How antibodies fold

Matthias J. Feige 1, Linda M. Hendershot 2,*, and Johannes Buchner 1,*

1 Center for Integrated Protein Science at the Department Chemie, Technische Universität München, Garching, Germany
2 Department of Genetics & Tumor Cell Biology, St Jude Children Research Hospital, Memphis, USA

Abstract

B cells use unusual strategies to enable the production of a seemingly unlimited number of antibodies from a very limited amount of DNA; however this approach also dramatically increases the likelihood that proteins will be produced that are unable to fold or assemble properly. Thus these cells are particularly dependent on quality control mechanisms to oversee the production of antibodies. Recent in vitro experiments demonstrate that Ig domains have evolved diverse folding strategies ranging from robust spontaneous folding to intrinsically disordered domains that require assembly with their partner domains in order to fold, and in vivo experiments reveal that these different folding characteristics form the basis for cellular checkpoints in Ig transport. Together these reports provide a detailed understanding of how B cells monitor and ensure the functional fidelity of Ig proteins.

A short overview of antibody biology

Immunoglobulin (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 (HCs) and two identical light chains (LCs) that are linked by disulfide bonds. Both chains are composed of multiple domains of ~100 amino acids each (Figure 1A). The N-terminal domains of both chains vary between antibodies, giving rise to their designation as variable domains (V_H and V_L), and contain particularly diverse stretches of amino acids (hypervariable regions) that provide the exquisite binding specificity of the antibody molecule. Together these two domains form the antigen binding site (Figure 1A). The remainder of the antibody sequence is conserved within antibody classes (constant domains) and is important for effector functions such as complement activation or recruitment of macrophages and natural killer cells. Five different classes of antibodies are made in most

* corresponding authors. Linda M. Hendershot, Dept. Genetics & Tumor Cell biology, St. Jude Children Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105-3678, USA, linda.hendershot@stjude.org, Johannes Buchner, Dept. Chemie, Technische Universität München, Lichtenbergstrasse 4, 85747 Garching, Germany, johannes.buchner@ch.tum.de.

Publisher’s Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
higher vertebrates, IgM, IgG, IgA, IgE, and IgD, that differ in the HC constant regions used. Only two types of LC (κ or λ) exist, which can assemble with all HC classes. However, in a given cell only one HC and one LC allele are expressed, so that antibodies with a single specificity are produced.1

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, a unique HC variable region is created by combining a single variable (VH) gene segment with one diversity (DH) and one joining (JH) gene segment at the DNA level on one allele2,3, which is initially spliced to the IgM constant region at the mRNA level4. However, later during development, class switching can occur to juxtapose the rearranged VH domain to other downstream constant regions. The DNA rearrangements that give rise to HC variable regions involve imprecise joining of these 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 exons5. Once a functional HC is made, similar gene rearrangements commence to form the VL domain of the LC6. 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. Therefore, B lineage cells are particularly dependent on the endoplasmic reticulum (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 were first identified by virtue of their association with antibody chains, and that Ig molecules were some of the earliest identified substrates of ER folding enzymes (Table 1).

HCs and LCs are co-translationally translocated into the ER, and folding begins even before the polypeptide chains are completely translated7. Most IgGs assemble first as HC dimers to which LCs are added covalently via a disulfide bond between the CH1 and CH2 domains8. IgG HC mutants with a deleted CH3 domain do not form HC dimers readily and are often secreted as HC-LC ‘hemimers’9. Indeed, Fc fragment dimerization is largely mediated by interactions between the CH3 domains and stabilized by disulfide bonds in the hinge region. The CH2 domains only interact via N-linked glycans (Figure 1A), which are covalently linked to this domain co-translationally10. They determine the orientation and spacing of the two CH2 domains, which is crucial for the binding of downstream effectors11–13. The heavily glycosylated HCs of IgM require the glycans for assembly and transport suggesting that they likely guide the folding of IgM μ heavy chains14, whereas the monoglycosylated γ HCs of IgG mature properly in their absence15. The analysis of the basic steps of antibody biosynthesis in cells, together with in vitro folding studies, now can provide a molecular understanding of Ig folding and assembly processes.

