
| $k_{I}^{\text{H}_2\text{O}} \text{ cis} \rightarrow \text{I trans} [\text{s}^{-1}]$ ($m_{I \text{ cis} \rightarrow \text{I trans}} [\text{M}^{-1}]$) | 0.011 (0.0) | |
| $k_{U}^{\text{H}_2\text{O}} \text{ cis} \rightarrow \text{N} [\text{s}^{-1}]$ ($m_{U \text{ cis} \rightarrow \text{N}} [\text{M}^{-1}]$) | | 0.090 (-1.80) |
| $k_{N}^{\text{H}_2\text{O}} \text{ cis} \rightarrow \text{U cis} [\text{s}^{-1}]$ ($m_{N \rightarrow \text{U cis}} [\text{M}^{-1}]$) | | 0.0027 (1.71) |
| $k_{U}^{\text{H}_2\text{O}} \text{ trans} \rightarrow \text{U cis} [\text{s}^{-1}]$ ($m_{U \text{ trans} \rightarrow \text{U cis}} [\text{M}^{-1}]$) | | 0.00090 (0.0) |
| $k_{U}^{\text{H}_2\text{O}} \text{ cis} \rightarrow \text{U trans} [\text{s}^{-1}]$ ($m_{U \text{ cis} \rightarrow \text{U trans}} [\text{M}^{-1}]$) | | 0.010 (0.0) |

Data were obtained from a fit of the chevron plot to a three-state on-pathway model for the oxidized C_L domain or a three-state off-pathway model for the reduced C_L domain (see Figure 6). Since the intermediate states build up independent of the Tyr₁₄₁-Pro₁₄₂ bond isomerization state, the kinetic parameters for their formation were regarded as independent of this reaction.

fast and intermediate phase observed. If reduced, the refolding limb of the intermediate phase not only shows a rollover but passes through a maximum (Figure 3(b)). This shape of the chevron plot makes the formation of a misfolded species within the fast phase likely which has to undergo unfolding before productive refolding reactions can occur.⁴⁴ If unfolding of this species becomes rate limiting, as observed for the reduced C_L domain at low GdmCl concentrations, an apparent deceleration of the productive folding reactions is expected which explains the shape of the intermediate phase in the chevron plot. For the oxidized domain, on the other hand, a linear GdmCl-dependency of the fast and intermediate phase is observed (Figure 3(a)). This implies that the reactions occurring in the fast phase are significantly different for the oxidized and the reduced C_L domain. To further analyze the underlying processes, interrupted refolding experiments and double jump (DJ) experiments were carried out. In interrupted refolding experiments, the protein is refolded for different times before changing to unfolding conditions again. The observed amplitudes of the different unfolding phases directly correspond to the amount of different species present. In DJ experiments, slow isomerization reactions

in the unfolded state are indirectly monitored by the amount of molecules following different pathways after varying denaturation times. These experiments allowed us to monitor the formation of intermediates and native molecules as well as slow isomerization reactions giving rise to the different observable phases. As can be seen in Figure 4(a) and Table 2, two exponential phases are necessary to describe the formation of native molecules in the case of the oxidized C_L domain in interrupted refolding experiments. Hence, refolding of the oxidized protein occurs on two parallel pathways. The slower one accounts for ~90% of the molecules after extended denaturation times (Figure 4(a) and (c) and Table 2) indicative of a proline-isomerization reaction. Only one *cis* peptide bond, between Tyr₁₄₁ and Pro₁₄₂, is found in the native state of the C_L domain. After equilibration, ~80 to 90% of the molecules are expected to populate the non-native *trans* conformation in the unfolded state⁴⁵ which corresponds very well to the experimentally observed amount of slow folding species (Table 2). Taking the rate constant for refolding and the amount of slow folding molecules into account, one of the parallel pathways can likely be attributed to the isomeriza-

Table 2. Kinetic parameters derived from the interrupted refolding assay

| Parameter | Value for oxidized C _L | Value for reduced C _L |
|---|-----------------------------------|----------------------------------|
| $k_{UI}^{cis/trans} \rightarrow I^{cis/trans}$ [s ⁻¹] | 1.02 (100%) | – |
| $k_I^{cis \rightarrow N}$ [s ⁻¹] | 0.069 (8%) | – |
| $k_I^{trans \rightarrow N}$ [s ⁻¹] | 0.0014 (92%) | – |
| $k_{UI}^{cis \rightarrow N}$ [s ⁻¹] | – | 0.051 (14%) |
| $k_{UI}^{trans \rightarrow N}$ [s ⁻¹] | – | 0.0013 (86%) |

The unfolding amplitudes obtained in the interrupted refolding assay were either fit to an irreversible three-state on-pathway model (oxidized C_L, equation (7)) or an irreversible two-state model (reduced C_L, equation (8)) with two parallel pathways in each case. The amount of molecules following a certain pathway is denoted in brackets. Values for the oxidized domain were determined in the presence of 140 mM residual GdmCl, values for the reduced domain in the presence of 90 mM residual GdmCl.

tion of the Tyr₁₄₁-Pro₁₄₂ bond to the native *cis* state. In the case of the oxidized C_L domain, an obligatory on-pathway folding intermediate seems to be populated in the fast phase as inferred from the lag time for the formation of native molecules (Figure 4(a), inset). The large extent of secondary structure, native fluorescence, and NMR signals which are already observed after the fast refolding

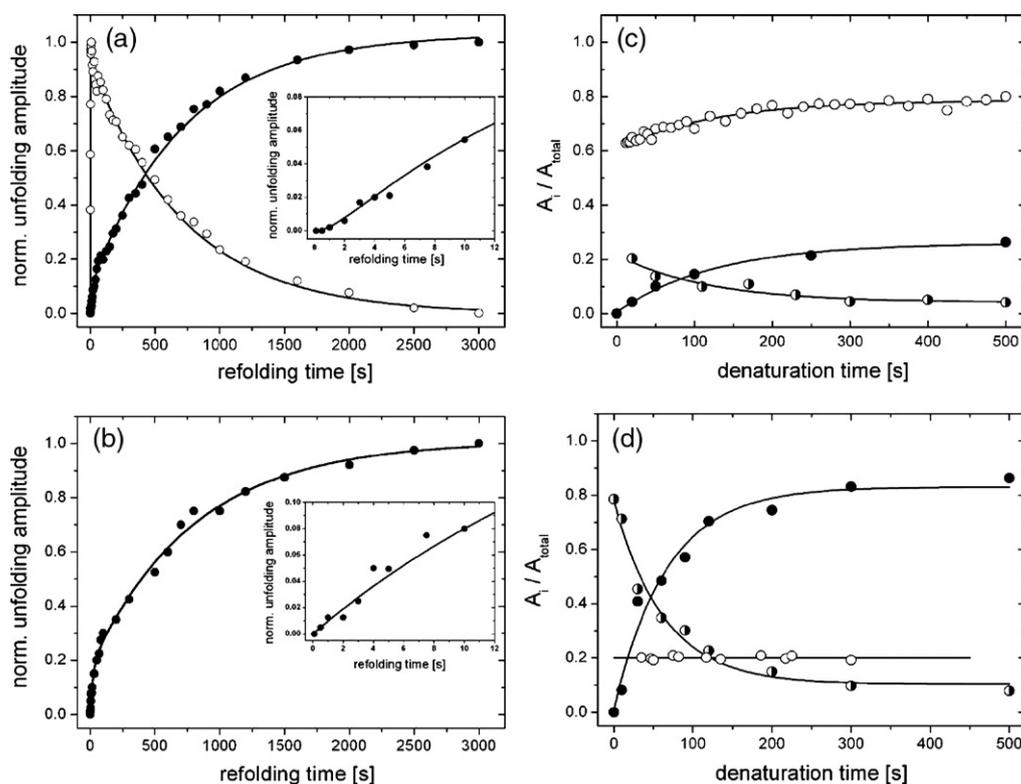


Figure 4. Interrupted refolding assay and DJ experiments of the C_L domain. The formation and decay of the folding intermediate were monitored for the oxidized domain (a, open circles). The formation of native molecules was monitored for the oxidized (a, filled circles) and the reduced (b, filled circles) C_L domain. The insets show the formation of native molecules within the first 10 s. The data were fit to an irreversible three-state on-pathway model (a), or an irreversible two-state model (b). In each case, two parallel pathways were assumed. DJ experiments were carried out for the oxidized (c) and the reduced C_L domain (d). A_i/A_{total} denotes the amplitude of each phase normalized in respect to the sum of all three phases. The interrupted refolding experiments as well as the fast phase in the DJ experiments (open circles) were followed by stopped-flow fluorescence spectroscopy in the sequential mixing mode. The intermediate phase (half-filled circles) as well as the slow phase (filled circles) of the DJ experiments were followed by manual mixing fluorescence spectroscopy.

phase also implies an on-pathway folding intermediate (Figures 4(c) and 5(a) and (d); far-UV CD data not shown). The persistence of the fast refolding phase after short- or long-term denaturation in the DJ experiments (Figure 4(c)) shows that this intermediate species is formed independent of the isomerization state of the proline residues. This finding is in agreement with the results from the interrupted refolding assay, where the decay of the intermediate species occurs in a slow and a fast reaction. The rate constants for the decay of the intermediate species are in good agreement with the ones for the formation of native molecules. In contrast to these results, an intermediate trapped by a *trans* Tyr₁₄₁-Pro₁₄₂ bond would only be populated after prolonged denaturation times, i.e. the amplitude of the fast phase would increase in DJ experiments, and would only decay with a single, isomerization limited reaction in the interrupted refolding assay.

For the reduced C_L domain, the interrupted refolding assay also shows that the protein folds on two parallel pathways. Again, ~90% of the

molecules follow the slow pathway after extended denaturation times with a similar rate constant as in the case of the oxidized domain (Figure 4(b) and Table 2). Accordingly, the isomerization of the Tyr₁₄₁-Pro₁₄₂ bond to the native *cis* state also slows down the folding of a major part of the reduced C_L domain as expected. The chevron plot for the reduced C_L domain exhibits three observable phases (Figure 3(b)), but in interrupted refolding experiments, the formation and decay of the misfolded species cannot be observed directly in the interrupted refolding assay due to its fast unfolding reaction (Figures 3(b) and 4(b)). Nevertheless, the presence of the fast refolding phase in DJ experiments, independent of the denaturation time and therefore the Pro-isomerization state, clearly demonstrates that the misfolded intermediate is populated on the slow and on the fast folding pathway (Figure 4(d)). The absence of a lag phase for the formation of native, reduced molecules strengthens the notion that the species formed is misfolded (Figure 4(b), inset). Furthermore, no

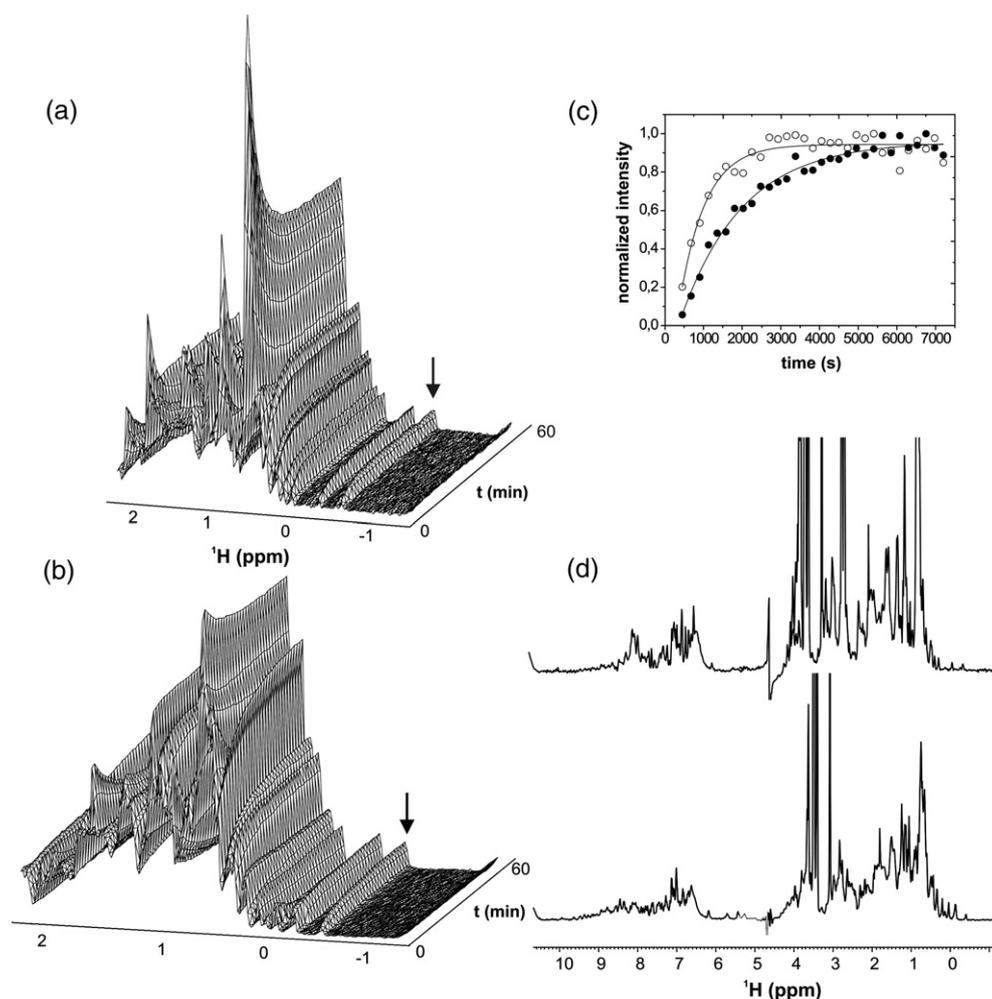


Figure 5. One-dimensional real-time NMR spectroscopy of the refolding of reduced (a) and oxidized (b) C_L . Refolding of the reduced domain was carried out in the presence of 60 mM GdmCl, the oxidized C_L domain was refolded in the presence of 100 mM GdmCl. The intensity of the methyl resonances marked by an arrow is plotted against refolding time in (c) and fitted to a single exponential ($k_{ox}=0.00062\text{ s}^{-1}$ (filled circles, oxidized protein) and $k_{red}=0.0015\text{ s}^{-1}$ (open circles, reduced protein)). The first 1D spectrum of each series is shown in (d) (upper panel: reduced C_L ; lower panel: oxidized C_L).

significant extent of native far-UV CD signal, native fluorescence, or NMR signals build up during the fast folding phase of the reduced C_L domain (Figures 4(d) and 5(b) and (d); far-UV CD data not shown).

Structural characterization of the different intermediates

The structural properties of the intermediate states formed during the folding of the oxidized and the reduced C_L domain were investigated by different spectroscopic techniques. Formation of secondary structure was monitored by far-UV CD spectroscopy. Spectra of the refolding protein were measured after 1 min incubation time within 2 min. In this period, the intermediate states have already formed but the majority of the molecules are still separated from the following folding reaction to the native state by the slow isomerization of the Tyr₁₄₁-Pro₁₄₂ bond. In the case of the oxidized protein, a large amount of the native secondary structure already forms in the 3 min the experiment takes whereas for the reduced protein, the spectrum after 3 min is still much more random coil-like (data not shown). 8-Anilino-1-naphthalene sulfonic acid (ANS) binding to the two intermediate states was not significantly different. Only a slightly higher ANS affinity was observed in the case of the oxidized protein showing that a few more hydrophobic residues are already buried in the intermediate state of the reduced protein (data not shown). To gain further insight into the similarities and differences of the intermediate states populated, real-time 1D ¹H NMR experiments were recorded (Figure 5). The reduced and oxidized C_L domains refold in a similar timescale (Figure 5(a)-(c)), but the initial states are significantly different (Figure 5(d)). After the ~5 min dead time of the mixing and first data acquisition step, the fast refolding phase is finished and the folding intermediate is populated. The spectrum of the oxidized form already shows a well dispersed ¹H spectrum which is indicative of a compact tertiary structure whereas the spectrum of the reduced form still shows poor dispersion. Therefore, this intermediate seems not to be compactly folded (Figure 5(d)). It should be noted that during the 5 min dead time of the experiment, ~20% of the molecules already fold to the native state due to amount of correct Pro-isomers present and the additionally occurring re-isomerization to the native conformer during that time. This likely explains the small amount of native methyl signals observed for the reduced C_L domain in the first spectrum but not the pronounced resonances observed for the oxidized protein.

Discussion

In the present study, we have analyzed the influence of the internal disulfide bridge on the

folding mechanism of the C_L antibody domain. This protein folds into an almost identical structure independent of its redox state. Therefore, it is a well suited model system to delineate the kinetic role of the disulfide bond in the folding of a representative of the immunoglobulin superfamily. This advantageous feature has already been recognized by Goto *et al.* who carried out groundbreaking studies on the stabilizing role of the internal disulfide bridge³⁹ and on its kinetic impact using a human type λ C_L domain.⁴⁰ The persistence of its structural integrity despite reduction is obviously a common feature for the human and the murine C_L domain of the κ and λ type.³⁹ The stability of the oxidized C_L domain was determined by Goto *et al.* to be ~21 kJ mol⁻¹ for a human κ C_L domain and to be ~24 kJ mol⁻¹ for a human λ C_L domain.^{39,46} The higher values than the ~15 kJ mol⁻¹ determined for the oxidized κ C_L domain in this study might be explained in two ways. On the one hand, a higher stability of the human κ and λ C_L domains compared to the murine κ C_L domain could be assumed. On the other hand, different models have been applied for determining the free energy of unfolding. The binding model of Tanford employed by Goto *et al.*⁴⁷ has been shown to overestimate protein stability, especially when GdmCl is used as a denaturant.⁴⁸ Both considerations might explain the differences in stability observed. The stabilizing effect of the internal disulfide bridge is very pronounced for the human and the murine C_L domain. In a simplistic approach, the stabilizing effect of a disulfide bond bridging *n* residues can be approximated to¹⁷:

$$\Delta\Delta G = -T\Delta\Delta S = -T(-2.1 - [1.5R\ln(n)]) \text{ cal mol}^{-1} \quad (1)$$

In the case of the C_L domain with *n*=60, ΔΔ*G* is calculated to ~17.5 kJ mol⁻¹ which clearly exceeds the stabilizing effect of ΔΔ*G*=8.2 kJ mol⁻¹ determined experimentally and even the entire stability of the domain. The above equation only takes the entropic destabilization of the unfolded state into account but neglects possibly reduced dynamics of the native state or induced strain. These factors may lead to a deviation between the experimentally observed and the theoretically expected stabilizing effect of the disulfide bridge. Only in very few cases, the stabilizing role of a disulfide bridge could be adequately predicted by this simple model.⁵ In the case of the C_L domain, the hydrodynamic radius increases upon reduction by ~10% as derived from pulsed field gradient diffusion NMR experiments. This points towards a certain strain the disulfide bridge imposes on the native state. Furthermore, the C_L domain showed increased flexibility upon reduction of its internal disulfide bridge.⁴⁹ Thus, more sophisticated models for analysis, e.g. including the reduced dynamics of the native state using a molecular dynamics approach, seem to be required.

Disulfide bridges have been shown to preserve a certain degree of residual structure in otherwise unfolded polypeptide chains in some cases.¹⁰ This might of course have a profound influence on the folding dynamics and pathway of a protein.¹³ Therefore, the denatured state of the reduced and oxidized C_L domain was characterized to detect residual structure. The *m*-value of denaturant-induced unfolding transitions is known to correlate with the amount of hydrophobic surface exposed upon unfolding.⁵⁰ For disulfide-bridged proteins, generally lower *m*-values are found than for disulfide-free proteins or their reduced counterparts.⁵⁰ This is also the case for the C_L domain. The *m*-value increases by ~20% from 17.4 kJ M⁻¹ mol⁻¹ to 20.9 kJ M⁻¹ mol⁻¹ upon reduction of the internal disulfide bond. Furthermore, the Stern-Volmer constant for iodide-induced tryptophane quenching increases by ~7% from 4.6 M⁻¹ to 4.9 M⁻¹ in the unfolded protein. On the other hand, far-UV and near-UV CD spectroscopy as well as 1D ¹H and 2D ¹H-¹H-NOESY NMR experiments rule out the persistence of significant residual structure in the case of the oxidized C_L domain after unfolding. In the primary sequence, Cys₁₃₅ is flanked by hydrophobic residues (VVC₁₃₅FL) and Cys₁₉₅ by residues of alternating hydrophobicity (YTC₁₉₅EA). Importantly, Val₁₃₃ and Leu₁₃₇ as well as Tyr₁₉₃ and Ala₁₉₇ are part of the hydrophobic core in the native state. Interaction of these residues might induce a small amount of hydrophobic clustering around the disulfide bridge, even in the otherwise unfolded polypeptide chain. From the fluorescence quenching experiments, which report on a slightly lower tryptophane solvent accessibility for the oxidized and unfolded C_L domain than for the reduced protein, the intermittent association of one or both tryptophane residues to this cluster seems possible.

Independent of the redox state, the isomerization of the Tyr₁₄₁-Pro₁₄₂ bond to the native *cis* state is the rate limiting step for the folding of the major part of the C_L molecules.^{27,40} The presence of the disulfide bridge has only a small effect on the isomerization rate of the Tyr₁₄₁-Pro₁₄₂ bond, either from its native *cis* to the non-native *trans* state upon unfolding or on the reverse reaction upon refolding (see Table 1 and Figure 5(c)). Both the reduced and the oxidized C_L domains populate an intermediate species upon refolding. This process is independent of their Pro-isomerization states as derived from DJ experiments. In the case of the oxidized C_L domain, a highly structured on-pathway intermediate is formed, which leads to a lag phase for the formation of native molecules (see Figure 4(a)). It already possesses a large amount of native secondary structure and native-like dispersion of all NMR signals (Figure 5(d)). Furthermore, its stability is rather high. Assuming a pre-equilibrium between the unfolded and the intermediate state and independence of the Pro-isomerization reaction, its stability is calculated to be ~11 kJ mol⁻¹ from the microscopic rate constants (Table 1), which is only

~25% lower than the stability of the native state. Surprisingly, in this highly structured and stable intermediate, a free isomerization of the Tyr₁₄₁-Pro₁₄₂ bond to the native *cis* state with a rate similar to that observed for unstructured peptides is possible.⁵¹ The Tyr₁₄₁-Pro₁₄₂ bond is found in the loop connecting strands B and C of the immunoglobulin fold. Its free isomerization is in agreement with the proposed small role the loops play as folding nuclei in the structure formation process of the Ig-topology.⁵² Furthermore, the presence of the intermediate not only on the folding but also on the unfolding pathway (Figure 3(a)) shows that unfolding of the C_L domain is likely to begin from the loops and expand to the hydrophobic core.

If the internal disulfide bridge of the C_L domain is reduced, its folding behavior changes significantly. For the intermediate phase of the chevron plot, corresponding to the formation of native molecules from the denatured state with the Tyr₁₄₁-Pro₁₄₂ bond in the *cis* conformation, a pronounced rollover with the rate constant passing through a maximum is observed (see Figure 3(b)). The same behavior has been reported for the tenth fibronectin type III domain of human fibronectin (FNfn10)⁵³ also belonging to the Ig-superfamily. As has been argued by Wildegger and Kiefhaber, this behavior of the observable rate constants points towards an off-pathway intermediate as part of a dead-end mechanism.⁴⁴ This is in agreement with the absence of a lag phase for the formation of native, reduced C_L molecules and the spectroscopic characteristics of the intermediate species. In contrast to the oxidized protein, in far-UV CD experiments, the intermediate is essentially random coil-like and 1D ¹H NMR experiments also show random coil chemical shift dispersion after the fast refolding phase (see Figure 5). Taken together, the experimental data allow the folding model proposed for the C_L domain^{27,40} to be refined, especially in the case of the reduced C_L domain (Figure 6).

Being very unstable (~2 kJ mol⁻¹), the kinetic *m*-values for the folding and unfolding of the misfolded species are only poorly defined. Nevertheless, compared to the oxidized domain, the kinetic *m*-values are clearly higher and therefore the transition states for the formation and unfolding of the off-pathway intermediate are less compact (see Table 1). ANS-binding experiments on the other hand argue for a pronounced clustering of hydrophobic residues in the misfolded species, even slightly higher than for the on-pathway intermediate in the case of the oxidized C_L domain. Consequently, in the folding process of reduced C_L, the population of a misfolded, aggregation-prone off-pathway species can be postulated. Almost completely devoid of native secondary or tertiary structure and mainly stabilized by the collapse of hydrophobic residues, this species has to unfold before productive refolding to the native state can occur.

In summary, our results imply a new role for the disulfide bridge in antibody folding: it facilitates the

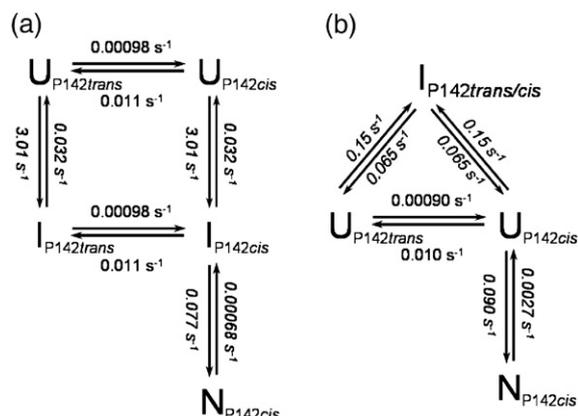


Figure 6. Folding mechanism for the C_L domain. Independent of the Tyr₁₄₁-Pro₁₄₂ bond isomerization state, the oxidized and the reduced C_L domain populate a folding intermediate. In the case of the oxidized protein (a), this intermediate is highly structured and on-pathway. The reduced C_L domain, on the other hand, populates a misfolded and unstable off-pathway species which has to undergo unfolding before productive folding reactions can occur (b). Folding of the major part of the molecules is limited by the isomerization of the Tyr₁₄₁-Pro₁₄₂ bond to the native *cis* state. Microscopic rate constants are denoted for the individual reactions. For the oxidized domain, isomerization of the Tyr₁₄₁-Pro₁₄₂ bond was assumed to occur in the unfolded state with the same rate constants as in the intermediate state.

efficient folding to the native state. In the case of the oxidized domain, irrespective of Pro-isomerization, a native-like intermediate is formed within a few hundred ms which then undergoes a second folding reaction to the native state. Importantly, the structured intermediate species shows no aggregation tendency at refolding concentrations up to a few hundred μ M (data not shown). Folding of the reduced C_L domain, on the other hand, is slowed down by the population of a misfolded off-pathway species. Furthermore, this species is much more aggregation-prone (data not shown). Moreover, this aggregation-prone species is present for up to a few thousand seconds if Pro-isomerization is rate limiting and if not, it still needs several seconds to fold to the native state (see Figure 3(b)). In the case of the oxidized domain, on the other hand, an aggregation-protected intermediate is formed within a few hundred ms for all molecules. It is therefore tempting to speculate that the buried disulfide bond of the immunoglobulins has not only been selected by evolution to stabilize the native state but also to direct the folding process. Connecting key residues which are part of the Ig-topology transition state, it guides the way to the native structure and allows unproductive side reactions to be circumvented. This might not only confer an advantage to the protein if unfolding occurs under the harsh extracellular conditions but may also be of relevance for the *de novo* folding process in the endoplasmic reticulum.

Materials and Methods

GdmCl (ultrapure) was purchased from MP Biomedicals (Eschwege, Germany). All other chemicals were from Merck (Darmstadt, Germany). The GdmCl concentrations were determined by the refractive indices of the solutions.⁵⁴ Unless otherwise stated, all experiments were carried out in 20 mM sodium phosphate (pH 7.5), 100 mM NaCl, 1 mM ethylene diamine tetra-acetic acid (EDTA), 1 mM DTT (if the reduced C_L domain was investigated) (standard buffer) at 20 °C.

Cloning, expression, and purification of the C_L domain

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-beta-D-thiogalactopyranoside at an OD₆₀₀ of 0.8. After 3 h, cells were harvested. The preparation of the C_L inclusion bodies was carried out as described previously.³⁰ Inclusion bodies were solubilized in 100 mM sodium phosphate (pH 7.5), 6 M GdmCl, 20 mM β -mercaptoethanol for 2 h at 20 °C. Insoluble components were removed by centrifugation (48,000g, 25 min, 20 °C). The supernatant was diluted fivefold in 50 mM sodium phosphate (pH 7.5), 4 M GdmCl and applied to a Ni-chelating column. 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), 5 mM EDTA, 1 mM oxidized glutathione at 4 °C overnight. Aggregates were removed by centrifugation (48,000g, 25 min, 4 °C) and 0.25 units Thrombin (restriction 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 (48,000g, 25 min, 4 °C), the concentrated supernatant was applied to a Superdex 75 pg 26/60 gel filtration column (Amersham Biosciences, Uppsala, Sweden) equilibrated in 20 mM sodium phosphate (pH 7.5), 100 mM NaCl, 1 mM EDTA.

The cloned C_L domain contains residues A₁₁₂-E₂₁₄ of the MAK33 κ light chain (labeled according to the crystal structure²⁴) and four additional N-terminal residues (GlySerHisMet) derived from the pET28a(+) vector after Thrombin cleavage of the His-tag. All vectors were sequenced and the mass of the purified proteins was controlled by matrix assisted laser desorption ionization-time of flight mass spectrometry.

Preparation of reduced C_L

The native C_L domain was unfolded and reduced in 20 mM sodium phosphate (pH 7.5), 100 mM NaCl, 1 mM EDTA, 2 M GdmCl, 50 mM DTT for 2 h at room temperature. Refolding of the reduced protein was performed overnight at 4 °C in 20 mM sodium phosphate (pH 7.5), 100 mM NaCl, 1 mM EDTA, 1 mM DTT *via* dialysis. Possibly present aggregates were removed by centrifugation (48,000g, 25 min, 4 °C). Successful reduction was confirmed by non-reducing SDS-PAGE, where the oxidized C_L domain runs faster, and loss in fluorescence quenching. To avoid re-oxidation, the reduced protein was always used for measurements within one day.

Fluorescence and CD measurements

Fluorescence measurements were carried out in a Spex FluoroMax II fluorimeter (Instruments SA, Edison, NJ). Equilibrium unfolding transitions (incubated for at least 12 h) as well as refolding kinetics were followed by the intrinsic tryptophane fluorescence of C_L excited at 280 nm (3 nm slit) and monitored at 370 nm (oxidized C_L) or 320 nm (reduced C_L), respectively. The emission slits were set to 5 nm. The protein concentration was 1 μM in 1 cm quartz cuvettes.

CD measurements were performed in a Jasco J-720 spectropolarimeter (Jasco, Grossumstadt, Germany). Far-UV spectra were recorded from 200 to 250 nm at a protein concentration of 10 μM in 1 mm quartz cuvettes, whereas for near-UV spectra a protein concentration of 50 μM in 5 mm quartz cuvettes was used. Spectra were accumulated 12 times and buffer-corrected. Equilibrium unfolding transitions and folding kinetics were monitored at 218 nm. In all cases, the reversibility of the GdmCl-induced unfolding transitions was confirmed. The method described by Santoro and Bolen was used to fit the data.⁵⁵

The CD spectra of the folding intermediates were measured at a protein concentration of 10 μM in a 1 mm quartz cuvette. Seven individually measured spectra were recorded after a refolding time of 1 min within 2 min and averaged in each case. The same refolding and measuring times were used for ANS-fluorescence spectroscopy, where always four individually measured spectra were averaged. Experiments were performed at a protein concentration of 1 μM and an ANS concentration of 100 μM in stirrable cuvettes. Spectra were excited at 365 nm and recorded from 430 nm to 600 nm (slits as denoted above).

NMR measurements

NMR samples were concentrated up to 250 μM. Spectra were recorded on a Bruker DMX750 spectrometer (Bruker, Rheinstetten, Germany) equipped with a triple resonance probe head at 293 K. For 1D ¹H spectra 2k data points were recorded with a spectral width of 12 ppm, zero filled to 4k data points and apodized with a squared sine window function. ¹H-¹H-nuclear Overhauser enhancement spectroscopy spectra⁵⁶ were recorded with 8k complex data points in the direct detected proton dimension and 128 complex data points in the indirect dimension. Data were zero filled and apodized with a squared sine window function. Double stimulated echo pulsed field gradient diffusion experiments⁵⁷ were recorded with 64 different gradient strengths ranging from 0.64 to 30.5 G/cm and 2k data points in the direct dimension. The obtained diffusion profiles of 30 separated resonances were analyzed as described by Stejskal and Tanner using Topspin Version 1.3 (Bruker, Rheinstetten, Germany) R1,R2 relaxation menu.⁵⁸ Hydrodynamic calculations were done with HydroNMR Version 5a.⁵⁹ ¹H real-time NMR experiments were conducted as described by Balbach *et al.*⁶⁰ Refolding was started by a dilution of denatured protein samples (in 2 M GdmCl) to a final protein concentration of ~160 μM. Oxidized C_L was refolded in the presence of 100 mM residual GdmCl, whereas for reduced C_L the residual GdmCl concentration was 60 mM. Due to manual mixing and succeeding equilibration and optimization of the magnetic field homogeneity, the dead time of the experiment was approximately 5 min. During refolding, 32 experiments were recorded with 128 transients per spectrum. Processing was done with Topspin Version 1.3 (Bruker, Rheinstetten, Germany).

Iodide quenching

The presence of heavy atoms like iodide or bromide or heavy atom containing molecules like alkyl halides is known to quench the intrinsic fluorescence of aromatic or heteroaromatic fluorophores.⁶¹ The collision-based dynamic quenching process can be described by the Stern-Volmer equation:

$$\frac{F_0}{F_Q} = 1 + K_{SV}[Q] \quad (2)$$

where F_0 denotes the fluorescence intensity in the absence of the quencher and F_Q the intensity in its presence which is linearly decreasing with the quencher concentration $[Q]$. The Stern-Volmer constant K_{SV} increases with increasing quencher accessibility of the chromophore. For quenching experiments, tryptophane fluorescence was excited at 295 nm and monitored at 350 nm. The protein concentration was 1 μM. NaI was added from a freshly prepared 5 M stock solution containing 0.1 mM of the anti-oxidant sodium thiosulfate.

Stopped-flow measurements

Rapid-mixing folding experiments were carried out in an Applied Photophysics SX18-MV stopped-flow instrument (Applied Photophysics, Leatherhead, UK). Fluorescence was excited at 280 nm and a 335 nm cutoff filter was used for detection. Dependent on the experiment, the stopped-flow or the sequential flow mode was used with mixing ratios of 1:1, 1:11, or 1:26 and averaging was performed for 1 to 7 traces. The final protein concentration in the 2 mm observation cuvette was always 1 to 2 μM. To monitor the unfolding rate of the intermediate in the case of the reduced C_L domain, it was first accumulated in a 15 s refolding pulse at 90 mM GdmCl before unfolding was carried out.

Double jump experiments

DJ experiments are well suited to follow isomerization reactions taking place in the unfolded state of a protein.⁶² To this end, the protein is denatured for different periods of time before refolding is carried out. The change in the amplitudes of different refolding phases directly corresponds to the amount of different species present in the unfolded state.

The C_L domain was unfolded for different times in 2 M GdmCl before refolding in the presence of 50 to 75 mM residual GdmCl was carried out. Slow refolding reactions were followed by manual mixing fluorescence experiments whereas for fast reactions, the sequential mixing mode of the stopped flow machine was used. It was confirmed that complete unfolding of the C_L domain is achieved within a few seconds at 2 M GdmCl, so that only isomerization processes in the unfolded C_L domain contribute to the observed changes in the refolding amplitudes.

Interrupted refolding experiments

Usually, the native conformation of a protein is not only thermodynamically but also kinetically the most stable state. This fact is exploited to follow the time-course of the formation of native or partially folded species in a protein folding reaction. In an interrupted refolding assay, *N*-test

or *I*-test, fully denatured protein is allowed to refold for different times before unfolding is carried out again.⁶³ Being kinetically most stable, native molecules unfold slower than all other species present. Importantly, the amplitudes of the different unfolding reactions directly mirror the amount of different species present. By selecting a final denaturant concentration where intermediates and the native state unfold with different rates, the simultaneous observation of their formation and decay becomes possible.

The C_L domain was unfolded in 1.5 M GdmCl (if oxidized) or 1.0 M GdmCl (if reduced) before refolding for different times at a residual GdmCl concentration of 140 mM or 90 mM, respectively, was carried out. Independent of the redox state, the following unfolding step was performed in 1.3 M GdmCl. For all measurements, the sequential mode of the stopped flow machine was used. The final protein concentration was always 1 to 2 μM.

