Section I

Principles and Analysis of Disulfide Bond Formation

CHAPTER 1.1

Disulfide Bonds in Protein Folding and Stability

MATTHIAS J. FEIGE*[a], INEKE BRAAKMAN*[b] AND LINDA M. HENDERSHOT*[c]

[a]Department of Chemistry and Institute for Advanced Study, Technische Universität München, 85748 Garching, Germany; [b]Cellular Protein Chemistry, Bijvoet Center for Biomolecular Research, Science for Life, Faculty of Science, Utrecht University, Utrecht, The Netherlands; [c]Department of Tumor Cell Biology, St Jude Children's Research Hospital, Memphis, TN 38105, USA
*E-mail: matthias.feige@tum.de, i.braakman@uu.nl, linda.hendershot@stjude.org

1.1.1 Stabilization of Proteins by Disulfide Bonds

Early protein folding studies were interpreted as suggesting that the native state of a protein corresponds to one well-defined conformation, whereas the unfolded state corresponds to a random coil.[1] If no other states than either the native or unfolded state are kinetically or thermodynamically stable, we speak of a two-state folding mechanism for a protein.[2] This scenario set the stage for early analyses of the role that disulfide bonds play in protein stability.

An ideal random coil is devoid of any long-range interactions except excluded volume effects. It behaves as a freely joined chain with segments of defined length.[3] In such a system, the impact of a covalent crosslink

Chemical Biology No. 9
Oxidative Folding of Proteins: Basic Principles, Cellular Regulation and Engineering
Edited by Matthias J. Feige

Published by the Royal Society of Chemistry, www.rsc.org

between two defined residues of the polypeptide chain, such as a disulfide bond, would be greatest on the unfolded state, significantly decreasing the conformational freedom of the random coil. Restricting the conformational space of the unfolded state reduces its entropy. Hence, in the presence of a disulfide bond, the entropy change for the reaction to the ordered native state is less negative, with net stabilization of the folded protein as a result. This model will be called the *chain-entropy model* in the following sections. A quantitative description was developed by Flory,[4] Schellman[5] and Poland and Scheraga.[6] 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). The mathematical description of the problem, based on polymer theory, can be found in the equation

$$\Delta S = -R\ln\left[\frac{3}{\left(2\pi l^2 N\right)^{\frac{3}{2}}}\right]v \qquad (1.1.1)$$

where R is the gas constant and 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. A major point of discussion has been the suitable choice of v. A value of 57.9 Å^3 based on the closest possible approach of two thiols is mostly in use.[7] Hence eqn (1.1.1) can be simplified to

$$\Delta S = -2.1 - \frac{3}{2}R\ln(n) \qquad (1.1.2)$$

where n is the number of amino acids bridged by the disulfide bond. Based on a study of ribonuclease (RNase) T_1 with no, one and two intact disulfide bonds, eqn (1.1.2) was developed by Pace *et al.*[7] They not only found a good correlation between n and $\Delta\Delta G$ upon removal of disulfide bonds in RNase T_1, but also observed agreement between the predictions from these equations and the experimental data for lysozyme, RNase A and the antibody C_L domain.

The above equations have two main consequences. Conceptually, the stabilization of a protein is thought to be an entirely entropy-driven process with an impact exclusively on the unfolded state. In theory, the stabilization achieved by a disulfide bond should therefore always increase with increase in the number of amino acids between the two cysteines. Despite its appealing simplicity, in practice this theory falls short in important aspects of real proteins. It treats the unfolded polypeptide chain as a system devoid of any intra- or intermolecular interactions. Furthermore, the water surrounding a protein is an important factor in shaping the free energy landscape of the polypeptide chain, and also in the native state, but is neglected in the equations.[8–10] The impact of these considerations regarding disulfide bonds was addressed by Doig and Williams in 1991.[11] They argued that disulfide bonds may significantly decrease the solvent-accessible surface in the unfolded state of a protein. As a consequence, hydrophobic residues and also hydrogen-bond donors and acceptors may become buried. Burial of hydrophobic

residues would lead to less ordering of water and thus a higher entropy of the solvent surrounding disulfide-containing proteins. Consequently, the hydrophobic effect, a major driving force in protein folding, should be less pronounced. On the other hand, hydrogen bonding with the solvent will be less extensive for a more compact unfolded state, thus reducing this competition. As a result, folding to the native state will be enthalpically more favorable. According to the authors, this enthalpic contribution must be considered as the major stabilizing factor of disulfide bonds. This model will therefore be called the *solvent-enthalpy model* in the following sections. As in the *chain-entropy model*, effects on the native state are also neglected in this model. Both models are summarized in Figure 1.1.1.

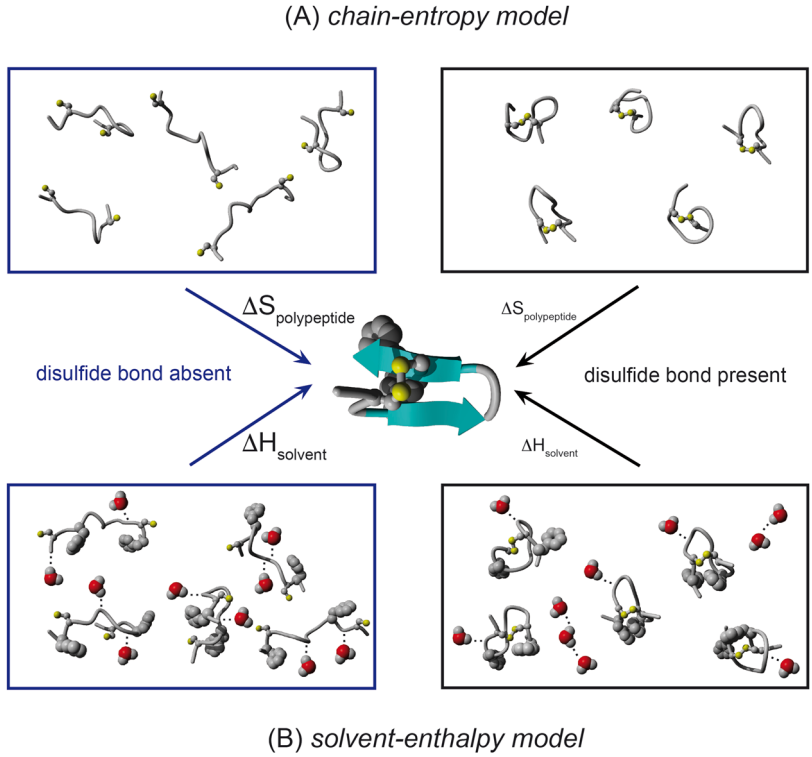

Figure 1.1.1 Models for the role of disulfide bonds on polypeptide stability. (A) The *chain-entropy model* predicts a smaller change in entropy (ΔS) upon folding of a polypeptide chain containing a disulfide bond than of one lacking a disulfide bond. This leads to a net stabilization of the native state. (B) The *solvent-enthalpy model* predicts fewer solvent–polypeptide interactions (water molecules are displayed in a CPK representation and hydrogen bonds as dashed lines) and less exposure of hydrophobic residues for a polypeptide chain containing a disulfide bond than for a polypeptide chain lacking one. This is assumed to reduce the enthalpy change (ΔH) upon loss of solvent–polypeptide interactions during folding and thus will lead to net stabilization of the native state.

A variety of experimental evidence argues either for or against these two theories, rendering the problem much more complex, but the data obtained from the various experiments also provide a chance for further insights. The advent of site-specific mutagenesis offered the possibility of introducing artificial disulfide bonds at defined positions within a protein and allowed investigators to explore the effects of disulfide bonds on protein folding more rigorously.[12–15] Many of the results obtained in 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 proteins by the introduction of a new disulfide bond that was not present in the original protein. In contrast to expectations, destabilization of the protein was the result of a significant number of these attempts.[16–18] High-resolution structures of the engineered proteins helped to explain some of these unexpected effects. It turned out that strain was imposed on the native state by the disulfide crosslinks in several cases.[19] Two particularly insightful examples are studies on barnase[16,20] and staphylococcal nuclease.[17] Clarke *et al.* introduced three artificial disulfide bonds into barnase: one destabilized the native state, one stabilized it to an extent predicted by the *chain-entropy model* and one had a much weaker effect on stabilization than predicted.[16] Furthermore, the bond connecting fewer residues was more stabilizing than the one encompassing more residues. An NMR H–D exchange analysis of the barnase mutants revealed altered dynamics of the native state, imposed by the presence of the covalent crosslinks.[20] For staphylococcal nuclease, no stabilization was found for any of the disulfide-introduced constructs, but in this case the *cis–trans* equilibrium of a peptidyl–prolyl bond was shifted in the native state and the catalytic activity of all the mutants was reduced.[17] Accordingly, strain on the native state, as reflected by the alteration of the *cis–trans* equilibrium, was evoked to explain the unexpected observations.