**Antibody structure and the evolution of the immunoglobulin fold**

In the case of IgG, for which most of the in vitro work has been done, the “Y”-shaped molecule is composed of two four-domain HCs and two two-domain LCs (Figure 1A).
However, the orientation of the two arms of the Y is flexible due to an unstructured hinge region between the first (C\textsubscript{H}1) and the second (C\textsubscript{H}2) constant domain of the heavy chain. The IgG molecule can be proteolytically cleaved in the hinge region, which subdivides it into three functional segments; each of which is a dimer\textsuperscript{16,17}. The two N-terminal fragments (composed of a LC associated with the V\textsubscript{H}-C\textsubscript{H}1 domains of the HC) are termed Fab fragments (for fragment antigen binding). The remaining Fc fragment (for fragment crystallizable) comprises two identical, two domain (C\textsubscript{H}2-C\textsubscript{H}3) segments that are covalently linked via disulfide bonds in the hinge region. The Fc fragment is important for connecting antigen binding to antibody effector functions. 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 instead composed of two sheets forming a sandwich-like structure. The variable domains (Figure 1B) comprise nine strands (abcc’c”defg) and the constant domains (Figure 1C) seven strands (abcdefg)\textsuperscript{18}. In most antibody domains, a buried disulfide bridge, which spans ~60–70 residues, connects strands b and f\textsuperscript{11,18}. It is orientated roughly perpendicular to the individual sheets and significantly stabilizes the folded domain\textsuperscript{19}. Another characteristic feature shared among antibody domains is a conserved tryptophan residue that is located in proximity to the internal disulfide bridge. As its fluorescence is 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\textsuperscript{19}. Additionally, proline residues are unusually abundant in antibody domains, contributing up to 10% of the amino acids. Particularly important is a conserved cis-proline residue in the loop connecting strands b and c of the constant domains (Figure 1B).

Although the Ig fold was first identified in antibodies, it is in fact one of the most widely used protein topologies in nature, giving rise to 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 as early as sponges\textsuperscript{20,21}; by contrast, the ability to produce antibodies is a more recent development (~500 million years), first appearing in cartilaginous fish, such as sharks, skates and rays\textsuperscript{21,22}. In vertebrates, the Ig fold is the major building block of extracellular recognition systems\textsuperscript{23,24}, whereas in invertebrates, expression of IgSF members is mostly limited to the neural system\textsuperscript{25}. The Ig fold has also been detected in prokaryotic and viral proteins, albeit less frequently, suggesting that it might have been acquired in these cases by horizontal gene transfer\textsuperscript{26}. The evolutionary success of the IgSF can likely be attributed to its robust fold, which provides stability against proteases and harsh environments, and the ability to build highly diverse binding loops or edge strands on this core structure.

**From the folding of antibody domains to complete IgG molecules**

Dissection of the Ig protein into individual domains or fragments was necessary to detect differences in the folding of structurally similar domains. The pioneering studies on antibody folding were performed on secreted LCs that were denatured and allowed to refold in vitro\textsuperscript{19}. Further studies examined isolated constant LC domains (C\textsubscript{L})\textsuperscript{27,28} and IgG HC C\textsubscript{H}3 domains\textsuperscript{29}, which revealed that these individual Ig domains can fold autonomously. A common characteristic in the folding of antibody domains are slow proline-isomerization...
reactions (Box 1)\(^{27,28,30–32}\), that often provide the rate limiting step\(^{33,34}\). Recent \textit{in vitro} studies have addressed the folding of each of the IgG domains in more detail\(^{30–39}\). These studies found that although all Ig domains are very similar in terms of their final structure, they can be grouped into three different folding categories (Figure 2).