Analysis of the kinetic data

A three-state on- or off-pathway model was used to fit the data derived from C_L folding experiments. For the analysis of chevron plots, the exact analytical solution was used to fit the fast and intermediate phase.^{64,44} For the three-state on-pathway model:



the characteristic equation is given by:

$$f(\lambda) = \lambda^2 - \lambda(k_{UI} + k_{IU} + k_{IN} + k_{NI}) + (k_{UI}k_{IN} + k_{UI}k_{NI} + k_{IU}k_{NI}) = 0 \quad (4)$$

whereas for the three-state off-pathway model:

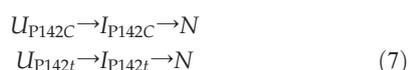


the characteristic equation is given by:

$$f(\lambda) = \lambda^2 - \lambda(k_{UI} + k_{IU} + k_{UN} + k_{NU}) + (k_{IU}k_{UN} + k_{IU}k_{NU} + k_{UI}k_{NU}) = 0 \quad (6)$$

The apparent rate constants λ_1 and λ_2 are the solutions of the quadratic characteristic equations. The logarithm of all microscopic rate constants was assumed to be linearly dependent on the denaturant concentration.⁴⁷ The slowest phase, i.e. Pro-isomerization rates, were fit as described by Goto and Hamaguchi.⁴⁰

For interrupted refolding experiments, on the other hand, irreversibility of the folding reactions was assumed to reduce the number of parameters, i.e. the data for the oxidized protein were numerically fit to the following model assuming two parallel pathways:



or in the case of the reduced protein:



where *c* and *t* stand for *cis* and *trans*. In equation (7), the rate for the formation of the intermediate was assumed to be equal on both parallel pathways. The Chemical

Reactions module from the Berkeley Madonna software package was used to fit all data (Berkeley, CA)†. All graphs were created with Origin Pro Version 7.5 (Origin Lab, Northampton, MA, USA).

Acknowledgements

The authors thank Stefan Walter, Titus M. Franzmann and Emma R. Simpson for helpful comments on the manuscript. Funding of Matthias J. Feige by the Studienstiftung des deutschen Volkes and of J. Buchner by the Fonds der chemischen Industrie is gratefully acknowledged. We are grateful to the Max-Buchner Forschungsstiftung for the funding of our work on antibodies.

References

1. Pace, C. N., Grimsley, G. R., Thomson, J. A. & Barnett, B. J. (Aug 1988). Conformational stability and activity of ribonuclease T1 with zero, one, and two intact disulfide bonds. *J. Biol. Chem.* **263**, 11820–11825.
2. Doig, A. J. & Williams, D. H. (Jan 1991). Is the hydrophobic effect stabilizing or destabilizing in proteins? The contribution of disulphide bonds to protein stability. *J. Mol. Biol.* **217**, 389–398.
3. van den Burg, B., Dijkstra, B. W., van der Vinne, B., Stulp, B. K., Eijssink, V. G. & Venema, G. (Jul 1993). Introduction of disulfide bonds into *Bacillus subtilis* neutral protease. *Protein Eng.* **6**, 521–527.
4. Hinck, A. P., Truckses, D. M. & Markley, J. L. (Aug 1996). Engineered disulfide bonds in staphylococcal nuclease: effects on the stability and conformation of the folded protein. *Biochemistry*, **35**, 10328–10338.
5. Mason, Jody M., Gibbs, Nicholas, Sessions, Richard B. & Clarke, Anthony R. (Oct 2002). The influence of intramolecular bridges on the dynamics of a protein folding reaction. *Biochemistry*, **41**, 12093–12099.
6. Vogl, T., Brengelmann, R., Hinz, H. J., Scharf, M., Lotzbeyer, M. & Engels, J. W. (Dec 1995). Mechanism of protein stabilization by disulfide bridges: calorimetric unfolding studies on disulfide-deficient mutants of the alpha-amylase inhibitor tendamistat. *J. Mol. Biol.* **254**, 481–496.
7. Balbach, J., Seip, S., Kessler, H., Scharf, M., Kashani-Poor, N. & Engels, J. W. (Nov 1998). Structure and dynamic properties of the single disulfide-deficient alpha-amylase inhibitor [C45A/C73A]tendamistat: an NMR study. *Proteins: Struct. Funct. Genet.* **33**, 285–294.
8. Vaz, Daniela C., Rodrigues, J. Rui, Sebald, Walter, Dobson, Christopher M. & Brito, Rui M. M. (Jan 2006). Enthalpic and entropic contributions mediate the role of disulfide bonds on the conformational stability of interleukin-4. *Protein Sci.* **15**, 33–44.
9. Schwalbe, H., Fiebig, K. M., Buck, M., Jones, J. A., Grimshaw, S. B., Spencer, A. *et al.* (Jul 1997). Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea. *Biochemistry*, **36**, 8977–8991.

† <http://www.berkeleymadonna.com>

10. Clarke, J., Hounslow, A. M., Bond, C. J., Fersht, A. R. & Daggett, V. (Dec 2000). The effects of disulfide bonds on the denatured state of barnase. *Protein Sci.* **9**, 2394–2404.
11. Mucke, M. & Schmid, F. X. (Jun 1994). Intact disulfide bonds decelerate the folding of ribonuclease T1. *J. Mol. Biol.* **239**, 713–725.
12. Eyles, S. J., Radford, S. E., Robinson, C. V. & Dobson, C. M. (Nov 1994). Kinetic consequences of the removal of a disulfide bridge on the folding of hen lysozyme. *Biochemistry*, **33**, 13038–13048.
13. Schonbrunner, N., Pappenberger, G., Scharf, M., Engels, J. & Kiefhaber, T. (Jul 1997). Effect of preformed correct tertiary interactions on rapid two-state tendamistat folding: evidence for hairpins as initiation sites for beta-sheet formation. *Biochemistry*, **36**, 9057–9065.
14. Yokota, A., Izutani, K., Takai, M., Kubo, Y., Noda, Y., Koumoto, Y. *et al.* (Feb 2000). The transition state in the folding-unfolding reaction of four species of three-disulfide variant of hen lysozyme: the role of each disulfide bridge. *J. Mol. Biol.* **295**, 1275–1288.
15. Clarke, J. & Fersht, A. R. (Apr 1993). Engineered disulfide bonds as probes of the folding pathway of barnase: increasing the stability of proteins against the rate of denaturation. *Biochemistry*, **32**, 4322–4329.
16. Roux, P., Ruoppolo, M., Chaffotte, A. F. & Goldberg, M. E. (Dec 1999). Comparison of the kinetics of S–S bond, secondary structure, and active site formation during refolding of reduced denatured hen egg white lysozyme. *Protein Sci.* **8**, 2751–2760.
17. Mason, Jody M., Cliff, Matthew J., Sessions, Richard B. & Clarke, Anthony R. (Dec 2005). Low energy pathways and non-native interactions: the influence of artificial disulfide bridges on the mechanism of folding. *J. Biol. Chem.* **280**, 40494–40499.
18. Frech, C. & Schmid, F. X. (Aug 1995). Influence of protein conformation on disulfide bond formation in the oxidative folding of ribonuclease T1. *J. Mol. Biol.* **251**, 135–149.
19. Weissman, J. S. & Kim, P. S. (Dec 1995). A kinetic explanation for the rearrangement pathway of BPTI folding. *Nat. Struct. Biol.* **2**, 1123–1130.
20. Creighton, T. E. (Aug 1997). Protein folding coupled to disulphide bond formation. *Biol. Chem.* **378**, 731–744.
21. Ruoppolo, M., Vinci, F., Klink, T. A., Raines, R. T. & Marino, G. (Oct 2000). Contribution of individual disulfide bonds to the oxidative folding of ribonuclease A. *Biochemistry*, **39**, 12033–12042.
22. Thies, Michael J. W., Talamo, Fabio, Mayer, Marcus, Bell, Stefan, Ruoppolo, Margherita, Marino, Gennaro & Buchner, Johannes (Jun 2002). Folding and oxidation of the antibody domain C(H)3. *J. Mol. Biol.* **319**, 1267–1277.
23. Abkevich, V. I. & Shakhnovich, E. I. (Jul 2000). What can disulfide bonds tell us about protein energetics, function and folding: simulations and bioinformatics analysis. *J. Mol. Biol.* **300**, 975–985.
24. Augustine, J. G., de La Calle, A., Knarr, G., Buchner, J. & Frederick, C. A. (Feb 2001). The crystal structure of the Fab fragment of the monoclonal antibody MAK33. Implications for folding and interaction with the chaperone bip. *J. Biol. Chem.* **276**, 3287–3294.
25. Bork, P., Holm, L. & Sander, C. (Sep 1994). The immunoglobulin fold. Structural classification, sequence patterns and common core. *J. Mol. Biol.* **242**, 309–320.
26. Isenman, D. E., Lancet, D. & Pecht, I. (Jul 1979). Folding pathways of immunoglobulin domains. The folding kinetics of the Cgamma3 domain of human IgG1. *Biochemistry*, **18**, 3327–3336.
27. Goto, Y. & Hamaguchi, K. (Apr 1982). Unfolding and refolding of the constant fragment of the immunoglobulin light chain. *J. Mol. Biol.* **156**, 891–910.
28. Jager, M. & Pluckthun, A. (Feb 1999). Folding and assembly of an antibody Fv fragment, a heterodimer stabilized by antigen. *J. Mol. Biol.* **285**, 2005–2019.
29. Thies, M. J., Mayer, J., Augustine, J. G., Frederick, C. A., Lilie, H. & Buchner, J. (Oct 1999). Folding and association of the antibody domain CH3: prolyl isomerization precedes dimerization. *J. Mol. Biol.* **293**, 67–79.
30. Feige, Matthias J., Walter, Stefan & Buchner, Johannes (Nov 2004). Folding mechanism of the CH2 antibody domain. *J. Mol. Biol.* **344**, 107–118.
31. Clarke, J., Hamill, S. J. & Johnson, C. M. (Aug 1997). Folding and stability of a fibronectin type III domain of human tenascin. *J. Mol. Biol.* **270**, 771–778.
32. Freund, C., Honegger, A., Hunziker, P., Holak, T. A. & Pluckthun, A. (Jun 1996). Folding nuclei of the scFv fragment of an antibody. *Biochemistry*, **35**, 8457–8464.
33. Hamill, S. J., Steward, A. & Clarke, J. (Mar 2000). The folding of an immunoglobulin-like Greek key protein is defined by a common-core nucleus and regions constrained by topology. *J. Mol. Biol.* **297**, 165–178.
34. Cota, E., Steward, A., Fowler, S. B. & Clarke, J. (Feb 2001). The folding nucleus of a fibronectin type III domain is composed of core residues of the immunoglobulin-like fold. *J. Mol. Biol.* **305**, 1185–1194.
35. Hoyer, Wolfgang, Ramm, Kathrin & Pluckthun, Andreas (May 2002). A kinetic trap is an intrinsic feature in the folding pathway of single-chain Fv fragments. *Biophys. Chemist.* **96**, 273–284.
36. Sanchez, Ignacio E. & Kiefhaber, Thomas (Jan 2003). Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J. Mol. Biol.* **325**, 367–376.
37. Vu, Ngoc-Diep, Feng, Hanqiao & Bai, Yawen (Mar 2004). The folding pathway of barnase: the rate-limiting transition state and a hidden intermediate under native conditions. *Biochemistry*, **43**, 3346–3356.
38. White, George W. N., Gianni, Stefano, Grossmann, J. Gunter, Jemth, Per, Fersht, Alan R. & Daggett, Valerie (Jul 2005). Simulation and experiment conspire to reveal cryptic intermediates and a slide from the nucleation-condensation to framework mechanism of folding. *J. Mol. Biol.* **350**, 757–775.
39. Goto, Y. & Hamaguchi, K. (Nov 1979). The role of the intrachain disulfide bond in the conformation and stability of the constant fragment of the immunoglobulin light chain. *J. Biochem. (Tokyo)*, **86**, 1433–1441.
40. Goto, Y. & Hamaguchi, K. (Apr 1982). Unfolding and refolding of the reduced constant fragment of the immunoglobulin light chain. Kinetic role of the intrachain disulfide bond. *J. Mol. Biol.* **156**, 911–926.
41. Williams, A. F. & Barclay, A. N. (1988). The immunoglobulin superfamily—domains for cell surface recognition. *Annu. Rev. Immunol.* **6**, 381–405.
42. Cantor, C. R. & Schimmel, P. R. (1980). *Biophysical Chemistry: Part II. Techniques for the Study of Biological Structure and Function*. W.H. Freeman, San Francisco.
43. Tyrrell, H. J. V. & Harris, K. R. (1984). *Diffusion in Liquids: A Theoretical, Experimental Study*. Butterworths, London.
44. Wildegger, G. & Kiefhaber, T. (Jul 1997). Three-state model for lysozyme folding: triangular folding

- mechanism with an energetically trapped intermediate. *J. Mol. Biol.* **270**, 294–304.
45. Kiefhaber, T., Kohler, H. H. & Schmid, F. X. (Mar 1992). Kinetic coupling between protein folding and prolyl isomerization: I. Theoretical models. *J. Mol. Biol.* **224**, 217–229.
 46. Goto, Y., Tsunenaga, M., Kawata, Y. & Hamaguchi, K. (Feb 1987). Conformation of the constant fragment of the immunoglobulin light chain: effect of cleavage of the polypeptide chain and the disulfide bond. *J. Biochem. (Tokyo)*, **101**, 319–329.
 47. Aune, K. C. & Tanford, C. (Nov 1969). Thermodynamics of the denaturation of lysozyme by guanidine hydrochloride: II. Dependence on denaturant concentration at 25 degrees. *Biochemistry*, **8**, 4586–4590.
 48. Schellman, John A. (Jul 2003). Protein stability in mixed solvents: a balance of contact interaction and excluded volume. *Biophys. J.* **85**, 108–125; (Evaluation Studies).
 49. Kikuchi, H., Goto, Y., Hamaguchi, K., Ito, S., Yamamoto, M. & Nishijima, Y. (Sep 1987). Fluorescence of the tryptophyl residues of the constant fragment of the immunoglobulin light chain. *J. Biochem. (Tokyo)*, **102**, 651–656.
 50. Myers, J. K., Pace, C. N. & Scholtz, J. M. (Oct 1995). Denaturant *m* values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4**, 2138–2148.
 51. Reimer, U., Scherer, G., Drewello, M., Kruber, S., Schutkowski, M. & Fischer, G. (Jun 1998). Side-chain effects on peptidyl-prolyl *cis/trans* isomerisation. *J. Mol. Biol.* **279**, 449–460.
 52. Wright, Caroline F., Christodoulou, John, Dobson, Christopher M. & Clarke, Jane (May 2004). The importance of loop length in the folding of an immunoglobulin domain. *Protein Eng. Des. Sel.* **17**, 443–453.
 53. Cota, E. & Clarke, J. (Jan 2000). Folding of beta-sandwich proteins: three-state transition of a fibronectin type III module. *Protein Sci.* **9**, 112–120.
 54. Pace, C. N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* **131**, 266–280.
 55. Santoro, M. M. & Bolen, D. W. (May 1992). A test of the linear extrapolation of unfolding free energy changes over an extended denaturant concentration range. *Biochemistry*, **31**, 4901–4907.
 56. Liu, M., Mao, X., He, C., Huang, H., Nicholson, J. K. & Lindon, J. C. (1998). Improved WATERGATE pulse sequences for solvent suppression in NMR spectroscopy. *J. Magn. Reson.* **132**, 125–129.
 57. Jerschow, A. M. & Mueller, N. (1997). Suppression of convection artifacts in stimulated-echo diffusion experiments. Double-stimulated-echo experiments. *J. Magn. Reson.* **125**, 372–375.
 58. Stejskal, E. O. & Tanner, J. E. (1965). Spin diffusion measurements-spin echoes in presence of a time-dependent field gradient. *J. Chem. Phys.* **42**, 288–292.
 59. de la Torre, J. G., Huertas, M. L. & Carrasco, B. (2000). HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J. Magn. Reson.* **147**, 138–146.
 60. Balbach, J., Forge, V., van Nul, N. A., Winder, S. L., Hore, P. J. & Dobson, C. M. (Oct 1995). Following protein folding in real time using NMR spectroscopy. *Nat. Struct. Biol.* **2**, 865–870.
 61. Zelent, B., Kusba, J., Gryczynski, I., Johnson, M. L. & Lakowicz, J. R. (Jul 1998). Time-resolved and steady-state fluorescence quenching of *N*-acetyl-L-tryptophanamide by acrylamide and iodide. *Biophys. Chemist.* **73**, 53–75.
 62. Schmid, F. X. & Baldwin, R. L. (Oct 1978). Acid catalysis of the formation of the slow-folding species of RNase A: evidence that the reaction is proline isomerization. *Proc. Natl Acad. Sci. USA*, **75**, 4764–4768.
 63. Kiefhaber, T., Quaas, R., Hahn, U. & Schmid, F. X. (Mar 1990). Folding of ribonuclease T1. 1. Existence of multiple unfolded states created by proline isomerization. *Biochemistry*, **29**, 3053–3061.
 64. Ikai, A. & Tanford, C. (Jan 1973). Kinetics of unfolding and refolding of proteins: I. Mathematical analysis. *J. Mol. Biol.* **73**, 145–163.

Edited by F. Schmid

(Received 2 May 2006; received in revised form 25 August 2006; accepted 16 October 2006)
Available online 21 October 2006

The structure of a folding intermediate provides insight into differences in immunoglobulin amyloidogenicity

Matthias J. Feige^{*†}, Sandra Groscurth^{*†}, Moritz Marcinowski^{*}, Zu Thur Yew[‡], Vincent Truffault^{*§}, Emanuele Paci[‡], Horst Kessler^{*}, and Johannes Buchner^{*¶}

^{*}Center for Integrated Protein Science Munich and Department Chemie, Technische Universität München, Lichtenbergstrasse 4, 85747 Garching, Germany; and [‡]Astbury Centre for Structural Molecular Biology, University of Leeds, LS2 9JT Leeds, United Kingdom

Edited by Alan R. Fersht, University of Cambridge, Cambridge, United Kingdom, and approved July 2, 2008 (received for review March 20, 2008)

Folding intermediates 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, they are poorly accessible to high-resolution techniques. Here, we made use of the intrinsically slow folding reaction of an antibody domain to characterize its major folding intermediate in detail. Furthermore, by a single point mutation we were able to trap the intermediate in equilibrium and characterize it at atomic resolution. The intermediate exhibits the basic β -barrel topology, yet some strands are distorted. Surprisingly, two short strand-connecting helices conserved in constant antibody domains assume their completely native structure already in the intermediate, thus providing a scaffold for adjacent strands. By transplanting these helical elements into β_2 -microglobulin, a highly homologous member of the same superfamily, we drastically reduced its amyloidogenicity. Thus, minor structural differences in an intermediate can shape the folding landscape decisively to favor either folding or misfolding.

amyloids | NMR | protein folding | antibodies | molecular dynamics

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 (1–6). The characteristics of folding intermediates are critical in determining whether a protein is able to fold robustly or has the tendency to misfold (6, 7). A detailed structural characterization of folding intermediates is thus key for the understanding of protein folding in general. Because they are transient, however, only very few folding intermediates have so far been described in atomic detail (8–13). In general, experimental data argue for near-native topology with incompletely folded or partially misfolded structural elements such as side-chain interactions (8–13). In this context it is of particular importance that partially folded states have recently been associated with a large variety of pathologies (14). In the case of amyloid diseases—e.g., transthyretin (TTR) familial amyloid polyneuropathy (15), light chain amyloidosis (16), or dialysis-related amyloidosis (17)—it is thought that folding intermediates are key precursors for the formation of amyloid fibrils (18). Amyloids are long-lived dissociation- and degradation-resistant structures. They are made up of β -strands that are arranged into sheets lying perpendicular to the long fiber axis and possess a core cross- β structure (19). Despite the large variety of native folds shown by amyloidogenic proteins, these structural features seem to be a recurring motif in amyloids suggesting a common assembly mechanism (20). A particularly well studied example in this respect is the MHCI component β_2 -microglobulin (β_2m), a member of the widespread immunoglobulin (Ig) superfamily (21, 22). β_2m has been shown to form amyloid fibrils if partially unfolded—e.g., at acidic pH (23), where the reaction is thought to be initiated by the population of a partially folded intermediate (23, 24). When such intermediates are populated for long periods, they are particularly susceptible to misfolding and misassembly reactions as has been shown for β_2m . Its productive folding from a native-like intermediate to the native state is

limited by an intrinsically slow *trans*-to-*cis* peptidyl-prolyl isomerization reaction (24, 25). Experiments in which the critical proline residue was held in a *trans* state confirmed that this intermediate is a major determinant in amyloid formation (24, 26). In this regard, several studies showed that the most probable amyloidogenic precursor already possesses a large part of the native β -sheet topology with only the outer strands and loop regions being distorted (24, 25, 27).

Bearing in mind that intermediates are a rather general aspect of a protein folding reaction and that most polypeptides are in principle susceptible to amyloid formation (28), the question arises of how proteins avoid aggregation in the majority of cases. To address this issue we set out to study the folding pathway of the constant domain of the antibody light chain (C_L) with high structural resolution. The C_L domain is a particularly instructive model system because it also belongs to the Ig superfamily and, like β_2m , forms a β -sandwich composed of seven strands stabilized by a single disulfide bond between strands B and F (29, 30). The *cis* proline residue associated with the amyloidogenic potential of β_2m is conserved in the C_L domain (29). Furthermore, the overall folding mechanisms of the two proteins are highly similar (24, 30), each populating an intermediate state en route to the native state. Nevertheless, the C_L domain has never been directly associated with amyloidogenic diseases even if present at much higher concentrations than β_2m in the blood (31). By the structural characterization of its major folding intermediate, we show how the C_L antibody domain might avoid such harmful misfolding reactions.

Results

The Major Kinetic Folding Intermediate of C_L is Highly Structured. The C_L domain folds via an obligatory intermediate on two parallel pathways to its native state, the slower one being limited by the isomerization of the Y34–P35 bond to the native *cis* conformation (30, 32). This bond is predominantly *trans* in the unfolded state. As a consequence, only $\approx 10\%$ of the molecules are able to fold to the native state within a few seconds (30, 32), and $\approx 90\%$ of the molecules have to undergo the intrinsically slow isomerization reaction before complete folding to the native state (30, 32). At 2°C

Author contributions: M.J.F., S.G., M.M., Z.T.Y., E.P., H.K., and J.B. designed research; M.J.F., S.G., M.M., Z.T.Y., and V.T. performed research; M.J.F., S.G., M.M., Z.T.Y., V.T., E.P., H.K., and J.B. analyzed data; and M.J.F., S.G., Z.T.Y., E.P., H.K., and J.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[†]M.J.F. and S.G. contributed equally to this work.

[§]Present address: Max Planck Institute for Developmental Biology, Spemannstrasse 35, 72076 Tübingen, Germany.

[¶]To whom correspondence should be addressed. E-mail: johannes.buchner@ch.tum.de.

This article contains supporting information online at www.pnas.org/cgi/content/full/0802809105/DCSupplemental.

© 2008 by The National Academy of Sciences of the USA

had only to be corrected for 10% of the C_L molecules possessing the correct Y34–P35 isomerization state (Fig. 1A; see *Materials and Methods* for details). 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. 1A). The HSQC spectra of the native C_L domain and the folding intermediate are superimposable for some residues but non-superimposable for others where significant differences in the chemical shifts are observed (Fig. 1A). To obtain more insights into the structural properties of the intermediate, the change in the peak intensity at the native chemical shift position was followed over time for each assigned residue. In every case the change in peak intensity could be well described by a single exponential function, if not already showing a native-like intensity after the dead-time of the experiment (Fig. 1B). As can be seen in Fig. 1C, 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, initial HSQC amplitudes in the folding intermediate show interesting patterns: almost native initial amplitudes are found in correspondence of the two short helices connecting strand A and B as well as E and F and adjacent β -sheet termini suggesting that they are already in a native environment in the intermediate whereas low initial amplitudes are observed for some of the β -strands, in particular strands C and D, which suggests a lack of native structure (Fig. 1D). In Fig. 1E regions of high or low initial amplitudes are mapped on the crystal structure of C_L revealing that the two helices and their local environment are highly structured in the intermediate.

An Intermediate Structure at Equilibrium Trapped by a Single Point Mutation. The initial HSQC amplitudes only provide hints on the structural properties of the intermediate. Equilibrium spectroscopic data could provide information more directly related to structure. Therefore, we tried to trap the intermediate at equilibrium by exploiting the isomerization reaction separating the intermediate from the native state. We hypothesized that mutating the P35 residue against another amino acid that preferentially adopts a *trans* peptide bond (33), such as Ala (C_L^{P35A}), might “trap” the kinetic intermediate making it populated at equilibrium. Indeed, far-UV and near-UV CD spectra of C_L^{P35A} were found to be very similar to the respective spectra of the kinetic intermediate (data not shown). To determine the stability of the mutant in comparison to the wild type (C_L^{wt}), denaturant-induced unfolding transitions were performed. The unfolding of both proteins, C_L^{wt} and C_L^{P35A} , was a two-state process because there was concurrent loss of secondary structure (monitored by far-UV CD-spectroscopy) and tertiary structure (monitored by the change in the intrinsic tryptophan fluorescence) (Fig. 2A). Consistent with a partially folded species being the major equilibrium state, the P35A mutation led to a stability reduction of the C_L domain by >50% from $\Delta G_{\text{unfolding}} = 13.4 \pm 0.9$ kJ·mol⁻¹ for the wild-type protein to $\Delta G_{\text{unfolding}} = 6.1 \pm 0.5$ kJ·mol⁻¹ for the mutant (Fig. 2A). The cooperativity parameter of the transition decreased by 20% from $m_{\text{eq}} = 15.9 \pm 0.9$ kJ·mol⁻¹·M⁻¹ for C_L^{wt} to $m_{\text{eq}} = 12.8 \pm 0.8$ kJ·mol⁻¹·M⁻¹ for C_L^{P35A} . If the kinetic intermediate is trapped by the Pro35Ala mutation, a simplified folding mechanism devoid of the kinetic phases leading to the native state is expected for C_L^{P35A} as compared to C_L^{wt} . To assess the folding mechanism of C_L^{P35A} , a chevron plot was determined. The mutant only showed one major folding/unfolding phase in the chevron plot (Fig. 2B) giving rise to one folding and one unfolding microscopic rate constant ($k_f = 4.4 \pm 0.1$ s⁻¹/ $k_u = 0.30 \pm 0.02$ s⁻¹ at 20°C and 0 M GdmCl). In contrast to this, two separate folding phases had been reported for the major folding pathway of C_L^{wt} (30). From a global fit of the overall folding mechanism, the rate constants for each individual folding process could be determined for C_L^{wt} (30). It folds to its intermediate state with a very similar rate

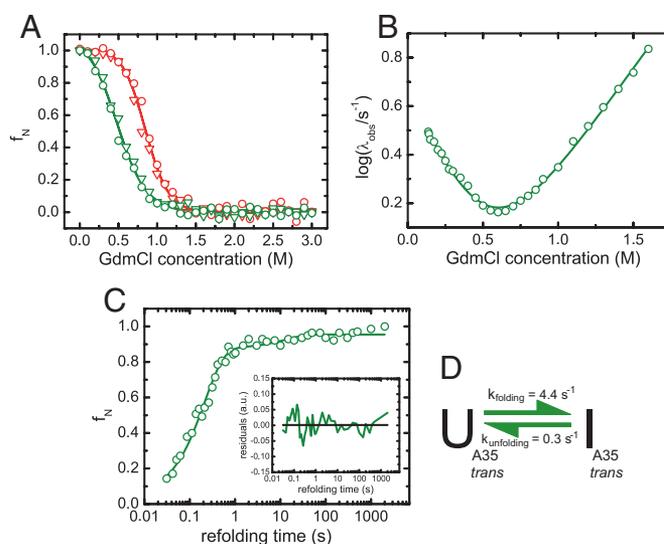


Fig. 2. Stability and folding mechanism of the C_L^{P35A} mutant. (A) Equilibrium unfolding transitions of C_L^{P35A} (green) and C_L^{wt} (red) determined by the intrinsic tryptophan fluorescence excited at 280 nm and detected at 360 nm (circles) as well as the far-UV CD-signal at 218 nm (inverted triangles). The data were fit to a two-state unfolding model. (B) Chevron plot for C_L^{P35A} determined by stopped-flow fluorescence spectroscopy. It could be described by a two-state folding reaction. (C) Formation of the final refolding species of C_L^{P35A} was followed by interrupted refolding experiments. The data were fit by a double exponential function (residuals shown as *inset*) with 93% of the molecules folding via the fast pathway. (D) The folding mechanism of C_L^{P35A} can be described by a simple two-state model neglecting the 7% slow folding species. Transitions were measured at 10 μ M protein concentration, and all kinetic experiments were performed at a final protein concentration of 2 μ M. All measurements were carried out at 20°C in PBS.

constant as C_L^{P35A} folds to its final state (30). From the derived rate constants, the stability of the C_L folding intermediate could be calculated to 11 ± 2 kJ·mol⁻¹, which is slightly higher than the stability of C_L^{P35A} ($\Delta G_{\text{unfolding}} = 6.1 \pm 0.5$ kJ·mol⁻¹). $\Delta G_{\text{unfolding}}$ for the intermediate of C_L^{wt} , however, is only an indirect estimate. Additionally, C_L^{P35A} unfolds faster than the C_L^{wt} intermediate (30) pointing towards a certain kinetic destabilization of the intermediate structure by the P35A exchange itself.

Notably, for C_L^{P35A} $\Delta G_{\text{kinetic}} = 6.6 \pm 0.3$ kJ·mol⁻¹, the stability calculated from the rate constants obtained from the chevron plot, is in very good agreement with the value from the equilibrium unfolding experiment ($\Delta G_{\text{unfolding}} = 6.1 \pm 0.5$ kJ·mol⁻¹) confirming a two-state kinetic process. This was corroborated by interrupted refolding experiments (Fig. 2C), which show that 93% of the molecules fold to their final state with a rate constant of 4.8 ± 0.4 s⁻¹ and only 7% with a rate constant of 0.08 ± 0.01 s⁻¹, which can likely be attributed to the isomerization of one of the four remaining Proline residues from a nonnative *cis* to the native *trans* state. All data argue for the C_L^{P35A} mutant populating the kinetic intermediate in equilibrium, which is trapped by the *trans* state of the bond preceding A35. The data are summarized in the folding model shown in Fig. 2D.

To confirm this, the structure of C_L^{P35A} was further characterized by NMR spectroscopy. ¹⁵N-¹H HSQC spectra of the C_L^{wt} kinetic intermediate and C_L^{P35A} were almost completely superimposable (Fig. 3A), demonstrating equivalent secondary and tertiary structure in both species. Crucially, the trapping of the intermediate state allowed an NMR assignment to be carried out (Fig. 3A), which was not feasible for the transiently populated kinetic folding intermediate observed in the folding of C_L^{wt} . To further compare the native and the intermediate states, a set of NMR measurements was recorded for both proteins (C_L^{wt} and C_L^{P35A}). These included, besides

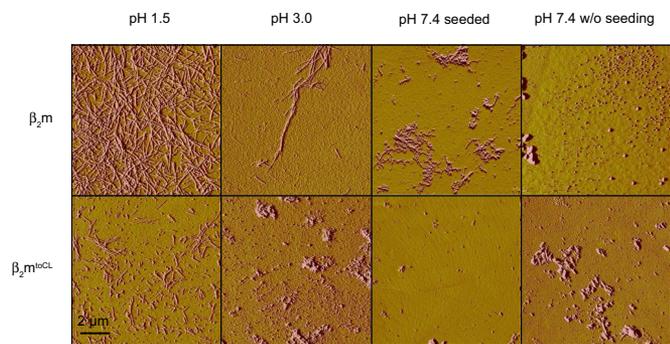


Fig. 4. Amyloidogenic properties of β_2m and the β_2m^{toCL} exchange mutant. Both proteins were incubated at pH 1.5 or 3.0 at 37°C for 7 days at a concentration of 50 μM . Additionally, the proteins were incubated under physiological conditions (PBS, 37°C) either seeded with β_2m or β_2m^{toCL} fibrils or not. Formation of amyloid fibrils was assessed by AFM measurements. Representative pictures of each sample are shown.

The data clearly show that transplanting the sequences corresponding to the C_L helices into the β_2m framework significantly reduces its amyloidogenicity.

Discussion

From a combination of NMR experiments and MD simulations we determined the ensemble of structures making up the major kinetic folding intermediate on the C_L folding pathway at atomic resolution. The structures provide hints on what sets an aggregation-prone folding intermediate apart from a structurally similar yet productive intermediate. On average, the overall β -sheet topology is well established for a major part of the protein, but most aromatic residues are still solvent-exposed or adjacent to dynamic structural elements. The only strand that seems to be highly disordered is strand D, despite native interactions of V53 and L54 with strand E. All other strands and in particular strands B, C, E, and F, which constitute the folding nucleus of Ig proteins (38, 39), exhibit some dynamics but are already well structured.

The most striking structural features of the intermediate are the two completely folded small helices. Although they are strongly conserved in constant antibody domains, their role in the folding process has not yet been recognized. These helices seem to fulfill a spacer and orienting function between strand pairs 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., Y80 in helix 2) so that they can participate in the formation of the hydrophobic core. Our data suggest that the two helices are able to fold efficiently and autonomously to their native structures in the context of the intermediate. 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.

An important protective role against amyloid formation has been attributed to the edge strands of β -sheet proteins (40, 41). Our finding that the edge strands A and G on one side of the C_L intermediate are highly structured provides one possible explanation for the marked difference in amyloidogenicity between C_L and β_2m . Grafting the sequences corresponding to the C_L helices onto the corresponding positions in β_2m significantly reduces the propensity of β_2m to form amyloids. This reduction is most pronounced at physiological pH, and the helices seem to be robust folding elements, which suggests a structural explanation for this effect but cannot completely rule out sequence effects of the transplanted elements themselves. Similar approaches to study the determinants of amyloid formation have been used previously but were targeted at increasing the amyloidogenicity of the “host” proteins rather than suppressing it as done here (42, 43). The edge strands A, D,

and G have been reported to be disturbed in the amyloidogenic intermediate of β_2m (24–27, 44). This leaves edge strands on both sides of the protein molecule unprotected, making a linear arrangement of monomers into fibrils more likely than when just one side of the protein is partially distorted, as observed for C_L . For the amyloidogenic TTR, one complete sheet of the native β -sandwich structure is destabilized in the amyloidogenic precursor (45), and for V_L , the major folding intermediate is less structured than the C_L intermediate (38). In this context, we also note that β_2m forms fibrils more readily than C_L even though it is significantly more stable than C_L (24, 30). Accordingly, it is not the stability of the native state *per se* that is the most important factor determining the amyloid forming propensity of a protein, but rather the sequence of the protein (46) and structural characteristics of partially folded species that may be populated along the folding pathway (37).

In conclusion, our data show how a high degree of local structuring in a protein folding intermediate can significantly influence the folding landscape and favor robust folding over harmful misfolding. The different characteristics 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 (47), whereas β_2m is found at much lower concentrations and usually associated with the MHCI complex (21). 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.

Materials and Methods

Protein Production and Purification. Proteins were expressed and purified as described in *SI Methods*.

Optical Spectroscopy. CD kinetics and spectra were measured as described in *SI Methods*. Equilibrium unfolding transitions, stopped-flow, and interrupted refolding experiments were performed as described in ref. 30. For the interrupted refolding experiments, C_L^{P35A} was unfolded in 1.5 M GdmCl, refolded in 136 mM GdmCl for different times, and finally unfolded again in 1.5 M GdmCl.

NMR Spectroscopy. If not stated otherwise, all spectra were recorded at 25°C on Bruker DMX600, DMX750, and AVANCE900 spectrometers as described in *SI Methods*. For folding studies, ^{15}N -labeled unfolded C_L in PBS containing 2 M GdmCl was diluted 10-fold by adding ice-cold PBS without GdmCl. Real-time ^{15}N - 1H HSQC spectra were recorded at 2°C every 14 min by using selective proton flip-back pulses (48). Identical processing of all of the spectra was performed by using the program TOPSPIN 1.3 (Bruker BioSpin). To obtain the ^{15}N - 1H HSQC of the intermediate, 10% of the final spectrum was subtracted from the first recorded HSQC spectrum. Peak intensities were analyzed by using the program SPARKY (www.cgl.ucsf.edu/home/sparky). For kinetic studies, the intensities of every amino acid during the folding process were corrected for 10% native molecules and normalized to the corresponding intensity in the final spectrum after 7 h. Backbone resonance assignments were transferred from 25°C to 2°C recording a temperature series of spectra referenced to the internal standard TSP.