One major conclusion can be drawn from these two studies and a variety of others:[12–14,19,21–25] the native state of a protein is often affected by the presence of a disulfide bond, in particular if it was artificially introduced and not naturally evolved. Thus, a major assumption of the *chain-entropy model* and the *solvent-enthalpy model* frequently does not hold. Covalent crosslinks often impose strain on the native state. This is then reflected in structural changes in some cases,[17,19] or in more subtle ones, such as variations in the dynamics of the native state. As a consequence, both the enthalpy and the entropy of the native state are likely altered when two residues in the polypeptide are crosslinked covalently. The alterations in dynamics and structure are not always global, but sometimes local and context specific. For example, 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. Changes in global dynamics, *e.g.* the vibrational normal modes of a protein, will also influence the entropy of the native state, as was shown in a molecular dynamics study by Tidor and Karplus.[26] Consequently,

contrary to simple models, the introduction of disulfide bonds into a protein may well lead to net destabilization due to a loss of native-state entropy. In addition, changes in the solvent-exposed hydrophobic surface, for example in disulfide mutants of interleukin 4, have a significant effect on the stability of the native state.[25] In summary, both experimental and theoretical insights clearly argue against the simple *chain-entropy model*, at least pertaining to one side of the equation, the native state. The same holds true for the *solvent-enthalpy model*, which is also at odds with thermodynamic parameters derived for some disulfide-bonded proteins, revealing that disulfide bonds do not necessarily stabilize the native state enthalpically.

But what about the unfolded state? Is it correct to assume that a random coil is either completely devoid of any interactions or is dominated only by hydrogen bonding to the solvent? Clearly this is not so in all cases. Residual structure has been detected in a variety of proteins.[27–31] In particular, residual hydrophobic interactions or fluctuating α-helical elements seem to be more of a general feature of proteins than an exceptional one. This is often the case under mildly denaturing conditions,[30–34] and therefore can be expected to be even more pronounced under physiological conditions in the absence of denaturants. Importantly, the structural features of the unfolded state, such as residual structure in the unfolded state of barnase, have in some cases been shown to be influenced by the presence or absence of disulfide bonds.[35] Hence disulfide bonds very likely not only influence the conformational freedom of the unfolded chain but can also introduce structure, which, for example, may be protective against irreversible aggregation.[36,37] Even apparently minor structural changes, such as the clustering of some hydrophobic residues, will influence the enthalpy and entropy of the unfolded state. These aspects – together with strain induced on the native state and an impact of disulfide bonds on the dynamics of the native structure, as outlined above – must all be considered to reflect the real complexity of proteins.

Is a comprehensive theory possible that quantitatively describes the effect of disulfide bonds on a protein's stability? Most likely it will remain an approximation, but many of the factors that need to be taken into account have been identified. The decreased entropy of an unfolded and crosslinked polypeptide chain needs to be considered as developed in the *chain-entropy model*. Effects on the hydrophobic effect and more prominently hydrogen bonding in the unfolded state as reflected in the *solvent-enthalpy model* should be included. Interactions induced in the unfolded state due to the disulfide bond, which may not be localized directly around the bond but can be present as long-range residual structure,[35] must also be included in a comprehensive model. This will clearly have an effect on the enthalpy and entropy balance for the reaction to the native state. The same holds true for decreased dynamics of the native state, locally or globally, and also enthalpically unfavorable strain or enthalpically favorable induced proximity of interacting residues. In summary, the effect of a disulfide bond on the stability of a protein may have been readily assessed experimentally, yet its molecular

explanation can be almost as diverse as the protein under investigation. The key factors giving rise to the net effect are most likely known in many cases, but the individual contributions of these factors to the overall effect are blurred in their sum, as is their influence on each other. Exact stability data, combined with structural data on the native and the unfolded states of the protein under investigation, are required for a detailed understanding of the effect of a disulfide bond. They should be complemented by theoretical approaches, such as molecular dynamics simulations of the native and unfolded states, to obtain a more complete picture. Although this will be unique in detail for different proteins, in summary these analyses should reveal general principles about the effects of disulfide bonds connected to their specific location within a protein.

1.1.2 Disulfide Bonds in Protein Folding Reactions: Biophysical Considerations

One of the most fundamental questions in biophysical chemistry is how a linear polypeptide chain specifically adopts its intricate three-dimensional structure within a reasonable amount of time. *In vitro*, a variety of mutational approaches,[38–46] high-resolution structural techniques[47–50] and ultrafast perturbation and analysis methods[51–55] have provided deep insight into this phenomenon of biological self-organization. It is best understood for two-state folders, proteins that populate only the completely folded or the completely unfolded state. Major concepts, as discussed below, have been developed by studies on these two-state folders.

While folding, proteins explore a multi-dimensional energy landscape (hypersurface), where loss of chain entropy is compensated by a gain in enthalpy-driven backbone or side-chain interactions, most prominently hydrogen bonds, ionic interactions and hydrophobic interactions.[56,57] Folding is not a random exploration of all possible conformers, because interactions formed early will restrict the conformational space for downstream exploration of the energy landscape. For naturally occurring proteins, this gives rise to a funnel-like energy landscape.[58,59] This funneled energy landscape, which allows protein folding to occur on biologically relevant time scales, is a product of evolution and not an intrinsic characteristic of a heteropolymer.[60] In many cases, readily foldable polypeptide sequences have been selected by evolution for minimal frustration, thus lessening competition between individual interactions and instead providing cooperativity in their formation, which reduces the ruggedness of the energy hypersurface.[61,62] Often, formation of the native topology, which can depend on a very small number of residues coming into contact, is the rate-limiting step and defines the overall nature of the transition state that separates the unfolded and native states.[63] Accordingly, the average sequence separation of amino acid residues that ultimately need to interact in the native state of a protein is a critical factor affecting the folding rate for two-state proteins.[64,65]

In agreement with the theory of protein-folding funnels, two-state folders can take multiple pathways to the native state – yet only the native state and the unfolded state will be kinetically and/or thermodynamically stable.[59,66]

Structure in the unfolded state is regarded as an important element in protein folding. As outlined before, interactions in the unfolded state will not only influence the net stability balance of a protein, but are also likely to have an impact on the pathways taken through the energy landscape on the way to the native state. Preformed interactions in the unfolded state may have opposing effects on a folding reaction. For instance, if these interactions are present in the native state and do not adversely influence the ability of the remaining polypeptide chain to explore the necessary conformational space, they may lead to faster and more efficient folding to the native state. Conversely, if they are non-native or too stable, they may thwart the folding process and acquisition of the native state. This is where disulfide bonds prominently come into play and examples for both cases have been reported. In the case of RNase T_1, the preservation of a disulfide bond during denaturation decelerated the refolding kinetics, which was attributed to an influence on peptidyl–prolyl isomerization, over-stabilization of partially folded states and decreased chain flexibility.[67] In this context, it is important to note that RNase T_1 possesses two disulfide bonds, one connecting a small N-terminal β-turn and the other connecting the C-terminus to this N-terminal β-turn (Figure 1.1.2). Accordingly, the protein will be an almost completely looped structure in its unfolded state if the long-range disulfide bond is preserved. This will clearly affect the overall dynamics of the polypeptide chain and may have an impact on the formation of certain topologies.