In the first category, exemplified by the well-studied C\(_L\) protein, domains are able to fold autonomously into a monomeric state. The chemically denatured protein exhibits no significant residual structure, regardless of whether or not its internal disulfide bridge is present\(^{34,36}\). However, once folding is initiated, the presence of an internal disulfide bridge exerts an important guiding impact on the folding pathway\(^{27,28,30,36}\). This is because the initial folding nucleus for Ig domains involves the clustering of hydrophobic residues in strands b, c, e and f, which establish the overall topology of the protein (Figure 2)\(^{35,40,41}\). The covalent linkage of the cysteines in strands b and f facilitates the establishment of this folding nucleus and the formation a structured, on-pathway intermediate thus preventing unproductive interactions (Figure 2). Indeed, a population of misfolded, aggregation-prone off-pathway folding intermediates for C\(_L\) was detected in the absence of the internal disulfide bridge\(^{36}\).

Although folding intermediates are usually transient and therefore elusive, they can be populated for longer times if a slow reaction limits the subsequent folding step thus allowing their characterization. This is the case for C\(_L\), where the major on-pathway folding intermediate is relatively long-lived due to the non-native \textit{trans} isomerization state of a proline residue between strands b and c, which must isomerize to its \textit{cis} state before folding can proceed (Figure 2)\(^{27,34}\). 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. In the intermediate, the core \(\beta\)-strands are almost completely formed, whereas the flanking strands, particularly strand d, remain highly flexible (Figure 2)\(^{34}\), in keeping with data on other IgSF members\(^{42–44}\). The analysis of the structural changes of individual amino acids in the course of folding unexpectedly revealed that two small helices linking strands a and b and strands e and f (Figure 1C, Figure 2) are important guiding elements in Ig domain folding. They become natively structured very early\(^{34}\) and can act as an organizing center, stabilizing the orientation and spacing of the \(\beta\)-strands of the Ig fold. Furthermore they correctly position bulky hydrophobic residues in the core of the protein\(^{34}\). Thus, these helices render the folding of the C\(_L\) domain more robust, as they stabilize the conformation of a highly-structured on-pathway intermediate that is poised for subsequent productive folding.

Consistent with their important role in folding, these small helices are highly conserved in most constant region domains as well as in other members of the Ig superfamily\(^{18,23,24}\). The structural insights gained for the major C\(_L\) folding intermediate seem to be readily transferable to other HC constant domains, in particular C\(_H\)\(^{231}\), where the effect of the sugar moieties on the folding reaction is currently unknown. Interestingly, no such helices are found in variable antibody domains (Figure 1B) or in several IgSF members that are prone to misfolding and amyloid formation\(^{43–45}\), suggesting that structural differences in folding intermediates might help determine whether or not IgSF members reliably fold (Box 2).
The second category of antibody domain folding pathways (Figure 2) is represented by the C_H3 domain of the IgG HC^{29,33}. In addition to folding slower than the domains of the first category, this domain forms an obligate homodimer with the internal disulfide bridge being dispensable for folding and self-association^{39}. As with the C_L domain, a partially folded species was observed, which was trapped by a non-native prolyl isomerization state^{33}. Interestingly, the monomeric intermediate could not dimerize until the native proline isomer was formed. Thus, proline switches can regulate not only folding but also dimerization in antibodies.

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 1A). Surprisingly, the isolated C_H1 domain is intrinsically disordered as determined by various spectroscopic techniques^{37}. This is in marked contrast to all antibody domains previously studied and completely unexpected from its structural similarity to other Ig domains^{11}. To induce its folding, the C_H1 domain strictly requires interaction with key residues in the dimerization interface of the folded C_L domain. Unlike antibody domains where the intramolecular disulfide bond only enhances the folding process, the covalent linkage of the two cysteines in the C_H1 domain is a prerequisite for its folding. Another unique twist is that the rate-limiting proline isomerization between strand b and c and subsequent productive folding can only occur after association with the C_L domain. These observations are in agreement with previous studies on the Fab fragment where C_H1 folding was proposed to be the slowest step, occurring after association of the HC and the LC^{32}.