MD Simulations. The C_L^{P35A} mutant was created from the crystal structure of the C_L domain (PDB entry 1FH5). The bond preceding A35 was set to *trans* and the protein was energy-minimized with the steepest-descent and the adopted Newton–Raphson methods (49). Distance and dihedral restraints derived from a comparison of the NMR chemical shifts for C_L^{wt} and the C_L^{P35A} mutant were used in conjunction with a simulated annealing (SA)-like protocol to derive a set of structures best representing the C_L^{P35A} mutant. One hundred cycles of SA were performed, and the final structures from the last 30 cycles were used for further analysis. Details about the restraints and the SA protocol are presented in *SI Methods*.

All simulations were performed with CHARMM (49), using the CHARMM19 force field, the EEF1 implicit solvent model (50), a Langevin dynamics scheme with a friction coefficient of 1 ps^{-1} , a time step of 2 fs, and holonomic constraints (SHAKE) on all bonds involving hydrogen atoms.

AFM Measurements. Fibrillization experiments and AFM measurements were performed as described in *SI Methods*.

ACKNOWLEDGMENTS. We thank Helmut Krause for performing the MS analysis and Emma R. Simpson for helpful comments on the manuscript. We are grateful for financial support of the Max-Buchner Forschungsstiftung, Deutsche Forschungsgemeinschaft Sonderforschungsbereich 749, and the

Fonds der Chemischen Industrie. M.J.F. and M.M. acknowledge a Ph.D. scholarship from the Studienstiftung des deutschen Volkes. Z.T.Y. acknowledges a Ph.D. scholarship from the Wellcome Trust and the University of Leeds.

1. Dill KA, Chan HS (1997) From Levinthal to pathways to funnels. *Nat Struct Biol* 4:10–19.
2. Sanchez IE, Kiefhaber T (2003) Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J Mol Biol* 325:367–376.
3. Korzhnev DM, et al. (2004) Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* 430:586–590.
4. Neuweiler H, Doose S, Sauer M (2005) A microscopic view of miniprotein folding: Enhanced folding efficiency through formation of an intermediate. *Proc Natl Acad Sci USA* 102:16650–16655.
5. Bai Y (2006) Energy barriers, cooperativity, and hidden intermediates in the folding of small proteins. *Biochem Biophys Res Commun* 340:976–983.
6. Brockwell DJ, Radford SE (2007) Intermediates: Ubiquitous species on folding energy landscapes? *Curr Opin Struct Biol* 17:30–37.
7. Jahn TR, Radford SE (2008) Folding versus aggregation: Polypeptide conformations on competing pathways. *Arch Biochem Biophys* 469:100–117.
8. Balbach J, et al. (1995) Following protein-folding in real-time using NMR-spectroscopy. *Nat Struct Biol* 2:865–870.
9. Mayor U, et al. (2003) The complete folding pathway of a protein from nanoseconds to microseconds. *Nature* 421:863–867.
10. Feng HQ, Zhou Z, Bai YW (2005) A protein folding pathway with multiple folding intermediates at atomic resolution. *Proc Natl Acad Sci USA* 102:5026–5031.
11. Religa TL, Markson JS, Mayor U, Freund SMV, Fersht AR (2005) Solution structure of a protein denatured state and folding intermediate. *Nature* 437:1053–1056.
12. Nishimura C, Dyson HJ, Wright PE (2006) Identification of native and non-native structure in kinetic folding intermediates of apomyoglobin. *J Mol Biol* 355:139–156.
13. Mizuguchi M, Kroon GJ, Wright PE, Dyson HJ (2003) Folding of a β -sheet protein monitored by real-time NMR spectroscopy. *J Mol Biol* 328:1161–1171.
14. Chiti F, Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* 75:333–366.
15. Hou X, Aguilar MI, Small DH (2007) Transthyretin and familial amyloidotic polyneuropathy—Recent progress in understanding the molecular mechanism of neurodegeneration. *FEBS J* 274:1637–1650.
16. Bellotti V, Mangione P, Merlini G (2000) Review: Immunoglobulin light chain amyloidosis—The archetype of structural and pathogenic variability. *J Struct Biol* 130:280–289.
17. Eakin CM, Miranker AD (2005) From chance to frequent encounters: Origins of β_2 -microglobulin fibrillogenesis. *Biochim Biophys Acta Proteins Proteomics* 1753:92–99.
18. Kelly JW (1998) The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. *Curr Opin Struct Biol* 8:101–106.
19. Fandrich M (2007) On the structural definition of amyloid fibrils and other polypeptide aggregates. *Cell Mol Life Sci* 64:2066–2078.
20. Dobson CM (2003) Protein folding and misfolding. *Nature* 426:884–890.
21. Guo HC, et al. (1992) Different length peptides bind to Hla-Aw68 similarly at their ends but bulge out in the middle. *Nature* 360:364–366.
22. Bork P, Holm L, Sander C (1994) The immunoglobulin fold—Structural classification, sequence patterns and common core. *J Mol Biol* 242:309–320.
23. McParland VJ, et al. (2000) Partially unfolded states of β_2 -microglobulin and amyloid formation in vitro. *Biochemistry* 39:8735–8746.
24. Jahn TR, Parker MJ, Homans SW, Radford SE (2006) Amyloid formation under physiological conditions proceeds via a native-like folding intermediate. *Nat Struct Mol Biol* 13:195–201.
25. Kameda A, et al. (2005) Nuclear magnetic resonance characterization of the refolding intermediate of β_2 -microglobulin trapped by non-native prolyl peptide bond. *J Mol Biol* 348:383–397.
26. Eakin CM, Berman AJ, Miranker AD (2006) A native to amyloidogenic transition regulated by a backbone trigger. *Nat Struct Mol Biol* 13:202–208.
27. McParland VJ, Kalverda AP, Homans SW, Radford SE (2002) Structural properties of an amyloid precursor of β_2 -microglobulin. *Nat Struct Mol Biol* 9:326–331.
28. Fandrich M, Fletcher MA, Dobson CM (2001) Amyloid fibrils from muscle myoglobin—Even an ordinary globular protein can assume a rogue guise if conditions are right. *Nature* 410:165–166.
29. Augustine JG, de la Calle A, Knarr G, Buchner J, Frederick CA (2001) The crystal structure of the Fab fragment of the monoclonal antibody MAK33—Implications for folding and interaction with the chaperone BiP. *J Biol Chem* 276:3287–3294.
30. Feige MJ, Hagn F, Esser J, Kessler H, Buchner J (2007) Influence of the internal disulfide bridge on the folding pathway of the CL antibody domain. *J Mol Biol* 365:1232–1244.
31. van Rhee F, et al. (2007) High serum-free light chain levels and their rapid reduction in response to therapy define an aggressive multiple myeloma subtype with poor prognosis. *Blood* 110:827–832.
32. Goto Y, Hamauchi K (1982) Unfolding and refolding of the constant fragment of the immunoglobulin light chain. *J Mol Biol* 156:891–910.
33. Pappenberger G, et al. (2001) Nonprolyl cis peptide bonds in unfolded proteins cause complex folding kinetics. *Nat Struct Biol* 8:452–458.
34. Aurora R, Rose GD (1998) Helix capping. *Protein Sci* 7:21–38.
35. Qin ZJ, Hu DM, Zhu M, Fink AL (2007) Structural characterization of the partially folded intermediates of an immunoglobulin light chain leading to amyloid fibrillation and amorphous aggregation. *Biochemistry* 46:3521–3531.
36. Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, Serrano L (2004) Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat Biotechnol* 22:1302–1306.
37. Smith DP, Jones S, Serpell LC, Sunde M, Radford SE (2003) A systematic investigation into the effect of protein destabilisation on β_2 -microglobulin amyloid formation. *J Mol Biol* 330:943–954.
38. Freund C, Honegger A, Hunziker P, Holak TA, Pluckthun A (1996) Folding nuclei of the scFv fragment of an antibody. *Biochemistry* 35:8457–8464.
39. Hamill SJ, Steward A, Clarke J (2000) The folding of an immunoglobulin-like Greek key protein is defined by a common-core nucleus and regions constrained by topology. *J Mol Biol* 297:165–178.
40. Richardson JS, Richardson DC (2002) Natural β -sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc Natl Acad Sci USA* 99:2754–2759.
41. Monsellier E, Chiti F (2007) Prevention of amyloid-like aggregation as a driving force of protein evolution. *EMBO Rep* 8:737–742.
42. Otzen DE, Kristensen O, Oliveberg M (2000) Designed protein tetramer zipped together with a hydrophobic Alzheimer homology: A structural clue to amyloid assembly. *Proc Natl Acad Sci USA* 97:9907–9912.
43. Ventura S, et al. (2004) Short amino acid stretches can mediate amyloid formation in globular proteins: The Src homology 3 (SH3) case. *Proc Natl Acad Sci USA* 101:7258–7263.
44. Hoshino M, et al. (2002) Mapping the core of the β_2 -microglobulin amyloid fibril by H/D exchange. *Nat Struct Mol Biol* 9:332–336.
45. Liu K, Cho HS, Lashuel HA, Kelly JW, Wemmer DE (2000) A glimpse of a possible amyloidogenic intermediate of transthyretin. *Nat Struct Mol Biol* 7:754–757.
46. Chiti F, Stefani M, Taddei N, Ramponi G, Dobson CM (2003) Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* 424:805–808.
47. Han JH, Batey S, Nickson AA, Teichmann SA, Clarke J (2007) The folding and evolution of multidomain proteins. *Nat Rev Mol Cell Biol* 8:319–330.
48. Diercks T, Daniels M, Kaptein R (2005) Extended flip-back schemes for sensitivity enhancement in multidimensional HSQC-type out-and-back experiments. *J Biomol NMR* 33:243–259.
49. Brooks BR, et al. (1983) CHARMM—A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* 4:187–217.
50. Lazaridis T, Karplus M (1999) Effective energy function for proteins in solution. *Proteins Struct Funct Genet* 35:133–152.

Supporting Information

Feige et al. 10.1073/pnas.0802809105

SI Methods

Protein Production and Purification. C_L^{wt} and C_L^{P35A} , the mutant generated by site-directed mutagenesis (Stratagene), were expressed, refolded, and purified as described in ref. 1. The refolding buffer was modified [250 mM Tris/HCl (pH 8.0), 100 mM L-Arg, 10 mM EDTA, 1 mM GSSG, 0.5 mM GSH]. β_2m was cloned from human cDNA (RZPD), the gene for β_2m^{10CL} was synthesized (GATC Biotech). Both proteins were cloned into the pET28a vector (Novagen) without a His-tag and expressed overnight (o.n.) at 37°C in *Escherichia coli* BL21 DE3 cells, and the inclusion bodies (IBs) were prepared as described in ref. 1. IBs were solubilized in 50 mM Tris/HCl (pH 7.5), 8 M urea, 50 mM 2-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 above. A Superdex 75 26/60 column (GE Healthcare) equilibrated in PBS was used as a final purification step for all proteins. Isotope-labeled proteins were expressed in M9 minimal medium containing either ^{15}N ammonium chloride as the only nitrogen source or additionally ^{13}C glucose as the only carbon source. All plasmids were sequenced, and the mass of each protein was confirmed by MALDI-TOF MS.

CD Spectroscopy. CD measurements were performed in a JASCO J-715 spectropolarimeter. CD spectra were accumulated 16 times and buffer-corrected. For far-UV CD measurements, 10 μM protein in a 1-mm quartz cuvette was used; near-UV CD measurements were carried out at 50 μM protein concentration in a 5-mm quartz cuvette. To obtain far-UV as well as near-UV CD spectra of the intermediate, three spectra were measured and averaged beginning after 2 min of refolding. At the chosen instrumental parameters this corresponds to a maximum measuring time of 10.5 min in which maximally 5% of the molecules fold to the native state at 2°C. The same experiment was repeated seven times independently for far-UV and three times for near-UV CD spectra and subsequently averaged. The individually averaged spectra were identical within $\pm 5\%$. The spectra of the intermediate were corrected for the 10% of native molecules that possess the correct Pro-isomerization state and hence refold to the native state within the dead time of the experiment (1). For pH transitions, 10 μM protein was incubated o.n. at the different pH values in Theorell–Stenhagen buffer (2) at 20°C before the far-UV CD signal at 218 nm was recorded. Temperature-induced unfolding transitions were monitored by the change in the far-UV CD signal at 205 nm with a heating rate of 20°C/h. Because all temperature-induced unfolding transitions were not completely reversible, an apparent melting temperature was derived by a Boltzmann fit.

NMR Spectroscopy. For C_L^{wt} as well as C_L^{P35A} , backbone sequential assignments were obtained by using standard triple resonance experiments implemented with selective proton flip-back techniques for fast pulsing (3). Aliphatic side chain assignments were

completed by using a combination of CCH-TOCSY and CCH-COSY experiments. Distance information was derived from a set of 3D-NOESY spectra, including NNH- and CNH-NOESY spectra (4), in addition to a ^{15}N -HSQC-NOESY spectrum. Dihedral angle restraints were determined for backbone Φ and Ψ angles based on C^α , C^β , C' , and H^α chemical shifts using the program TALOS (5).

MD Simulations. The restraints for simulations of C_L^{P35A} were derived from a comparison of the NMR chemical shifts of C_L^{wt} and C_L^{P35A} . (1) C^α , NH, C' , and C^β atoms of residues with chemical shifts similar to C_L^{wt} were restrained by minimizing the RMSD of the interatomic distances of the four atoms with respect to the crystal structure. (2) C^α , C' , and C^β atoms were restrained as described for restraint 1 for residues where the NH chemical shift could not be determined but otherwise had similar chemical shifts to the wild type. (3) Residues that were unassigned in the C_L^{P35A} mutant alone were restrained toward zero native contacts. In this regard, native contacts were computed by using only C^α , NH, C' , and C^β atoms, a cut-off of 0.65 nm, and for pairs of residues separated by at least five other residues in the primary sequence. (4) Φ and Ψ dihedral angles were restrained toward values determined by TALOS if available.

To minimize the number of restraints, restraints 1–3 were only applied to residues located in secondary structure elements. The ensembles of structures obtained with and without the dihedral restraints were largely similar. The restraint potential, V , used was $V(\rho, t) = (\alpha_i/2)(\rho - \rho_i)^2$, where $i = 1, 2, 3, 4$ was the type of restraint used and corresponded to 1–4 described above, ρ was defined as the mean squared difference between the calculated and target quantities to be restrained, ρ_i was the target value of the restrained quantity, and α_i was the value of the bias (kcal/mol) used to drive the restrained quantities toward the target values. A simulated annealing protocol was used to enhance the sampling of the conformational space. It was composed of six stages lasting a total of 452 ps: (i) 520 K for 62 ps, (ii) 650 K for 62 ps, (iii) 500 K for 62 ps, (iv) 420 K for 82 ps, (v) 350 K for 82 ps, and (vi) 300 K for 102 ps. A total of 100 cycles were performed.

AFM Measurements. For fibrillization experiments, a 100 μM protein solution in PBS was mixed 1:1 with buffer A (25 mM sodium acetate, 25 mM sodium phosphate) at pH 1.5 or 2.5 (final pH: 1.5 or 3.0 respectively, pH-meter reading) or PBS. The solution was incubated for 7 days, under slight shaking at 37°C. Seeds were prepared from β_2m protein fibrils (grown at pH 1.5 for 7 days) by sonication for 15 min in a sonication bath. Two microliters of seeds were used to seed 100 μl of solution, which corresponds to 1 μM seeds for 50 μM protein. For β_2m^{10CL} , seeds were prepared analogously. To prepare AFM samples, 20 μl of protein solution was spotted on freshly cleaved mica, incubated for 2 min at room temperature, washed three times with 100 μl of MilliQ water (Millipore), and dried overnight at room temperature. AFM measurements were done in contact mode with a Digital Instruments multimode scanning probe microscope (Veeco) at a scanning speed of 1.5 $\mu m/min$. DNP-S20 tips were used for all measurements (Veeco). Each experiment was repeated seven times.

1. Feige MJ, Hagn F, Esser J, Kessler H, Buchner J (2007) Influence of the internal disulfide bridge on the folding pathway of the CL antibody domain. *J Mol Biol* 365:1232–1244.
2. Theorell T, Stenhagen E (1938) Ein Universalpuffer für den pH-Bereich 2.0 bis 12.0. *Biochem Z* 299:416.
3. Diercks T, Daniels M, Kaptein R (2005) Extended flip-back schemes for sensitivity enhancement in multidimensional HSQC-type out-and-back experiments. *J Biomol NMR* 33:243–259.

4. Diercks T, Coles M, Kessler H (1999) An efficient strategy for assignment of cross-peaks in 3D heteronuclear NOESY experiments. *J Biomol NMR* 15:177–180.
5. Cornilescu G, Delaglio F, Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* 13:289–302.

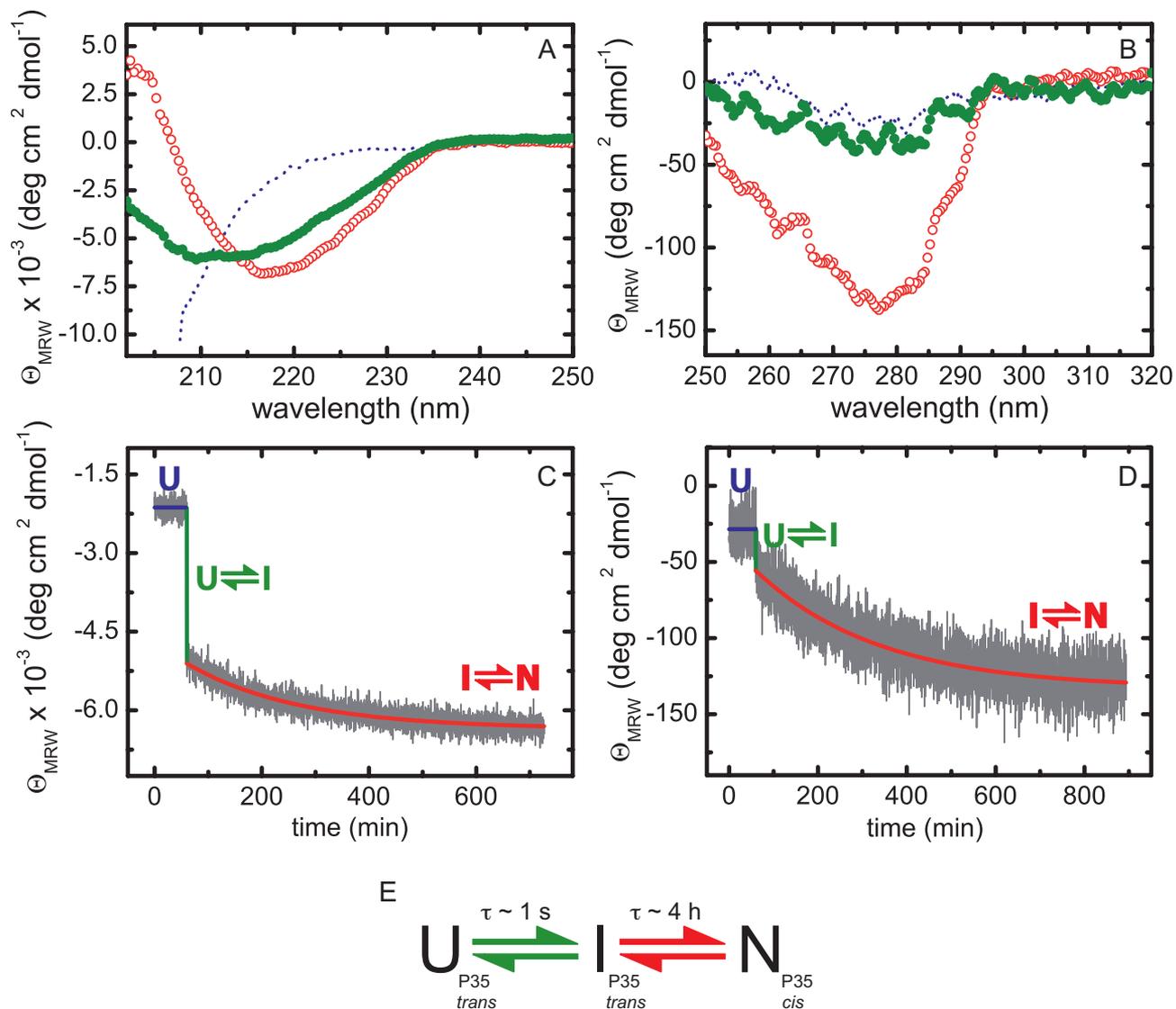


Fig. S1. Structural and kinetic characterization of the C_L domain and its major folding intermediate by CD spectroscopy. The secondary structure of the native (red open circles) and the unfolded C_L domain (blue dots) was investigated by far-UV CD spectroscopy (A) and its tertiary structure by near-UV CD-spectroscopy (B). During refolding, spectra of the major folding intermediate were measured (green, filled circles) (A and B). The spectra of the intermediate were corrected for 10% of the molecules that already possess the correct Y34–P35 isomerization state and therefore refold to the native state during the dead time of the experiment. Refolding kinetics were followed by far-UV CD spectroscopy at 218 nm (C) as well as by near-UV CD spectroscopy at 280 nm (D). The signal of the unfolded domain is shown in blue, refolding to the intermediate in green, and folding from the intermediate to the native state as a single exponential fit in red (C and D). (E) The overall folding mechanism on the slow C_L folding pathway can be described as a three-state process. Time constants for each reaction are indicated. For simplicity, the folding mechanism does not show the parallel folding pathway of the C_L domain with the Y34–P35 bond in the correct isomerization state. For far-UV CD measurements, 10 μ M protein was used, for near-UV CD measurements 50 μ M protein was used. All measurements were carried out at 2°C in PBS in the presence of a GdmCl concentration of 2 M for the unfolded protein and 0.2 M for the folded protein.

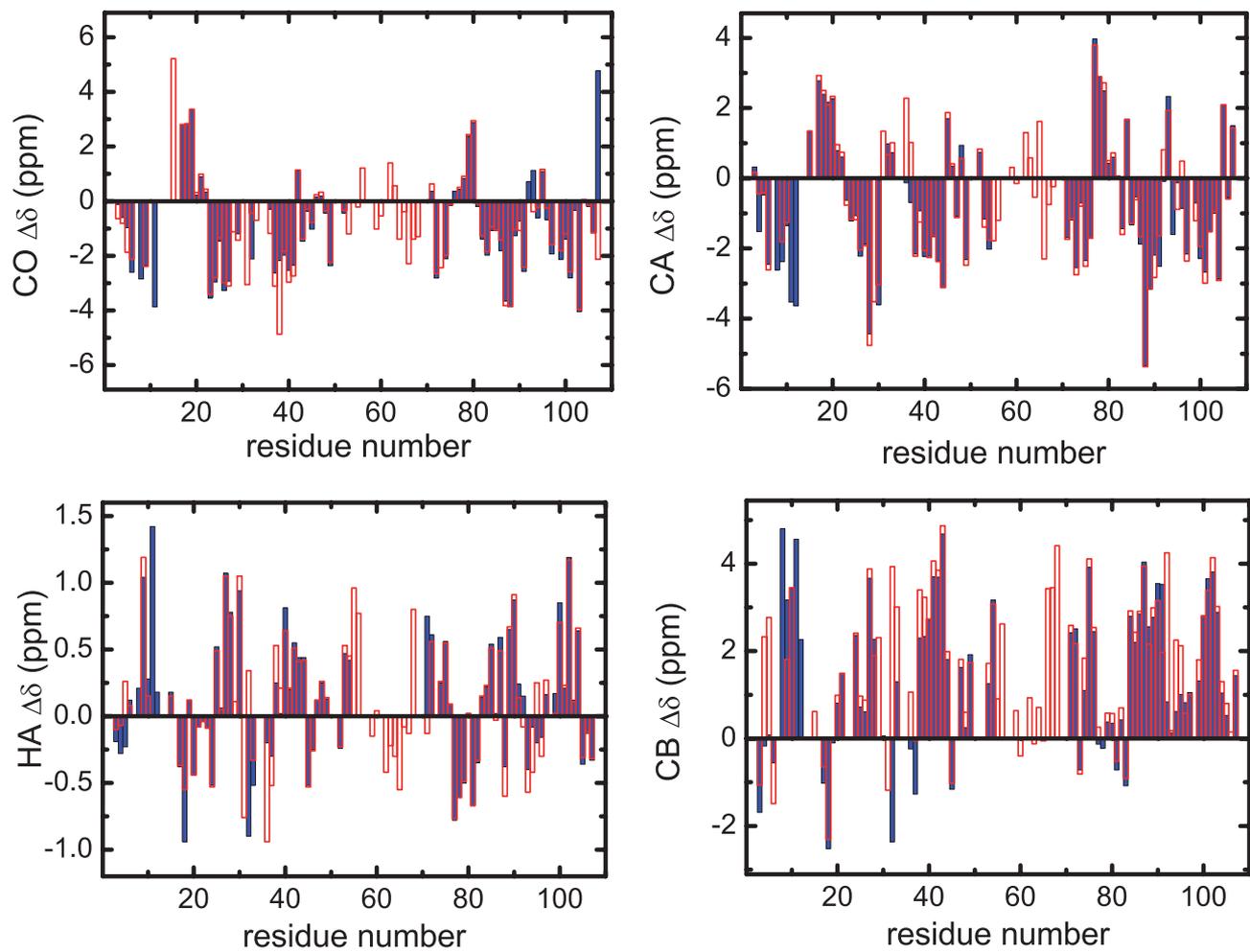


Fig. S2. Chemical-shift deviations of the nuclei C' , C^α , C^β , and H^α from the random coil values published by Wishart and colleagues [Wishart DS, Sykes BD, Richards FM (1992) *Biochemistry* 31:1647–1651; Wishart DS, Sykes BD (1994) *J Biomol NMR* 4:171–180] for C_L^{wt} shown in blue and for the mutant C_L^{P35A} shown in red.

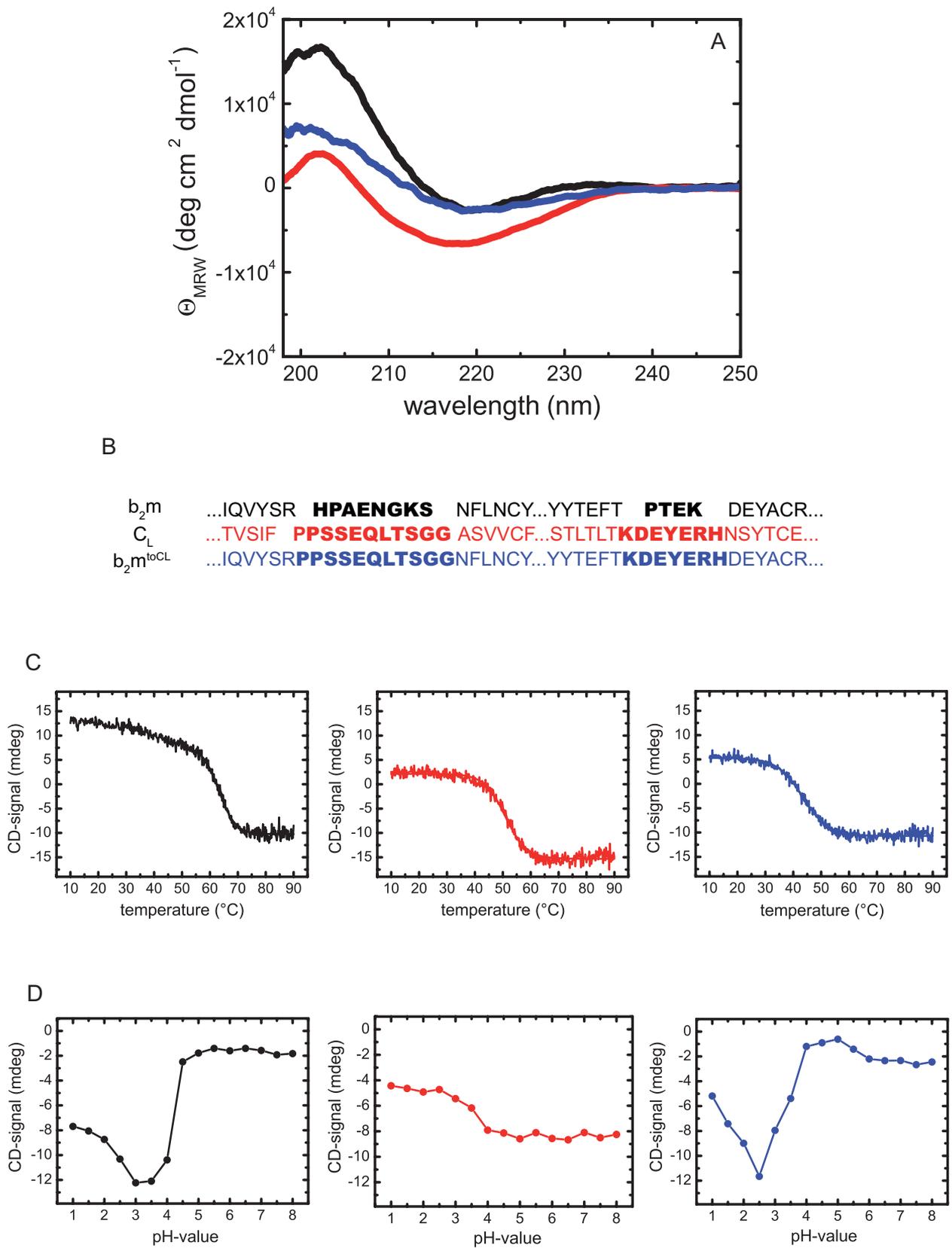


Fig. S3. Secondary structure, sequence and stability of $\beta_2\text{m}$, C_L and the $\beta_2\text{m}^{\text{toCL}}$ exchange mutant. (A) Far-UV CD spectra were measured at 20°C in PBS at 10 μM protein concentration. $\beta_2\text{m}$ is shown in black, C_L in red, and $\beta_2\text{m}^{\text{toCL}}$ in blue. (B) Sequence elements from $\beta_2\text{m}$ (black), C_L (red), and $\beta_2\text{m}^{\text{toCL}}$ (blue). Exchanged elements and the corresponding wild-type sequences are depicted in bold. The temperature-induced unfolding of all three proteins is shown in C. It was followed by far-UV CD spectroscopy at 205 nm. $\beta_2\text{m}$ (black) shows a transition midpoint of 63°C, for C_L (red) the midpoint is at 51°C, and for $\beta_2\text{m}^{\text{toCL}}$ (blue) at 43°C. (D) The pH stability of $\beta_2\text{m}$ (black), C_L (red), and $\beta_2\text{m}^{\text{toCL}}$ (blue) at 20°C was investigated at 218 nm by far-UV CD spectroscopy.

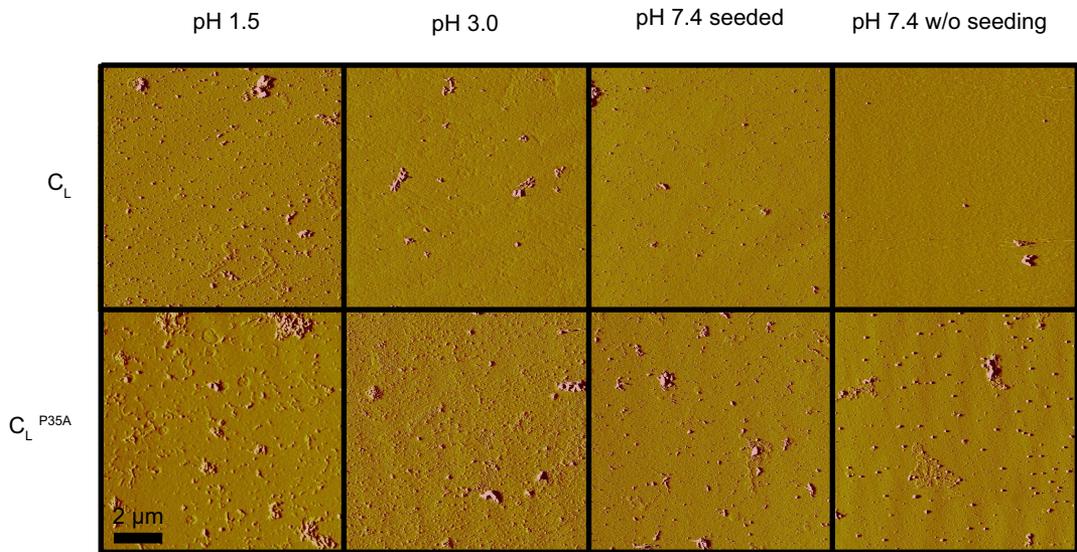


Fig. S4. Amyloidogenic properties of C_L and C_L^{P35A} . The proteins were incubated at pH 1.5 or 3.0 at 37°C for 7 days at a concentration of 50 μ M. Additionally, both proteins were incubated under physiological conditions (PBS, 37°C) either seeded with β_2m fibrils or not. Formation of amyloid fibrils was assessed by AFM measurements, and representative pictures of each sample are shown.

Structure of the Murine Unglycosylated IgG1 Fc Fragment

Matthias J. Feige¹†, Susanne Nath²†, Silvia R. Catharino¹,
Daniel Weinfurtner¹, Stefan Steinbacher¹ and Johannes Buchner¹*

¹Department Chemie and Center for Integrated Protein Science Munich, Technische Universität München, Lichtenbergstr. 4, 85747 Garching, Germany

²Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany

Received 10 March 2009;
received in revised form
16 June 2009;
accepted 17 June 2009
Available online
24 June 2009

A prototypic IgG antibody can be divided into two major structural units: the antigen-binding fragment (Fab) and the Fc fragment that mediates effector functions. The IgG Fc fragment is a homodimer of the two C-terminal domains (C_H2 and C_H3) of the heavy chains. Characteristic of the Fc part is the presence of a sugar moiety at the inner face of the C_H2 domains. The structure of this complex branched oligosaccharide is generally resolved in crystal structures of Fc fragments due to numerous well-defined sugar–protein interactions and a small number of sugar–sugar interactions. This suggested that sugars play an important role in the structure of the Fc fragment. To address this question directly, we determined the crystal structure of the unglycosylated Fc fragment of the murine IgG1 MAK33. The structures of the C_H3 domains of the unglycosylated Fc fragment superimpose perfectly with the structure of the isolated MAK33 C_H3 domain. The unglycosylated C_H2 domains, in contrast, approach each other much more closely compared to known structures of partly deglycosylated Fc fragments with rigid-body motions between 10 and 14 Å, leading to a strongly “closed” conformation of the unglycosylated Fc fragment. The glycosylation sites in the C’E loop and the BC and FG loops are well defined in the unglycosylated C_H2 domain, however, with increased mobility and with a significant displacement of about 4.9 Å for the unglycosylated Asn residue compared to the glycosylated structure. Thus, glycosylation both stabilizes the C’E-loop conformation within the C_H2 domain and also helps to ensure an “open” conformation, as seen upon Fc receptor binding. These structural data provide a rationale for the observation that deglycosylation of antibodies often compromises their ability to bind and activate Fcγ receptors.

© 2009 Elsevier Ltd. All rights reserved.

Keywords: IgG1; antibody structure; glycosylation; antibody stability; antibody effector functions

Edited by I. Wilson

Introduction

Antibodies not only are potent mediators and effectors within the arsenal of humoral immune

response but also constitute a link to cellular immunity. Among them, the IgG class is the most abundant one. This heterotetrameric glycoprotein is composed of two light chains and two heavy chains and possesses two antigen-binding

*Corresponding author. E-mail address: johannes.buchner@ch.tum.de.

† M.J.F. and S.N. contributed equally to this work.

Present addresses: S. Nath, Roche Diagnostics, Nonnenwald 2, 82377 Penzberg, Germany; S. R. Catharino, Institute for Neurodegenerative Diseases, University of California, 513 Parnassus Avenue, San Francisco, CA, USA; D. Weinfurtner, MorphoSys AG, Lena-Christ-Str. 48, 82152 Martinsried/Planegg, Germany; S. Steinbacher, Proteros Biostructures GmbH, Am Klopferspitz 19, 81252 Martinsried, Germany.