Conversely, in the case of the constant domain (C_L) of the antibody light chain (Figure 1.1.2), formation of its single disulfide bond accelerated folding up to ~100-fold.[68,69] In this case, the disulfide bond is found in the hydrophobic core of the protein and is part of the immunoglobulin folding nucleus.[70,71] For other proteins that have been investigated, the presence of their natural intrinsic disulfide bonds can either enhance or reduce their respective folding reactions.[18,72–77] Hence disulfide bonds are in general far from being inert in kinetic terms. Very often, different disulfide bonds within the same protein have distinct effects on the folding rates. A comprehensive study in this respect has been carried out for the all-β-sheet protein CD2, where 13, artificially introduced, disulfide bonds showed markedly different effects on the folding behavior,[18] and also the unfolding kinetics of the altered proteins.[18,78] The experimental findings are in agreement with simple lattice-based simulations, where disulfide bonds inside the folding nucleus were found to accelerate the folding reaction, whereas those outside decelerated folding.[79]

Despite their heterogeneity, the effects of disulfide bonds on protein folding/unfolding kinetics can be more easily rationalized than their effects on native-state stability. The key lies in the transition states for folding and unfolding. As outlined above, establishment of the native state topology is often rate limiting for protein folding. Hence acceleration of folding is

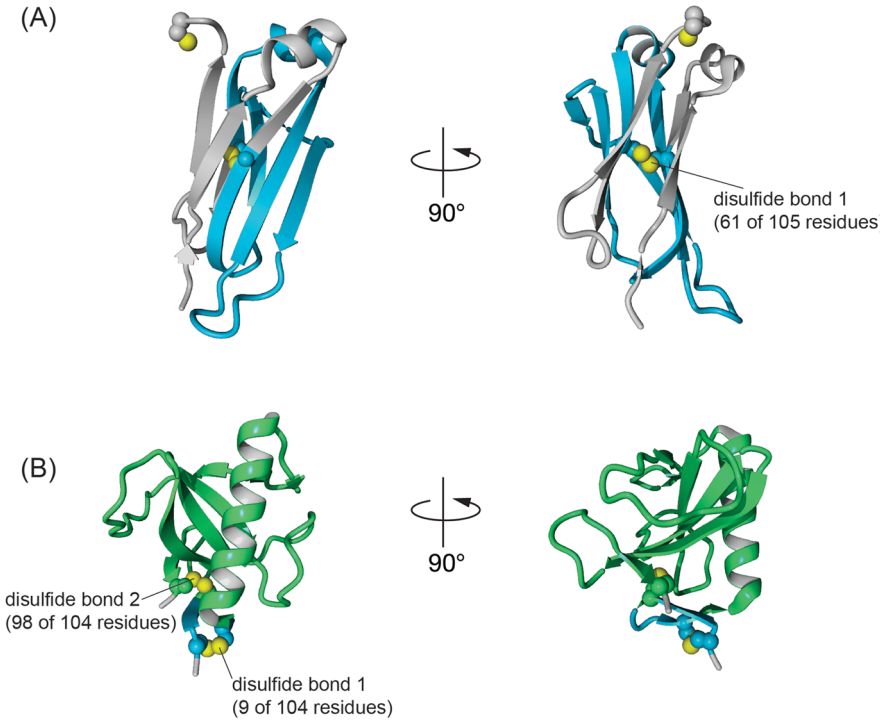

Figure 1.1.2 Location and range of disulfide bonds in proteins. (A) In the human antibody κ C_L domain (PDB code: 2R8S), its single internal disulfide bond is located in the hydrophobic core and connects ~60% of the residues (marked in blue). The cysteine that will covalently link the Ig light chain to the Ig heavy chain in order to form an antibody molecule is shown unpaired at the top of this model. (B) In the *Aspergillus oryzae* RNase T_1 (PDB code: 3RNT), disulfide bond 1 connects ~10% of its residues (marked in blue), whereas disulfide bond 2 links ~90% of its residues (marked in green). Cysteines are shown in a CPK representation with the sulfur atoms highlighted in yellow.

expected when residues that need to come into contact early in a protein-folding reaction are crosslinked, if not outweighed by entropy/enthalpy compensation in the transition state. Analogously, deceleration of unfolding is expected when two residues, whose interactions need to be broken in the transition state for unfolding, are covalently linked. Both effects can also provide structural information about the otherwise hardly accessible transition states. Examples are the immunoglobulin domain-containing proteins C_L and CD2, where the transition states for folding are stabilized by disulfide bonds,[18,68] and barnase, where the transition state for unfolding is destabilized by a disulfide bond.[78] Often, disulfide-bonded proteins are found to fold less cooperatively.[76] This may be caused by the population of disulfide-stabilized, partially folded intermediates. If stabilization of partially folded structures becomes too strong, by either native or non-native interactions,

this can even result in a net deceleration of protein folding, as was reported for CD2, where attempts to increase the folding rate by multiple disulfide bonds had the opposite effect, namely deceleration of the folding reaction by over-stabilization of a partially folded state.[75] This highlights the aforementioned role of cooperativity for efficient protein refolding, which can be beneficially or detrimentally influenced by maintaining disulfide bonds during the denaturation process. In addition, if erroneous disulfide bonds form during the refolding of a reduced substrate, they may trap and stabilize non-native folding intermediates, thus inducing covalent frustration, so to speak, which is more difficult to overcome. When multiple disulfide bonds can form in a protein, these may lead to parallel folding pathways, further complicating protein folding. These scenarios are of particular relevance for protein folding *in vivo* and will be discussed below.

Insights into the impact of disulfide bonds on protein folding are necessarily biased by the available experimental studies. As described before, these mostly rely on using unfolded proteins with disulfide bonds already formed – or studying folding under reducing conditions to prevent the formation of disulfide bonds (Figure 1.1.3). Of particular relevance for protein folding in the cell, however, is the more complex scenario: folding of proteins while all possible disulfide bonds can form, break and isomerize (Figure 1.1.3). Recent single-molecule,[80] theoretical[81] and *in vivo* studies[82] have provided insight into these processes and argue that conformational folding, which brings correct cysteines into proximity, drives disulfide bond formation. However, these studies also rely on a limited set of proteins and many *in vivo* studies argue that these are only part of the picture.

1.1.3 Distinctions Between *In vitro* Refolding Assays and Protein Biosynthesis in a Cell

The same biophysical principles outlined in the preceding sections also form the basis of oxidative protein folding *in vivo*, which in eukaryotic cells occurs primarily in the endoplasmic reticulum (ER). However, a number of major differences exist between how protein folding is studied *in vitro* and what we have learned about protein maturation in the ER. For instance, *in vitro* folding reactions usually employ small, single-domain, two-state folders that are not modified with glycans. Conversely, in the mammalian ER, most proteins are large, comprised of multiple domains, often oligomeric, and glycosylated. In cases where their folding trajectory has been elucidated, multiple intermediates have been identified. Perhaps one of the most prominent difference is that folding reactions *in vitro* generally start by transferring a denatured protein into a solution that will allow it to refold to a native conformation (Figure 1.1.3). As such, the full-length polypeptide chain will be present from the beginning of the protein folding reaction. All cysteines are available and free to react with each other. To deal with this complexity, many *in vitro* studies start with either a denatured protein with disulfide