Together, the data reveal that the β-barrel topology of antibody domains is reached by an overall conserved mechanism, although strikingly different pathways are employed. The biological significance of these differences becomes clear when one considers that different domains of antibodies are used for significantly different purposes in the cell such as guiding dimerization, like C_H3, or as a quality control sensor for assembly, like C_H1.

**Quality control of antibody folding and assembly in vivo**

As the secretion of incompletely folded or assembled antibodies would be deleterious to the immune response, a number of quality control checkpoints are required during B cell development and differentiation that monitor the integrity of the antibody (Figure 3). One of the first quality control measures after HC variable gene rearrangements centers on the ability of the HC protein to associate with the “surrogate LC”, which is assembled from the V_{preB} protein (contributing the “variable domain”)^{46} and the λ_5 protein (supplying the “constant domain”)^{47}. This LC-mimetic tests the ability of HC to properly fold and assemble with a LC-like protein. If this occurs correctly, the preB cell receptor is transported to the cell surface along with signaling proteins, and provides the stimulus for the further development of the pre-B cell^{48}. The C_H1 domain constitutes a crucial aspect of this quality control step, as HCs that lack the C_H1 domain can be transported to the cell surface and signal without assembling with a surrogate LC^{49}, and remains a critical focus of Ig quality control efforts throughout B cell development and plasma cell differentiation. Once the HC is judged functional by successfully completing these steps, conventional LC gene
rearrangements commence. Unlike LC, which can be secreted alone, HCs are retained in the ER and eventually degraded unless they assemble with LC (Figure 3). LC loss variants of plasmacytomas are very rarely observed, whereas HC loss variants occur much more frequently. This was argued to be due to the “toxicity” of free HC, which could be neutralized by LC. 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 (Box 2). 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 VH and CH1 domains. Similarly, mouse plasmacytoma lines expressing HC with deletions of the CH1 domain can secrete free HC, whereas deletion of any of the other constant region domains does not permit this. Finally, the serum of Camelidae contains a significant fraction of antibodies that are naturally devoid of LC. These “HC-only” antibodies do not possess a CH1 domain further underscoring the evolutionarily conserved importance of this domain in regulating Ig transport and quality control.

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 correctly 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. BiP is the ER orthologue of the Hsp70 family of chaperones and is retained in the ER, along with any associated proteins, by virtue of its C-terminal KDEL tetrapeptide. Similar to the differences detected by *in vitro* folding experiments, the folding requirements and dependence on BiP are quite different for the various Ig domains in cells. BiP binds transiently to some Ig domains (i.e., VL, VH, and some CH domains), but other domains, such as CL, appear to fold rapidly without ever interacting with BiP, even though they possess potential BiP binding sites. Only the CH1 domain interacts stably with BiP in the absence of LC (Figure 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. In *vivo* studies determine the folding status of proteins possessing intrachain disulfide bonds rely on the analysis of the oxidation status of cysteine residues, as disulfide bonds make proteins more compact and faster migrating on non-reducing SDS PAGE gels. Unlike other Ig domains, the CH1 domain remains reduced in the absence of LC, suggesting that it is not completely folded. However, in contrast to the *in vitro* studies described above, there is no evidence that BiP can associate oxidized CH1 domains in cells, it might well be that *in vivo* the association with LC, oxidation, and folding of CH1 are more tightly coupled.

In keeping with *in vitro* studies, the ATP-mediated release of BiP from isolated HC resulted in CH1 domain oxidation, but this is not sufficient for proper folding. This finding argues that LC association is also required for the folding of the CH1 domain *in vivo*. Only
LC in which both domains (V_L and C_L) were folded could assemble with HC and induced C_H1 domain oxidation and secretion from cells. In vivo experiments also confirmed the requirement for proline isomerization in the C_H1 domain in these processes, as a mutation of the critical proline inhibited oxidation, assembly with LC and secretion\(^\text{37}\). Thus, the evolution of a unique C_H1 domain that absolutely requires assembly with a C_L domain for its folding, allows the cell to ensure that newly rearranged HCs in preB and cells will be retained unless they are able to pass an important test: their ability to combine with a surrogate or conventional LC respectively. In addition, it also ensures that plasma cells, which have been estimated to synthesize up to 10^3 antibody molecules a second\(^\text{68}\), are not releasing partially assembled subunits that cannot properly bind to the selected antigen or perform effector functions.