Abbreviations used: Fab, antigen-binding fragment; GlcNAc, N-acetylglucosamine; Man, mannose; GdmCl, guanidinium chloride; PDB, Protein Data Bank; PBS, phosphate-buffered saline.

sites that are made up of the N-terminal domains of the light chain and the heavy chain, respectively.¹ Based on proteolytically accessible fragments, the IgG antibody can be divided into two identical antigen-binding fragments (Fabs) and the so-called Fc fragment. Fabs are connected by a flexible linker (the hinge region) to the Fc fragment. The Fc part itself is composed of two C_H3 domains that form a homodimer at the C-terminal end of the antibody molecule and two glycosylated C_H2 domains with sugar moieties located between the C_H2 domains.¹ An important role of the Fc fragment is the mediation of effector functions such as antibody-dependent cellular cytotoxicity, phagocytosis, oxidative burst, and regulation of antibody production.^{2–4} All effector functions mediated through Fc γ receptors, as well as complement factor C1q-dependent reactions, have been shown to be compromised by deglycosylation of the C_H2 domains.^{5–12} Since Fc γ receptors are constitutively expressed in different leukocytes such as macrophages, B-cells, and natural killer cells, they are an essential part of the connection between specific antibody–antigen binding and cellular responses (i.e., between humoral and cellular immune defense mechanisms).¹³ Therefore, the role of Fc glycosylation has gained much attention in molecular immunology. It was shown that a complex biantennary oligosaccharide is attached to a conserved Asn residue of the C_H2 domain.^{1,12} The oligosaccharide possesses a defined tertiary structure, as it could be resolved in crystal structures of glycosylated antibodies. Stabilization of the sugar structure is achieved primarily *via* sugar–protein contacts with the C_H2 domain and, in some cases, also *via* some sugar–sugar contacts. The core heptasaccharide is made up of four *N*-acetylglucosamine (GlcNAc) residues and three mannose (Man) residues. It is often extended by fucose at the core heptasaccharide and by galactose or galactose and *N*-acetylneuraminic acid residues at the branches.¹² The branching point of the biantennary oligosaccharide is found at the third sugar residue after the Asn attachment, which is a Man.¹⁴ The first crystal structure of a complex between an IgG Fc and a type III Fc γ receptor suggested that almost no direct interactions between the sugar moieties and the Fc γ receptor exist¹⁵ as in that structure binding mainly takes place at the C_H2 domains and lower hinge regions.¹⁵ However, a strong correlation between trimming of the Fc oligosaccharide, compaction of the Fc fragment, and affinity for Fc receptors, as well as the stability of the C_H2 domain, was observed^{5,14} to be mediated by protein–sugar contacts and, occasionally, sugar–sugar contacts. Removing the last four sugar residues resulted in a complete loss of Fc–Fc γ receptor interaction.^{5,14} A molecular explanation for this drastic effect (i.e., the structure of the completely unglycosylated Fc part) is still missing. Within this work, we solved this structure of the unglycosylated Fc fragment from the murine IgG1 MAK33, an established model for antibody folding,^{16–20} and provide a structural rationale for the pronounced effect of deglycosylation on Fc structure and stability.

Results

Biophysical characterization of the unglycosylated Fc fragment

Oligosaccharides attached to the antibody Fc part are special because they fulfill essential structural and functional roles in antibody molecules. It has been argued that sugars serve as a spacer between the C_H2 domains determining the shape of the Fc part,²¹ thereby influencing its affinity for Fc γ receptors and other Fc fragment-binding proteins. To investigate the structural consequences of the complete absence of oligosaccharides, we produced the Fc fragment of the murine IgG1 MAK33 in *Escherichia coli*. The Fc fragment was refolded from inclusion bodies in an established manner.²² Its secondary and tertiary structures in solution, based on far-UV and near-UV CD spectra, respectively, were highly similar to that of the authentic glycosylated Fc fragment and the enzymatically deglycosylated Fc fragment produced in eukaryotic cell cultures (data not shown).

To address the influence of glycosylation on the overall stability of the MAK33 Fc fragment, we measured temperature and guanidinium chloride (GdmCl)-induced unfolding transitions. Thermal melting curves were monitored by changes in far-UV CD signals, and GdmCl-induced unfolding transitions were followed by far-UV CD and fluorescence spectroscopy. For the glycosylated Fc fragment (Fig. 1a), an equilibrium stability of $\Delta G_{N \rightarrow U} = 50 \pm 2$ kJ/mol and a cooperativity parameter of $m_{eq} = 15 \pm 1$ kJ/mol M were obtained. For the enzymatically deglycosylated (Fig. 1b) and recombinant Fc fragment (Fig. 1c), stability was reduced to $\Delta G_{N \rightarrow U} = 48 \pm 1$ and $\Delta G_{N \rightarrow U} = 47 \pm 1$ kJ/mol, respectively, and the cooperativity parameter was reduced to 13 ± 1 kJ/mol M. Thus, the apparent overall stability was only reduced by 4–6%, whereas the cooperativity of the overall transition was decreased by $\sim 13\%$ upon deglycosylation. The same tendency is observed in temperature-induced unfolding transition, where a similar midpoint is observed for all Fc variants (Table 1) yet the sugar-free Fc forms show a lower apparent cooperativity of the transition (Fig. 1).

Structure determination

We determined the crystal structure of the unglycosylated Fc fragment by X-ray crystallography and molecular replacement at 2.5 Å, with an *R*-factor of 24.2% (*R*_{free} = 30.5%) (Table 2). The crystals belong to space group *P*6₁22, with one Fc monomer in the asymmetric unit.

The structure comprises all residues from Ser241 to Ser445 of MAK33. Residues Met238–Val240 at the N-terminus and residues His445–Glu450 were disordered and not defined by electron density; this corresponds to the Fc fragment, with the connection between the C_H2 domain and the C_H3

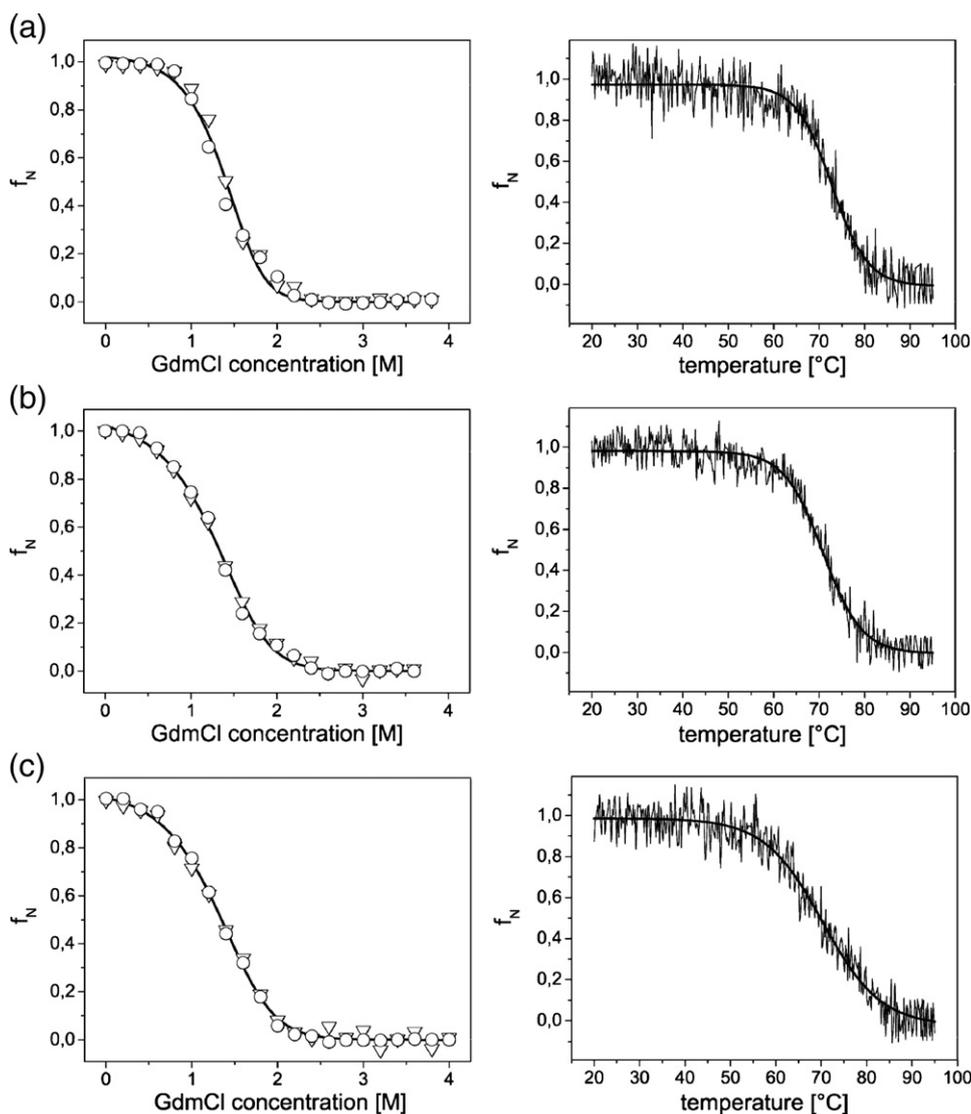


Fig. 1. Stability of the Fc fragment. The stability of the glycosylated (a), enzymatically deglycosylated (b), and recombinant unglycosylated (c) Fc fragments was assessed by GdmCl-induced (left) and temperature-induced (right) unfolding experiments. In chemical denaturation experiments, unfolding was monitored by the far-UV CD signal at 218 nm (down triangles), as well as by intrinsic tryptophan fluorescence at 354 nm (circles). Temperature melts were followed by the far-UV CD signal at 218 nm at a heating rate of 30 °C/h. GdmCl-induced unfolding data were fitted by a two-state dimer unfolding model, whereas temperature melts were fitted by a Boltzmann function. Measurements were carried out in PBS at an Fc monomer concentration of 5 μ M.

domain formed by the amino acid stretch Ser340-Ala348. The electron density of all amino acid residues of the domains was well defined; however, regions with higher temperature factors and, therefore, higher mobility were observed in the

C_H2 domain. These include the regions around the C'E, BC, and FG loops. The quality of the refined atomic model structure was confirmed with the program PROCHECK²³ by analyzing the binding geometry of the backbone of the Fc fragment.

Table 1. Stability parameters of the Fc fragment

| Sample | $\Delta G_{\text{unfolding}}$ [kJ/mol] | m_{eq} [kJ mol ⁻¹ M ⁻¹] | T_{melt} [°C] |
|-------------------|---|--|---------------------------|
| Fc glycosylated | 50±2 | 15±1 | 72.8±0.2 |
| Fc deglycosylated | 48±1 | 13±1 | 70.8±0.2 |
| Fc recombinant | 47±1 | 13±1 | 70.6±0.4 |

Thermodynamic parameters derived from GdmCl-induced unfolding transitions, as well as melting points from CD (T_{melt}) measurements, are shown.

Structure of the unglycosylated Fc fragment

The domains of the unglycosylated MAK33 Fc fragment show a typical immunoglobulin fold: In both domains, C_H2 and C_H3, two layers of anti-parallel β -sheets enclose the hydrophobic core of each domain, as shown in Fig. 2. Disulfide bridges are formed between amino acid residues Cys264 and Cys324 (C_H2 domain), and amino acid residues Cys370 and Cys428 (C_H3 domain), respectively. As

Table 2. Data collection and refinement statistics for the unglycosylated MAK33 Fc fragment

| | |
|--|---------------------|
| <i>Data collection</i> | |
| Space group | $P6_122$ |
| Unit cell dimensions a, b, c [Å] | 96.45, 96.45, 90.63 |
| Resolution range [Å] | 25–2.5 (2.5–2.6) |
| Unique reflections | 9063 (966) |
| Multiplicity | 11.4 (10.3) |
| Completeness [%] | 99.9 (100) |
| R_{merge} [%] | 5.6 (48.3) |
| $I/\sigma(I)$ | 25.7 (2.5) |
| <i>Refinement</i> | |
| Number of protein atoms | 1648 |
| Number of water molecules | 62 |
| Resolution range [Å] | 10–2.5 |
| Reflections [n (%)] | |
| Working set | 8342 (93.6) |
| Test set | 485 (5.4) |
| Completeness [%] | 99.9 (100) |
| R_{cryst} [%] ^a | 24.2 |
| R_{free} [%] ^b | 30.5 |
| rmsd from ideal geometry | |
| Bond length [Å] | 0.011 |
| Bond angle [°] | 1.66 |
| Ramachandran plot [%] | |
| Residues in most favored regions | 80.6 |
| Residues in additionally allowed regions | 14.4 |
| Residues in generously allowed regions | 3.9 |
| Residues in disallowed regions (Asn387 and Lys417) | 1.1 |
| Nonglycine and nonproline residues [n (%)] | 180 (100.0) |
| Average B -factor [Å ²] | |
| Protein | 55.3 |
| Water | 47.2 |
| Ions | 70.2 |

Values for the outer-resolution shell are given in parentheses.

$$^a R_{\text{cryst}} = \frac{\sum_{hkl \subset W} ||F_{\text{obs}}| - k|F_{\text{calc}}||}{\sum_{hkl \subset W} |F_{\text{obs}}|}$$

$$^b R_{\text{free}} = \frac{\sum_{hkl \subset T} ||F_{\text{obs}}| - k|F_{\text{calc}}||}{\sum_{hkl \subset T} |F_{\text{obs}}|}$$

judged by electron density, the disulfide bridges are fully formed. These are oriented perpendicular to the β -sheets and are located in the hydrophobic core of the protein.

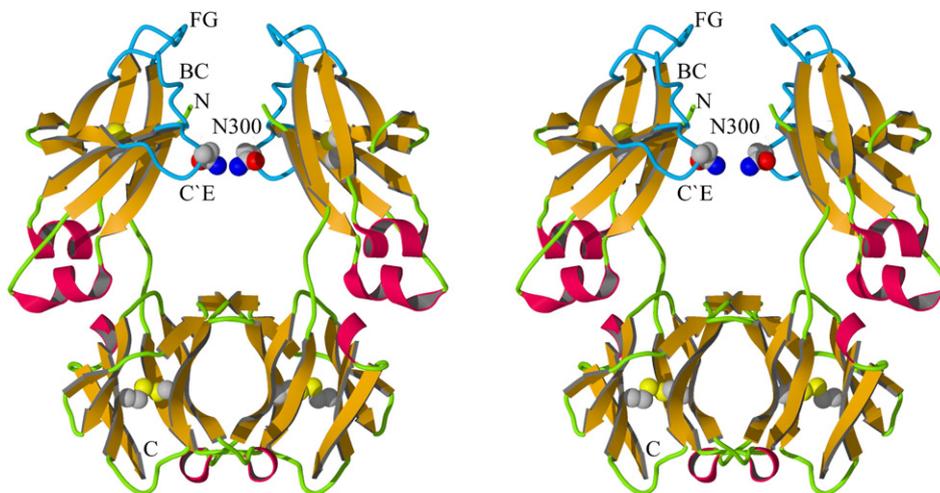


Fig. 2. Crystal structure of the unglycosylated MAK33 IgG Fc fragment. The internal disulfide bridges and Asn300 (marking the N-glycosylation site) are shown. The N-terminus and C-terminus of one Fc fragment are labeled. The C'E, BC, and FG loops, which show increased flexibility in the unglycosylated C_{H2} domain, are depicted in blue.

A common feature of many members of the immunoglobulin superfamily is the presence of at least one *cis*-proline residue in the native structure.²⁴ Peptidyl–prolyl *cis/trans* isomerization reactions have been found as a rate-limiting step in the folding of a number of proteins.²⁵ In agreement with the already solved crystal structure of the C_{H3} domain of MAK33 from mouse,¹⁸ Pro377, which is located in a loop structure connecting two β -strands of the C_{H3} domain, shows a *cis*-conformation. Previous folding studies suggested the absence of any *cis*-proline residues in the MAK33 C_{H2} domain,¹⁹ which is confirmed by the crystal structure.

Figure 2 shows the physiological dimer of the MAK33 Fc fragment. It is obvious that the dimer is formed by noncovalent interactions of the C_{H3} domains. Both C_{H2} domains show no contacts with each other as they are at least 5 Å apart. The C_{H2} loop C'E, ranging from Gln293 to Phe303, contains the N-glycosylation site consisting of the recognition sequence Asn-Ser-Thr.²⁶ In the unglycosylated form of the Fc fragment of MAK33, this loop is more flexible than in structures of glycosylated Fc fragments.^{1,21,24} This seems to be due to the absence of the sugar moiety, which may have a stabilizing effect on this region.

A superposition of the unglycosylated IgG Fc fragment of MAK33 from mice with a glycosylated IgG Fc fragment from humans [Protein Data Bank (PDB) ID 1H3X²¹] is shown in Fig. 3. Due to rigid-body motions of the C_{H2} domains, significant shifts between about 10 and 14 Å are observed for the FG, BC, and C'E loops, respectively (indicated by the glycosylation site Asn297 or Asn300, respectively, in Fig. 3). These rigid-body motions result in a strongly closed conformation of the Fc fragment. The C_{H3} domains, including the loop structures, superimpose very well. When superimposing only the C_{H2} domains (data not shown), it is observed that they retain their structural integrity in the absence of oligosaccharides. However, the location and precise

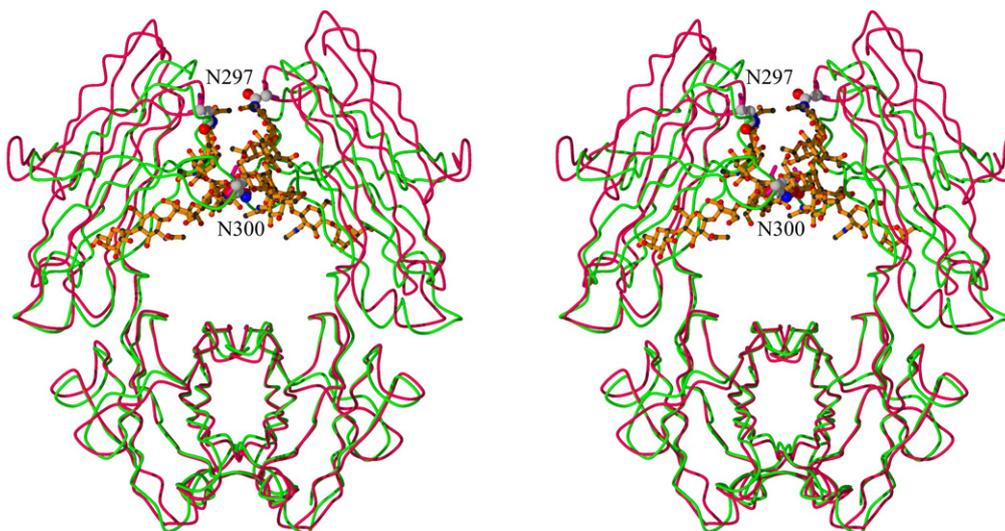


Fig. 3. Comparison of the unglycosylated murine Fc fragment with a glycosylated human counterpart. Superposition of the unglycosylated IgG Fc fragment of MAK33 (in green) with a glycosylated IgG Fc fragment from humans (PDB ID 1H3X,²¹ in magenta). The human glycosylation site Asn297, as well as the corresponding murine residue Asn300, is shown to highlight its displacement upon deglycosylation. The sugar moiety of the human Fc fragment is shown in a ball-and-stick representation.

conformations of the C'E, BC, and FG loops are changed to some extent. A major shift (4.9 Å) is observed for the C^α position of Asn300, the

glycosylation site, in the C'E loop compared to the equivalent Asn297 in the human Fc. For the BC-loop position (Lys271 in murine Fc/His268 in human Fc),

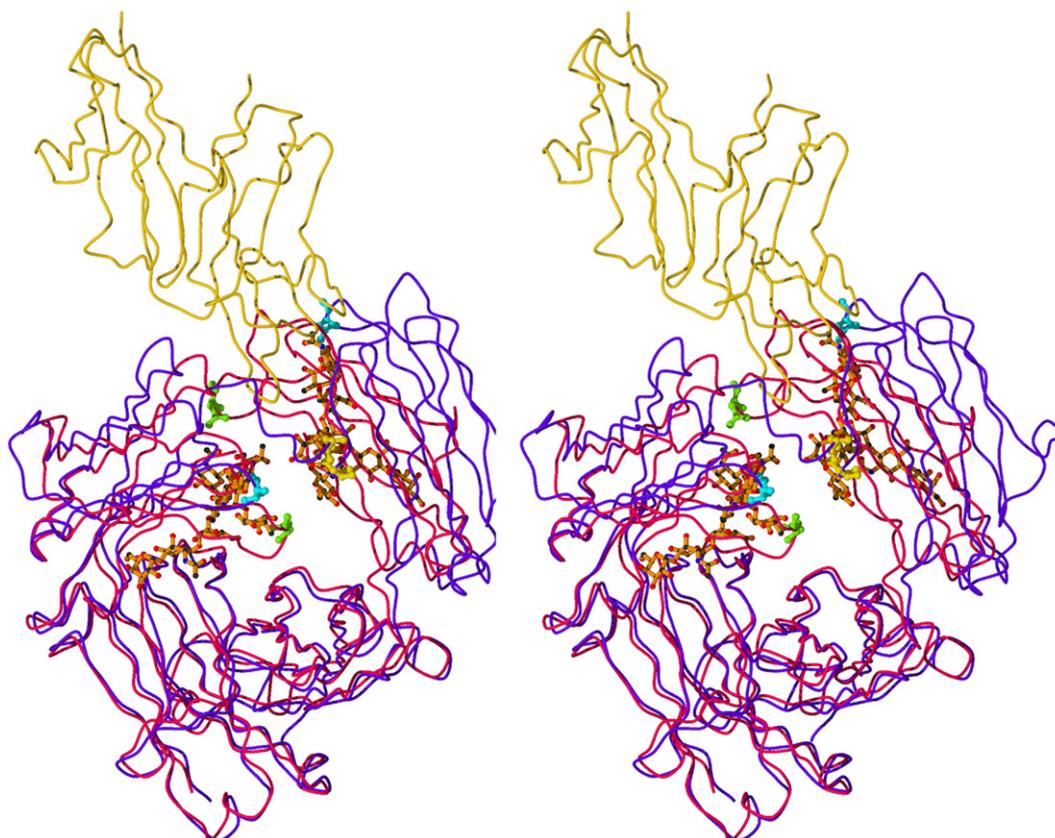


Fig. 4. Superposition of the unglycosylated IgG Fc fragment of MAK33 shown in red with a glycosylated IgG Fc fragment from humans bound to the Fc_γ receptor III given in blue and yellow, respectively (PDB ID 1E4K¹⁵). Oligosaccharides of the IgG Fc fragment are depicted in a ball-and-stick model. To illustrate the “super-closed” conformation of the unglycosylated murine MAK33 Fc fragment, we show the Asn residues of the glycosylation sites as ball-and-stick models in green (Asn300, MAK33) and blue (Asn297, human IgG Fc), respectively.

a shift of 2.6 Å is observed; in the FG loop, a displacement of 1.3 Å occurs. The overall motion is towards the binding site of the oligosaccharide. We conclude that especially the C'E loop, which has already been shown to show different conformations in different crystal structures,^{21,27,28} is supported and stabilized in the fully glycosylated antibody.

Figure 4 shows a superposition of the unglycosylated IgG Fc fragment of murine MAK33 with a glycosylated human IgG Fc fragment bound to an Fcγ receptor III (PDB ID 1E4K¹⁵). Here as well, no significant differences in the C_H3 domains are found, in contrast to the C_H2 domains, where huge movements are observed. The binding of the Fc fragment to the Fcγ receptor III is based on the interactions of five regions of the C_H2 domain with the receptor. These interactions involve rather few side chains, including the highly conserved amino acid residue Pro329, which represents one main interaction contact with the Fcγ receptor III. It is just as well ordered in the Fc fragment of MAK33 from mouse. The regions from Asp265 to Glu269 (Val268-Ile272 in mouse) and those from Asn297 to Thr299 (Asn300-Thr302 in mouse) that interact with the Fcγ receptor III are found in loops in the unglycosylated Fc fragment of the MAK33. These residues show rather small changes in their position within each domain. Although Fcγ receptor III binding induces a structural adaptation of domain positions of C_H2 domains in Fc, including a prominent deviation from 2-fold symmetry,¹⁵ the highly closed conformation observed in the present structure would require a significantly larger rearrangement that will have a prominent effect on binding affinity.

Discussion

Effect of deglycosylation on Fc overall conformation

It has been noticed that the Fc fragment assumes an "open" conformation in the complex with Fcγ receptor III¹⁵ and that stepwise deglycosylation is associated with the formation of more "closed" conformations and structural variations in the C'E loop with shifts of up to 1.5 Å.²¹ The distance between the C^α atom of Pro329A and the C^α atom of Pro329B has been used as reference, as they point towards the 2-fold axis of the Fc fragment and are located in a rigid part of the molecule. These distances vary between 33.7 Å for the (G2F)₂ glycoform with 10 monosaccharides per chain and 21.9 Å for the (MN2F)₂ glycoform with 4 monosaccharides per chain.²¹ In the present study, the distance between the corresponding Pro332A and Pro332B is only 11.6 Å, which indicates that the fully unglycosylated Fc fragment can adopt the most "closed" structure observed so far. It should be noted, however, that effects of the crystallization conditions cannot be ruled out completely. Additionally, the absence of hinge disulfide bonds in this

work might have an impact on the overall Fc conformation. However, the first resolved residue of the C_H2 domain is only 13.6 Å apart from its counterpart. Five flexible amino acids would separate the last hinge Cys residue from this first resolved residue and, hence, the observed structure would, in principle, be compatible with the presence of hinge disulfide bonds.

By the overall motions of the C_H2 domains, the C^α atom of Asn300 (the sugar attachment site) moves inwards and considerably downwards to the C_H3 platform. As a result, for assuming the Fcγ-receptor-binding-competent conformation, drastic structural rearrangements would be required from that starting conformation. For other glycoforms with known structures, at least some contacts between C_H2 domains are maintained *via* oligosaccharides, which may guide the molecules towards an active conformation.

In the absence of oligosaccharides, such a rearrangement will be entropically unfavorable. In addition, there is no support for the C'E loop and, to a lesser degree, for the BC and FG loops to adopt the binding-competent conformation. It should be noted that Fcγ receptors bind in a highly nonsymmetric fashion, requiring an induced fit.¹⁵ It involves the C'E loop (with contacts between Fcγ receptor III and GlcNAc1), the BC loop, residues from the hinge region (around Leu235) on one C_H2 domain, and residues from the hinge region and the FG loop from the other C_H2 domain. Most significantly, the disulfide-linked (Cys229A-Cys229B) hinge region connecting the Fc and Fab parts of the antibody is pushed to the direction opposite to that of the Fcγ receptor III binding site. Nevertheless, both C_H2 domains retain very similar loop conformations.¹⁵ In this complex, the Man4 moieties of the minimally required oligosaccharides have a minimum distance of about 13 Å. Shortening of the Pro329A and Pro329B distance below 20 Å would be prohibited by van der Waals contacts between the core units of the oligosaccharides.

Effect of deglycosylation on the C_H2 domain

A comparison of native stepwise deglycosylated Fc structures and our completely unglycosylated Fc structure, as well as a recently solved crystal structure of a human unglycosylated C_H2 domain,²⁹ reveals that the C'E loop becomes more flexible with decreasing sugar content. In the latter structure, the BC and FG loops are also significantly displaced compared to the native glyco form. A shift in the C'E loop has been observed in glycoforms (M3N2F)₂, which lacks the outer GlcNAc6 and GlcNAc9 of the core heptasaccharide, and (MN2F)₂, which additionally lacks Man5 and Man8, the first branched sugars of the core heptasaccharide.²¹ It is accompanied by a concomitant movement of the entire oligosaccharide moiety and a slight reduction in affinity for Fcγ receptor IIb, which was attributed to an unfavorable change in entropy.⁶ It has been suggested that the higher Fcγ receptor IIb affinities of all glycoforms containing GlcNAc sugar residues on the α(1-6) and

$\alpha(1-3)$ arms result from a rigid C'E-loop conformation that has also been observed in Fc γ receptor complexes.²¹ A conformational change in the BC loop has been also observed upon deglycosylation using ¹H NMR with His268 as reporter group.³⁰

Large conformational differences are observed for sugars of the $\alpha(1-6)$ arm in glycoform (G0F)₂, which only comprises the core heptasaccharide and usually fucose attached to GlcNAc1, with only GlcNAc6 remaining in an invariable position where it shields Phe243 (Phe246 in murine Fc) from the solvent. Strikingly, its absence in the glycoform (M3N2F)₂ is accompanied by a major decrease in melting temperature and unfolding enthalpy.⁵ In addition, longer oligosaccharide chains result in lower temperature factors for oligosaccharide moieties and C'E loops.²¹

Taking these observations together, the oligosaccharides, which are bound by multiple weak contacts, have a significant conformational variability and, by a shift of the C'E loop bearing the covalent attachment site, can also slide along the protein surface within certain limits. The overall stabilizing effect appears to involve shielding of key residues such as Phe243 from the solvent.

Stability of the Fc fragment

The C_H2 and C_H3 domains in the context of the Fc fragment are mostly regarded as independent units in terms of stability and folding. The stability of the C_H3 domain is assumed to be unaffected by the C_H2 glycosylation status, whereas C_H2 stability itself decreases with trimming of its oligosaccharide.^{6,14} We find that the enzymatically deglycosylated and unglycosylated Fc fragments are slightly less stable against GdmCl-induced denaturation than the glycosylated counterpart. The observed cooperativity is reduced more significantly. This finding can most likely be attributed to two underlying transitions: one for the C_H2 domain and one for the C_H3 domain, which cannot be dissected under the applied conditions due to very similar stabilities.^{18,19} The observed chemical stabilities for the different Fc glycoforms correspond to the ones of the isolated C_H2 or C_H3 domain, arguing against a significant impact of the structure of the Fc fragment on the chemical stability of individual domains.^{18,19} A somehow different behavior is observed in temperature-induced unfolding experiments. Again, the midpoint of the transition is only slightly affected by an unglycosylated C_H2 domain, yet cooperativity seems to be significantly reduced due to the two underlying transitions.^{6,14} The thermal stability of the unglycosylated C_H2 domain itself seems to be higher in the Fc fragment than in the isolated C_H2 domain, where unfolding around physiological temperatures has been reported.¹⁹ Even though sequence effects at the termini of the C_H2 domain, which were slightly different in a previous study and this study, cannot be ruled out completely,¹⁹ stabilization by a weak interaction with C_H3 (the partner C_H2 domain) or due to excluded volume effects in the context of the

dimeric Fc fragment is more likely to explain the differences.

Conclusion

In the absence of a short core tetrasaccharide, which comprises all sugar moieties until the branching point at Man4, the C'E loop, including Asn297 as the glycosylation site, changes its conformation. This leads to a shift of the neighboring BC loop and, to a lesser degree, also of the FG loop by about 2.2 and 1.7 Å relative to the position of the fully glycosylated Fc. In addition, this allows the Fc fragment to adopt an extremely "closed" state, indicated by a distance of only 11.6 Å between Pro332A and Pro332B (murine) compared to a distance of 30.3 Å between the corresponding residues Pro329A and Pro329B in the human Fc-Fc γ receptor III complex.¹⁵ The latter value corresponds well to that observed in native Fc.^{1,21} Stepwise truncation of the oligosaccharides induces progressively "closed" conformations with Pro329A-Pro329B distances of between 26.6 and 21.9 Å, resulting in a rather small loss of affinity (from 1 to 3 μ M).^{5,6} The absence of the four core oligosaccharide moieties obviously results mainly in the destabilization of the C'E loop within a C_H2 domain and a "super-closed" conformation of the Fc fragment. Together, these effects explain the often observed inability of unglycosylated Fc to bind and activate Fc γ receptors.

Materials and Methods

Materials

GdmCl (ultrapure) was purchased from MP Biomedicals (Eschwege, Germany). All other chemicals were obtained from Merck (Darmstadt, Germany). GdmCl concentrations were determined by the refractive indices of the solutions. Unless otherwise stated, all experiments were carried out in phosphate-buffered saline (PBS) at 25 °C.

Proteins

The glycosylated MAK33 Fc part was a kind gift from Dr. Helmut Lenz (Roche, Penzberg, Germany). For complete deglycosylation, 10 mg of the native Fc fragment was incubated with 100 U of N-glycosidase F (Roche, Mannheim, Germany) for 72 h at 37 °C. The native Fc fragment and the deglycosylated Fc fragment were purified by size-exclusion chromatography on a Superdex75pg 26/60 gel-filtration column (Amersham Biosciences, Uppsala, Sweden) equilibrated in PBS. Complete deglycosylation was confirmed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Recombinant Fc comprised MAK33 residues Glu239-Glu450 and was purified as described previously.²² All proteins used in this study were devoid of hinge disulfide bridges.

Crystallization and structure determination

The deglycosylated form of the human Fc fragment could not be crystallized in the past by dialyzing against

water or different buffers.²¹ Therefore, an extensive screening for crystallization conditions was performed. Under the final conditions using the sitting-drop vapor-diffusion method, each drop containing 2 μl of the recombinant Fc fragment with a concentration of 15 mg/ml in 50 mM potassium phosphate buffer (pH 8.0) was mixed with an equal volume of the crystallization buffer and incubated in the presence of a 300- μl crystallization buffer reservoir. The reservoir solution contained 0.05 M calcium acetate, 0.1 M imidazole/HCl (pH 8.0), and 35% 2-ethoxyethanol (Emerald Cryo I, 18; Emerald BioSystems, Bainbridge Island, WA) as crystallization buffer. Hexagonal crystals of space group $P6_122$ (cell constants of $a=96.45 \text{ \AA}$, $b=96.45 \text{ \AA}$, and $c=90.63 \text{ \AA}$), which grew to their maximum size of approximately $300 \mu\text{m} \times 300 \mu\text{m} \times 100 \mu\text{m}$ at $20 \text{ }^\circ\text{C}$ in about 10 days, were obtained after several days.

Diffraction data collection and data reduction

Because cryoprotection was given under crystallization conditions, unglycosylated Fc fragment crystals were directly frozen in a nitrogen stream of 100 K (Oxford Cryosystems, Oxford, UK). Data were collected at a wavelength of 1.5418 \AA on an image plate system (MAR Research, Norderstedt, Germany) using monochromators (Osmic mirrors) for Cu $K\alpha$ radiation from a RU200b rotating anode generator (Rigaku, Kent, UK) operated at 50 kV and 100 mA. Processing and scaling of diffraction data sets were performed using the HKL³¹ and CCP4³² program suites.

The unglycosylated Fc fragment structure was solved by molecular replacement using the program MOLREP³² with the already determined structure of the C_{H3} domain of the MAK33 from mouse¹⁸ (PDB ID 1CQK) being used as Patterson search model.

Model building and refinement

Model building was performed with the program MAIN.³³ The structure was refined with CNS.³⁴ During refinement, a rigid-body routine was employed using the Fc domains as individual rigid domains. The dihedral angles of all amino acid residues were refined to allowed values in the Ramachandran plot.³⁵ The model was further improved in successive cycles of model building and refinement. Individual isotropic B -factors were refined.

Analysis and graphical representation

Structural figures were prepared with MOLSCRIPT³⁶ and BOBSCRIPT³⁷ and were rendered with Raster3D.³⁸

Stability measurements

Fluorescence measurements were performed in a Spex FluoroMax II fluorometer (Instruments SA, Edison, NJ), and CD measurements were performed in a Jasco J-715 spectropolarimeter (Jasco, Grossumstadt, Germany). GdmCl-induced unfolding transitions were followed either by the intrinsic tryptophan fluorescence of the Fc fragment at 354 nm and excited at 280 nm or by the far-UV CD signal at 218 nm. Fluorescence measurements were carried out in a 1-cm quartz cuvette; for far-UV CD measurements, a cuvette with 1-mm pathlength was used.

At each GdmCl concentration, 50 individual data points were recorded with an integration time of 1 s at the observation wavelength and averaged. Samples were incubated for 10 days prior to measurements. GdmCl-induced unfolding transitions were completely reversible. Data were evaluated according to a two-state unfolding model for a dimeric protein:

$$N_2 \leftrightarrow 2U$$

$$S_{\text{obs}} = y_N(1 - F_U) + y_U F_U$$

$$F_U = \frac{[U]}{c_P} = \frac{\sqrt{K_U^2 + 8K_U c_P} - K_U}{4c_P}$$

$$K_U = \frac{[U]^2}{[N_2]} = e^{-\frac{\Delta G_U}{RT}} = e^{-\frac{(\Delta G_{H_2O} - m[\text{GdmCl}])}{RT}}$$

where S_{obs} denotes the observed CD or fluorescence signal at a certain GdmCl concentration, y_N is the signal of the native protein, and y_U is the signal of the unfolded protein (both signals are assumed to be linearly dependent on GdmCl concentration). The fraction of unfolded protein is termed F_U , and the total protein concentration is termed c_P . The protein stability ΔG was assumed to be linearly dependent on GdmCl concentration with the cooperativity parameter m .