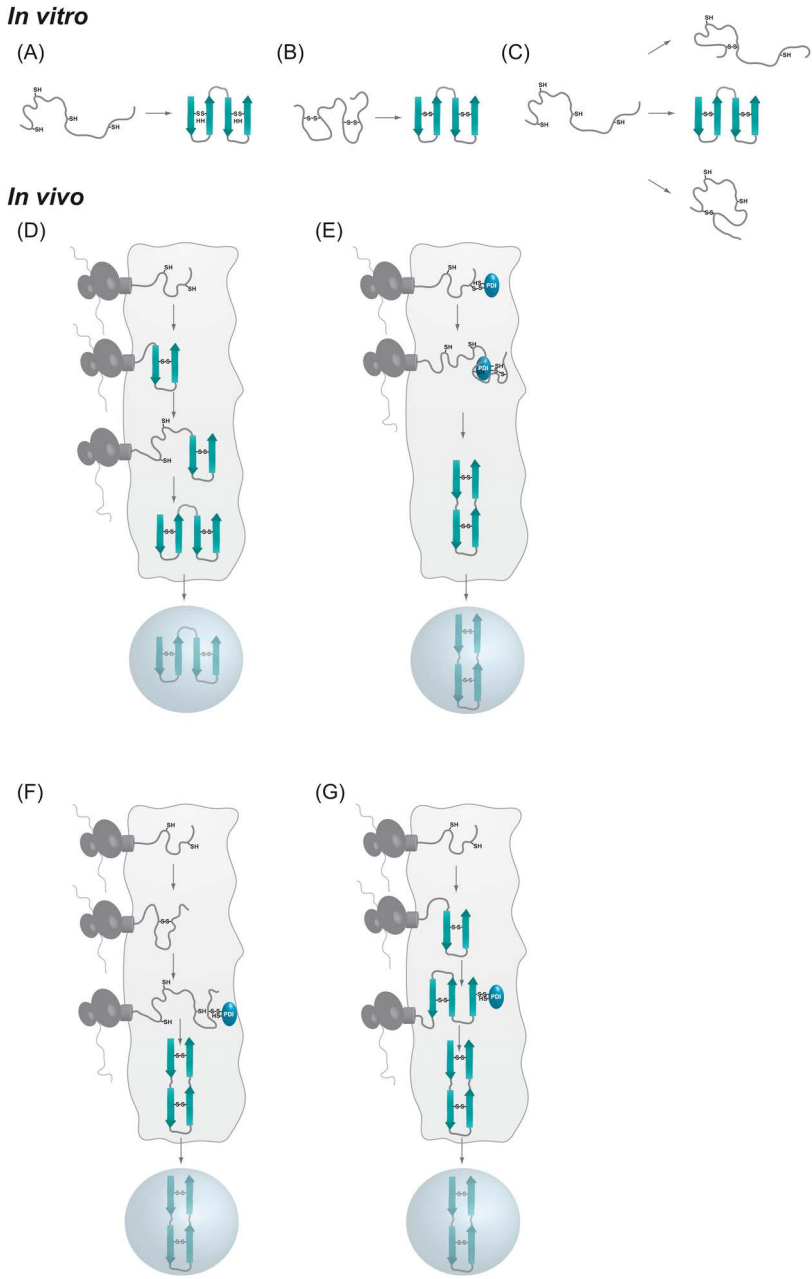

Figure 1.1.3 Folding of disulfide-bonded proteins *in vitro* and *in vivo*. *In vitro*: (A) under reducing conditions, which begin with a reduced, unfolded protein and include a reductant such as β-ME; (B) under non-reducing conditions, starting from an oxidized, unfolded protein; (C) under oxidizing conditions, starting from a reduced, unfolded protein. Possible outcomes are shown. *In vivo*: (D) formation of a disulfide

bonds intact or a denatured and completely reduced protein. In the latter case, reducing agents are included in the folding reaction. In contrast, most proteins that will ultimately form disulfide bonds *in vivo* will emerge from the Sec61 translocon co-translationally into the disulfide-promoting environment of the ER lumen and cysteines will be exposed vectorially during biosynthesis. Co-translational folding therefore may simplify disulfide bond formation when sequential cysteines are the native ones to bond. However, sequential emergence of cysteines can also complicate folding when cysteines that form native disulfide bonds are far apart in the linear sequence or are separated by interjacent cysteine residues (Figure 1.1.3). Both scenarios occur and will be discussed in the following sections.

A second major difference between *in vitro* and *in vivo* oxidative folding reactions is that the speed and conditions at which disulfide bonds form and the native structure is achieved can differ vastly. Early experiments by Anfinsen's group revealed that urea-denatured ribonuclease (RNase) spontaneously oxidizes to form four intramolecular disulfide bonds over a protracted period, but this did not represent the native form of the protein, as the oxidized RNase had no enzymatic activity.[83] In contrast, if a molar excess of β-mercaptoethanol was added to the reaction, nearly 80% of activity could be spontaneously achieved within 24 h, or considerably faster when much lower concentrations of RNase were used.[84] In both cases, however, non-physiological conditions of pH and temperature were required. This is in dramatic contrast to RNase folding in cells, where it has been estimated to occur within about 3 min of entry into the ER.[85] Although many proteins synthesized in the ER achieve a folded state sufficient to pass ER quality control within 30–60 min, it is important to note that not all proteins fold this rapidly *in vivo*.[86–88] The binding of molecular chaperones and complex folding pathways can lead to very long maturation times. For instance, a portion of the HIV-1 envelope glycoprotein gp160 is still folding after 24 h of synthesis, resulting in a significant steady-state ER localization. Folding is still very efficient, leading to high yields of the mature protein.[89] Anfinsen and co-workers found that the addition of purified microsomes from a variety of tissue sources dramatically increased the rate at which activity was restored to the denatured RNase in their *in vitro* refolding assays and allowed this to be achieved under more physiological conditions.[90] Isolation of the molecule(s) responsible for the increased folding activity led to the identification of the enzyme protein disulfide isomerase (PDI), which is the founding member of the large PDI family of enzymes found in the mammalian ER and

bond between sequential cysteines; (E) formation of a disulfide bond between non-sequential cysteines, with a PDI retaining them in a folding-competent state; (F) formation of an erroneous disulfide bond between sequential cysteines – these can be isomerized by PDI to allow formation of the correct bond between non-sequential cysteines; (G) initial formation of non-native disulfide bonds that are needed to form native structure with the support of a PDI as indicated.

is discussed in detail in Sections 3 and 4 of this book. Subsequent studies using an *in vitro* translation system coupled with dog pancreas microsomes demonstrated that microsomes depleted of PDI were dramatically deficient in supporting the formation of intramolecular disulfide bonds in the wheat storage protein γ-gliadin.[91]

1.1.4 Disulfide Bonds in ER Protein Folding

The formation of disulfide bonds in proteins is largely restricted to proteins of the secretory pathway and specifically those portions of the protein that reside within the lumen of secretory pathway organelles. In addition, disulfide bonds are critical to the biogenesis of a number of proteins present in the intermembrane space of mitochondria,[92] and their formation is dependent on Mia40, an oxidoreductase, and Erv1, a sulfhydryl oxidase (see Chapter 3.2 for details).[93–95] Although the reducing environment of the cytosol is likely to be largely responsible for the scarcity of disulfide bonds in proteins expressed there, it is noteworthy that one study examined the cysteine content of intracellular and extracellular mammalian proteins and found that cysteines are much rarer in intracellular proteins.[96] Cysteines therefore appear to be selected against even in the face of a reducing cytosolic environment owing to their high potential reactivity. The intricate connection between protein folding and disulfide bond formation that exists *in vivo* is highlighted by the fact that reducing agents such as dithiothreitol (DTT), which inhibit disulfide bond formation, are among the most potent inducers of the ER unfolded protein response (UPR), which is activated by the accumulation of unfolded proteins in the ER.[97] This argues that for at least some proteins, their folding/stability in the cell is driven by or dependent on disulfide bonds.

Whereas exquisite methods have been developed to study the refolding of proteins *in vitro* in great detail, owing to the complex and crowded environment of cells that express a large number of different proteins in various states of folding, the methods for examining folding *in vivo* are more limited and often rather indirect. These include the failure of a protein to be expressed in the correct cellular location,[98] the unusually rapid degradation of a protein[99] or prolonged association of the protein with molecular chaperones.[100,101] More direct methods for studying protein folding *in vivo* include detergent solubility experiments,[102] interaction with conformation-specific antibodies,[103,104] and sensitivity to proteolysis.[105,106] However, even if a protein fails to fold and as a result, for instance, becomes insoluble in non-ionic detergents, the portion of a protein that has failed to fold or the effects of unfolded segments on other regions of a protein are not readily revealed. By far the most direct method of examining oxidative protein folding in a cell is the determination of the redox status of cysteines in various regions of a protein.