It appears that a comparable folding-based retention mechanism also operates on some LCs. Many LCs fold readily and can be secreted by themselves as either monomers\(^\text{69}\) or dimers\(^\text{70}\), suggesting that the V_L domains of these LCs are likely to belong to folding categories 1 and 2 described above (Fig. 2). However, LCs exist that are not secreted without HC\(^\text{71}\) due to a failure of the V_L domain to fold properly by itself\(^\text{72}\), suggesting that these V_L domains might belong to category 3. The requirement for assembly-assisted folding of some V_L, and presumably V_H domains, would limit which V_L and V_H pairings could pass ER quality control. It is likely that only self-folding V regions on one chain could complement and fold the assembly-dependent V regions on the other chain, in much the same way that a folded C_L domain is required to induce folding of the C_H1 domain. This possibility could explain the fact that only certain possible HC and LC pairings are observed in cell lines and immune responses\(^\text{73}\).

**Concluding remarks and future perspectives**

In summary, the combination of biophysical and in vivo studies on individual Ig domains, antibody fragments and complete antibodies have identified common themes that together now provide a detailed picture of IgG folding (Figure 4). Once the internal disulfide bridge is formed, most domains will autonomously fold in at least a three step reaction. The first observable step is the formation of an on-pathway folding intermediate whose lifetime is increased by incorrect peptidyl-prolyl isomerization states. Subsequent peptidyl-prolyl isomerization reactions control folding to the native state, assembly and formation of interchain disulfide bridges and might also play a role in inhibiting aggregation. In the case of most IgG subclasses, once the C_H3 domain folds it induces dimerization of the HC, which is further stabilized by the formation of disulfide bonds in the hinge region (Figure 4). At this stage, all the constant region domains, except C_H1, are folded. The folding of the LCs will generally occur independently and in parallel. Association of a folded LC with HC will induce C_H1 domain folding, and once C_H1 is completely folded the assembly of HCs and LCs will be stabilized by an interchain disulfide bond (Figure 4). In the cell, the individual steps are attended by the ER chaperone machinery, which associates co-translationally with precursor HC and LC and allows high concentrations of unfolded domains to exist without aggregating and might be further supported by the co-translational folding of the individual domains\(^\text{74}\). Future work must therefore focus on further integrating the role of the complex

*Trends Biochem Sci. Author manuscript; available in PMC 2016 January 18.*
ER folding network in modulating, synchronizing and controlling the folding and assembly of antibodies and other IgSF members.

Acknowledgments

We thank Moritz Marcinowski and Dr. Roger Müller for helpful comments on the manuscript and Julia Behnke for help with preparing Box 1. Funding of MJF by the Studienstiftung des deutschen Volkes, of JB by the DFG SFB 749 and the Fonds der chemischen Industrie and of LMH by by NIH Grant GM54068, the Cancer Center CORE Grant CA21765, and the American Lebanese Syrian Associated Charities of St. Jude Children's Research Hospital is gratefully acknowledged.

Glossary

**Ig domain**
a folding unit of approximately 100 amino acids with a highly conserved twisted barrel-like β-sheet structure that is in most cases stabilized by a buried intrachain disulfide bond.

**Hypervariable regions**
highly diverse portions of the heavy and light chain variable regions that form the antigen binding site. They arise in part due to the genetic mechanisms used to produce variable regions.

**Heavy/light chain**
constituent polypeptide chains of antibody molecules. Light chains are made up of two Ig domains and therefore possess a lower molecular weight than the heavy chains which are made up of a minimum of three Ig domains.

**Isotypes**
refers to the antibody class, IgM, IgG, IgA, IgD, and IgE, which are named for the heavy chain constant region used; μ, γ, α, δ, and ε respectively. A single heavy chain variable region can be sequentially associated with different constant regions via a process known as class switching.