Temperature-induced unfolding was followed by CD spectroscopy. Fc was unfolded at a heating rate of $30 \text{ }^\circ\text{C}/\text{h}$ in a 1-mm quartz cuvette, and the signal change at 218 nm was recorded. Melts were fitted by a Boltzmann function. In all cases, temperature melts were not completely reversible; however, in each case, three individual experiments overlaid well. Hence, no thermodynamic data could be derived from the temperature melts.

All spectroscopic measurements were carried out at a concentration of $5 \mu\text{M}$ Fc monomer.

PDB accession number

The PDB accession number is 3HKF.

Acknowledgements

We are grateful to Dr. Helmut Lenz (Roche) for the kind gift of the MAK33 Fc fragment produced in eukaryotic cells and to Helmut Krause for performing mass spectrometry.

Funding by the Studienstiftung des deutschen Volkes (M.J.F.), the Deutsche Forschungsgemeinschaft and SFB 749 (S.S. and J.B.), the Fonds der chemischen Industrie (J.B.), and the Max-Buchner-Forschungstiftung (our work on antibodies) is gratefully acknowledged.

References

1. Huber, R., Deisenhofer, J., Colman, P. M., Matsushima, M. & Palm, W. (1976). Crystallographic structure studies of an IgG molecule and an Fc fragment. *Nature*, **5585**, 415–420.
2. Ravetch, J. V. & Clynes, R. A. (1998). Divergent roles for Fc receptors and complement *in vivo*. *Annu. Rev. Immunol.* **16**, 421–432.

3. Gergely, J., Sarmay, G. & Rajnavolgyi, E. (1992). Regulation of antibody-production mediated by Fc-gamma receptors, IgG binding-factors, and IgG Fc-binding autoantibodies. *Crit. Rev. Biochem. Mol. Biol.* **3**, 191–225.
4. Morgan, E. L., Hobbs, M. V., Thoman, M. T. & Weigle, W. O. (1986). Lymphocyte-activation by the Fc region of immunoglobulins. *Immunol. Invest.* **7**, 625–687.
5. Mimura, Y., Sondermann, P., Ghirlando, R., Lund, J., Young, S. P., Goodall, M. & Jefferis, R. (2001). Role of oligosaccharide residues of IgG1-Fc in Fc gamma RIIIb binding. *J. Biol. Chem.* **276**, 45539–45547.
6. Mimura, Y., Church, S., Ghirlando, R., Ashton, P. R., Dong, S., Goodall, M. & Jefferis, R. (2000). The influence of glycosylation on the thermal stability and effector function expression of human IgG1-Fc: properties of a series of truncated glycoforms. *Mol. Immunol.* **37**, 697–706.
7. Lund, J., Tanaka, T., Takahashi, N., Sarmay, G., Arata, Y. & Jefferis, R. (1990). A protein structural change in aglycosylated IgG3 correlates with loss of huFc gamma R1 and huFc gamma R111 binding and/or activation. *Mol. Immunol.* **27**, 1145–1153.
8. Leatherbarrow, R. J., Rademacher, T. W., Dwek, R. A., Woof, J. M., Clark, A., Burton, D. R. *et al.* (1985). Effector functions of a monoclonal aglycosylated mouse IgG2A—binding and activation of complement component C1 and interaction with human monocyte Fc-receptor. *Mol. Immunol.* **4**, 407–415.
9. Dwek, R. A., Lellouch, A. C. & Wormald, M. R. (1995). Glycobiology—the function of sugar in the IgG molecule. *J. Anat.* **187**, 279–292.
10. Galon, J., Robertson, M. W., Galinha, A., Mazieres, N., Spagnoli, R., Fridman, W. H. & Sautes, C. (1997). Affinity of the interaction between Fc gamma receptor type III (Fc gamma RIII) and monomeric human IgG subclasses. Role of Fc gamma RIII glycosylation. *Eur. J. Immunol.* **8**, 1928–1932.
11. Tao, M. H. & Morrison, S. L. (1989). 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.* **8**, 2595–2601.
12. Arnold, J. N., Wormald, M. R., Sim, R. B., Rudd, P. M. & Dwek, R. A. (2007). The impact of glycosylation on the biological function and structure of human immunoglobulins. *Annu. Rev. Immunol.* **25**, 21–50.
13. Nimmerjahn, F. & Ravetch, J. V. (2008). Fc[gamma] receptors as regulators of immune responses. *Nat. Rev. Immunol.* **1**, 34–47.
14. Ghirlando, R., Lund, J., Goodall, M. & Jefferis, R. (1999). Glycosylation of human IgG-Fc: influences on structure revealed by differential scanning calorimetry. *Immunol. Lett.* **1**, 47–52.
15. Sondermann, P., Huber, R., Oosthuizen, V. & Jacob, U. (2000). The 3.2-Å crystal structure of the human IgG1 Fc fragment–FcRIII complex. *Nature*, **406**, 267–273.
16. Lilie, H., Rudolph, R. & Buchner, J. (1995). Association of antibody chains at different stages of folding: prolyl isomerization occurs after formation of quaternary structure. *J. Mol. Biol.* **1**, 190–201.
17. Lilie, H., McLaughlin, S., Freedman, R. & Buchner, J. (1994). Influence of protein disulfide isomerase (PDI) on antibody folding *in vitro*. *J. Biol. Chem.* **19**, 14290–14296.
18. Thies, M., Mayer, J., Augustine, J., Frederick, C., Lilie, H. & Buchner, J. (1999). Folding and association of the antibody domain CH3: prolyl isomerization precedes dimerization. *J. Mol. Biol.* **293**, 67–79.
19. Feige, M. J., Walter, S. & Buchner, J. (2004). Folding mechanism of the CH2 antibody domain. *J. Mol. Biol.* **1**, 107–118.
20. Feige, M. J., Groscurth, S., Marcinowski, M., Yew, Z. T., Truffault, V., Paci, E. *et al.* (2008). The structure of a folding intermediate provides insight into differences in immunoglobulin amyloidogenicity. *Proc. Natl Acad. Sci. USA*, **36**, 13373–13378.
21. Krapp, S., Mimura, Y., Jefferis, R., Huber, R. & Sondermann, P. (2003). Structural analysis of human IgG-Fc glycoforms reveals a correlation between glycosylation and structural integrity. *J. Mol. Biol.* **5**, 979–989.
22. Vinci, F., Catharino, S., Frey, S., Buchner, J., Gennaro, M., Pucci, P. & Ruoppolo, M. (2004). Hierarchical formation of disulfide bonds in the immunoglobulin Fc fragment is assisted by protein-disulfide isomerase. *Biol. Chem.* **279**, 15059–15066.
23. Laskowski, R. A., McArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK—a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.
24. Deisenhofer, J. (1981). 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*, **20**, 2361–2370.
25. Schmid, F. X. (1993). Prolyl isomerase: enzymatic catalysis of slow protein-folding reactions. *Annu. Rev. Biophys. Biomol. Struct.* **22**, 123–142.
26. Kornfeld, R. & Kornfeld, S. (1985). Assembly of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.* **54**, 631–664.
27. Matsumiya, S., Yamaguchi, Y., Saito, J., Nagano, M., Sasakawa, H., Otaki, S. *et al.* (2007). Structural comparison of fucosylated and nonfucosylated Fc fragments of human immunoglobulin G1. *J. Mol. Biol.* **3**, 767–779.
28. Yamaguchi, Y., Nishimura, M., Nagano, M., Yagi, H., Sasakawa, H., Uchida, K. *et al.* (2006). Glycoform-dependent conformational alteration of the Fc region of human immunoglobulin G1 as revealed by NMR spectroscopy. *Biochim. Biophys. Acta*, **4**, 693–700.
29. Prabakaran, P., Vu, B. K., Gan, J., Feng, Y., Dimitrov, D. S. & Ji, X. (2008). Structure of an isolated unglycosylated antibody CH3 domain. *Acta Crystallogr. Sect. D*, **64**, 1062–1067.
30. Matsuda, H., Nakamura, S., Ichikawa, Y., Kozai, K., Takano, R., Nose, M. *et al.* (1990). Proton nuclear magnetic resonance studies of the structure of the Fc fragment of human immunoglobulin G1: comparisons of native and recombinant proteins. *Mol. Immunol.* **27**, 571–579.
31. Otwinowski, Z. & Minor, W. (1997). Proceedings of X-ray diffraction data collected in oscillations mode. *Methods Enzymol.* **276**, 307–326.
32. CCP4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. Sect. D*, **50**, 760–763.
33. Turk, D. (1992). Weiterentwicklung eines Programmes für Molekülgraphik und Elektronendichte-Manipulation und seine Anwendung auf verschiedene Protein-Strukturaufklärungen. PhD Thesis.
34. Brünger, A. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–475.

-
35. Ramachandran, G. N. & Sasisekharan, V. (1968). Conformation of polypeptides and proteins. *Adv. Protein Chem.* **23**, 283–437.
 36. Kraulis, P. J. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *Appl. Crystallogr.* **24**, 946–950.
 37. Esnouf, R. M. (1997). An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph. Model.* **15**, 132–134.
 38. Merritt, E. A. & Bacon, D. J. (1997). Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524.

An Unfolded C_H1 Domain Controls the Assembly and Secretion of IgG Antibodies

Matthias J. Feige,¹ Sandra Groscurth,¹ Moritz Marcinowski,¹ Yuichiro Shimizu,² Horst Kessler,¹ Linda M. Hendershot,² and Johannes Buchner^{1,*}

¹Center for Integrated Protein Science Munich and Department Chemie, Technische Universität München, Lichtenbergstrasse 4, 85747 Garching, Germany

²Department of Genetics and Tumor Cell Biology, St. Jude Children's Research Hospital, Memphis, TN 38105, USA

*Correspondence: johannes.buchner@ch.tum.de

DOI 10.1016/j.molcel.2009.04.028

SUMMARY

A prerequisite for antibody secretion and function is their assembly into a defined quaternary structure, composed of two heavy and two light chains for IgG. Unassembled heavy chains are actively retained in the endoplasmic reticulum (ER). Here, we show that the C_H1 domain of the heavy chain is intrinsically disordered *in vitro*, which sets it apart from other antibody domains. It folds only upon interaction with the light-chain C_L domain. Structure formation proceeds via a trapped intermediate and can be accelerated by the ER-specific peptidyl-prolyl isomerase cyclophilin B. The molecular chaperone BiP recognizes incompletely folded states of the C_H1 domain and competes for binding to the C_L domain. *In vivo* experiments demonstrate that requirements identified for folding the C_H1 domain *in vitro*, including association with a folded C_L domain and isomerization of a conserved proline residue, are essential for antibody assembly and secretion in the cell.

INTRODUCTION

In eukaryotic cells, proteins destined for secretion mature within the endoplasmic reticulum (ER) and are subject to rigorous quality control prior to their transport to the Golgi (Helenius *et al.*, 1992). This usually involves surveillance of the folding status, the correct posttranslational modifications, and proper oligomerization (Helenius *et al.*, 1992; Ellgaard and Helenius, 2003; Christis *et al.*, 2008). A prominent example of this is immunoglobulin G (IgG), the most abundant antibody in the blood. It is a heterotetrameric glycoprotein assembled from two light and two heavy chains that are comprised of two and four, respectively, compact Ig domains that are structurally almost identical (Huber *et al.*, 1976). Each domain shows a β barrel topology, a two-layer sandwich structure composed of between seven and nine antiparallel β strands (Huber *et al.*, 1976; Amzel and Poljak, 1979). The fold is stabilized by an internal disulfide bridge (Goto and Hamaguchi, 1979) that is located in the hydrophobic core and lies perpendicular to the β sheets (Huber *et al.*, 1976). Most of these structural characteristics are shared by the ubiqui-

tous members of the immunoglobulin (Ig) superfamily (Bork *et al.*, 1994), which perform a broad variety of extracellular recognition functions (Williams and Barclay, 1988; Rougon and Hobert, 2003; Aricescu and Jones, 2007). The evolutionary success of the Ig superfamily has fueled a vast scope of investigations on its biophysical properties. In particular, the folding pathways of diverse members of the Ig superfamily have been studied in detail (Goto and Hamaguchi, 1982; Freund *et al.*, 1996; Thies *et al.*, 1999; Cota *et al.*, 2001; Paci *et al.*, 2003; Feige *et al.*, 2004) and have provided insights into determinants of robust folding (Hamill *et al.*, 2000; Feige *et al.*, 2008) as well as potentially harmful misfolding of this class of proteins (Kameda *et al.*, 2005; Jahn *et al.*, 2006; Qin *et al.*, 2007).

To become secreted from the cell and fulfill their biological functions, the individual domains of an antibody not only have to fold into their native tertiary structure, but furthermore must assemble into a defined quaternary structure (Porter, 1973; Huber *et al.*, 1976). Whereas isolated antibody light chains can be secreted from the ER (Coffino *et al.*, 1970; Melchers, 1971), unpaired heavy chains are actively and efficiently retained in the ER (Bole *et al.*, 1986; Hendershot *et al.*, 1987). Antibody heavy-chain and light-chain synthesis occur asynchronously during B cell development (Burrows *et al.*, 1979), and only completely assembled molecules can both bind to antigen and carry out effector functions. Therefore, tight quality control of their assembly prior to secretion is vital. It is known that the first constant domain of the heavy chain, the C_H1 domain, plays an important role in this retention process (Hendershot *et al.*, 1987; Kaloff and Haas, 1995). If deleted or replaced with another antibody domain, isolated heavy chains can be secreted, as occurs in the case of the rare heavy-chain diseases (Wolfenstein-Todel *et al.*, 1974; Adetugbo, 1978; Hendershot *et al.*, 1987), or naturally in camelid antibodies, which do not contain light chains (Hamers-Casterman *et al.*, 1993). In the context of the whole IgG molecule (Figure 1A), the C_H1 domain is associated with the constant domain of the light chain (C_L) and shows the typical immunoglobulin fold (Huber *et al.*, 1976). *In vivo*, C_H1 is the only antibody domain that is stably bound to the molecular chaperone BiP and remains in a reduced form before assembly with light chain (Vanhove *et al.*, 2001). The basis for the unusual behavior of the C_H1 domain has remained enigmatic.

Here, we set out to study the role of the C_H1/C_L association for correct antibody assembly and secretion. To our surprise, we found that C_H1 is an unfolded protein in isolation that gains

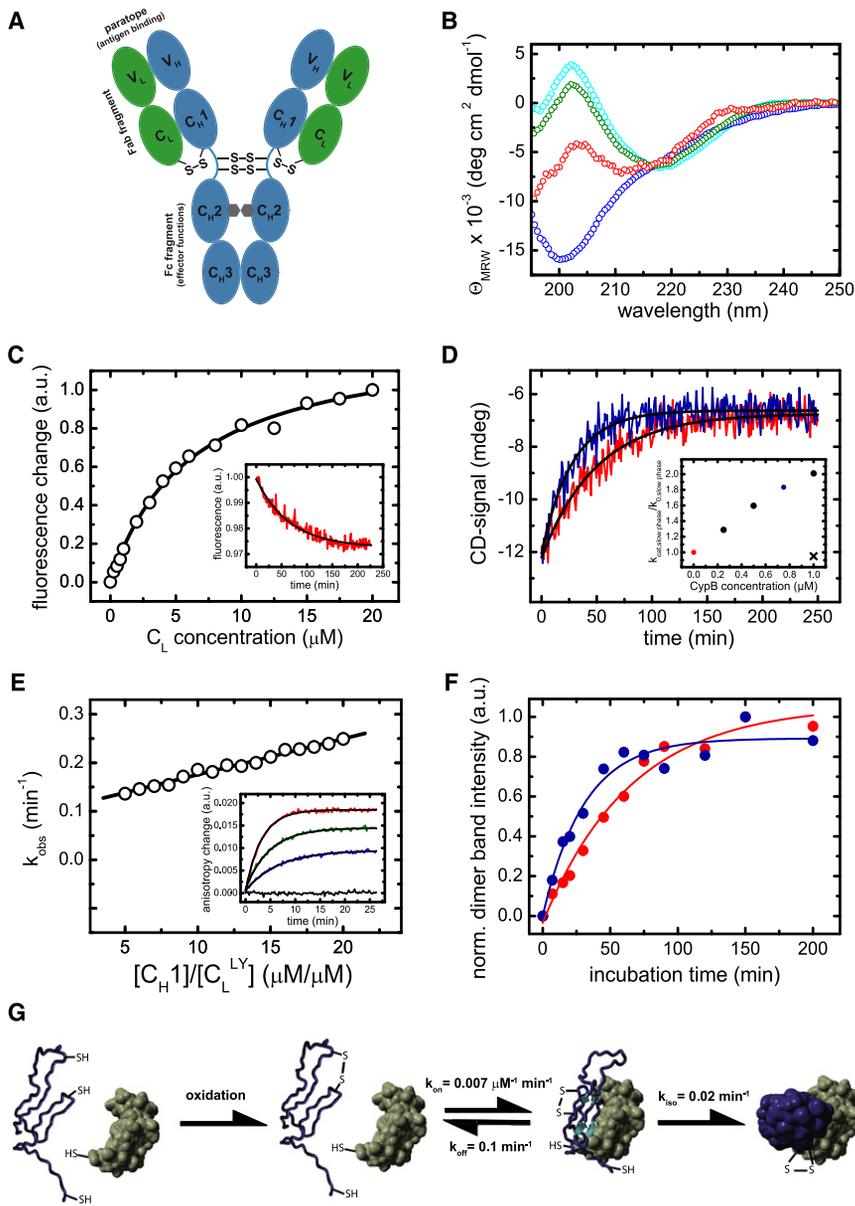


Figure 1. Structural Characteristics of the C_H1 Domain and Its Assembly Mechanism with the C_L Domain

(A) A schematic representation of the IgG molecule. The IgG molecule is composed of two heavy chains (blue) and two light chains (green). The heavy chain consists of the V_H, C_H1, C_H2, and C_H3 domains, and the light chain consists of the V_L and the C_L domains (listing from the N to the C terminus of each chain). The C_H2 domains of the heavy chains are glycosylated with complex biantennary oligosaccharides (depicted in gray). Each domain possesses an internal disulfide bridge (omitted for clarity), and additional disulfide bonds link the two heavy chains in the flexible hinge region. A single disulfide bond covalently connects C_H1 with the C_L domain. The two identical antigen-binding sites (paratopes) are made up by the two variable domains, V_H and V_L. The overall IgG molecule can be divided into two Fab fragments (composed of V_H, C_H1, V_L, and C_L) and one Fc fragment (composed of two C_H2 and two C_H3 domains).

(B) The isolated C_L domain (cyan) displays a typical all-β far-UV CD spectrum, whereas the isolated C_H1 domain (blue) shows a random coil spectrum. To assess if C_L induces structural changes in C_H1, the spectrum of coincubated C_L and C_H1 was recorded (green). From this spectrum and the far-UV CD spectrum of the isolated C_L domain, the spectrum of the C_H1 domain in the presence of C_L was calculated (red), which shows the characteristics of β sheet secondary structure.

(C and D) (C) The affinity between C_H1 and C_L was determined by the change in the intrinsic fluorescence upon C_L-induced folding of the C_H1 domain, recorded before and after a 4 hr equilibration step. A one-site binding model was used to fit the data. The inset shows a representative single exponential trace observed after the addition of 1 μM C_L to 2 μM C_H1. A single exponential reaction with a very similar rate was observed by far-UV CD spectroscopy ([D], red trace). The folding reaction could be accelerated by the PPIase CypB (red trace: 10 μM C_L and 10 μM C_H1 alone; blue trace: in the presence of 0.75 μM CypB). The inset shows the dependence of the slow reaction on CypB concentration. The observed rate in the absence and presence of CypB is denoted as k₀ and k_{cat}, respectively. If 1 μM CypB was inhibited by 2 μM cyclosporine A, no acceleration was observed (black cross).

(E) Association of the C_H1 domain with a lucifer yellow-labeled C_L domain was followed by the change in the lucifer yellow anisotropy signal. The observed rate constants were fitted with a linear function to yield the k_{on} value and the k_{off} value of 0.007 ± 0.0002 μM⁻¹ min⁻¹ and 0.1 ± 0.01 min⁻¹, respectively. The inset shows individual single exponential traces after the addition of 0 μM (black), 5 μM (blue), 10 μM (green), and 20 μM (red) C_H1 to 1 μM labeled C_L.

(F) To assess the formation of the C_L/C_H1 interchain disulfide bridge, SDS-PAGE experiments were carried out, and the dimer band intensity was quantified. 25 μM of each domain was used. In the absence of CypB (red), a time constant of τ = 63 ± 7 min was observed for the formation of covalent C_L/C_H1 dimers. In the presence of 5 μM CypB, a time constant of τ = 31 ± 5 min was obtained.

(G) The overall C_H1/C_L assembly mechanism. Only after formation of the internal disulfide bridge in the C_H1 domain (blue) is the fast formation of a dimeric intermediate with the C_L domain (green) observed. Subsequently, prolyl isomerization limits complete folding and formation of the interchain disulfide bridge. All measurements were carried out at 25°C in PBS.

structure only upon interaction with its cognate partner, C_L. Based on this finding, we analyzed the association-coupled C_H1 folding pathway and its modulation by the chaperone BiP in detail and provide a comprehensive picture for the control of antibody secretion in the cell.

RESULTS

The Murine IgG1 C_H1 Domain Is Intrinsically Disordered

Antibodies are modular structures composed of a series of structurally highly homologous domains (Figure 1A). These domains

can usually be produced and studied separately, as they represent independent structural units (Goto and Hamaguchi, 1982; Lilie et al., 1995). Surprisingly, analysis of the murine IgG1 C_H1 domain revealed that, in marked contrast to all antibody domains studied thus far (Goto and Hamaguchi, 1982; Thies et al., 1999; Feige et al., 2004; Rothlisberger et al., 2005), the isolated C_H1 domain is an unfolded protein, irrespective of whether its internal disulfide bridge is formed or not (Figure 1B; Figures S1 and S2, available online). To further characterize the unfolded state of the C_H1 domain under physiological conditions, iodide fluorescence quenching experiments were carried out. The experiments indicate no significant differences in the burial of tryptophan residues between C_H1 in PBS and in 3 M GdmCl (data not shown). Additionally, NMR experiments were recorded on a highly deuterated C_H1 sample. Deuteration enables detection of long-range NOEs even if they belong to only a subset of conformers. In these experiments, no long-range NOEs could be determined, and, consequently, no preferential conformation of C_H1 seems to persist. Taken together, these data argue against the presence of a significant amount of stable structure in the isolated C_H1 domain. However, the pattern changed completely when the C_L domain, the cognate association partner of C_H1 in the antibody, was added. Only then did we observe folding of the C_H1 domain to a well-defined β sheet structure (Figure 1B). Thus, the C_L domain is necessary and sufficient to induce structure formation in C_H1. This folding process was observed only if the internal disulfide bridge in the C_H1 domain was present (Figure S2). Based on these findings, a key role of this association-coupled folding reaction for correct antibody assembly can be anticipated, which we aimed to elucidate in more detail.

The Mechanism of Induced Folding of the C_H1 Domain

To understand how binding to C_L and folding of C_H1 are coupled, the thermodynamic and kinetic parameters for this reaction were established. The dissociation constant of the two domains was determined to be $6.2 \pm 0.4 \mu\text{M}$ (Figure 1C) by the change in intrinsic fluorescence emission upon C_L-induced C_H1 folding (Figure 1C, inset). A moderate affinity is expected because C_H1 has to fold upon binding to C_L. One should bear in mind that the observed dissociation constant is orders of magnitude lower than the antibody concentration in the ER of plasma cells (Cenci and Sitia, 2007); thus, association will readily occur in vivo. The analysis of the kinetics of secondary, tertiary, and quaternary structure formation by using far-UV CD spectroscopy (Figure 1D), near-UV CD spectroscopy, and analytical high-performance liquid chromatography (HPLC) (Figure S1) showed that all three processes occur with virtually identical time constants of $\tau = 60 \pm 10 \text{ min}$ at 25°C. Hence, all of these processes are likely rate limited by the same slow reaction. This slow folding reaction could be accelerated by the ER-specific peptidyl-prolyl isomerase cyclophilin B (Figure 1D). Thus, the slow folding phase can be attributed to the isomerization of peptidyl-prolyl bonds within the C_H1 domain, which possesses an unusually high number of three *cis* prolines in the native state (Augustine et al., 2001). Prior to the slow folding to the native structure, the C_H1 domain forms an intermediate with the C_L domain in a concentration-dependent reaction (Figure 1E). As this complex could be detected by fluorescence anisotropy measurements, but not by

the other techniques outlined above, it is likely a dynamic species with an only marginally folded C_H1 domain. In the complete antibody, a disulfide bridge covalently links the C_H1 domain with the C_L domain (Figure 1A). If the bridge-forming Cys residues were included in the C_H1 domain as well as the C_L domain, no change in the folding state of the isolated domains and the C_L-induced folding of the C_H1 domain was observed (Figure S3), but formation of covalent dimers could be readily followed by SDS-PAGE. As covalent dimers were formed with the same rate as the slow C_H1 folding reaction and the reaction could be accelerated by cyclophilin B (Figure 1F), it is clearly limited by proline isomerization and hence complete folding of the C_H1 domain.

Taken together, the C_L-induced folding of the C_H1 domain can be dissected into three reactions: first, oxidation of the internal C_H1 disulfide bridge has to take place. Then, a transient heterodimeric intermediate is formed, and, subsequently, peptidyl-prolyl isomerization is required to allow folding to the native state and covalent assembly with C_L (Figure 1G).

An Atomic Level Description of the C_H1 Folding Pathway

To resolve the specific recognition and the folding pathway of the intrinsically disordered C_H1 domain at the level of atomic resolution, NMR experiments were performed. The ¹⁵N-¹H HSQC spectrum of the isolated C_H1 domain is characteristic of an unfolded protein (Figure 2A, red spectrum), confirming the results described above. In contrast, after induced folding by C_L, the C_H1 domain shows well-dispersed spectra (Figure 2A, blue spectrum). The backbone assignment of the C_H1 domain in the complex was achieved by a combination of triple-resonance experiments and NH residual dipolar couplings (RDCs). All obtained NMR data, the carbon chemical shifts, the NH RDCs, and the MEXICO (Gemmecker et al., 1993) water exchange rates (data not shown) agree with an all- β structure for the C_H1 domain in the presence of C_L, similar to that observed in the crystal structure of IgG antibodies (Augustine et al., 2001). Because complete folding of the C_H1 domain is limited by proline isomerization, and hence associated with a high activation energy, the final folding step is significantly decelerated at low temperatures. This allowed us to characterize the trapped intermediate and resolve the association-coupled folding process by using real-time ¹⁵N-¹H HSQC experiments. For each assigned residue, changes of the amplitudes over time could be described by a single exponential function (Figure 2A, inset). Notably, some residues already exhibit significant intensities in the first spectrum recorded after 20 min (Figure 2B, red bars). These residues are likely to already adopt a native-like backbone conformation prior to the slow peptidyl-prolyl isomerization reaction. Altogether, 10 residues, which are part of the β sheets that form the mature structure, were found to be in a native-like environment in the intermediate. Mapping these residues on the crystal structure of the C_H1 domain of a murine IgG1 Fab fragment revealed how the association-coupled folding reaction of this antibody domain might proceed. Residues Thr22, His49, Ser65, and Thr67 in the C_H1 domain, which form part of the C_L interface, seem to be already correctly positioned in the intermediate (Figure 2C, left panel). Importantly, His49 and Ser65 are involved in hydrogen bonds with the C_L domain in the native state. The interaction with C_L apparently initiates the formation of a hydrophobic cluster

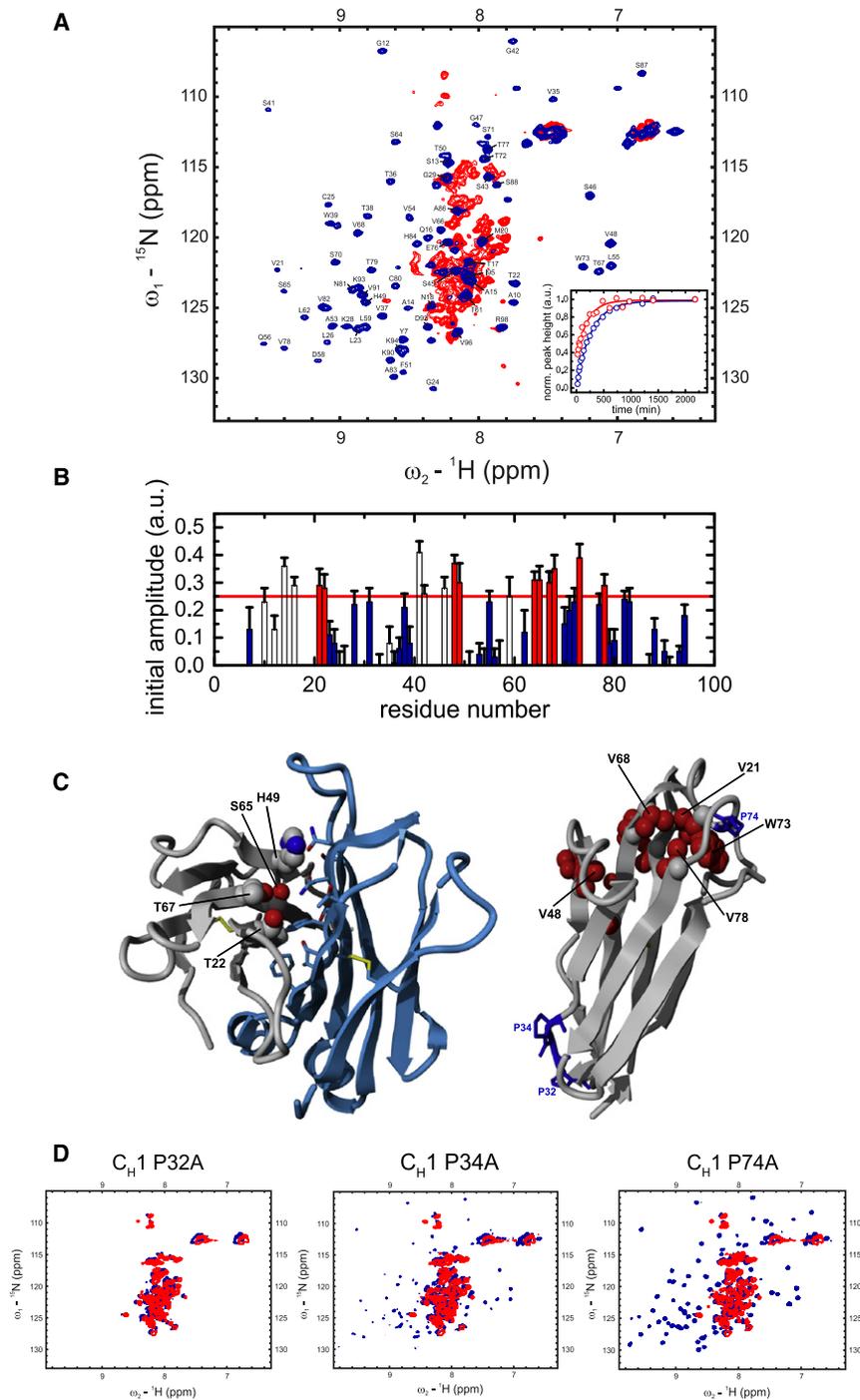


Figure 2. NMR Spectroscopic Characterization of C_L-Induced C_{H1} Folding

(A) ¹⁵N-¹H HSQC spectra of the isolated C_{H1} domain (red) and the assigned C_{H1} domain in complex with the C_L domain (blue) are shown. In order to characterize the folding pathway of the intrinsically disordered C_{H1} domain, time-dependent HSQC intensities upon addition of unlabeled C_L to ¹⁵N-labeled C_{H1} were measured for each assigned residue at the native chemical shift position and fitted by a single exponential function. Two representative traces for Val68 (red) and Lys90 (blue) are shown in the inset.

(B) Initial amplitudes for each assigned C_{H1} residue were derived from the fitted exponential functions. Residues with an initial HSQC amplitude below a threshold of 25% native intensity are colored in blue, and residues above the threshold are red (open bars, residues in loop regions; filled bars, residues in structured regions). Errors indicate standard deviations from single exponential fits.

(C) C_{H1} residues with intensities above the threshold in the intermediate are mapped on the crystal structure of the C_{H1}/C_L dimer (PDB code: 1Q9K). The dimerization interface between C_{H1} (gray) and C_L (blue) is shown on the left; only residues of C_{H1} that are involved in this interaction and are above the HSQC amplitude threshold are indicated. On the right, internal C_{H1} residues above the 25% threshold are shown in red; the three *cis* proline residues in C_{H1} are depicted and labeled in blue.

(D) The HSQC spectra of the C_{H1} Pro32Ala, Pro34Ala, and Pro74Ala mutants show that only the Pro32Ala mutant is not able to fold in the presence of C_L anymore (red spectrum). For the other two mutants, Pro34Ala and Pro74Ala, well dispersed spectra, and hence folding are observed in the presence of C_L (blue spectra). In the absence of C_L (red spectra), all three mutants show typical HSQC spectra of unfolded proteins. All measurements were carried out in PBS at 25°C, except for the folding kinetics that were recorded at 12.5°C.

in the C_{H1} domain including Val21, Val68, Trp73, and Val78 (Figure 2C, right panel). Additionally, interaction of Val48 and Val66 might also be involved in this hydrophobic cluster, although this could not be directly addressed due to peak overlap for Val66. Thus, a few key interactions between C_L and C_{H1} establish an interface between the two domains in the intermediate, which allows for the formation of a hydrophobic folding nucleus in the C_{H1} domain, and subsequent prolyl isomerization paves the path to the native state.

mutants in the absence and in the presence of C_L. All three mutants showed almost indistinguishable spectra compared to the wild-type C_{H1} domain in the absence of C_L (Figure 2D). Importantly, two of the mutants, Pro34Ala and Pro74Ala, displayed well dispersed HSQC spectra in the presence of C_L that are very similar to that of wild-type C_{H1} in the presence of C_L (Figure 2D), arguing that isomerization of these two prolines is not essential for C_{H1} domain folding. However, when the Pro32Ala mutant was similarly examined, identical spectra

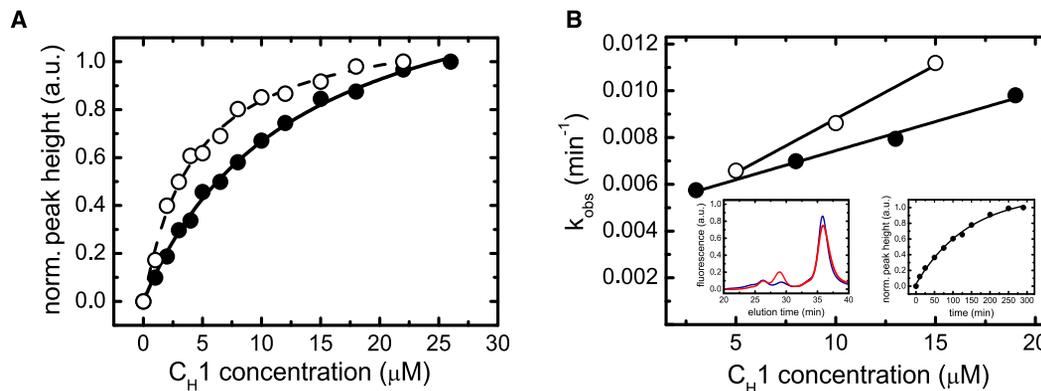


Figure 3. Characterization of the Interaction between BiP and C_H1 In Vitro

(A) The affinity between BiP and oxidized C_H1 (filled circles, straight line) as well as reduced C_H1 (open circles, dashed line) was determined by analytical HPLC experiments. The data were fitted to a one-site binding model to determine the K_D.

(B) The association kinetics between 1 μM BiP and varying concentrations of oxidized C_H1 (filled circles) and reduced C_H1 (open circles) were measured to determine the rate constants of the reaction. For oxidized C_H1, $k_{on} = 0.00026 \pm 0.00002 \mu\text{M}^{-1} \text{min}^{-1}$ and $k_{off} = 0.0050 \pm 0.0002 \text{min}^{-1}$ were obtained. For the reduced C_H1 domain, the corresponding values were $k_{on} = 0.00041 \pm 0.00003 \mu\text{M}^{-1} \text{min}^{-1}$ and $k_{off} = 0.0047 \pm 0.0003 \text{min}^{-1}$. The left inset shows single HPLC runs of 8 μM oxidized C_H1 and 1 μM BiP after 10 min (blue) and 200 min (red) of incubation. The right inset shows the overall observed single exponential association kinetics between 1 μM BiP and 8 μM oxidized C_H1.

were obtained in the presence and in the absence of C_L (Figure 2D). Thus, isomerization of Pro32 from *trans* to *cis* is a prerequisite for C_L-induced folding of the C_H1 domain.