As a protein unfolds, its hydrodynamic radius often increases, which can result in a slower mobility on sodium dodecyl sulfate (SDS)-polyacrylamide gels.[107] Correspondingly, in some cases, the formation of intramolecular

disulfide bonds can confer increased mobility to a protein on SDS-polyacrylamide gels, if the cysteine pairs are a sufficient distance from each other, owing to the stabilization of more compact conformers of the protein.[108] This realization provided the basis for two methods to study the folding of a protein during translation and translocation into the ER lumen. The first employs cDNA constructs of proteins in which ribosome-stalled intermediates are generated by introducing restriction sites at various locations along the reading frame.[109] The cDNA is linearized with the restriction enzyme and used to generate mRNA transcripts that are translated in the presence of ER microsomes to the end of the mRNA generated by the cut. The absence of an encoded stop codon stalls and stabilizes the bound ribosome at this site, allowing the N-terminal region of the protein to enter the ER lumen, while ~20 amino acids span the microsomal membrane and ~40–60 amino acids remain in the ribosome channel. The stalled, labeled construct can be examined directly on non-reducing polyacrylamide gels, or the addition of either puromycin or RNase can be used to force the release of ribosomes, allowing an additional ~60–80 amino acids to enter the ER before analysis by SDS-PAGE.[110] A second method relies on pulse-labeling conditions that are shorter than the time required to synthesize the full-length protein, which is followed by chase periods of varying length. The isolated polypeptide is then analyzed by two-dimensional SDS-PAGE (polyacrylamide gel electrophoresis) in which the first dimension is run under non-reducing conditions and the second under reducing conditions. The formation of a disulfide bond on the nascent chain is visualized by a shift in migration below the diagonal due to the induced increase in mobility.[111] The demonstration that DTT could act as a cell-permeable reductant[112] paved the way for studies by Braakman *et al.* to use this small molecule to synthesize metabolically labeled, full-length proteins in a completely reduced state. Subsequent removal of the DTT resulted in reoxidation of the ER and allowed oxidative folding of the protein of interest in the natural environment of the ER.[113,114] This method provided an *in vivo* correlate to some of the *in vitro* methods that employed full-length, reduced proteins. Together, these methods have allowed fairly detailed delineations of both rather simple and amazingly complex pathways of oxidative folding. These include the folding of proteins in which disulfide bonds occur through oxidation of sequential cysteines, ones in which long-range, non-sequential cysteines are used, and even to pathways that rely on the initial production of non-native disulfide bonds that are subsequently reduced during the folding process to allow the formation of native bonds in order to achieve the mature, functional structure (Figure 1.1.3). Examples of each type of oxidative folding are discussed below.

Some of the earliest attempts to examine the formation of disulfide bonds in nascent chains as they entered the ER lumen came from the work of Bergman and Kuehl,[115–117] who used heavy and light chains of antibody molecules for their studies. This client is particularly interesting to discuss, as it has continued to be used for *in vitro* and *in vivo* folding studies as described in more detail below. The early studies by this group utilized a mouse plasmacytoma

making very large quantities of an immunoglobulin (Ig) light chain (LC) that is composed of two ~100 amino acid Ig domains organized in tandem, each of which folds independently into a twisted β-barrel structure that is secured with a single disulfide bond as already described for the C_L domain (Figure 1.1.2). Nascent chains that remained associated with a tRNA were biochemically isolated and tryptic fragments were examined. They found that Ig LC in which the entire first domain (V_L) had entered the ER, but in which only a small portion of the second domain (C_L) had been synthesized, had already formed the V_L disulfide bond and they calculated that this occurred within ~1 s after the second cysteine of the bond entered the ER lumen.[115,116] Additional studies by this group revealed similar co-translational oxidation of the N-terminal V_H domain of an Ig γ heavy chain and even an interchain disulfide-linked assembly of Ig subunits before full translation of the proteins was completed.[117]

1.1.5 Formation of Disulfide Bonds Between Sequential Cysteines

In their simplest form, antibodies are heteromeric proteins covalently assembled from two Ig heavy chains (HC) and two Ig light chains (LC). Each chain is comprised of multiple Ig domains (four and two, respectively) that are ~100 amino acids in length and form a twisted β-barrel structure, which is stabilized by a single intradomain disulfide bond between two sequential cysteines (Figure 1.1.2). It is noteworthy that antibodies were one of the very first proteins that PDI was shown to bind in cells,[118] and the addition of recombinant PDI to purified Ig HC and LC enhanced the formation of heteromeric HC_2LC_2 antibodies.[119,120] *In vitro* studies performed with single Ig domains[69,121] and full-length LC[122] provided evidence that the isolated Ig domains are able to fold independently of each other, that the presence of a slower folding domain does not impede folding of other domains and that the maintenance of the intradomain disulfide bond during unfolding can accelerate refolding of a domain. These studies also revealed that the oxidized form of the isolated C_L domain could be readily distinguished from a reduced form on non-reducing gels. DTT washout experiments, as described above, revealed that LC synthesized in cells in the presence of this reducing agent rapidly formed the intrachain disulfide bond in the C_L domain upon reoxidation of the ER, with the V_L forming its bond more slowly.[123] This is contrary to *in vitro* studies with a Bence Jones protein, in which the V_L domain was found to fold more rapidly than the C_L domain,[122] and points to the fact that the sequences of V_L domains can vary considerably. In fact, some V_L domains are unable to fold to a native structure in isolation and instead become substrates for ER-associated degradation in cells.[124] The *in vitro* refolding of most domains of the Ig HC have been studied individually, which revealed that the intradomain disulfide bond significantly increased

the stability of the folded state of the C_H3 domain,[125] or for the less stable C_H2 domain, needed to be left intact in order to allow refolding.[126] However, the C_H1 domain, which forms the cellular basis for Ig quality control, is unique in that it remains unoxidized in cells when expressed without LC and is bound to BiP (immunoglobulin-binding protein).[127] The post-lysis release of BiP with ATP allows the C_H1 domain to oxidize, but it is not properly folded and aggregates quickly.[128] *In vitro* studies performed on an oxidized C_H1 domain confirmed that it binds to BiP and thus is still unfolded, and NMR analyses established that the C_H1 domain is unstructured in isolation.[129] The introduction of LC into cells expressing BiP:HC complexes led to dissociation of BiP and the concomitant native oxidation of the C_H1 domain.[127,128] More refined NMR analyses provided a molecular mechanism for this observation, where it was observed that interaction of the C_L domain of the LC with several residues in the C_H1 domain initiated a folding nucleus in this domain.[129] The Ig domain represents a basic structural module that is widely used in a large number of functionally diverse proteins, because it contains a core with an anti-parallel β-sheet structure in which the multiple loops emanating from this very stable fold are modified to encode other functions. Hence the rules for oxidative folding of Ig domains elucidated for antibody molecules are likely to apply to many other proteins. A study employing a combination of *in vitro* refolding assays and *in vivo* expression data examined the oxidative folding of the extracellular portions of the T-cell receptor (TCR) α and β chains, each of which possesses two Ig-like domains in tandem. Indeed, for the A6 TCR, a single domain in each of the chains folded well in isolation and formed its intradomain disulfide bond, whereas the complementary domain in the partner chain remained unstructured and reduced prior to assembly. In both cases, the well-folded domain induced folding of its unstructured partner domain and formation of the intradomain disulfide bond upon heteromeric assembly.[130]

1.1.6 Disulfide Bonds Between Non-sequential, Often Long-range, Cysteines

The second group of secretory pathway proteins that undergo oxidative folding includes those that require the formation of long-range disulfide bonds occurring between non-sequential cysteines (Figure 1.1.3). This group of proteins by necessity is likely to require the involvement of molecular chaperones/PDIs to prevent the reduced cysteines from interacting prematurely with other cysteines, leading to the formation of non-native bonds. The oxidative folding of several proteins representing this group have been analyzed in detail and three examples are highlighted in the following. The HIV-1 envelope glycoprotein gp160 possesses 10 highly conserved disulfide bonds, five of which involve non-consecutive cysteines, and the formation of these bonds begins co-translationally with final bonds forming after a

full-length translation product has been made.[89,131] Cell-based studies have shown that gp160 is a client of both BiP[132] and calnexin/calreticulin,[133] and the additional activity of PDI is required to establish proper oxidative maturation of this envelope protein.[134,135] A second member of this group is the cytokine IL-12, which is composed of an α and a β subunit assembled into a covalent heterodimer. The α chain possesses three intramolecular disulfide bonds, all of which occur between non-consecutive cysteines, one of which is particularly long distance.[136] Their establishment during folding was recently determined through cell-based studies, although the role of molecular chaperones in protecting these vulnerable cysteines was not directly assessed in this study.[137] However, prior to assembly with the β subunit, the α chain populated multiple oxidation states, one that remained partially reduced and several in which incorrect disulfide bonds formed, arguing that chaperones are likely to play a role in protecting the unassembled IL-12 α chain. In support of this possibility, an independent study found that several ER chaperones, including PDI, associate with the IL-12 α and β subunits.[138,139] Interaction with the β subunit triggered the oxidative folding of the α subunit,[137] providing a further example of assembly-mediated oxidative folding, similar to that observed for antibodies and T-cell receptors. Lastly, the oxidative folding of the influenza hemagglutinin (HA) ectodomain, which requires the formation of six intrachain disulfide bonds (four involving consecutive cysteines and two that link very distant, non-sequential cysteines), has been determined in exquisite detail. Studies using a combination of techniques, including very short (~1 min) pulse-labeling conditions,[103] *in vitro* translation/translocation assays,[140] DTT washout studies,[113] and the production of stalled translation intermediates[110] have provided a detailed understanding of disulfide bond formation during HA biosynthesis. These studies revealed that first relatively short-range disulfide bonds form co-translationally between consecutive cysteines 64/76 and 97/139. Once a full translation product has been made, a long-range disulfide bond forms first between cysteines 52/277, stabilizing the folded head domain. A second bond, connecting even more distant cysteines 14/466, forms after the mid- and stalk regions fold, snapping the native monomeric structure together in a conformation suitable for trimerization and transport to the Golgi. This folding trajectory immediately reveals several N-terminal cysteines that remain reduced and vulnerable until very late in the maturation of HA. Correspondingly, HA maturation relies on the lectin chaperones calnexin and calreticulin to protect these regions,[141,142] and also their dedicated oxidoreductase, ERp57, which is required specifically for the post-translational oxidative events.[143] Whereas the complex maturation of HA has been almost entirely elucidated through cell- or microsome-based assays, on-column oxidative refolding of the recombinant HA1 top domain to a state that can be recognized by a native conformation-specific antibody has been obtained using GSH/GSSG-containing buffers.[144]