**Ig superfamily**
refers to a large group of proteins that are composed of Ig domains. It is one of the most widespread protein topologies observed in nature and is often involved in extracellular binding and recognition processes.

**Pre-B cell**
an early B cell developmental stage that is characterized by the production of heavy chain proteins but not light chains.

**B cell**
refers to a mature developmental stage in which both heavy and light chains are synthesized and expressed at the cell surface via a transmembrane region at the C-terminus of the heavy chain that is produced by alternative splicing of the heavy chain mRNA.

**Plasma cell**
the terminal stage of B cell differentiation that occurs after mitogen or antigen stimulation of B cells. These normally short-lived cells produce tremendous quantities of a single type of antibody.

References


Trends Biochem Sci. Author manuscript; available in PMC 2016 January 18.

98. Hayashi T, Su TP. Sigma-1 receptor chaperones at the ER-mitochondrion interface regulate Ca(2+) signaling and cell survival. Cell. 2007; 131:596–610. [PubMed: 17981125]


Within proteins, the individual amino acids are covalently linked by the planar peptide bond, the product of the reaction between the carboxyl group of the amino acid number i and the amine group of the amino acid number i+1. For 19 out of the 20 natural amino acids, this bond populates almost exclusively a *trans* state, i.e. the angle between the CO group of amino acid i and the NH group of amino acid i+1 is $\omega=180^\circ$ and the C$\alpha$ atoms of amino acid i and i+1 are on opposite sides of the CO-NH bond (Figure I). The only exception is proline, where, due to its cyclic side chain, the *cis* state ($\omega=0^\circ$) of the peptide bond is energetically only slightly less favourable than the *trans* state (Figure I). In mature protein structures, ~10% of all bonds between proline and its preceding amino acid (Xaa-Pro bond) are found in the *cis* state. The conversion to the *cis* state is influenced by the side chain of the amino acid preceding the proline residue. The *cis* conformation is particularly pronounced in bends and turns of proteins suggesting a structural role. As the activation energy of the peptidyl-prolyl isomerization reaction is rather high, ~80 kJ/mol, it is an intrinsically slow reaction taking several minutes at room temperature. Furthermore, all *cis* Xaa-Pro bonds leave the ribosome in a *trans* state after polypeptide synthesis. Therefore, peptidyl-prolyl isomerization reactions are important rate limiting steps in protein folding and were among the first to be identified. Due to their unusual structural properties, prolines can be used as molecular timers and switches in protein conformational changes. In the cell, peptidyl-prolyl isomerization reactions are catalyzed by the diverse and ubiquitous family of peptidyl-prolyl isomerases (PPIases).

Box 1 / Figure I – Schematic representation of an alanyl-prolyl isomerization reaction.
Box 2

Alternatively folded states and antibody deposition diseases

For most proteins, the accessible conformations in equilibrium include the native state, the unfolded state and unspecific aggregates; some polypeptides might additionally adopt oligomeric fibrillar structures. For antibodies, the situation is different. Antibodies can adopt a specific additional conformation at low pH (below pH 3) where many other proteins would be largely unfolded. The low pH antibody conformation had been termed “alternatively folded state” as it exhibits characteristics of the folded state, such as remarkable stability against unfolding, but the available spectroscopic information suggests that it is structurally significantly different from the native state. It was first described for a complete IgG antibody \(^83\), but single domains, such as C\(_{H3}\), are also able to adopt this state\(^84\). The biological significance of this process remains enigmatic. However, there are biotechnological implications for this state, as antibody manufacturing processes often include low pH steps which can induce the alternatively folded state.