The Role of the Molecular Chaperone BiP in C_H1/C_L Assembly

Although association-coupled folding of the C_H1 domain is an intrinsic feature of this protein, the folding, assembly, and subsequent secretion of IgG molecules in the cell involve additional factors (Meunier et al., 2002). To retain the unassembled heavy chain in the ER, a protein recognizing the unfolded C_H1 domain is needed. The molecular chaperone BiP, a member of the Hsp70 chaperone family that is present at high concentrations in the ER, plays a crucial role in this process in vivo (Haas and Wabl, 1983; Lee et al., 1999). After synthesis and prior to assembly with the light chain, the C_H1 domain of the heavy chain remains in a reduced state (Vanhove et al., 2001). In line with this finding, BiP formed stable complexes with the reduced C_H1 domain in vitro with a dissociation constant of $4.2 \pm 0.4 \mu\text{M}$ (Figure 3A). Oxidation of C_H1 only slightly reduced the affinity for BiP to a K_D of $12.6 \pm 0.7 \mu\text{M}$ (Figure 3A), which is consistent with this domain remaining unfolded. This is further supported by our finding that reduced C_H1 could form its intradomain disulfide bridge while in the BiP-bound state (data not shown). Oxidized and reduced C_H1 both bind to BiP in a two-state, concentration-dependent reaction guaranteeing a fast association in the ER (Figure 3B). To assess whether C_L can form stable triple complexes with BiP and C_H1, FRET experiments were carried out with ATTO594-labeled BiP and ATTO532-labeled C_L to which unlabeled C_H1 was added (Figure S4). No FRET signal was detected even though BiP readily associates with C_H1 under these conditions (Figure S4) and C_H1 interacts with C_L (Figure 1). Thus, the presence of stable BiP:C_H1:C_L complexes is very unlikely, even though the light chain can trigger the dissociation of BiP:C_H1 complexes in vivo (Lee et al., 1999).

The Antibody Domain Folding Status Controls Binding to BiP and Secretion from the ER In Vivo

It has long been appreciated that the C_H1 domain is central to correct assembly and transport of IgG molecules and other immunoglobulin isotypes (Wolfenstein-Todel et al., 1974; Adetugbo, 1978; Hendershot et al., 1987; Shaffer and Schlissel, 1997). Deletion of this domain allows for secretion or surface expression of free heavy chains and various Ig assembly intermediates (Hendershot et al., 1987), which shows that quality control is focused on this domain. Our data put the unexpected unfolded nature of the C_H1 domain at the center of the secretion control mechanism of IgG antibodies. To test this notion in a cellular context, we first expressed the MAK33 κ light chain, which contained the C_L domain that was used in the in vitro experiments, in COS-1 cells and performed metabolic labeling and immunoprecipitation assays. As expected, this wild-type light chain (LC_{wt}) was detected not only in the cells, but also in the medium, indicating that LC_{wt} was secreted efficiently (Figure 4A, lanes 1 and 3). When we replaced the C_L domain of the light chain with the C_H1 domain, this light chain (LC_{CH1}) now behaved like a heavy chain in terms of retention in the ER and interaction with BiP, as demonstrated by an increase in the amount of the altered light chain that coprecipitated with BiP and its absence in the medium (Figure 4A, lanes 4–6). 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, we exchanged the small helical elements of the C_L domain, which have been reported to play a crucial role in the folding of this domain (Feige et al., 2008), against the corresponding elements of the C_H1 domain. This exchange transformed the C_L domain into an unfolded protein in vitro (data not shown). When a light chain containing this altered C_L domain (LC_{CLmut}) was expressed in COS-1 cells, it strongly interacted with BiP in vivo and was no longer secreted

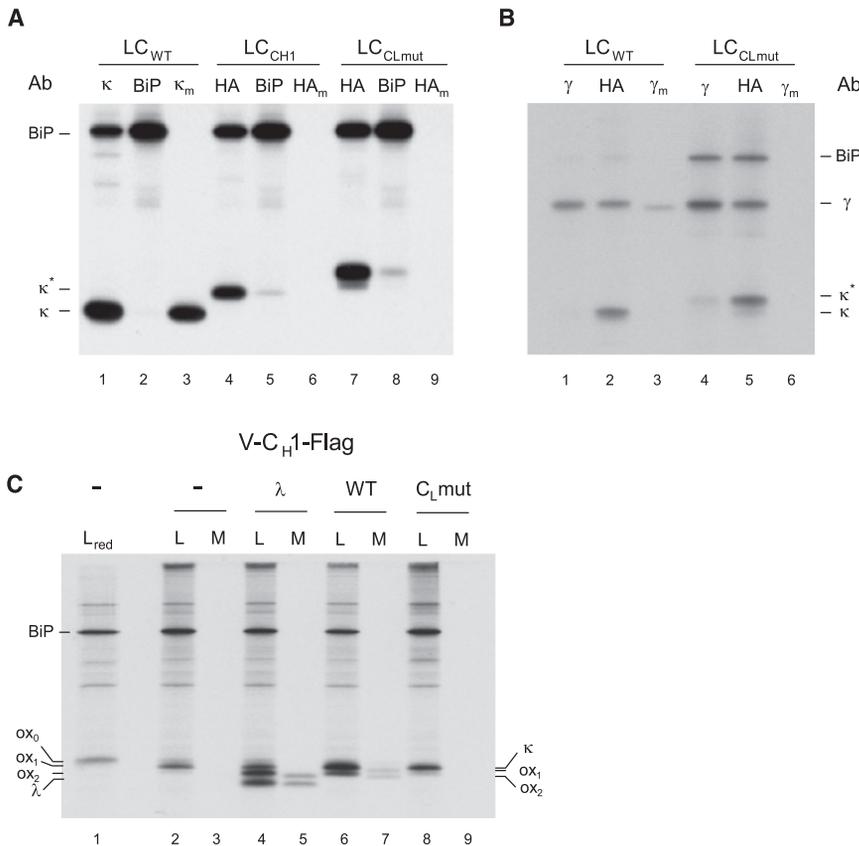


Figure 4. The Folding Status of an Antibody Domain Controls IgG Secretion In Vivo

(A) COS-1 cells were cotransfected with vectors encoding BiP and either a wild-type light chain (LC_{wt}), a light chain containing the C_H1 domain instead of the C_L domain (LC_{CH1}), or a light chain containing an unfolded C_L domain (LC_{CLmut}). Cells were metabolically labeled for 3 hr, and both cell lysates (no subscript) and culture supernatants (subscript m) were immunoprecipitated with the indicated antisera (Ab). Precipitated proteins were separated by SDS-PAGE under reducing conditions and visualized by autoradiography.

(B) COS-1 cells were cotransfected as in (A), except that a vector encoding a chimeric humanized mouse heavy chain was also included. Cells were labeled and analyzed as in (A).

(C) COS-1 cells were cotransfected with vectors encoding BiP, a Flag-tagged truncated heavy chain consisting of only the V_H and C_H1 domains (Lee et al., 1999), and the indicated light-chain constructs (i.e., λ, wild-type κ, or C_L mutant κ). Cells were metabolically labeled, and both cell lysates (L) and culture supernatants (M) were immunoprecipitated with the anti-Flag antibody. Precipitated proteins were separated by SDS-PAGE under nonreducing conditions (except the first lane, which included 2-ME in the sample buffer and is indicated as red) and visualized by autoradiography. Mobilities of completely reduced (ox₀), partially oxidized (ox₁), and fully oxidized (ox₂) forms of truncated heavy chain, as well as those of λ and κ light chains, are indicated. The tagged forms of the κ light-chain constructs comigrate with the ox₁ form of the truncated heavy chain. These data demonstrate that whereas both wild-type λ and κ light chains induced oxidation of the C_H1 domain, assembly, and secretion, the light chain containing the C_L mutant was defective in these processes.

from the cell (Figure 4A, lanes 7–9), which argues that the folding status of an antibody domain is key for its retention.

To add support to the key role played specifically by the interactions between C_H1 and C_L domains, we first performed additional in vivo experiments with two different full-length heavy chains (the MAK33 γ1 heavy chain and a humanized mouse IgG heavy chain [Liu et al., 1987]) in combination with both the wild-type (LC_{wt}) and the mutated MAK33 light chain (LC_{CLmut}). The wild-type MAK33 light chain assembled with both heavy chains and allowed their secretion from cells, whereas the mutant light chain was unable to assemble or induce secretion of the heavy chains (Figure 4B; Figure S5). This is in keeping with our in vitro data showing that the mutant C_L domain was unfolded and therefore unable to induce the folding of the C_H1 domain. To determine if this was also the case in vivo, we used a truncated version of the chimeric heavy chain consisting of only the V_H and C_H1 domains, which allows us to monitor oxidation of the C_H1 domain based on an increase in its mobility (Lee et al., 1999). Indeed, we found that the wild-type MAK33 light chain induced oxidation and secretion comparable to a different λ light chain that was used in previous studies (Lee et al., 1999) (Figure 4C, compare lanes 4 and 6 and lanes 5 and 7). However, the mutant light chain was unable to induce either oxidation or

secretion of this truncated heavy chain (Figure 4C, lanes 8 and 9). These data clearly show that the correct folding of the C_L domain is absolutely required to induce the folding and oxidation of the C_H1 domain as well as their covalent assembly in vivo, which is in line with our in vitro data. Furthermore, these data demonstrate that the template-assisted folding of C_H1 is a general and interchangeable control element functioning in different murine and human antibodies.

Isomerization of a Conserved Proline Residue in the C_H1 Domain Is Essential for Assembly and Secretion of IgG Molecules In Vivo

To directly assess the role of proline isomerization in the association-coupled folding reaction of the C_H1 domain in vivo, each single *cis* proline residue in the C_H1 domain was exchanged against alanine in the context of the full-length heavy chain, and metabolic labeling experiments were conducted. For the wild-type heavy chain and two of the heavy chains mutated in the C_H1 domain, Pro34Ala and Pro74Ala, all Ig assembly intermediates could be detected (Figure 5A, lanes 1, 2, 4, 5, 10, and 11), and, importantly, completely assembled Ig molecules were secreted (Figure 5A, lanes 3, 6, and 12), demonstrating that isomerization of neither of these prolines was critical for

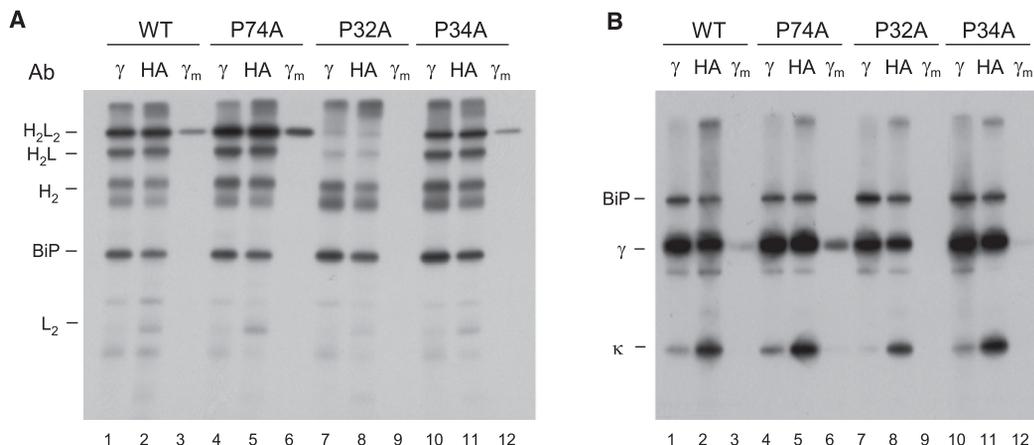


Figure 5. The Isomerization of a Single Proline Residue Controls the Assembly and Secretion of Heavy Chains In Vivo

(A) COS-1 cells were cotransfected with vectors encoding wild-type MAK33 heavy chain (WT) or one of the three Pro to Ala mutants (P74A, P32A, or P34A) together with wild-type light chain and BiP. Cells were metabolically labeled, and both cell lysates (no subscript) and culture supernatants (subscript m) were immunoprecipitated with the indicated antisera (Ab). Precipitated proteins were separated by SDS-PAGE under nonreducing conditions and visualized by autoradiography.

(B) The cell lysates and culture supernatants from (A) were divided in half, immunoprecipitated with the indicated antibodies, analyzed by SDS-PAGE under reducing conditions, and visualized with autoradiography. These data demonstrate that the failure of the P32A mutant to induce assembly and secretion is not because it is expressed poorly, as the signal for this mutant heavy chain is very similar to that of the wild-type heavy chain.

C_H1 domain folding in vivo. In fact, mutation of Pro74 to alanine actually increased the assembly and secretion of the heavy chain. In contrast, when Pro32 in the C_H1 domain was mutated to alanine, no significant amount of heavy-chain and light-chain assembly were detected, nor was any heavy chain secreted (Figure 5A, lanes 7–9). This is not due to poor expression of either the light chain or mutant heavy chain in this experiment (Figure 5B). These results are in excellent agreement with our in vitro data (Figure 2D) and reveal a key role of the *trans* to *cis* isomerization of Pro32 in the C_H1 domain for the assembly, inter-chain disulfide bridge formation, and secretion of IgG molecules. It should be noted that this critical proline residue is highly conserved in the various murine Ig isotypes as well as in the immunoglobulins of different species.

DISCUSSION

In this work, we show that the C_H1 domain of the murine IgG1 antibody is an intrinsically disordered protein. As C_H1 does not possess an unusual number or distribution of charged or hydrophobic residues, it can still form a well defined globular structure once folded and does not show the typical sequence signature usually associated with intrinsically disordered proteins (Ward et al., 2004; Yang et al., 2005; Fink, 2005; Dunker et al., 2008). The large number of proline residues, also in a small helical part of the C_H1 domain that was recently identified as important for antibody domain folding (Feige et al., 2008), might be one determinant contributing to its unfolded nature. Our data thus suggest that the C_H1 domain is a representative of a distinct class of intrinsically disordered proteins.

We demonstrate that the C_L domain, which is the cognate partner of the C_H1 domain in the complete IgG molecule, is required to fold C_H1 to the structure observed in IgG antibodies.

The detailed analysis of the underlying pathway suggests that the reaction is initiated by the recognition of a few key interface residues between C_L and C_H1, which then promotes the formation of a hydrophobic core in C_H1. Both reactions render each other energetically more favorable and thereby allow the entropically demanding structuring of an unfolded polypeptide chain. In vitro, this folding reaction requires the presence of the internal disulfide bridge in C_H1. Therefore, one might assume that a roughly preformed topology or residual structure, which could not be detected in this study, might play a role in the folding process. Folding of the C_H1 domain is rate limited by proline isomerization, as observed for most isolated antibody domains (Goto and Hamaguchi, 1982; Thies et al., 1999; Feige et al., 2004). The presence of several *cis* proline residues in the native state and the overall very large number of proline residues, however, sets the C_H1 domain apart from most other antibody domains and may reflect the special role played by this domain. The key step in C_H1 folding, the isomerization of a single highly conserved proline residue in the loop between strands B and C, is likely to be essential for the folding and secretion of most Ig classes and potentially of other Ig superfamily members.

We show that BiP, a major ER chaperone, strongly binds to C_H1 in vitro, in agreement with previous in vivo studies (Hendershot et al., 1987). Oxidation of the internal C_H1 disulfide bridge was possible in the BiP-bound state in vitro, although this form has not been detected in vivo. After release from BiP, C_H1 can complete folding upon association with C_L and, if successful, form an interchain disulfide bridge with C_L. Because heavy chains of most isotypes that are devoid of the C_H1 domain can be secreted (Coffino et al., 1970; Hendershot et al., 1987), the essential steps controlling IgG assembly described here are likely to be general for all antibodies. Our data allow us to propose a possible order of events for this quality control mechanism in vivo. First,

the C_H1 domain binds to BiP as it enters the ER cotranslationally in the reduced state. Then, likely triggered by association with light chain (Lee et al., 1999), the oxidation of the internal disulfide bridge between Cys25 and Cys80 takes place, which brings at least two of the residues (Val21 and Val78) that are involved in the formation of the hydrophobic folding nucleus in close proximity to each other. Only after release from BiP can the oxidized C_H1 domain complete its folding in association with C_L. This scenario is in agreement with the fact that some residues in the C_H1 domain found to be involved in the initial interaction with C_L were among those identified as putative BiP binding sequences in this domain in a previous study (Knarr et al., 1995). Even though most antibody domains possess BiP binding sequences (Knarr et al., 1995), they interact with BiP in the cell only transiently or not at all due to the competing, rapid folding reaction (Hellman et al., 1999). In contrast, the continued unfolded status of the C_H1 domain in the absence of light-chain association allows it to permanently expose binding sites for BiP, predisposing it for a stable interaction in the ER. In the complete antibody, the C_L and the C_H1 domain are covalently crosslinked via a disulfide bridge. We found that once this intermolecular disulfide bridge is formed, BiP no longer associates with the C_H1 domain *in vitro* (data not shown), because formation of this interchain disulfide bridge is rate limited by proline isomerization and hence depends on the complete folding of C_H1. Thus, folding-dependent, covalent assembly provides yet another checkpoint for monitoring the proper maturation of Ig molecules in the ER and allows the assembled heavy chain to escape thiol-mediated retention mechanisms in cells (Sitia et al., 1990). It is conceivable that in the ER, association, oxidation, and folding of C_H1 are tightly coupled by the immunoglobulin assembly machinery, as no oxidized C_H1 was found to be bound to BiP *in vivo* and ATP-induced release of BiP from unassembled heavy chains results in the formation of disulfide-linked heavy-chain aggregates (Vanhove et al., 2001).

Taken together, our data provide a detailed mechanism by which BiP and an intrinsically disordered antibody domain control the secretion of IgG antibodies (Figure 6). The comprehensive model incorporates previous *in vitro* (Lilie et al., 1995; Mayer et al., 2000) and *in vivo* findings (Bole et al., 1986; Hendershot et al., 1987; Lilie et al., 1995; Lee et al., 1999; Mayer et al., 2000; Vanhove et al., 2001) and suggests regulatory hubs at which additional components may come into play *in vivo* (Vanhove et al., 2001; Elkabetz et al., 2005) to orchestrate folding, oxidation, assembly, and secretion.

A rigorous assembly control mechanism is particularly important for antibody molecules, which undergo a developmentally asynchronous expression of heavy- and light-chain genes. In pre-B cells, heavy-chain genes are rearranged first, and the resultant proteins are largely retained in the cell (Burrows et al., 1979), except for a limited number that assemble with the surrogate light chain (Pillai and Baltimore, 1987). The developmentally more mature B cell expresses light chains, which assemble with heavy chains and allow their transport to the cell surface. Finally, the terminally differentiated plasma cell produces enormous quantities of Ig molecules (Cenci and Sitia, 2007), which must be appropriately assembled to bind specifically to antigens and fulfill their effector functions. Accordingly, the nature of the

reactions that govern C_H1 folding and assembly with C_L allow for the efficient and accurate assembly of antibodies prior to secretion and might hint toward the coevolution of substrates and folding factors in the ER as well as a general mechanism of quality control for oligomeric proteins.

EXPERIMENTAL PROCEDURES

Protein Production

All antibody domains were produced in a manner similar to published protocols (Feige et al., 2004, 2007). Details can be found in the [Supplemental Data](#). Numbering of the C_H1 domain begins with 1 in this work.

Hamster BiP (Wei and Hendershot, 1995) was mutated by site-directed mutagenesis to the murine sequence. Expression was carried out in HB101 cells for 3 hr at 37°C. After cell disruption, Ni-NTA affinity purification was performed in 50 mM HEPES/KOH (pH 7.5), 400 mM NaCl, 50 mM Imidazole. BiP was eluted with an Imidazole gradient from 0.05 M to 1 M. BiP containing fractions were applied to a Superdex 200 pg (26/60) gel-filtration column (GE Healthcare, München, Germany) equilibrated in HKM buffer (50 mM HEPES/KOH [pH 7.5], 150 mM KCl, 10 mM MgCl₂) and finally to a Superdex 200 10/300GL HPLC column (GE Healthcare, München, Germany) equilibrated in the same buffer. The cyclophilin B (CypB) gene was amplified without the signal sequence from the murine cDNA (imaGenes, Berlin, Germany) and was inserted into the pET28a vector. Expression was carried out for 3 hr at 37°C in BL21-DE3 cells. The cell pellet was dissolved in 50 mM HEPES/KOH (pH 7.0), 10 mM EDTA, and the cleared lysate was applied to a SP-Sepharose column equilibrated in the same buffer. The protein was eluted with a 0–1 M NaCl gradient. Subsequently, CypB-containing fractions were applied to a Superdex 75 pg (26/60) gel-filtration column equilibrated in HKM buffer. All vectors were sequenced, and protein masses were verified by mass spectrometry.

Optical Spectroscopy

A Jasco J-720 spectropolarimeter was used for all CD measurements (Jasco, Gross-Umstadt, Germany). Far-UV CD spectra were recorded in a 0.2 mm quartz cuvette; far-UV kinetics were recorded in a 1 mm quartz cuvette. Spectra of the isolated domains were recorded at 45 μM protein concentration; for the spectrum of the complex, 15 μM C_H1 in the presence of 45 μM C_L was used. Far-UV CD kinetics were recorded at 10 μM protein concentration of each domain and followed at 205 nm. Spectra of the C_H1 domain in the complex were calculated by subtraction of the spectrum of the isolated C_L domain from the spectrum of the complex, measured after a 4 hr equilibration step at 25°C. All spectra were averaged 16 times and buffer corrected. Fluorescence measurements were carried out in a Spex FluoroMaxIII spectrofluorimeter (Jobin Yvon, München, Germany) in a stirred 1 cm quartz cuvette. Kinetics and titrations were measured by the change in the intrinsic tryptophan fluorescence, excited at 280 nm, and detected at 350 nm. For titrations, varying concentrations of C_L were added to 2 μM C_H1, and immediately as well as after a 4 hr equilibration step at 25°C, the fluorescence of the same samples was determined. The difference between initial and final fluorescence emission was normalized and analyzed according to a one-site binding model. Iodide quenching experiments were performed as published (Feige et al., 2007).

For anisotropy measurements, 1 μM lucifer yellow-labeled C_L Ala113Cys and varying concentrations of C_H1 were used. Lucifer yellow fluorescence was excited at 430 nm and detected at 525 nm. The change in quantum yield of the chromophore was less than 5% upon association of the labeled C_L domain with C_H1. Individual traces were fitted by single exponential functions. The obtained rate constants were fitted to a linear equation to derive the k_{on} and the k_{off} value.

NMR Spectroscopy

Spectra of the C_H1 domain in complex with the C_L domain were recorded at 25°C on Bruker DMX600 and DMX750 spectrometers (Bruker, Rheinstetten, Germany), whereas spectra for the assignment of the unfolded C_H1 domain were measured at 12.5°C on a Bruker AVANCE900 spectrometer (Bruker, Rheinstetten, Germany). Backbone sequential assignment of the isolated C_H1 domain was obtained by standard triple-resonance experiments implemented

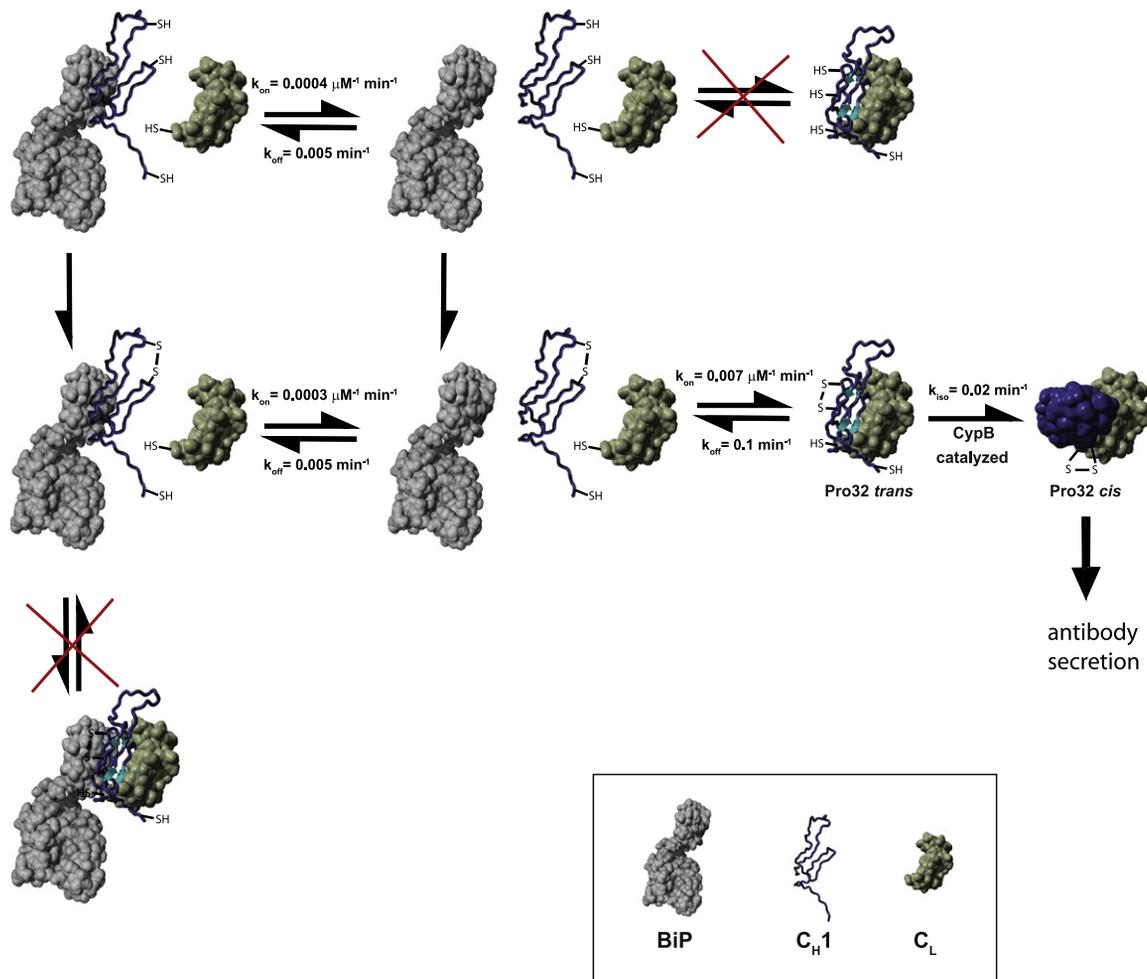


Figure 6. A Model for the Overall IgG Secretion Control Mechanism

A schematic indicating the possible pathways for the C_H1 domain (blue); its folding and assembly in association with C_L (green) and BiP (gray) is shown. C_H1 has to form its internal disulfide bridge and to be released from BiP before it can associate with C_L. In vivo, these processes are tightly coupled and thus cannot be dissected kinetically. Prior to complete folding and irreversible formation of the C_L/C_H1 interchain disulfide bridge, the proline residue 32 has to isomerize from *trans* to *cis* and otherwise traps a partially folded C_H1 intermediate. The isomerization reaction can be accelerated by cyclophilin B. All rate constants were determined at 25°C.

with selective proton flip-back techniques for fast pulsing (Diercks et al., 2005). To gain information about any preferential conformations present in the disordered C_H1 domain, an NNH-NOESY spectrum and a ¹⁵N-HSQC-NOESY spectrum were recorded on a highly deuterated sample with a mixing time of 600 ms in order to detect long-range H^N-H^N NOEs (Mok et al., 1999).

For all measurements of the folded C_H1 domain in association with C_L, a 2-fold excess of unlabeled C_L was added to ¹⁵N or ¹⁵N, ¹³C-labeled C_H1. Prior to steady-state measurements, samples were incubated for at least 6 hr at room temperature to ensure complete folding of the C_H1 domain. Backbone sequential assignment of the assembled C_H1 domain was achieved with standard triple-resonance experiments with selective proton flip-back techniques for fast pulsing. The assignment of the carbon chemical shifts was limited to the C' and C² chemical shifts due to the relaxation properties of the whole protein complex. To verify the backbone resonance assignment, NH residual dipolar couplings (RDCs) were determined. The sample was prepared as described above and aligned with nonionic liquid crystalline media (Ruckert and Otting, 2000). NH RDC values were extracted from IPAP-HSQC spectra by using Bruker pulse sequences. The sequential information based on the C' and C² chemical shifts as well as the NH RDC values and the crystal

structure of the folded C_H1 domain (PDB code: 1ORS) served as input for the software MARS (Jung and Zweckstetter, 2004a, 2004b).

In order to characterize the folding pathway of the intrinsically disordered C_H1 domain, ¹⁵N-HSQC spectra were recorded at 12.5°C every 14 min immediately after adding unlabeled C_L to ¹⁵N-labeled C_H1 by using selective proton flip-back techniques for fast pulsing. Identical processing of all spectra measured during the folding process was performed with the software TOPSPIN 1.3 (Bruker Biospin). Peak intensities were analyzed by using the software SPARKY (www.cgl.ucsf.edu/home/sparky) and normalized to the corresponding intensities in the final spectrum after 36 hr. The backbone resonance assignment was transferred from 25°C to 12.5°C recording a HSQC temperature series.

Analytical HPLC Experiments

For all experiments, a Shimadzu HPLC system (Shimadzu, München, Germany) was used. Complex formation between BiP and C_H1 was analyzed on a Superdex 200 10/300GL column in HKM buffer at a flow rate of 0.5 ml/min. For the determination of the dissociation constant respectively binding kinetics, peak intensities at the retention time of 28.4 min corrected for baseline drifts were

plotted against the C_H1 concentration or the incubation time, respectively, and were normalized. The rate constants, k_{obs} , of the binding reaction were determined from single exponential fits and evaluated with a linear equation to derive k_{on} and k_{off} . Detection of all proteins was performed by the intrinsic fluorescence excited at 280 nm and monitored at 350 nm. Incubation steps were performed in HKM buffer with 1 mM ADP.

Cell Culture Experiments

The murine IgG1 MAK33 light-chain (LC_{WT}) and heavy-chain (HC_{WT}) cDNAs were obtained with an intact signal sequence for expression in mammalian ER. Two LC mutants, one in which the C_H1 domain was substituted for the C_L domain (LC_{CH1}) and the other in which structural features of the C_H1 domain were substituted for the corresponding regions of the C_L domain (LC_{CLmut}) were produced, and all constructs were inserted into the pSVL vector (GE Healthcare, München, Germany). An HA epitope tag was engineered at the C terminus of the wild-type light chain and the two mutants for immunoprecipitation purposes. Heavy-chain proline exchange mutants were generated by site-directed mutagenesis. The cDNA for a chimeric humanized heavy chain was used as published (Liu et al., 1987), and a truncated version of this heavy chain containing only the V_H and C_H1 domains was produced previously (Lee et al., 1999), as was a mouse lambda light-chain cDNA (Hellman et al., 1999). The recombinant plasmids, along with a pMT vector encoding hamster BiP (Lee et al., 1999), were introduced into COS-1 cells (Gluzman, 1981) that were cultured as described (Lee et al., 1999) by using FuGENE 6 transfection reagent (Roche, Indianapolis, USA) following the manufacturer's protocol. Metabolic labeling, cell lysis, immunoprecipitation, and visualization of the proteins were performed as described previously (Lee et al., 1999). Anti-rodent BiP antiserum (Hendershot et al., 1995), a monoclonal anti-HA (12CA5) antibody (kindly provided by Dr. Al Reynolds, Vanderbilt University, USA), goat anti-mouse Ig κ and goat anti-mouse γ antibodies (Southern Biotech, Birmingham, AL, USA), and Protein A Sepharose beads were used for immunoprecipitations. For metabolic labeling experiments, cells were cultured in DMEM lacking methionine and cysteine and labeled with ³⁵S Translabel (MP Biomedicals, Irvine, CA, USA) for the indicated times.

SUPPLEMENTAL DATA

Supplemental Data include Supplemental Experimental Procedures and five figures and are available at [http://www.cell.com/molecular-cell/supplemental/S1097-2765\(09\)00304-9](http://www.cell.com/molecular-cell/supplemental/S1097-2765(09)00304-9).

ACKNOWLEDGMENTS

We thank Frank Striebel for preliminary work on the in vivo experiments and Helmut Krause for performing mass spectrometry experiments. Funding of M.J.F. and M.M. by the Studienstiftung des deutschen Volkes, of L.M.H. by National Institutes of Health Grant GM54068, the Cancer Center CORE Grant CA21765, and the American Lebanese Syrian Associated Charities of St. Jude Children's Research Hospital and of J.B. by the SFB749, the Fonds der chemischen Industrie, and the Bayerische Forschungsstiftung is gratefully acknowledged.

Received: November 10, 2008

Revised: March 5, 2009

Accepted: April 28, 2009

Published: June 11, 2009

REFERENCES

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

Amzel, L.M., and Poljak, R.J. (1979). 3-Dimensional structure of immunoglobulins. *Annu. Rev. Biochem.* 48, 961–997.

Aricescu, A.R., and Jones, E.Y. (2007). Immunoglobulin superfamily cell adhesion molecules: zippers and signals. *Curr. Opin. Cell Biol.* 19, 543–550.

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

Bole, D.G., Hendershot, L.M., and Kearney, J.F. (1986). Posttranslational association of immunoglobulin heavy-chain binding-protein with nascent heavy-chains in nonsecreting and secreting hybridomas. *J. Cell Biol.* 102, 1558–1566.

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

Burrows, P., LeJeune, M., and Kearney, J.F. (1979). Evidence that murine pre-B cells synthesize μ heavy chains but no light chains. *Nature* 280, 838–840.

Cenci, S., and Sitia, R. (2007). Managing and exploiting stress in the antibody factory. *FEBS Lett.* 581, 3652–3657.

Christis, C., Lubsen, N.H., and Braakman, I. (2008). Protein folding includes oligomerization: examples from the endoplasmic reticulum and cytosol. *FEBS J.* 275, 4700–4727.

Coffino, P., Laskov, R., and Scharff, M.D. (1970). Immunoglobulin production: method for quantitatively detecting variant myeloma cells. *Science* 167, 186–188.

Cota, E., Steward, A., Fowler, S.B., and Clarke, J. (2001). The folding nucleus of a fibronectin type III domain is composed of core residues of the immunoglobulin-like fold. *J. Mol. Biol.* 305, 1185–1194.

Diercks, T., Daniels, M., and Kaptein, R. (2005). Extended flip-back schemes for sensitivity enhancement in multidimensional HSQC-type out-and-back experiments. *J. Biomol. NMR* 33, 243–259.

Dunker, A.K., Silman, I., Uversky, V.N., and Sussman, J.L. (2008). Function and structure of inherently disordered proteins. *Curr. Opin. Struct. Biol.* 18, 756–764.

Elkabetz, Y., Argon, Y., and Bar-Nun, S. (2005). Cysteines in CH1 underlie retention of unassembled ig heavy chains. *J. Biol. Chem.* 280, 14402–14412.

Elgaard, L., and Helenius, A. (2003). Quality control in the endoplasmic reticulum. *Nat. Rev. Mol. Cell Biol.* 4, 181–191.

Feige, M.J., Walter, S., and Buchner, J. (2004). Folding mechanism of the C(H)2 antibody domain. *J. Mol. Biol.* 344, 107–118.

Feige, M.J., Hagn, F., Esser, J., Kessler, H., and Buchner, J. (2007). Influence of the internal disulfide bridge on the folding pathway of the CL antibody domain. *J. Mol. Biol.* 365, 1232–1244.

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

Fink, A.L. (2005). Natively unfolded proteins. *Curr. Opin. Struct. Biol.* 15, 35–41.

Freund, C., Honegger, A., Hunziker, P., Holak, T.A., and Pluckthun, A. (1996). Folding nuclei of the scFv fragment of an antibody. *Biochemistry* 35, 8457–8464.

Gemmecker, G., Jahnke, W., and Kessler, H. (1993). Measurement of fast proton-exchange rates in isotopically labeled compounds. *J. Am. Chem. Soc.* 115, 11620–11621.

Gluzman, Y. (1981). SV40-transformed simian cells support the replication of early SV40 mutants. *Cell* 23, 175–182.

Goto, Y., and Hamaguchi, K. (1979). Role of the intrachain disulfide bond in the conformation and stability of the constant fragment of the immunoglobulin light chain. *J. Biochem. (Tokyo)* 86, 1433–1441.

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

Haas, I.G., and Wabl, M. (1983). Immunoglobulin heavy-chain binding-protein. *Nature* 306, 387–389.