1.1.7 Non-native Disulfide Bonds as a Prerequisite to Correct Protein Maturation

The third paradigm in oxidative folding is exhibited by proteins in which non-native disulfide bonds are a mandatory intermediate in maturation and must be broken to allow native bonds to form as folding progresses (Figure 1.1.3). As such, this group of proteins should rely on PDI family members with oxidoreductase activity. Because these intermediate bonds cannot be inferred from structural data, it is unclear how many proteins possessing intramolecular disulfides belong to this group. The detailed and complex studies required to detect these intermediates *in vivo* have been performed for only a handful of proteins. Data on the maturation pathway of the low-density lipoprotein receptor (LDLR) is the most complete. The ligand-binding domain of LDLR is comprised of seven cysteine-rich repeats (LR1–LR7) that ultimately will be stabilized by three intra-repeat disulfide bonds, all of which are formed between non-consecutive cysteines.[145,146] The ligand-binding domain is followed by three EGF-like domains, each with three disulfide bonds that are arranged differently than in the LR repeats, around a β-propeller structure. However, to achieve this native state, the full-length LDLR protein first collapses into a compact structure with many non-native disulfide bonds occurring between cysteines in various LRs and with cysteines in the EGF-like domains.[147,148] The molecular chaperone BiP binds transiently to nascent LDLR[149] and ERdj5, an ER reductase, is required to reduce the non-native bonds so that the correct mature bonds can be formed.[150] The LRs each have a number of conserved amino acid residues that coordinate a calcium ion.[146] Mutation of these acidic residues affects disulfide bonding within the LRs and has been linked to familial hypercholesterolemia.[151] Correspondingly, *in vitro* refolding of isolated LRs can be achieved in the presence of calcium under conditions that allow disulfide exchange.[152] The *Amaranthus* α-amylase inhibitor, a member of the cysteine-knot family of proteins that possess three intramolecular disulfide bonds, is a second likely member of this group. Although *in vivo* studies of the type performed on LDLR have not been conducted on this protein, *in vitro* refolding studies revealed the formation of non-native disulfides between vicinal cysteines that were ultimately resolved to form the native bonds, all of which occur between non-sequential cysteines.[153] To promote native bond formation, *in vitro* refolding was performed in a buffer containing guanidinium with 1 mM cysteine/0.05 mM cystine. Although the aforementioned IL-12 α chain also forms non-native disulfide bonds before heterodimerization, these do not seem to be a prerequisite for correct folding, as mutational studies have shown,[137] hence care must be taken in interpreting the role of non-native disulfide bonds in protein folding. Lastly, evidence for the presence of non-native disulfide intermediates in protein maturation extend all the way to bacteria. The *Escherichia coli* lipopolysaccharide transport protein D (LptD) forms a complex with LptE that is critical for transporting LPS from

the periplasm to the cell surface of Gram-negative bacteria. *In vivo* studies found that LptD forms a number of non-native disulfide bonds through the action of the oxidase DsbA and that subsequent formation of native bonds is triggered upon association with LptE.[154] The fact that examples of this class of oxidative folding have been identified from bacteria to humans and in proteins that are members of large families strongly suggests that non-native disulfide bonds may be required for the maturation of many proteins.

1.1.8 Disulfide Bonds, Protein Misfolding and Human Disease

Mutations in tryptophans and cysteines have the highest probability of causing diseases in humans among all possible amino acid mutations.[155] This can be attributed to the regulatory roles of cysteines (see, *e.g.*, Chapter 2.3 and ref. 156), but most prominently to their structural roles in proteins of the secretory pathway. The broad spectrum of human diseases associated with cysteine mutations can be seen in Table 1.1.1, which can show only a subset of mutations involving cysteine residues that have been associated with human pathologies.

Although very heterogeneous in the types of proteins affected, and also the resulting disease phenotypes, some general characteristics can be deduced from the analysis of the proteins in Table 1.1.1:

- Mutations involving cysteines generally introduce or delete a single cysteine residue, leading to an uneven number of cysteines in the protein. On the one hand, if the cysteine was part of a disulfide bond, this may destabilize the protein under investigation. On the other hand, this will lead to an unpaired cysteine residue in the mutated protein, which becomes free to interact with other cysteines either in the same protein, another copy of the same protein or other proteins in the ER. If mutations in disulfide bonds lead to partial folding, the problem may be further exacerbated as partially folded states of proteins are generally particularly prone to misfolding and mis-assembly.
- Most of the disulfide mutants listed in Table 1.1.1 are retained in the ER, and in many cases targeted for degradation, consistent with ER quality control recognizing the aberrant proteins. This quality control is exerted by molecular chaperones, which retain their clients by binding to exposed hydrophobic sites in non-native proteins,[157,158] and by PDI family members, which retain proteins in the ER by binding to partnerless cysteines (due to an odd number or misfolding) (see Sections 3 and 4 of this book).
- Mutated proteins can be roughly divided into two classes: loss of function or gain of function. For each, two major scenarios exist. Loss of function is caused by either ER retention and degradation, as outlined above, or by a failing quality control, which releases dysfunctional

Table 1.1.1 Mutations involving cysteines giving rise to human disease – a few selected references are given in each case.

Classification	Disease	Affected proteins	Mutations	Cellular effects	Ref.
Loss of cysteine residue(s)	Hyper-IgM syndrome	CD40	C83R in cysteine-rich extracellular domain	ER retention and slow degradation of $CD40^{C83R}$ UPR induction by $CD40^{C83R}$	159
	Achromatopsia 2	α subunit of cone cyclic nucleotide-gated channel	C191Y/S in TM region 1	ER retention of C191Y	160
	Medullary cystic kidney disease/familial juvenile hyperurecemic nephropathy	Uromodulin	EGF-like domains: C148W, C315R, C317Y Central part of the protein: C150S, C217R Zona pellucida domain: C347G	ER retention and trafficking defects	161
	Von Willebrand disease	Von Willebrand factor	Intrachain: C1130F, C2671Y Interchain: C2773S	ER retention and impaired secretion of C1130F and C2671Y Multimerization defects in C2773S	162
	MEGF10 myopathy	MEGF10	6th EGF-like domain: C326R 16th EGF-like domain: C774R	Changes in *N*-glycosylation of C774R but normal cellular trafficking of both mutants Weak (C326R)/strong (C774R) impairment of Tyr phosphorylation	163
	Glanzmann thrombasthenia	β subunit of $\alpha_{IIb}\beta_3$ integrin (platelet glycoprotein)	C435A	Aberrant activation of $\alpha_{IIb}\beta_3$ integrin	164
	Multiple endocrine neoplasia type 2A/B (MEN2A/B)/familial medullary thyroid carcinoma (FMTC)/hirschsprung (HSCR) disease	RET receptor tyrosine kinase	MEN2A: C364G, C380R/G/Y/S/F, C634R/Y/W; HSCR disease: C142S, C609Y	Increased ER retention (C142S, C609Y, C634R) Failure in ligand binding (C142S) Aberrant receptor dimerization and activation in MEN2A (C634R/Y/W)	165–169
	TNFR1-associated periodic fever syndrome (TRAPS)	Tumor necrosis factor receptor 1 (TNFR1)	C30R/S, C33G, C43S, C52F, C88R	Covalent oligomerization ER retention Failure in interaction with wild-type TNFR1 Reduced signaling	170
	Familial hypercholesterolemia	Low-density lipoprotein (LDL) receptor	Cysteine-rich repeats, C297F	Failures in LDL binding ER retention	171 and 172