Another accessible state for some antibodies is the fibrillar amyloid structure which is associated with a number of protein folding diseases. In this cross-beta structure, fibrils are formed by \(\beta\)-strand exchange of the individual subunits. Whereas antibodies certainly evolved to robustly form and maintain their \(\beta\)-sheet structure in the human body, both secreted isolated LCs and also truncated HC s have been found to form fibrils which are deposited in organs, such as kidneys. As antibodies are produced and can thus be deposited in large quantities, these deposits can strongly interfere with physiological functions and thus these diseases can have fatal consequences. The most prevalent of the fatal diseases is light chain amyloidosis (AL), resulting from the over-production of monoclonal LCs which are prone to misfolding and the formation of amyloid deposits. In the case of AL, certain \(V_L\) domains seem to be particularly susceptible to amyloid formation\(^85\), consistent with the idea that the CL domain is protected against misfolding by helical elements which are missing in \(V_L\)\(^34,38,45\) and the fact that all CL domains are the same, whereas all VL domains are different. Although the precise fibrilization mechanism remains incompletely understood, it seems that the mechanisms that give rise to the production of variable domains might, at times, generate less stable domains that are able to pass ER quality control but that have a propensity outside the cell to misfold.
Figure 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 oligosaccharides between the C_H2 domains are depicted as grey hexagons. 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 (V_L) (B) and constant (C_L) domains (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.
Figure 2. three pathways of antibody domain folding
(i) C_L and C_H2 fold 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). (ii) The obligate dimer C_H3 folds via two intermediates, both most likely similar in structure to those of the C_L and C_H2 domains. In a first, rapidly formed intermediate, a critical proline residue (highlighted in yellow) must isomerize to its native cis state, leading to a second intermediate which can dimerize and thereby complete folding. (iii) C_H1 is intrinsically disordered in isolation. Upon association with C_L, it forms a loosely folded intermediate. In this complex, 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.
Figure 3. Immunoglobulin quality control checkpoints at various stages in B cell development

After HC gene rearrangements, preB cells produce IgM HCs (μ HCs) (blue) bound to BiP (red). If their association with the surrogate LC, which is assembled from the V_{preB} (deep purple) and λ_{5} (light purple) proteins, induces BiP release and folding of the C_{H1} domain, and if the other Ig domains fold properly, the HC can traffic to the plasma membrane and engage signalling molecules (HC membrane anchor shown in yellow). If there is a failure in any of these steps, the μ HCs become substrates for ER associated degradation (ERAD) and are retrotranslocated to the cytosol for degradation by the 26S proteasome. Once conventional LCs (green) are produced in the B cell, they assemble with μ HCs, displace BiP from the C_{H1} domain, and induce its folding. As the ability of all domains of the HC to fold properly upon assembly was tested at the preB cell stage, quality control at this stage monitors the pairing and folding of the V domains. Plasma cell differentiation leads to the synthesis of extremely high levels of antibodies. Because the ability of the specific HC and LC combination to assemble and fold properly was verified at the B cell stage of development, 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. There is a shift to production of 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. Thiol-mediated retention mechanisms monitor the redox state of this cysteine and prevent IgM monomers from being secreted.\(^6\)
Figure 4. A comprehensive view of IgG folding and assembly

Folding, formation of disulfide bridges and glycosylation of the HC (blue) and LC (green) begins cotranslationally in the ER. The molecular chaperone BiP (red) interacts with most of the domains transiently before folding is completed. All constant domains except C_H1 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 the important C_H1 prolines are in the correct isomerization state and C_H1 is folded, a disulfide bridge between the LC and the HC forms rendering the IgG molecules ready for secretion. Most of these steps are likely to hold for other Ig classes. Chaperones and folding catalysts, such as Grp94, protein disulfide isomerase (PDI) and the peptidyl-prolyl isomerase CyclophilinB contribute to the individual steps in immunoglobulin biogenesis.
Members of the ER protein folding machinery.