Hamers-Casterman, C., Atarhouch, T., Muyldermans, S., Robinson, G., Hamers, C., Songa, E.B., Bendahman, N., and Hamers, R. (1993). Naturally-occurring antibodies devoid of light-chains. *Nature* 363, 446–448.

- Hamill, S.J., Steward, A., and Clarke, J. (2000). The folding of an immunoglobulin-like Greek key protein is defined by a common-core nucleus and regions constrained by topology. *J. Mol. Biol.* *297*, 165–178.
- Helenius, A., Marquardt, T., and Braakman, I. (1992). The endoplasmic reticulum as a protein-folding compartment. *Trends Cell Biol.* *2*, 227–231.
- Hellman, R., Vanhove, M., Lejeune, A., Stevens, F.J., and Hendershot, L.M. (1999). 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.* *144*, 21–30.
- Hendershot, L., Bole, D., Kohler, G., and Kearney, J.F. (1987). Assembly and secretion of heavy-chains that do not associate posttranslationally with immunoglobulin heavy-chain binding-protein. *J. Cell Biol.* *104*, 761–767.
- Hendershot, L.M., Wei, J.Y., Gaut, J.R., Lawson, B., Freiden, P.J., and Murti, K.G. (1995). In vivo expression of mammalian BiP ATPase mutants causes disruption of the endoplasmic reticulum. *Mol. Biol. Cell* *6*, 283–296.
- Huber, R., Deisenhofer, J., Colman, P.M., Matsushima, M., and Palm, W. (1976). Crystallographic structure studies of an IgG molecule and an Fc fragment. *Nature* *264*, 415–420.
- Jahn, T.R., Parker, M.J., Homans, S.W., and Radford, S.E. (2006). Amyloid formation under physiological conditions proceeds via a native-like folding intermediate. *Nat. Struct. Mol. Biol.* *13*, 195–201.
- Jung, Y.S., and Zweckstetter, M. (2004a). Backbone assignment of proteins with known structure using residual dipolar couplings. *J. Biomol. NMR* *30*, 25–35.
- Jung, Y.S., and Zweckstetter, M. (2004b). Mars: robust automatic backbone assignment of proteins. *J. Biomol. NMR* *30*, 11–23.
- Kaloff, C.R., and Haas, I.G. (1995). Coordination of immunoglobulin chain folding and immunoglobulin chain assembly is essential for the formation of functional IgG. *Immunity* *2*, 629–637.
- Kameda, A., Hoshino, M., Higurashi, T., Takahashi, S., Naiki, H., and Goto, Y. (2005). Nuclear magnetic resonance characterization of the refolding intermediate of $\beta(2)$ -microglobulin trapped by non-native prolyl peptide bond. *J. Mol. Biol.* *348*, 383–397.
- Knarr, G., Gething, M.J., Modrow, S., and Buchner, J. (1995). BiP binding sequences in antibodies. *J. Biol. Chem.* *270*, 27589–27594.
- Lee, Y.K., Brewer, J.W., Hellman, R., and Hendershot, L.M. (1999). BiP and immunoglobulin light chain cooperate to control the folding of heavy chain and ensure the fidelity of immunoglobulin assembly. *Mol. Biol. Cell* *10*, 2209–2219.
- Lilie, H., Rudolph, R., and Buchner, J. (1995). Association of antibody chains at different stages of folding: prolyl isomerization occurs after formation of quaternary structure. *J. Mol. Biol.* *248*, 190–201.
- Liu, A.Y., Robinson, R.R., Murray, E.D., Jr., Ledbetter, J.A., Hellstrom, I., and Hellstrom, K.E. (1987). Production of a mouse-human chimeric monoclonal antibody to CD20 with potent Fc-dependent biologic activity. *J. Immunol.* *139*, 3521–3526.
- Mayer, M., Kies, U., Kammermeier, R., and Buchner, J. (2000). BiP and PDI cooperate in the oxidative folding of antibodies in vitro. *J. Biol. Chem.* *275*, 29421–29425.
- Melchers, F. (1971). The secretion of a Bence-Jones type light chain from a mouse plasmacytoma. *Eur. J. Immunol.* *1*, 330–335.
- Meunier, L., Usherwood, Y.K., Chung, K.T., and Hendershot, L. (2002). A subset of chaperones and folding enzymes form multi-protein complexes in the ER to bind nascent proteins. *Mol. Biol. Cell* *13*, 4456–4469.
- Mok, Y.K., Kay, C.M., Kay, L.E., and Forman-Kay, J. (1999). NOE data demonstrating a compact unfolded state for an SH3 domain under non-denaturing conditions. *J. Mol. Biol.* *289*, 619–638.
- Paci, E., Clarke, J., Steward, A., Vendruscolo, M., and Karplus, M. (2003). Self-consistent determination of the transition state for protein folding: application to a fibronectin type III domain. *Proc. Natl. Acad. Sci. USA* *100*, 394–399.
- Pillai, S., and Baltimore, D. (1987). Formation of disulphide-linked $\mu 2 \omega 2$ tetramers in pre-B cells by the 18K ω -immunoglobulin light chain. *Nature* *329*, 172–174.
- Porter, R.R. (1973). Structural studies of immunoglobulins. *Science* *180*, 713–716.
- Qin, Z.J., Hu, D.M., Zhu, M., and Fink, A.L. (2007). Structural characterization of the partially folded intermediates of an immunoglobulin light chain leading to amyloid fibrillation and amorphous aggregation. *Biochemistry* *46*, 3521–3531.
- Rothlisberger, D., Honegger, A., and Pluckthun, A. (2005). 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.* *347*, 773–789.
- Rougon, G., and Hobert, O. (2003). New insights into the diversity and function of neuronal immunoglobulin superfamily molecules. *Annu. Rev. Neurosci.* *26*, 207–238.
- Ruckert, M., and Otting, G. (2000). Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *J. Am. Chem. Soc.* *122*, 7793–7797.
- Shaffer, A.L., and Schlissel, M.S. (1997). A truncated heavy chain protein relieves the requirement for surrogate light chains in early B cell development. *J. Immunol.* *159*, 1265–1275.
- Sitia, R., Neuberger, M., Alberini, C., Bet, P., Fra, A., Valetti, C., Williams, G., and Milstein, C. (1990). Developmental regulation of IgM secretion: the role of the carboxy-terminal cysteine. *Cell* *60*, 781–790.
- Thies, M.J.W., Mayer, J., Augustine, J.G., Frederick, C.A., Lilie, H., and Buchner, J. (1999). Folding and association of the antibody domain C(H)3: prolyl isomerization precedes dimerization. *J. Mol. Biol.* *293*, 67–79.
- Vanhove, M., Usherwood, Y.K., and Hendershot, L.M. (2001). Unassembled Ig heavy chains do not cycle from BiP in vivo but require light chains to trigger their release. *Immunity* *15*, 105–114.
- Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F., and Jones, D.T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* *337*, 635–645.
- Wei, J.Y., and Hendershot, L.M. (1995). Characterization of the nucleotide-binding properties and atpase activity of recombinant hamster BiP purified from bacteria. *J. Biol. Chem.* *270*, 26670–26676.
- Williams, A.F., and Barclay, A.N. (1988). The immunoglobulin superfamily-domains for cell surface recognition. *Annu. Rev. Immunol.* *6*, 381–405.
- Wolfenstein-Todel, C., Mihaesco, E., and Frangione, B. (1974). “Alpha chain disease” protein def: internal deletion of a human immunoglobulin A1 heavy chain. *Proc. Natl. Acad. Sci. USA* *71*, 974–978.
- Yang, Z.R., Thomson, R., McNeil, P., and Esnouf, R.M. (2005). RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics* *21*, 3369–3376.

Supplemental Data

An Unfolded C_H1 Domain Controls the Assembly and Secretion of IgG Antibodies

Matthias J. Feige, Sandra Groscurth, Moritz Marcinowski, Yuichiro Shimizu, Horst Kessler, Linda M. Hendershot, and Johannes Buchner

Supplemental Experimental Procedures

Protein production. The C_H1 domain (Thr123-Arg215 of the pdb file 1FH5; numbering starts with one in this work) and an Ala113Cys mutant of the C_L domain were amplified from the murine IgG1 MAK33 cDNA. Additionally, the murine IgG1 MAK33 cDNA was used as a template to amplify the gene segments encoding the C_L domain with the native C-terminal cysteine residue (amino acids Ala112-Cys215 of the pdb file 1FH5) and the C_H1 domain with the native C-terminal cysteine residue (Thr123-Cys217 of the pdb file 1FH5). All genes were cloned into the pET28a (Novagen, Gibbstown, NJ, USA) vector without a tag, except for the wild type C_L domain which was essentially purified as published (Feige et al., 2007). Isotope labeled C_H1 for NMR experiments was expressed in M9 minimal medium. For FRET measurements, the Ala113Cys mutant of the C_L domain was used. In the case of C_H1, the native C-terminal cysteine residue was used for labeling in FRET experiments. Proteins were expressed as inclusion bodies overnight at 37°C in BL21-DE3 cells in selective LB medium. Inclusion bodies were isolated as described (Feige et al., 2004). Inclusion bodies were solubilized in 50 mM Tris/HCl, pH 7.5, 10 mM β-mercaptoethanol, 10 mM EDTA, 8 M urea and subsequently applied to a Q-

Sepharose column equilibrated in 50 mM Tris/HCl, pH 7.5, 10 mM EDTA, 5 M urea. All proteins of interest did not bind to the column under these conditions. Refolding was carried out by dialysis as published (Feige et al., 2007). After refolding, all proteins were applied to a Superdex 75pg (26/60) gel filtration column (GE Healthcare, München, Germany) equilibrated in PBS. Labeling of C_L Ala113Cys with lucifer yellow iodoacetamide (Invitrogen, Carlsbad, CA, USA) was performed according to the manufacturer's protocol. For FRET experiments between BiP, C_L and C_{H1}, C_L Ala113Cys was labeled with the ATTO532 maleimide dye (ATTO-TEC, Weidenau, Germany), and murine BiP, which possesses two free cysteines, was labeled with ATTO594 maleimide. For FRET experiments between BiP and C_{H1}, C_{H1} was labeled with ATTO532 at its native C-terminal cysteine residue. All proteins were separated from the free dye *via* a Superdex200 10/300GL HPLC column equilibrated in HKM buffer. All vectors in this study were sequenced and protein masses were verified by mass spectrometry.

Preparation of reduced C_{H1}. To reduce the internal disulfide bond of the C_{H1} domain, the protein was unfolded in 2 M GdmCl in PBS at 25°C for 2 h in the presence of 10 mM TCEP. Subsequently, the protein was applied to a Superdex75 10/300GL HPLC column equilibrated in PBS supplemented with 0.1 mM TCEP.

Optical spectroscopy. A Jasco J-720 spectropolarimeter was used for all measurements (Jasco, Gross-Umstadt, Germany). A 2 mm quartz cuvette was used for near-UV CD spectra and kinetics (observed at 280 nm). 100 μM protein respectively 100 μM of each domain were used. Spectra of the C_{H1} domain in the complex were calculated by subtraction of the spectrum of the isolated C_L domain

from the spectrum of the complex, measured after a 4 h equilibration step at 25°C. All spectra were averaged 16 times and buffer corrected.

Analytical HPLC experiments. For all experiments, a Shimadzu HPLC system (LC-20AT, SPD-20A, RF-10A_{XL}, SIL-20AC) (Shimadzu, München, Germany) was used. To assess complex formation between C_L and C_{H1}, both proteins were incubated at 25°C with a domain concentration of 25 µM each. After varying times, 100 µl of the sample were applied to a Superdex75 10/300GL HPLC column (GE Healthcare, München, Germany) in PBS at a flow rate of 0.75 ml/min. The peak height corresponding to the dimer was analyzed over incubation time. To assess whether C_{H1} can be oxidized if bound to BiP, C_{H1} was reduced in HKM buffer with 2 mM TCEP for 1 hour. Complex formation between 30 µM reduced C_{H1} and 50 µM BiP was performed at 37°C for 2 hours in HKM buffer supplemented with 1 mM ADP and 1 mM TCEP. The sample was applied at a flow rate of 0.25 ml/min to a Superdex 200 10/300GL column equilibrated in HKM buffer supplemented with 4 mM GSSG and 1 mM TCEP. Only the peak of BiP-bound C_{H1} was collected and free sulfhydryls were reacted with 10 mM NEM for 15 minutes at room temperature. The samples were subsequently analyzed by MALDI-TOF/TOF mass spectrometry which showed peaks for oxidized as well as NEM-reacted C_{H1}. As a control, reduced, BiP-bound C_{H1} was reacted under the same conditions with NEM and only showed peaks for NEM bound C_{H1}.

SDS-PAGE experiments. To assess the formation of the disulfide-bridged C_L/C_{H1} dimer, both proteins were co-incubated at a concentration of 25 µM each in PBS at 25°C. If present, the concentration of CypB was 5 µM. At the indicated times, a 20 µl aliquot was withdrawn, supplemented with 10 µl 3x Laemmli sample buffer (without

β -mercaptoethanol) and boiled for 1 min at 95°C. Subsequently, 20 μ l of each sample was applied to a 17.5% SDS-PAGE gel and run for 75 min at 25 mA. The bands were quantified with ImageQuant TL (GE Healthcare, München, Germany).

Supplemental References

Feige, M.J., Hagn, F., Esser, J., Kessler, H., and Buchner, J. (2007). Influence of the internal disulfide bridge on the folding pathway of the C-L antibody domain. *Journal of Molecular Biology* 365, 1232-1244.

Feige, M.J., Walter, S., and Buchner, J. (2004). Folding mechanism of the C(H)2 antibody domain. *Journal of Molecular Biology* 344, 107-118.

Supplemental Figure Legends

Figure S1. Formation of tertiary and quaternary structure upon C_{H1}/C_L association. (A) The isolated C_L domain shows a well defined near-UV CD spectrum (cyan), whereas the isolated C_{H1} domain shows the featureless spectrum of an unfolded protein (blue). The complex between C_L and C_{H1} is shown in green and the spectrum of the C_{H1} domain in the complex in red. The change in near-UV CD signal at 280 nm upon the co-incubation of C_L and C_{H1} is shown in (B). The trace could be fitted with a single exponential function with a time constant of $\tau = 50 \pm 2$ min (C) Formation of stable C_{H1}/C_L quaternary structure was assessed by analytical HPLC-experiments using 25 μ M of each protein. The elution profiles after 1 min of co-incubation of C_L and C_{H1} (dashed line) and after 5 h of co-incubation (straight line) are shown. In (D), the normalized intensity of the peak corresponding to the C_L/C_{H1} dimer (15 min) is plotted over time. It could be fitted with a single exponential function ($\tau = 65 \pm 5$ min). All measurements were performed in PBS at 25°C.

Figure S2. Influence of the redox status of the C_{H1} domain on its folding status and its association-coupled folding process. (A) The far-UV CD spectrum of the reduced C_{H1} domain is characteristic of an unfolded protein. (B) 5 μ M oxidized (straight line) or reduced C_{H1} (dashed line) were added to 1 μ M lucifer yellow labeled C_L. The change in anisotropy over time could be well described as a single exponential function for oxidized C_{H1} ($k_{\text{obs}} = 0.19 \text{ min}^{-1}$), for reduced C_{H1} no change was observed. (C) For oxidized C_{H1} (straight line), folding in the presence of C_L could

be observed by far-UV CD-spectroscopy at 205 nm ($\tau = 44 \pm 4$ min) whereas no structure formation was detectable for the reduced C_H1 domain (dashed line). A concentration of 10 μM was used for each protein. All measurements were carried out at 25°C in PBS in the presence of 30 μM TCEP.

Figure S3. C_L/C_H1 association coupled to intermolecular disulfide bridge formation. (A) The isolated C_L domain with the C-terminal cysteine residue shows an all- β far-UV CD spectrum (cyan), whereas the isolated C_H1 domain with the C-terminal cysteine residue displays a random coil spectrum (blue). The complex of both proteins is shown in green. The spectrum of the C_H1 domain, calculated from the individual spectra of C_L and the C_H1/C_L complex, reflects folding of the C_H1 domain in the presence of C_L (red). (B) Folding of the C_H1 domain, followed by far-UV CD spectroscopy at 205 nm, occurred with a time constant of $\tau = 53 \pm 5$ min in the absence of CypB (red trace). It could be accelerated to $\tau = 38 \pm 4$ min in the presence of 2 μM CypB (blue trace). The acceleration could be inhibited by 4 μM CspA (not shown). All measurements were carried out in PBS at 25°C with C_L respectively C_H1 containing their native C-terminal cysteine residue.

Figure S4. Assessment of the formation of triple complexes between BiP, C_H1 and C_L . To investigate the possibility that either C_L could associate with BiP bound C_H1 or if the C_L/C_H1 complex binds to BiP, 1 μM ATTO532 labeled C_L was incubated with 10 μM ATTO594 labeled BiP (A). 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). After 30 min, where no binding of C_L to BiP could be detected as expected (data not shown), 10 μM

unlabeled C_H1 was added. Under these conditions, C_H1 binds to BiP within several minutes. If either C_L could bind stably to the BiP:C_H1 complex or if BiP could bind to the C_L/C_H1 complex, a FRET signal between BiP and C_L is expected. However, this was not observed arguing against the presence of BiP:C_H1:C_L triple complexes. As a control of the FRET system, the association between 10 μM ATTO594 labeled BiP and 1 μM ATTO532 labeled C_H1 was measured (B). The association reaction occurred with a rate of $k_{\text{obs}} = 0.01 \pm 0.005 \text{ min}^{-1}$. In (C, left) a putative triple complex between BiP (grey, model based on the pdb code 3C7N, chain B), C_H1 (purple) and C_L (blue) is shown. The labels are schematically shown in green (ATTO532) respectively in red (ATTO594) and distances are indicated. The Förster radius of the dye pair is 6.6 nm (www.atto-tec.de). On the right, a putative complex between BiP and C_H1 is shown. All measurements were performed at 25°C in HKM buffer with 1 mM ADP and donor fluorescence was linearly corrected for bleaching.

Figure S5. A folded C_L domain is required for the proper maturation and secretion of heavy chains from cells. COS-1 cells were co-transfected with vectors encoding BiP, the HA-tagged versions of either wild type light chain (LC_{wt}) or the mutant light chain containing an unfolded C_L domain (LC_{C_Lmut}) and either a humanized chimeric γ heavy chain or the MAK33 γ heavy chain. After 24 h, cells were metabolically labeled and both cell lysates (no subscript) and culture supernatants (subscript m) were immunoprecipitated with the indicated antisera. Precipitated proteins were separated by SDS-PAGE under non-reducing condition and visualized by autoradiography.

Supplemental Figures

Figure S1

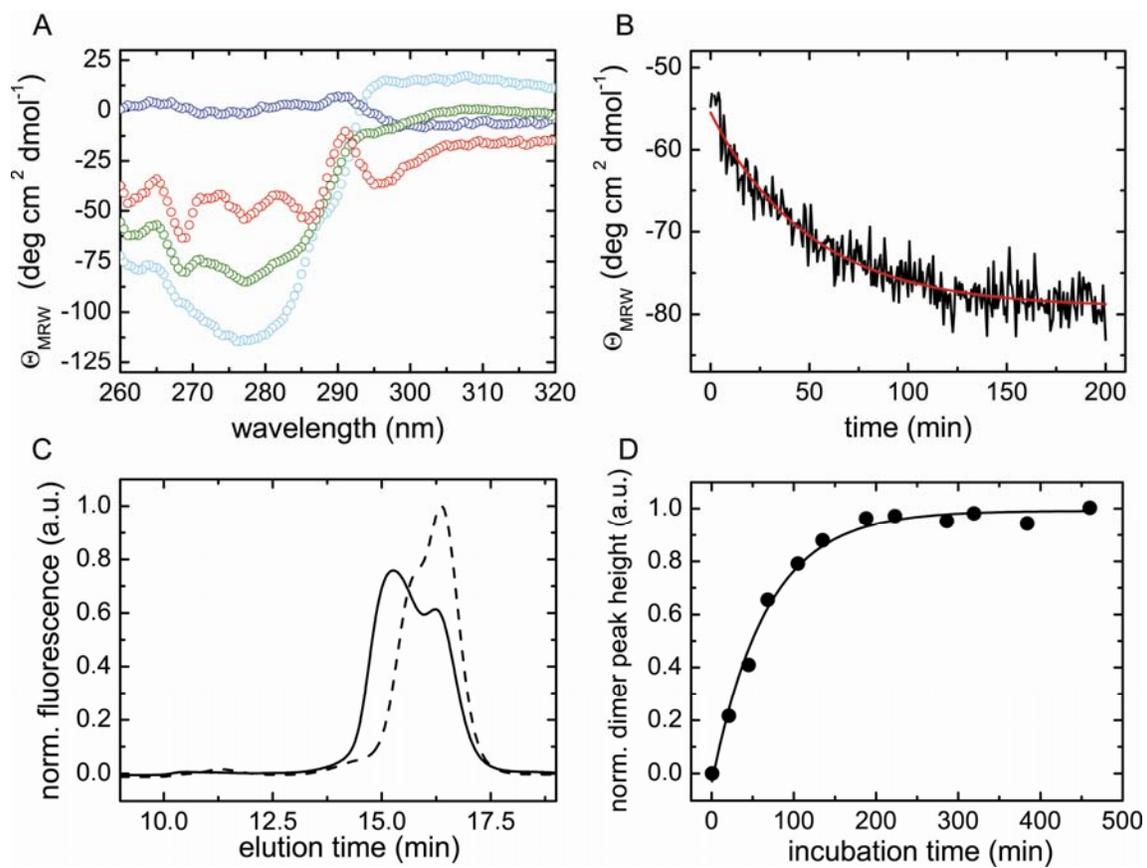


Figure S2

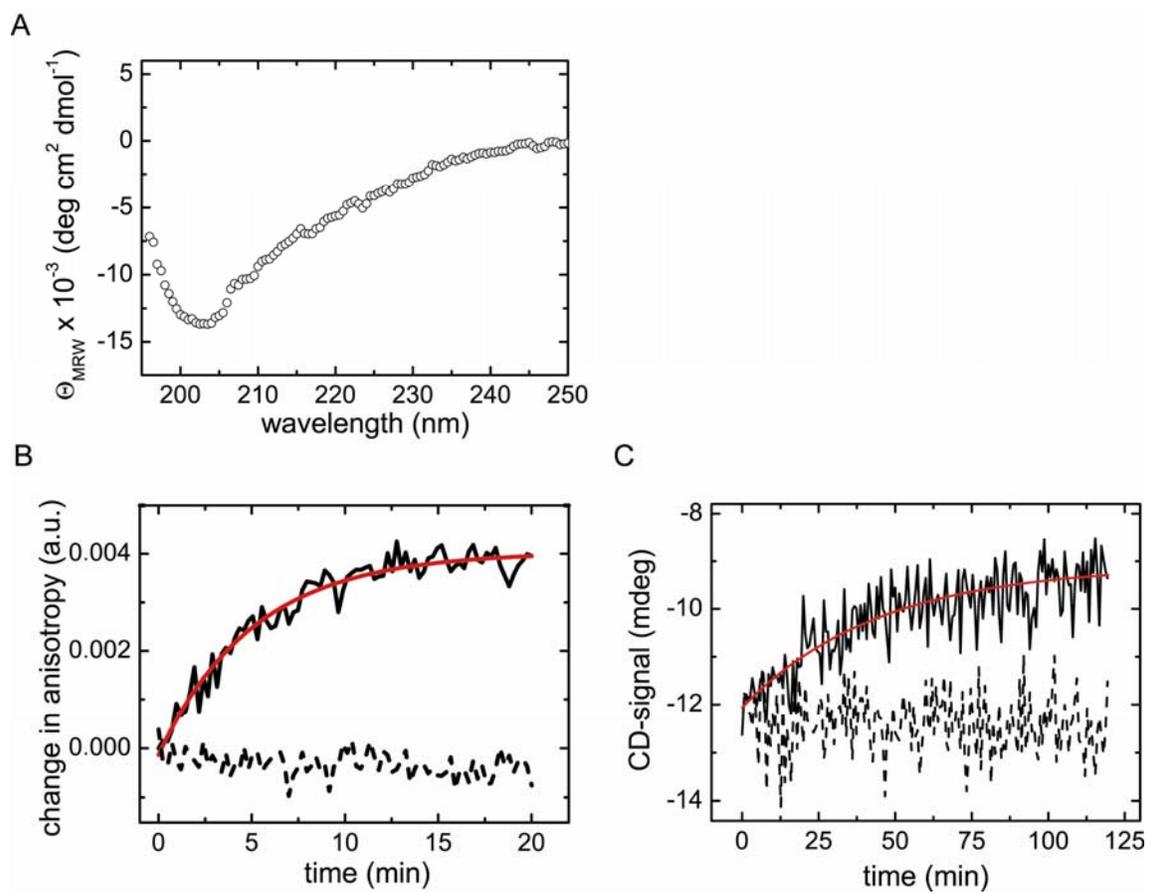


Figure S3

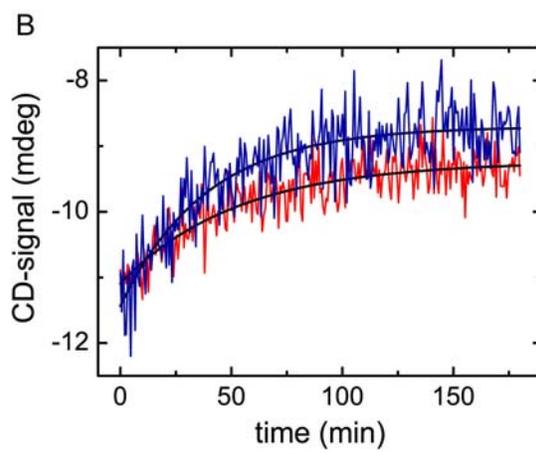
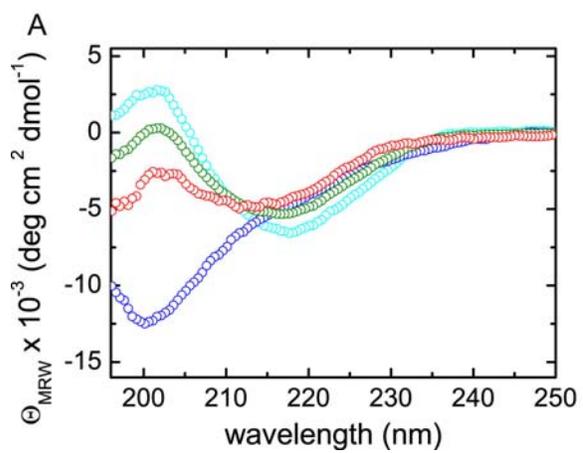


Figure S4

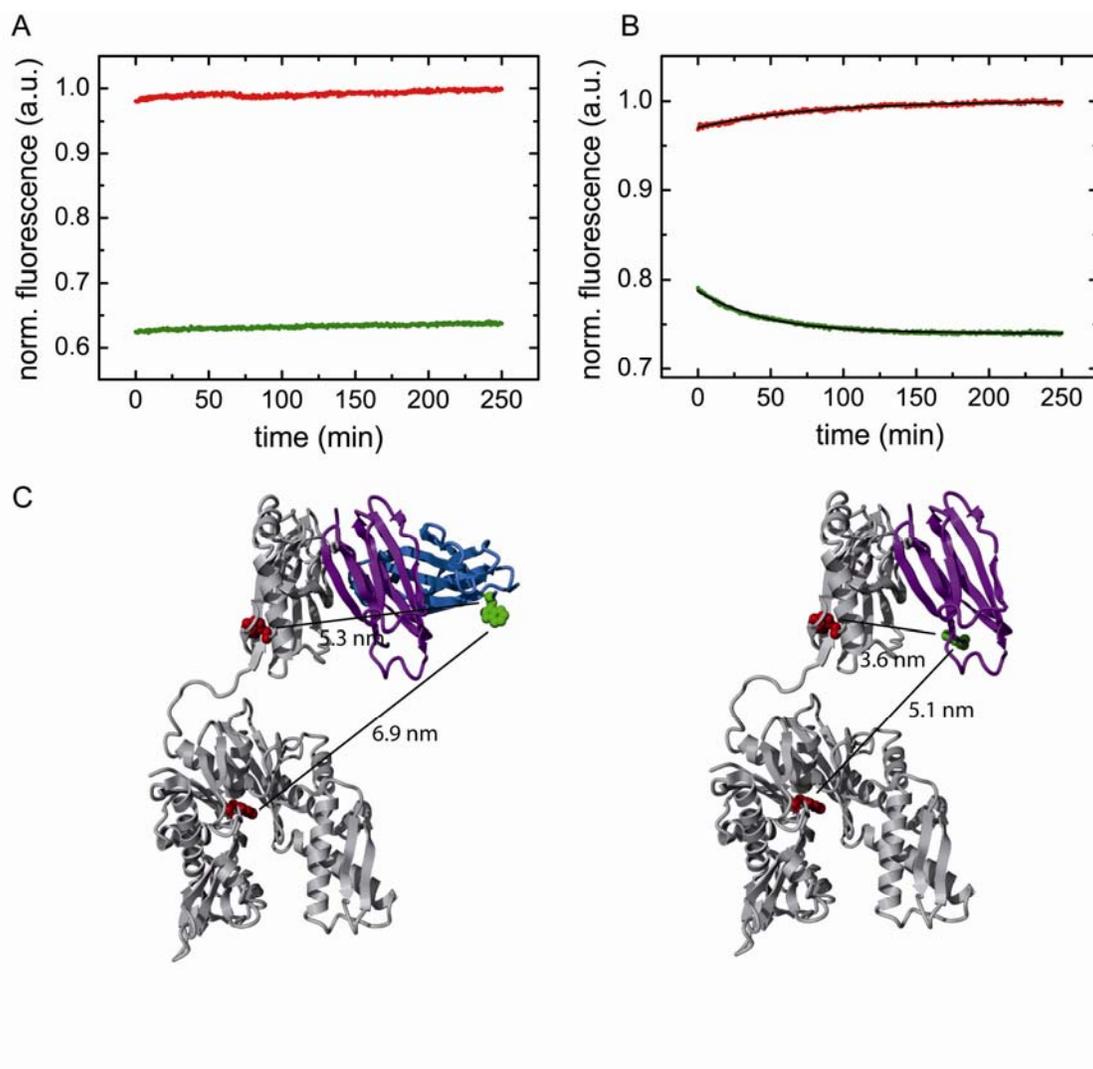
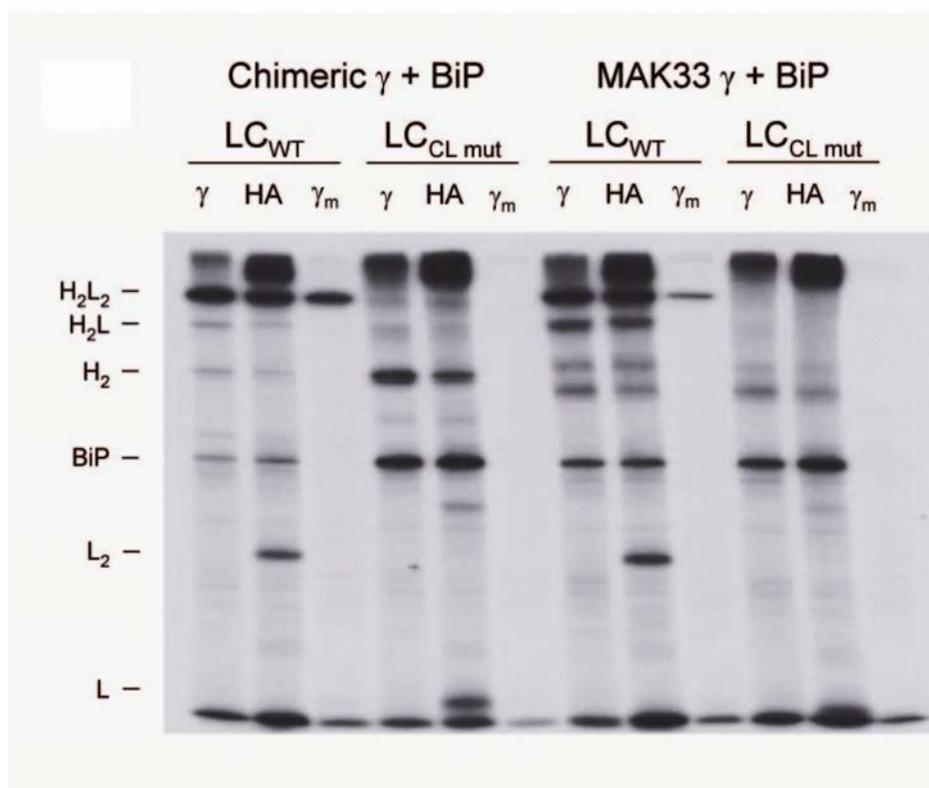


Figure S5



VI. Published book chapters

CHAPTER 2

Oxidative Folding of Proteins in vitro

CHAPTER 2.1

The Role of Disulfide Bonds in Protein Folding and Stability

MATTHIAS JOHANNES FEIGE AND
JOHANNES BUCHNER

Department Chemie, Technische Universität München, Lichtenbergstrasse 4,
85747 Garching, Germany

2.1.1 Introduction

Disulfide bonds add covalent cross-links to the linear polypeptide chain. It is therefore intuitive to assume that this posttranslational modification has a pronounced impact on the folding and stability of proteins. These covalent linkages are found mostly in extracellular proteins where they contribute significantly to the stability of the respective protein. An extension to the intrinsic stabilization of extracellular proteins by disulfide bridges is the stabilization of quaternary structure by intermolecular cystine links. This is an effective strategy to prevent dissociation in the extracellular environment which, due to the often low concentrations, would in many cases be irreversible. However, in the context of discussing the effects on protein structure and folding, these disulfides are set apart from the intra-chain disulfides.

RSC Biomolecular Sciences
Oxidative Folding of Peptides and Proteins
Edited by Johannes Buchner and Luis Moroder
© Royal Society of Chemistry 2009
Published by the Royal Society of Chemistry, www.rsc.org

Interestingly, disulfide bonds in proteins are highly conserved; only Trp residues are even more conserved.¹ The natural selection for disulfide bonds in secreted proteins seems to be due to the simple and reversible redox chemistry involved in their formation and breaking and their intrinsic stability. Only at extreme conditions such as 100 °C and alkaline pH values do they spontaneously break or rearrange.² Disulfide bridges come in different flavors, reflecting the different purposes they can fulfill in a protein. In addition to their effect on the global stability of a polypeptide chain, they may have important effects on the structure, stability and dynamics of local structural elements. This is especially evident for the so-called cysteine knots, in which disulfide bridges cross each other and thus determine the local topology³ and for “allosteric” disulfide bonds which serve to regulate protein function.⁴ In this review we will focus exclusively on the dissection of the effects of structural, in particular non-local, disulfide bonds on protein stability and folding.

The effect a disulfide bond will have on a given protein will of course depend on the position of the disulfide bond in the structure of the protein. The complexity of the influence of disulfide bonds is reflected by the finding that stabilizing disulfide bonds can be found either in the core of proteins as such or surface-localized. An important issue in this context is the distance between two Cys residues in the sequence of a polypeptide chain. As will be pointed out in this chapter, the stabilizing effect depends to a significant extent on the number of amino acids spanned between the two residues. In agreement with this notion the average distance of two Cys residues in proteins was found to be 15 amino acids.¹

Disulfide-bonded proteins have been studied early on in the history of protein folding. Notably, one of the first proteins analyzed extensively, RNase A, contains several disulfide bridges. In seminal work, Christian Anfinsen showed that the completely reduced and denatured protein could spontaneously regain its native structure including the correct disulfide bonds.⁵ Efforts to analyze the impact of disulfide bonds on protein folding and stability mechanistically have contributed enormously to the generation of concepts for these processes. Furthermore, engineered disulfide bonds have become valuable tools to explore them in detail. In this chapter we will discuss the basic models resulting from these studies and we will outline the thermodynamic and kinetic implications the presence of a disulfide bond has on unfolded, folding and folded proteins.