(*continued*)

Table 1.1.1 (*continued*)

Classification	Disease	Affected proteins	Mutations	Cellular effects	Ref.
Loss or introduction of cysteine residue(s)	Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL)	Notch3	>200 mutations in extracellular EGF-like repeats of notch3; deletion or introduction of single cysteines, *e.g.* C49Y, R90C	Formation of detergent–resistant complexes Slower degradation of notch variants	173 and 174
	Familial frontotemporal dementia	Progranulin	C105R, C139R, C521Y, R432C	Covalent misfolding Increased ER retention and degradation Aberrant proteolytic processing by elastase	175–178
	Neonatal diabetes	Insulin	A-chain: R89C, G90C, C96Y, Y108C B-chain: C43G, F48C	ER retention and ERAD (C96Y) Dominant-negative effect on wild-type insulin transport (C96Y) UPR induction	179–181
	Lysosomal storage diseases	Various, *e.g.* aspartylglucosaminidase (AGA), protective protein/cathepsin A (PPCA)	AGA: C163S PPCA: Y395C	Missing disulfide bond (AGA) Reduced enzymatic activity (AGA) Absence of protein (PPCA)	182 and 183

Introduction of cysteine residue(s)	Ehlers–Danlos syndrome (EDS)	Proα_1 type I collagen	R134C, R396C, R915C	Formation of destabilized, disulfide-bonded collagen dimers Delayed collagen processing by *N*-proteinase (R134C, R396C)	184
	Spondyloarthropathy/ stickler dysplasia/spondyloepiphyseal dysplasia	α_1 Type II collagen	Spondyloarthropathy: R75C, R519C, R1076C Stickler dysplasia: R365C, R704C Spondyloepiphyseal dysplasia: R789C	Altered collagen II/IX affinity (R519C)	185 and 186
	Systemic α_1-antitrypsin deficiency subtypes	α_1-Antitrypsin (AAT)	F35C, R39C, R223C	Aberrant covalent oligomerization ER retention, reduced secretion	187 and 188
	Autism	Neuroligin 3	R451C	Misfolding of extracellular domain Increased ER retention and degradation UPR induction Less ligand binding	189–192
	Leukemia	Granulocyte colony-stimulating factor receptor (CSF3R)	W341C	Aberrant receptor dimerization Aberrant receptor activation	193

mutants to their native location (as, for example, patient mutants of classes 3 and higher in cystic fibrosis and familial hypercholesterolemia). Gain of function is caused either by a conformer with aberrant signaling properties (as, for example, growth factor receptors in cancer), or by (disulfide) mutants that are severely misfolded and not easily degraded because of aggregation, leading to UPR induction and possibly cell death, or amyloid formation, leading to (neurodegenerative) disease.

It will be important to study which of these general features are relevant for each particular human disease that involves cysteine mutations. This will provide a basic idea of in which direction therapies may be developed.

1.1.9 Concluding Thoughts

A variety of temporal and structural high-resolution *in vitro* techniques have provided an exquisite understanding of the principles that govern protein folding. By necessity, however, most of the proteins that have been examined in detail are purified, small, full-length and often single-domain proteins, although several more complex folding scenarios have been elucidated, as discussed in this chapter. The folding of a protein in the cell, and particularly in the ER, occurs under vastly different conditions (*i.e.* time scales, concentrations and protein composition) and begins even before the full-length protein has been translated. Nonetheless, it is important to reiterate that the principles identified *in vitro* are obeyed by proteins that fold in a cell, although the minute details in their folding trajectories have remained fairly obscure owing to a lack of similar high-resolution *in vivo* methods. A number of relatively large proteins translated to full length in the presence of DTT have been shown to fold to their native structure when DTT is removed from the cell, despite the fact that these proteins normally fold co-translationally. This argues that the information encoded in the polypeptide chains remains a critical element of protein folding in the cell, similar to and further validating conclusions derived from *in vitro* studies. Finally, the question remains of whether disulfide bond formation contributes to folding or whether it merely reflects the acquisition of a folded state. We discussed cases where the retention of disulfide bonds during denaturation greatly enhanced *in vitro* refolding, and also cases where the presence of these bonds hampered the ability to reach a native state. Similarly, there are examples of proteins synthesized in the ER in which disulfide bond formation indicates that a natively folded state has been achieved, but there are also proteins in which disulfides can occur in a domain that does not fold correctly or even proteins with non-native bonds as an intermediate in the folding pathway. Hence examples exist for each scenario in the cell: Folding may drive disulfide formation – but also *vice versa*. And even disulfide bonding coupled to temporary misfolding may drive the formation of native structure. A combination of complementary *in vitro* and *in vivo* assays and the development of higher resolution methods

to study the maturation of proteins in a cell will be required to understand oxidative folding pathways fully and how these are compromised in many disease states.

Acknowledgements

The early sections of this chapter are an update from the book *Oxidative Folding of Peptides and Proteins*, published by the Royal Society of Chemistry in 2009. M. J. F. is a Rudolf Mössbauer Tenure Track Professor and as such gratefully acknowledges funding through the Marie Curie COFUND program and the Technical University of Munich Institute for Advanced Study, funded by the German Excellence Initiative and the European Union Seventh Framework Program under Grant Agreement 291763. I. B. gratefully acknowledges funding by the Netherlands Science Organization (NWO), Chemical Sciences, and the NWO Graduate Program. L. M. H. is grateful for support from National Institutes of Health Grant R01 GM54068 and by the American Lebanese Syrian Associated Charities of St Jude Children's Research Hospital.

References

1. C. Tanford, K. Kawahara and S. Lapanje, *J. Biol. Chem.*, 1966, **241**, 1921–1923.
2. S. E. Jackson, *Folding Des.*, 1998, **3**, R81–R91.
3. L. J. Smith, K. M. Fiebig, H. Schwalbe and C. M. Dobson, *Folding Des.*, 1996, **1**, R95–R106.
4. P. J. Flory, *J. Am. Chem. Soc.*, 1956, **78**, 5222–5235.
5. J. A. Schellman, *C. R. Trav. Lab. Carlsberg, Ser. Chim.*, 1955, **29**, 230–259.
6. D. C. Poland and H. A. Scheraga, *Biopolymers*, 1965, **3**, 379–399.
7. C. N. Pace, G. R. Grimsley, J. A. Thomson and B. J. Barnett, *J. Biol. Chem.*, 1988, **263**, 11820–11825.
8. R. Lumry and S. Rajender, *Biopolymers*, 1970, **9**, 1125–1227.
9. B. P. Schoenborn, A. Garcia and R. Knott, *Prog. Biophys. Mol. Biol.*, 1995, **64**, 105–119.
10. M. C. Bellissent-Funel, A. Hassanali, M. Havenith, R. Henchman, P. Pohl, F. Sterpone, D. van der Spoel, Y. Xu and A. E. Garcia, *Chem. Rev.*, 2016, **116**, 7673–7697.
11. A. J. Doig and D. H. Williams, *J. Mol. Biol.*, 1991, **217**, 389–398.
12. C. M. Johnson, M. Oliveberg, J. Clarke and A. R. Fersht, *J. Mol. Biol.*, 1997, **268**, 198–208.
13. R. T. Sauer, K. Hehir, R. S. Stearman, M. A. Weiss, A. Jeitler-Nilsson, E. G. Suchanek and C. O. Pabo, *Biochemistry*, 1986, **25**, 5992–5998.
14. J. E. Villafranca, E. E. Howell, S. J. Oatley, N. H. Xuong and J. Kraut, *Biochemistry*, 1987, **26**, 2182–2189.
15. T. Vogl, R. Brengelmann, H. J. Hinz, M. Scharf, M. Lotzbeyer and J. W. Engels, *J. Mol. Biol.*, 1995, **254**, 481–496.