<table>
<thead>
<tr>
<th>Folding helper</th>
<th>Additional names</th>
<th>Molecular mass (kDa)</th>
<th>Function</th>
<th>Involved in Ig biosynthesis</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BiP</td>
<td>Grp78, Kar2 (yeast)</td>
<td>78</td>
<td>Molecular chaperone, Hsp70 homologue</td>
<td>+</td>
<td>58, 59</td>
</tr>
<tr>
<td>GRP94</td>
<td>Gp96, Hsp90b1</td>
<td>94</td>
<td>Molecular chaperone, Hsp90 homologue</td>
<td>+</td>
<td>87</td>
</tr>
<tr>
<td>Calnexin/Calreticulin</td>
<td></td>
<td>65/47</td>
<td>Quality control, ER retention</td>
<td>+</td>
<td>88</td>
</tr>
<tr>
<td>ERdj1</td>
<td>Mrj1</td>
<td>64</td>
<td>Co-chaperone of BiP, J-protein, Hsp40 protein membrane-bound</td>
<td>?</td>
<td>89</td>
</tr>
<tr>
<td>ERdj2</td>
<td>Sec63</td>
<td>83</td>
<td>Co-chaperone of BiP, J-protein, Hsp40 protein membrane-bound</td>
<td>?</td>
<td>90</td>
</tr>
<tr>
<td>ERdj3</td>
<td>DnaJB11</td>
<td>43</td>
<td>Co-chaperone of BiP, J-protein, Hsp40 protein</td>
<td>+</td>
<td>91</td>
</tr>
<tr>
<td>ERdj4</td>
<td>DnaJB9 Mrj1</td>
<td>25</td>
<td>Co-chaperone of BiP, J-protein, Hsp40 protein membrane-bound</td>
<td>+</td>
<td>92</td>
</tr>
<tr>
<td>ERdj5</td>
<td>DnaJC10</td>
<td>91</td>
<td>Co-chaperone of BiP, J-protein, reductase</td>
<td>?</td>
<td>93</td>
</tr>
<tr>
<td>ERdj7</td>
<td>Gng10, DnaJC25</td>
<td>40</td>
<td>Co-chaperone of BiP, J-protein, Hsp40 protein</td>
<td>?</td>
<td>95</td>
</tr>
<tr>
<td>GRP170</td>
<td>Lhs1 (yeast)</td>
<td>110</td>
<td>Nucleotide exchange factor for BiP</td>
<td>+</td>
<td>96</td>
</tr>
<tr>
<td>BAP</td>
<td>S11</td>
<td>50</td>
<td>Nucleotide exchange factor for BiP</td>
<td>?</td>
<td>97</td>
</tr>
<tr>
<td>Sig1R</td>
<td></td>
<td></td>
<td>Cofactor of BiP</td>
<td>?</td>
<td>98</td>
</tr>
<tr>
<td>PDI</td>
<td></td>
<td>55</td>
<td>Oxido-reductase, disulfide bond formation, isomerisation</td>
<td>+</td>
<td>99</td>
</tr>
<tr>
<td>Perp1</td>
<td></td>
<td>20</td>
<td>Oxido-reductase, chaperone?</td>
<td>+</td>
<td>100, 101</td>
</tr>
<tr>
<td>ERP72</td>
<td>CaBP2</td>
<td>72</td>
<td>Oxido-reductase</td>
<td>+</td>
<td>102</td>
</tr>
<tr>
<td>PS</td>
<td>CaBP1</td>
<td>55</td>
<td>Oxido-reductase</td>
<td>+</td>
<td>102</td>
</tr>
<tr>
<td>ERP57</td>
<td></td>
<td>60</td>
<td>Oxido-reductase</td>
<td>+</td>
<td>103</td>
</tr>
<tr>
<td>CypB</td>
<td></td>
<td>20</td>
<td>Peptidyl prolyl isomerase</td>
<td>+</td>
<td>91</td>
</tr>
<tr>
<td>ERP29</td>
<td></td>
<td>25</td>
<td>Chaperone?</td>
<td>?</td>
<td>104</td>
</tr>
<tr>
<td>UDP-GT</td>
<td></td>
<td>53</td>
<td>Glucosyl-Transferase</td>
<td>?</td>
<td>105</td>
</tr>
</tbody>
</table>

*The references refer either to the first description of the protein, or to the first demonstration of a role in antibody biosynthesis.*