2.1.2 Stabilization of Proteins by Disulfide Bonds

In the early days of protein science, the native state of a protein was believed to be one, rather fixed, conformation and the unfolded state to be a random coil.⁶ An ideal random coil is devoid of any long-range interactions except excluded volume effects. It behaves as a freely joined chain of segments of defined length.⁷ In this description, the impact of a covalent cross-link between two defined residues of the polypeptide chain, such as a disulfide bridge, was thought to be exclusively on the unfolded state. The fixed geometry of the native state should be left essentially unaltered by a bond between two residues which are in proximity, yet the freedom

of the random coil polypeptide should be significantly decreased. Restricting the conformational space of the unfolded state clearly reduces its entropy. Hence, the entropy change for the reaction to the ordered native state was thought to be less negative with a net stabilization of the folded protein as a result. A quantitative description of the phenomenon was developed by Schellman, Flory, Poland and Scheraga.⁸⁻¹⁰ The decrease in entropy of the unfolded state is derived from the probability that two otherwise free elements of the chain are now found in a defined volume element (v). Mathematically and based on polymer theory, the problem is described in Equation (2.1.1) (where R is the gas constant, l the average length of a statistical segment of the chain composed of N segments; in proteins l is assumed to be 3.8 Å corresponding to one amino acid):

$$\Delta S = -R \ln[3/(2\pi l^2 N)^{3/2}]v \quad (2.1.1)$$

A major point of discussion has been the adequate choice of v . A value of 57.9 \AA^3 based on the closest possible approach of two thiols is still mostly in use.¹¹ Hence, Equation (2.1.1) can be simplified to Equation (2.1.2) where n is the number of amino acids bridged by the disulfide bond:

$$\Delta S = -2.1 - 3/2 R \ln(n) \quad (2.1.2)$$

Based on a study of Ribonuclease T1 with zero, one and two intact disulfide bonds, Equation (2.1.2) was developed by Pace *et al.*¹¹ The authors did not only find a good correlation between n and $\Delta\Delta G$ upon removal of disulfide bonds in RNase T1, but additionally agreement between experimental data for lysozyme, RNaseA and the antibody C_L domain.¹¹

The above equations have two consequences. Conceptually, the stabilization of a protein is thought to be an entirely entropy-driven process with an impact exclusively on the random coil unfolded state. Practically, the stabilization exerted by a disulfide bridge should increase with the number of amino acids between the two cysteines. This theory treats the unfolded polypeptide chain as a system devoid of any intra- or intermolecular interactions. However, the water surrounding a protein is an important factor shaping the free energy landscape of the polypeptide chain during folding and in the native state.¹² Sometimes it can be regarded as being a part of the native structure.¹³ The impact of this scenario with respect to disulfide bonds has been addressed by Doig and Williams in a widely recognized publication in 1991.¹⁴ The authors argued that disulfide bonds might lead to a significantly decreased solvent accessible surface in the unfolded state. Hence, hydrophobic residues as well as hydrogen-bond donors and acceptors might become buried. This is supposed to lead on the one hand to a larger entropy of the solvent in disulfide-containing proteins due to buried hydrophobic residues. Consequently, the hydrophobic effect should be less pronounced for these proteins. On the other hand, hydrogen bonding to water in the more compact unfolded state will be possible to a lesser extent, and hence folding to the native state will be enthalpically more favorable. According to the authors, the enthalpic contribution has to be considered as the major stabilizing factor of

disulfide bridges. The model will therefore be called *solvent enthalpy model* in the subsequent paragraphs. As in the *chain entropy model*, eventually occurring differences in the native state such as induced strain or reduced dynamics are also neglected in this model. Both models are summarized in Figure 2.1.1 where the reduced conformational freedom of a disulfide-linked polypeptide chain is visualized as well as hydrophobic clustering in the unfolded state or reduced solvent-protein interactions which might be present for a disulfide-linked protein.

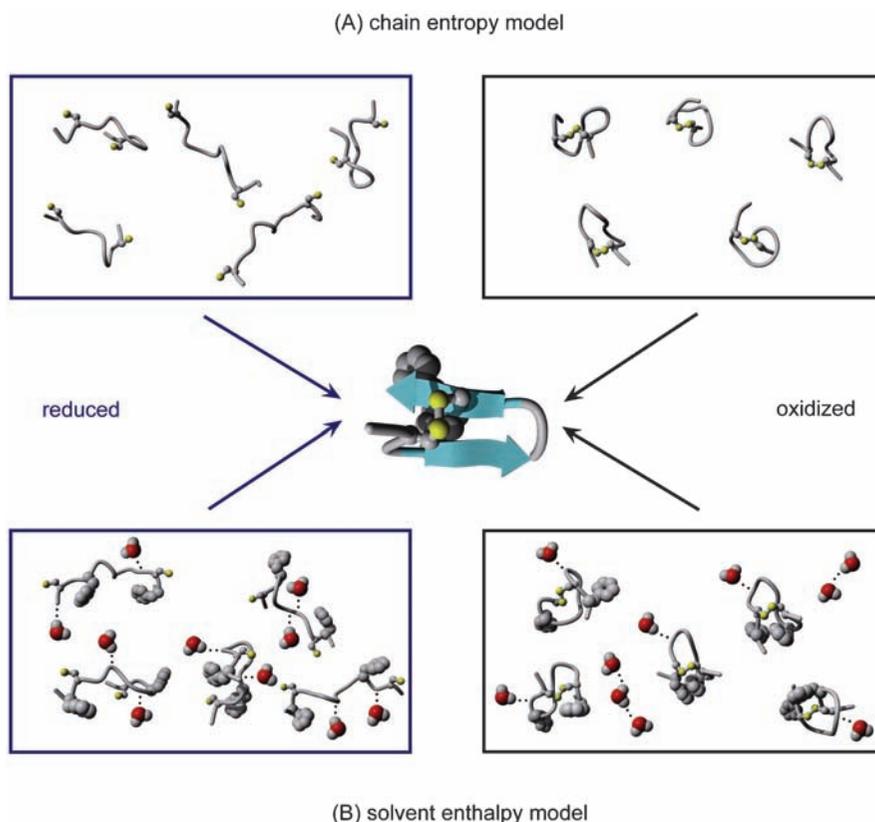


Figure 2.1.1 Graphical representation of the chain entropy and the solvent enthalpy model. In the chain entropy model, the major stabilizing factor exerted by a disulfide bridge is believed to be the conformational restriction of the unfolded state which renders the folding reaction entropically less unfavorable (A). In the solvent enthalpy model (B), two major effects are attributed to disulfide bridges. First, hydrophobic interactions are believed to be more pronounced in unfolded, disulfide-bridged proteins decreasing the hydrophobic effect for a protein folding reaction. Second, and numerically more important, disulfide bridges are assumed to inhibit hydrogen bonding to water in the unfolded state, making the folding reaction enthalpically more favorable. In both theories, the native state is believed to be unaffected. To illustrate both models, a hypothetical 12-residue β -hairpin structure with a single disulfide bridge was designed.

A variety of experimental evidence argues against both theories rendering the problem much more complex yet more insightful. Site-specific mutagenesis offered the possibility to introduce artificial disulfide bonds at defined positions within a protein and allowed to explore the effects of disulfide bonds rigorously.¹⁵⁻¹⁸ Many results of these studies were not compatible with predictions based on the prevailing theories concerning the effects of disulfide bonds on protein stability. Most discrepancies arose from efforts to stabilize a variety of proteins by the introduction of new disulfide bridges. In contrast to expectations, destabilization was the result of a significant proportion of these attempts.¹⁹⁻²¹ High-resolution structures of the engineered proteins helped to rationalize these unexpected effects. It turned out that strain was imposed on the native state by the disulfide cross-links in several cases.²² Two particularly insightful examples are studies on barnase^{19,23} and staphylococcal nuclease.²¹ Clarke and Fersht introduced three artificial disulfide bonds into barnase. One was found to destabilize the native state, one stabilized it to an extent predicted by the chain entropy model and one to a much lesser extent.¹⁹ Furthermore, the bridge connecting fewer residues was found to be more stabilizing than the one encompassing more residues. For staphylococcal nuclease, no stabilization was found for all disulfide constructs, but in this case a peptide bond *cis/trans* equilibrium was shifted in the native state and the catalytic activity of all mutants was reduced.²¹ The NMR H/D exchange analysis of the barnase mutants revealed an altered dynamics of the native state imposed by the presence of the covalent cross-links.²³ In the study on staphylococcal nuclease, strain on the native state, as reflected by the alteration of the *cis/trans* equilibrium, was evoked to explain the observed unexpected alterations.²¹ A more general comparison between experimental data for disulfide-bridged proteins with native or engineered disulfide bonds and the chain entropy model, which predicts that disulfide bonds stabilize proteins in a strictly loop length dependent manner, can be found in Figure 2.1.2. A clear deviation between experimental data and predictions is evident from Figure 2.1.2 for the major part of the proteins.

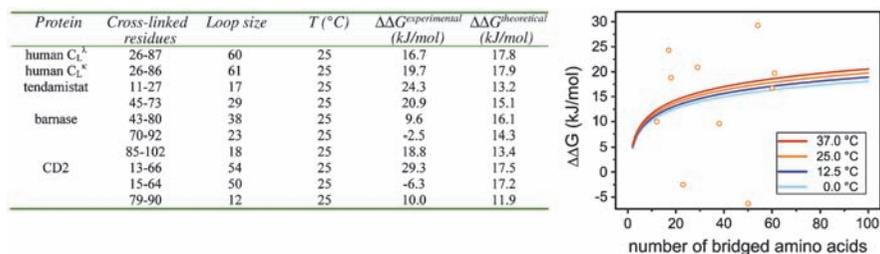


Figure 2.1.2 Comparison of experimental data and theory for the chain entropy model. Experimental data on the stability of proteins in the reduced and oxidized state are summarized in the table on the left.^{11,20,70,75} A comparison to the predicted stabilization based on Equation (2.1.2) is shown on the right. The chain entropy model predicts a larger stabilization *via* disulfide bridges at higher temperatures and longer loop lengths.

One clear conclusion can be drawn from the two mentioned and a variety of other studies:^{15,16,18,22,24-29} the native state is not left unaffected by a disulfide bridge. A major assumption underlying the chain entropy model as well as the solvent enthalpy model hence does clearly not hold in all cases. Strain is often found to be imposed on the native state by the presence of covalent cross-links. This can be reflected in structural changes^{21,22} or more subtle changes, like alterations of the dynamics of the native state.³⁰ The important consequence is that the enthalpy as well as the entropy of the native state are not unlikely to be altered if two residues in the polypeptide are covalently cross-linked. Furthermore, the alterations in dynamics and structure are not always a global but sometimes a local context-specific effect. β -sheets and loops are thought to be more suited to dissipate induced strain than α -helical elements, and more dynamic parts of the structure are influenced to a greater extent. In addition, changes in the solvent-exposed hydrophobic surface which have been observed for disulfide mutants of interleukin-4 have a significant effect on the enthalpy of the native state.²⁶ Changes in the dynamics, *e.g.* the vibrational normal modes of a protein, will also influence the entropy of the native state as has been shown in a molecular dynamics study by Karplus and co-workers.³⁰ Consequently, a net destabilization due to a loss of native state entropy can be expected in some cases if disulfide bonds are introduced into a protein.³⁰ In summary, novel experimental as well as theoretical insights clearly argue against the simple chain entropy model, at least concerning one side of the equation, the native state. The same holds for the solvent enthalpy model by Doig and Williams, which is additionally at odds with thermodynamic parameters derived for some disulfide-bridged proteins which showed that disulfide bridges do not necessarily stabilize the native state enthalpically.²⁶ But what about the unfolded state? Do the assumptions underlying the theory hold? In other words, is it correct to assume a random coil almost devoid of any interactions except for hydrogen bonding to the solvent? Clearly not in all cases. Residual structure which is believed to be important for folding pathways has been detected in a variety of proteins.³¹⁻³⁵ In particular, residual hydrophobic interactions or fluctuating α -helical elements seem to be a rather general than an exceptional feature of proteins, in particular under mildly denaturing conditions.^{32,34,36-38} Importantly, the structural features of the unfolded state like residual structure, which has been reported for the unfolded state of barnase,³⁹ or the tendency towards irreversible inactivation have in some cases been shown to be influenced by the presence or absence of disulfide bridges.^{40,41} Hence, disulfide bonds very likely not only influence the conformational freedom of the unfolded chain but can also introduce structure which might also be protective against irreversible aggregation.⁴⁰ Even apparently minor structural changes like clustering of some hydrophobic residues will influence the enthalpy and entropy of the unfolded state. In addition, as outlined above, dynamic phenomena, strain induced on the native state and an impact on the native structure are completely omitted in these theories. Although some proteins could be adequately described by one of the theories, it comes as no surprise that the chain entropy model as well as the

solvent enthalpy model fail to describe the major part of experimental data (Figure 2.1.2).

How can the different findings on the divergent effects disulfide bridges have on the stability of different proteins be summarized in a comprehensive way? Hardly by any theory describing the unfolded polypeptide chain as a construct composed of identical elements and devoid of structural features. The decreased entropy of an unfolded and cross-linked polypeptide chain has to be taken into account as developed in the chain entropy model. Furthermore, the solvent enthalpy model based on the decreased hydrophobic effect and hydrogen bonding in the unfolded state should be included. And to be added to these models are interactions eventually present in the unfolded state due to the disulfide bridge, which may not be localized directly around the bond but can be present as long-range residual structure.³⁹ This will clearly have an effect on the enthalpy and entropy balance for the reaction to the native state. The same holds for decreased dynamics of the native state, locally or globally, as well as enthalpically unfavorable strain or enthalpically favorable induced proximity of interacting residues. In summary, the effect of a disulfide bridge on the stability of a protein can be easily assessed experimentally, yet its molecular explanation is almost as diverse as the protein under investigation. The key factors giving rise to the net effect are most likely all known but, as for the prediction of the native state of a protein, their contribution to the overall effect are blurred in their sum as well as their mutual influence on each other. It seems therefore highly rewarding to use a combined empirical and theoretical approach. Exact stability data, if possible together with structural data on the native as well as the unfolded state of the protein under investigation, are a prerequisite for the detailed understanding of the effect of a disulfide bridge. They should be complemented by theoretical approaches like molecular dynamics simulations of the native and the unfolded state to obtain a more complete picture. This might be different in detail for different proteins but in summary these analyses will most likely reveal general principles.

2.1.3 Disulfide Bonds in Protein Folding Reactions

One of the major questions in biophysical chemistry is how a linear polypeptide chain specifically adopts its intricate three-dimensional structure in a reasonable amount of time. A variety of mutational approaches,^{42–50} high-resolution structural techniques^{51–54} and ultrafast perturbation methods^{55–58} have recently provided deep insight into this phenomenon of biological self-organization. Nowadays, proteins are believed to fold with a certain heterogeneity *via* multiple individual pathways to the native state,⁵⁹ yet distinct features of the protein dictate the general trajectories of the folding process. In particular, residues whose interactions define the overall topology of a protein are assumed to be in contact in the transition state and are believed to be key elements in a protein folding reaction in general.^{60–62} While there seems to be consensus concerning the general scheme of events, a variety of details are still under intense debate, like

the role of residual structure in the unfolded state,³⁵ the role of folding intermediates on the way to the native state⁶³ and the heterogeneity of the transition states.⁶¹ In all these questions, disulfide bridges have played a pronounced role as factors influencing these processes as well as tools to elucidate them.

Independent of its pathway, every protein begins its folding reaction in the unfolded state, be it at the ribosome or under denaturing conditions. Therefore, the nature of the unfolded state of proteins has gained much attention again recently, in part fuelled by the development of novel experimental techniques.^{34,64} In this context, deviations from a random coil are regarded as important elements of protein folding in general. They will not only have an impact on the net stability balance of a protein, as outlined above, but furthermore likely shape the energy landscape on the way to the native state. Preformed interactions in the unfolded state might therein have a rather controversial effect on a folding reaction. If native and not detrimentally influencing the ability of the remaining polypeptide chain to explore the necessary conformational space, they might lead to faster and more efficient folding to the native state. Yet if non-native or too stable, the opposite might hold. Examples for both cases have been reported for disulfide-bridged proteins, which *per se* possess preformed correct tertiary interactions. In the case of RNase T1 (Figure 2.1.3), deceleration of the folding kinetics was observed which has been attributed to decreased chain flexibility in the presence of a disulfide bridge.⁶⁵ In this context it is important to note that RNase T1 possesses two disulfide bridges, one connecting a small N-terminal β -turn and one connecting the C-terminus to this N-terminal β -turn (Figure 2.1.3). Accordingly, the protein will almost be a completely looped structure in its unfolded state, which will clearly have an effect on the dynamics of the whole polypeptide chain and might interfere with the establishment of a folding nucleus, if not in proximity of the disulfide bridges, which is unlikely due to the solvent exposure of the disulfide bridges in the native state. This has been confirmed for one of the disulfide bonds connecting residues 2 and 10.

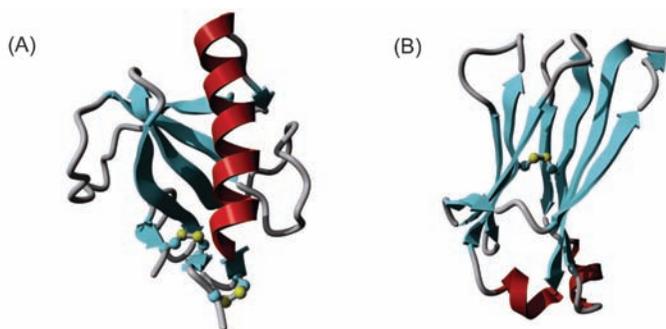


Figure 2.1.3 Structural comparison of RNase T1 and the murine C_L domain. RNaseT1 (A) possesses two solvent-exposed disulfide bridges. The murine immunoglobulin G C_L domain has one disulfide bridge buried in the hydrophobic core of the protein (B).

Its deletion led to a stability reduction of RNase T1 but had no impact on the folding mechanism of the protein.⁶⁶

For C_L, the constant domain of the antibody light chain, on the other hand, its single intrinsic disulfide bridge accelerated folding up to ~100-fold (Figure 2.1.3).^{67,68} In the case of this protein, the disulfide bond is found in the hydrophobic core of the protein and part of the folding nucleus.⁶⁸ Accelerated folding due to preformed interactions which facilitate the way to the native state were hence expected and experimentally observed. For many other proteins, either a deceleration or an acceleration of the respective folding reactions have been observed in the presence of their natural intrinsic disulfide bonds.^{20,69–74} Hence, disulfide bonds are in general far from being inert in kinetic terms. Very often, different disulfide bridges within the same protein had different effects on the folding rates. A comprehensive study in this respect has been carried out for the all- β protein CD2, where thirteen different artificially introduced disulfide bridges showed a markedly different impact on the folding behaviour.²⁰ Similar results were obtained for unfolding reactions which were either left unaffected by disulfide bonds or their rates were reduced.^{20,75} The experimental findings are in agreement with simple lattice-based simulations, where disulfide bonds inside the folding nucleus were found to be accelerating yet decelerating if outside.⁷⁶ In some simulations folding has been found to be influenced by a disulfide bond to a lesser extent than expected, revealing a larger dynamics of a real polypeptide chain in comparison to a chain moving on a lattice.⁷⁶ Despite the heterogeneous effects of a disulfide bridge on protein folding/unfolding kinetics at first glance, it can be more easily rationalized than the effect on native state stability. The key lies in the transition state for folding and unfolding. Acceleration of folding is expected when residues are cross-linked which have to come into contact early in a protein folding reaction, if not weighed out by entropy/enthalpy compensation in the transition state. Analogously, deceleration of unfolding is expected when two residues, whose interactions are broken in the transition state for unfolding are covalently linked. Both effects can provide structural information about the otherwise hardly accessible transition state as exemplified for the immunoglobulin proteins C_L and CD2 where the transition state for folding could be stabilized by disulfide bridges^{20,68} or for barnase, where the transition state for unfolding could be destabilized by a disulfide bond.⁷⁵ Often, disulfide-bridged proteins are found to fold less cooperatively.⁷⁰ This might be caused by the population of disulfide-stabilized intermediates in agreement with the stabilization of the transition state and subsequent partially folded states. If stabilization of partially folded structures becomes too strong, either by native or non-native interactions, this can even result in a net deceleration of a protein folding reaction as has been observed for CD2.⁷¹ Attempts to increase the folding rate by multiple disulfide bonds had the opposite effect: deceleration of the folding reaction by the over-stabilization of a partially folded state.⁷¹ This highlights the role of cooperativity for efficient protein folding which can be beneficially or detrimentally influenced by preformed tertiary interactions like disulfide bridges.

2.1.4 Conclusions

Disulfide bonds are one of the most widespread structural elements stabilizing the native state of a protein. Their stabilizing role likely arises from a variety of effects imposed on the unfolded as well as the native state. A destabilization of the unfolded state due to the restriction of the conformational freedom as well as decreased protein-solvent interaction will favor folding to the native state which, additionally, can be stabilized by induced proximity of energetically favorable interactions. In most cases, in particular for naturally selected disulfide-bond positions, the stabilizing effects of this covalent bond are more pronounced than eventually occurring destabilizing effects. These can include a stabilization of the unfolded state due to residual or non-native structure. Additionally, strain as well as reduced dynamics imposed on the native state can significantly decrease the stability of the native state. The sum of all these factors will be the net stabilizing effect of a disulfide bridge.

Having a closer look at the kinetic impact of disulfide bridges is highly insightful. By accelerated folding or decelerated unfolding, entropic, enthalpic and structural conclusions about otherwise almost inaccessible transitions states can be drawn. Furthermore, how disulfide bonds are positioned in natural proteins might not only help in the design of proteins with improved characteristics but also hold a clue for the specific structural features a protein has been selected for. In this respect, it is particularly revealing to look at disulfide bridge positions in an evolutionary perspective. For many extracellular proteins which do not have to undergo a large variety of binding reactions requiring large scale dynamics, loop structures and flexible parts are often found to be disulfide-linked.⁷⁷ This, on the one hand, directly reduces susceptibility to protease digestion and, on the other hand, might generally reduce unfolding rates by linking parts which are likely to come apart early in an unfolding reaction. If large flexibility is needed, as *e.g.* in members of the immunoglobulin superfamily, which are optimized for molecular recognition processes, another evolutionary strategy has prevailed. Here, disulfide bridges are found in the folding nucleus,⁷⁷ where they seem to accelerate refolding if unfolding occurs. At the same time this strategy permits flexibility where needed and also results in a net stabilization of the native state. Positions of evolutionary selected disulfide bonds might hence not only provide insight into folding nuclei but additionally provide a basis for the rational stabilization of homologous proteins.

The elucidation of the versatile effects disulfide bonds have on protein structure, stability and folding has significantly extended our knowledge about how proteins fold and function and it paved the path to influence their properties in desirable ways.

References

1. J. M. Thornton, *J. Mol. Biol.*, 1981, **151**, 261–287.
2. T. J. Ahern and A. M. Klibanov, *Science*, 1985, **228**, 1280–1284.

3. N. W. Isaacs, *Curr. Opin. Struct. Biol.*, 1995, **5**, 391–395.
4. B. Schmidt and P. J. Hogg, *BMC Struct. Biol.*, 2007, **7**, 49.
5. C. B. Anfinsen and E. Haber, *J. Biol. Chem.*, 1961, **236**, 1361–1363.
6. C. Tanford, K. Kawahara and S. Lapanje, *J. Biol. Chem.*, 1966, **241**, 1921–1923.
7. L. J. Smith, K. M. Fiebig, H. Schwalbe and C. M. Dobson, *Fold. Des.*, 1996, **1**, R95–106.
8. J. A. Schellman, *C.R. Trav. Lab Carlsberg. (Chim.)*, 1955, **29**, 230–259.
9. P. J. Flory, *J. Am. Chem. Soc.*, 1956, **78**, 5222–5235.
10. D. C. Poland and H. A. Scheraga, *Biopolymers*, 1965, **3**, 379–399.
11. C. N. Pace, G. R. Grimsley, J. A. Thomson and B. J. Barnett, *J. Biol. Chem.*, 1988, **263**, 11820–11825.
12. R. Lumry and S. Rajender, *Biopolymers*, 1970, **9**, 1125.
13. B. P. Schoenborn, A. Garcia and R. Knott, *Prog. Biophys. Mol. Biol.*, 1995, **64**, 105–119.
14. A. J. Doig and D. H. Williams, *J. Mol. Biol.*, 1991, **217**, 389–398.
15. R. T. Sauer, K. Hehir, R. S. Stearman, M. A. Weiss, A. Jeitlernilsson, E. G. Suchanek and C. O. Pabo, *Biochemistry*, 1986, **25**, 5992–5998.
16. J. E. Villafranca, E. E. Howell, S. J. Oatley, N. H. Xuong and J. Kraut, *Biochemistry*, 1987, **26**, 2182–2189.
17. T. Vogl, R. Brengelmann, H. J. Hinz, M. Scharf, M. Lotzbeyer and J. W. Engels, *J. Mol. Biol.*, 1995, **254**, 481–496.
18. C. M. Johnson, M. Oliveberg, J. Clarke and A. R. Fersht, *J. Mol. Biol.*, 1997, **268**, 198–208.
19. J. Clarke, K. Henrick and A. R. Fersht, *J. Mol. Biol.*, 1995, **253**, 493–504.
20. J. M. Mason, N. Gibbs, R. B. Sessions and A. R. Clarke, *Biochemistry*, 2002, **41**, 12093–12099.
21. A. P. Hinck, D. M. Truckses and J. L. Markley, *Biochemistry*, 1996, **35**, 10328–10338.
22. B. A. Katz and A. Kossiakoff, *J. Biol. Chem.*, 1986, **261**, 5480–5485.
23. J. Clarke, A. M. Hounslow and A. R. Fersht, *J. Mol. Biol.*, 1995, **253**, 505–513.
24. H. Kikuchi, Y. Goto, K. Hamaguchi, S. Ito, M. Yamamoto and Y. Nishijima, *J. Biochem.*, 1987, **102**, 651–656.
25. R. Kuroki, K. Inaka, Y. Taniyama, S. Kidokoro, M. Matsushima, M. Kikuchi and K. Yutani, *Biochemistry*, 1992, **31**, 8323–8328.
26. D. C. Vaz, J. R. Rodrigues, W. Sebald, C. M. Dobson and R. M. M. Brito, *Protein Sci.*, 2006, **15**, 33–44.
27. B. Asgeirsson, B. V. Adalbjornsson and G. A. Gylfason, *Biochim. Biophys. Acta*, 2007, **1774**, 679–687.
28. A. Shimizu-Ibuka, H. Matsuzawa and H. Sakai, *Biochemistry*, 2004, **43**, 15737–15745.
29. A. McAuley, J. Jacob, C. G. Kolvenbach, K. Westland, H. J. Lee, S. R. Brych, D. Rehder, G. R. Kleemann, D. N. Brems and M. Matsumura, *Protein Sci.*, 2008, **17**, 95–106.
30. B. Tidor and M. Karplus, *Protein. Struct. Funct. Genet.*, 1993, **15**, 71–79.

31. A. Lapidot and C. S. Irving, *J. Am. Chem. Soc.*, 1977, **99**, 5489–5490.
32. D. Neri, M. Billeter, G. Wider and K. Wüthrich, *Science*, 1992, **257**, 1559–1563.
33. V. L. Arcus, S. Vuilleumier, S. M. Freund, M. Bycroft and A. R. Fersht, *Proc. Natl. Acad. Sci. USA*, 1994, **91**, 9412–9416.
34. K. H. Mok, L. T. Kuhn, M. Goez, I. J. Day, J. C. Lin, N. H. Andersen and P. J. Hore, *Nature*, 2007, **447**, 106–109.
35. E. R. McCarney, J. E. Kohn and K. W. Plaxco, *Crit. Rev. Biochem. Mol. Biol.*, 2005, **40**, 181–189.
36. D. Eliezer, J. Chung, H. J. Dyson and P. E. Wright, *Biochemistry*, 2000, **39**, 2894–2901.
37. J. Yao, J. Chung, D. Eliezer, P. E. Wright and H. J. Dyson, *Biochemistry*, 2001, **40**, 3561–3571.
38. C. S. Le Duff, S. B. Whittaker, S. E. Radford and G. R. Moore, *J. Mol. Biol.*, 2006, **364**, 824–835.
39. J. Clarke, A. M. Hounslow, C. J. Bond, A. R. Fersht and V. Daggett, *Protein Sci.*, 2000, **9**, 2394–2404.
40. L. J. Perry and R. Wetzel, *Science*, 1984, **226**, 555–557.
41. R. Wetzel, L. J. Perry, W. A. Baase and W. J. Becktel, *Proc. Natl. Acad. Sci. USA*, 1988, **85**, 401–405.
42. A. R. Fersht, A. Matouschek and L. Serrano, *J. Mol. Biol.*, 1992, **224**, 771–782.
43. A. Matouschek, L. Serrano, E. M. Meiering, M. Bycroft and A. R. Fersht, *J. Mol. Biol.*, 1992, **224**, 837–845.
44. A. Matouschek, L. Serrano and A. R. Fersht, *J. Mol. Biol.*, 1992, **224**, 819–835.
45. A. Matouschek, J. T. Kellis Jr., L. Serrano, M. Bycroft and A. R. Fersht, *Nature*, 1990, **346**, 440–445.
46. A. Matouschek, J. T. Kellis Jr., L. Serrano and A. R. Fersht, *Nature*, 1989, **340**, 122–126.
47. L. Serrano, J. T. Kellis Jr., P. Cann, A. Matouschek and A. R. Fersht, *J. Mol. Biol.*, 1992, **224**, 783–804.
48. L. Serrano, A. Matouschek and A. R. Fersht, *J. Mol. Biol.*, 1992, **224**, 847–859.
49. L. Serrano, A. Matouschek and A. R. Fersht, *J. Mol. Biol.*, 1992, **224**, 805–818.
50. D. P. Raleigh and K. W. Plaxco, *Protein Pept. Lett.*, 2005, **12**, 117–122.
51. D. M. Korzhnev, X. Salvatella, M. Vendruscolo, A. A. Di Nardo, A. R. Davidson, C. M. Dobson and L. E. Kay, *Nature*, 2004, **430**, 586–590.
52. T. L. Religa, J. S. Markson, U. Mayor, S. M. V. Freund and A. R. Fersht, *Nature*, 2005, **437**, 1053–1056.
53. U. Mayor, N. R. Guydosh, C. M. Johnson, J. G. Grossmann, S. Sato, G. S. Jas, S. M. V. Freund, D. O. V. Alonso, V. Daggett and A. R. Fersht, *Nature*, 2003, **421**, 863–867.
54. H. Q. Feng, Z. Zhou and Y. W. Bai, *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 5026–5031.

55. T. L. Religa, C. M. Johnson, D. M. Vu, S. H. Brewer, R. B. Dyer and A. R. Fersht, *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 9272–9277.
56. H. Ma, C. Wan and A. H. Zewail, *J. Am. Chem. Soc.*, 2006, **128**, 6338–6340.
57. W. A. Eaton, V. Munoz, S. J. Hagen, G. S. Jas, L. J. Lapidus, E. R. Henry and J. Hofrichter, *Annu. Rev. Biophys. Biomol. Struct.*, 2000, **29**, 327–359.
58. J. Kubelka, T. K. Chiu, D. R. Davies, W. A. Eaton and J. Hofrichter, *J. Mol. Biol.*, 2006, **359**, 546–553.
59. K. A. Dill and H. S. Chan, *Nat. Struct. Biol.*, 1997, **4**, 10–19.
60. K. Lindorff-Larsen, P. Rogen, E. Paci, M. Vendruscolo and C. M. Dobson, *Trends Biochem. Sci.*, 2005, **30**, 13–19.
61. K. Lindorff-Larsen, M. Vendruscolo, E. Paci and C. M. Dobson, *Nat. Struct. Mol. Biol.*, 2004, **11**, 443–449.
62. M. Vendruscolo, E. Paci, C. M. Dobson and M. Karplus, *Nature*, 2001, **409**, 641–645.
63. D. J. Brockwell and S. E. Radford, *Curr. Opin. Struct. Biol.*, 2007, **17**, 30–37.
64. K. Sugase, H. J. Dyson and P. E. Wright, *Nature*, 2007, **447**, 1021–1025.
65. M. Mucke and F. X. Schmid, *J. Mol. Biol.*, 1994, **239**, 713–725.
66. L. M. Mayr, D. Willbold, O. Landt and F. X. Schmid, *Protein Sci.*, 1994, **3**, 227–239.
67. Y. Goto and K. Hamaguchi, *J. Mol. Biol.*, 1982, **156**, 911–926.
68. M. J. Feige, F. Hagn, J. Esser, H. Kessler and J. Buchner, *J. Mol. Biol.*, 2007, **365**, 1232–1244.
69. S. R. K. Ainaravaru, J. Brujic and J. M. Fernandez, *Biophys. J.*, 2007, **92**, 225–233.
70. N. Schönbrunner, G. Pappenberger, M. Scharf, J. Engels and T. Kiefhaber, *Biochemistry*, 1997, **36**, 9057–9065.
71. J. M. Mason, M. J. Cliff, R. B. Sessions and A. R. Clarke, *J. Biol. Chem.*, 2005, **280**, 40494–40499.
72. M. E. Denton, D. M. Rothwarf and H. A. Scheraga, *Biochemistry*, 1994, **33**, 11225–11236.
73. S. H. Lin, Y. Konishi, B. T. Nall and H. A. Scheraga, *Biochemistry*, 1985, **24**, 2680–2686.
74. A. Yokota, K. Izutani, M. Takai, Y. Kubo, Y. Noda, Y. Koumoto, H. Tachibana and S. Segawa, *J. Mol. Biol.*, 2000, **295**, 1275–1288.
75. J. Clarke and A. R. Fersht, *Biochemistry*, 1993, **32**, 4322–4329.
76. V. I. Abkevich and E. I. Shakhnovich, *J. Mol. Biol.*, 2000, **300**, 975–985.
77. L. A. Mirny and E. I. Shakhnovich, *J. Mol. Biol.*, 1999, **291**, 177–196.

Hiermit erkläre ich, dass die vorliegende Arbeit selbstständig verfasst wurde und keine anderen als die hier angegebenen Quellen und Hilfsmittel verwendet wurden. Die Arbeit wurde noch keiner Prüfungskommission vorgelegt.

Matthias J. Feige
München, September 2009

Es gibt zu viele Menschen, denen ich gerne am Ende meiner Doktorarbeit meinen Dank aussprechen möchte, als dass ich sie alle hier erwähnen könnte. Mein besonderer Dank gilt jedoch Johannes, meinem Doktorvater. Für die Freiheit, das Vertrauen, die immerwährende Unterstützung bei all den gemeinsam eingeschlagenen Wegen; für bereichernde Gespräche über und fern der Biochemie, für eine Promotionszeit, wie ich sie genau ebenso wieder wählen würde und der ich durchaus mit Wehmut gedenken werde. Danke.

Mich bewogen nach München zurückzukehren hat nicht zuletzt die Atmosphäre im Labor. Frau Hilber, dem Geist und der rechten Hand des gesamten Lehrstuhls möchte ich aufrichtig danken. Für Ihre Selbstlosigkeit, Ihre Erfahrung, Ihre stete Geduld, Hilfe in allen Belangen und ehrlich Anteilnahme. Aus Kollegen wurden Freunde und Freunde wurden zu Kollegen. Ohne sie wäre die Zeit während der Promotion ganz sicher nicht diejenige gewesen, die ich nun in allerbesten Erinnerung behalten werde. Wir hatten unglaublich viel Spaß, gute Diskussionen, hervorragende Zusammenarbeiten und sind uns bewusst, dass zwar die Promotion, jedoch nicht die Freundschaft enden wird.

Emanuele in Leeds, Alan in Toronto – ihnen danke ich für persönliche, freundschaftliche Zusammenarbeiten und große Gastfreundschaft. Beides hoffe ich auch in Zukunft nicht missen zu müssen. Auch all den vielen helfenden Händen und Geistern unseres Lehrstuhls sowie der anderen Lehrstühle, mit denen wir so gut zusammengearbeitet haben und zusammenarbeiten danke ich für ebendies.

Geld bringt keine Freiheit, kann sie aber nehmen, deshalb danke ich der Studienstiftung des deutschen Volkes nicht nur für die finanzielle Freiheit während meines Studiums und der Promotion sondern auch die geistige auf all den gemeinsamen Veranstaltungen.

Meinen Eltern und meinen Brüdern kann ich schwerlich für etwas Bestimmtes danken. Ohne meine so wichtige Familie wäre ich schlicht und einfach nicht der Mensch geworden der ich heute bin. Sie prägen mein ganzes Leben, im Guten. Ihnen allen bin ich zutiefst verbunden nicht nur für all die Freiheit und Unterstützung, sondern vor allem dafür, dass sie immer für mich da waren und sind. Meiner Julia, danken kann ich ihr kaum und kaum könnte ich dankbarer sein. Erst durch sie bin ich ein Ganzes und weiß was es heißt, wie es sich anfühlt, einen Menschen, den Menschen gefunden zu haben, den man nie wieder missen möchte. Euer Matthias.