16. J. Clarke, K. Henrick and A. R. Fersht, *J. Mol. Biol.*, 1995, **253**, 493–504.
17. A. P. Hinck, D. M. Truckses and J. L. Markley, *Biochemistry*, 1996, **35**, 10328–10338.
18. J. M. Mason, N. Gibbs, R. B. Sessions and A. R. Clarke, *Biochemistry*, 2002, **41**, 12093–12099.
19. B. A. Katz and A. Kossiakoff, *J. Biol. Chem.*, 1986, **261**, 15480–15485.
20. J. Clarke, A. M. Hounslow and A. R. Fersht, *J. Mol. Biol.*, 1995, **253**, 505–513.
21. B. Asgeirsson, B. V. Adalbjornsson and G. A. Gylfason, *Biochim. Biophys. Acta*, 2007, **1774**, 679–687.
22. H. Kikuchi, Y. Goto, K. Hamaguchi, S. Ito, M. Yamamoto and Y. Nishijima, *J. Biochem.*, 1987, **102**, 651–656.
23. R. Kuroki, K. Inaka, Y. Taniyama, S. Kidokoro, M. Matsushima, M. Kikuchi and K. Yutani, *Biochemistry*, 1992, **31**, 8323–8328.
24. A. Shimizu-Ibuka, H. Matsuzawa and H. Sakai, *Biochemistry*, 2004, **43**, 15737–15745.
25. D. C. Vaz, J. R. Rodrigues, W. Sebald, C. M. Dobson and R. M. Brito, *Protein Sci.*, 2006, **15**, 33–44.
26. B. Tidor and M. Karplus, *Proteins*, 1993, **15**, 71–79.
27. V. L. Arcus, S. Vuilleumier, S. M. Freund, M. Bycroft and A. R. Fersht, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 9412–9416.
28. A. Lapidot and C. S. Irving, *J. Am. Chem. Soc.*, 1977, **99**, 5489–5490.
29. E. R. McCarney, J. E. Kohn and K. W. Plaxco, *Crit. Rev. Biochem. Mol. Biol.*, 2005, **40**, 181–189.
30. 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.
31. D. Neri, M. Billeter, G. Wider and K. Wuthrich, *Science*, 1992, **257**, 1559–1563.
32. D. Eliezer, J. Chung, H. J. Dyson and P. E. Wright, *Biochemistry*, 2000, **39**, 2894–2901.
33. C. S. Le Duff, S. B. Whittaker, S. E. Radford and G. R. Moore, *J. Mol. Biol.*, 2006, **364**, 824–835.
34. J. Yao, J. Chung, D. Eliezer, P. E. Wright and H. J. Dyson, *Biochemistry*, 2001, **40**, 3561–3571.
35. J. Clarke, A. M. Hounslow, C. J. Bond, A. R. Fersht and V. Daggett, *Protein Sci.*, 2000, **9**, 2394–2404.
36. L. J. Perry and R. Wetzel, *Science*, 1984, **226**, 555–557.
37. R. Wetzel, L. J. Perry, W. A. Baase and W. J. Becktel, *Proc. Natl. Acad. Sci. U. S. A.*, 1988, **85**, 401–405.
38. A. R. Fersht, A. Matouschek and L. Serrano, *J. Mol. Biol.*, 1992, **224**, 771–782.
39. A. Matouschek, J. T. Kellis Jr, L. Serrano, M. Bycroft and A. R. Fersht, *Nature*, 1990, **346**, 440–445.
40. A. Matouschek, L. Serrano and A. R. Fersht, *J. Mol. Biol.*, 1992, **224**, 819–835.
41. A. Matouschek, L. Serrano, E. M. Meiering, M. Bycroft and A. R. Fersht, *J. Mol. Biol.*, 1992, **224**, 837–845.

42. D. P. Raleigh and K. W. Plaxco, *Protein Pept. Lett.*, 2005, **12**, 117–122.
43. K. L. Maxwell, D. Wildes, A. Zarrine-Afsar, M. A. De Los Rios, A. G. Brown, C. T. Friel, L. Hedberg, J. C. Horng, D. Bona, E. J. Miller, A. Vallee-Belisle, E. R. Main, F. Bemporad, L. Qiu, K. Teilum, N. D. Vu, A. M. Edwards, I. Ruczinski, F. M. Poulsen, B. B. Kragelund, S. W. Michnick, F. Chiti, Y. Bai, S. J. Hagen, L. Serrano, M. Oliveberg, D. P. Raleigh, P. Wittung-Stafshede, S. E. Radford, S. E. Jackson, T. R. Sosnick, S. Marqusee, A. R. Davidson and K. W. Plaxco, *Protein Sci.*, 2005, **14**, 602–616.
44. L. Serrano, J. T. Kellis Jr, P. Cann, A. Matouschek and A. R. Fersht, *J. Mol. Biol.*, 1992, **224**, 783–804.
45. L. Serrano, A. Matouschek and A. R. Fersht, *J. Mol. Biol.*, 1992, **224**, 847–859.
46. L. Serrano, A. Matouschek and A. R. Fersht, *J. Mol. Biol.*, 1992, **224**, 805–818.
47. H. Feng, Z. Zhou and Y. Bai, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 5026–5031.
48. 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.
49. U. Mayor, N. R. Guydosh, C. M. Johnson, J. G. Grossmann, S. Sato, G. S. Jas, S. M. Freund, D. O. Alonso, V. Daggett and A. R. Fersht, *Nature*, 2003, **421**, 863–867.
50. T. L. Religa, J. S. Markson, U. Mayor, S. M. Freund and A. R. Fersht, *Nature*, 2005, **437**, 1053–1056.
51. 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.
52. J. Kubelka, T. K. Chiu, D. R. Davies, W. A. Eaton and J. Hofrichter, *J. Mol. Biol.*, 2006, **359**, 546–553.
53. H. Ma, C. Wan and A. H. Zewail, *J. Am. Chem. Soc.*, 2006, **128**, 6338–6340.
54. M. B. Prigozhin, S. H. Chao, S. Sukenik, T. V. Pogorelov and M. Gruebele, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 7966–7971.
55. T. L. Religa, C. M. Johnson, D. M. Vu, S. H. Brewer, R. B. Dyer and A. R. Fersht, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 9272–9277.
56. R. L. Baldwin, *J. Mol. Biol.*, 2007, **371**, 283–301.
57. A. R. Fersht, *Nat. Rev. Mol. Cell Biol.*, 2008, **9**, 650–654.
58. K. A. Dill and J. L. MacCallum, *Science*, 2012, **338**, 1042–1046.
59. P. G. Wolynes, *Biochimie*, 2015, **119**, 218–230.
60. A. L. Watters, P. Deka, C. Corrent, D. Callender, G. Varani, T. Sosnick and D. Baker, *Cell*, 2007, **128**, 613–624.
61. J. N. Onuchic and P. G. Wolynes, *Curr. Opin. Struct. Biol.*, 2004, **14**, 70–75.
62. F. O. Tzul, D. Vasilchuk and G. I. Makhatadze, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, E1627–E1632.
63. M. Vendruscolo, E. Paci, C. M. Dobson and M. Karplus, *Nature*, 2001, **409**, 641–645.
64. K. W. Plaxco, K. T. Simons, I. Ruczinski and D. Baker, *Biochemistry*, 2000, **39**, 11177–11183.

65. V. Grantcharova, E. J. Alm, D. Baker and A. L. Horwich, *Curr. Opin. Struct. Biol.*, 2001, **11**, 70–82.
66. K. A. Dill and H. S. Chan, *Nat. Struct. Biol.*, 1997, **4**, 10–19.
67. M. Mucke and F. X. Schmid, *J. Mol. Biol.*, 1994, **239**, 713–725.
68. M. J. Feige, F. Hagn, J. Esser, H. Kessler and J. Buchner, *J. Mol. Biol.*, 2007, **365**, 1232–1244.
69. Y. Goto and K. Hamaguchi, *J. Mol. Biol.*, 1982, **156**, 911–926.
70. E. Cota, A. Steward, S. B. Fowler and J. Clarke, *J. Mol. Biol.*, 2001, **305**, 1185–1194.
71. I. Lappalainen, M. G. Hurley and J. Clarke, *J. Mol. Biol.*, 2008, **375**, 547–559.
72. S. R. Ainavarapu, J. Brujic, H. H. Huang, A. P. Wiita, H. Lu, L. Li, K. A. Walther, M. Carrion-Vazquez, H. Li and J. M. Fernandez, *Biophys. J.*, 2007, **92**, 225–233.
73. M. E. Denton, D. M. Rothwarf and H. A. Scheraga, *Biochemistry*, 1994, **33**, 11225–11236.
74. S. H. Lin, Y. Konishi, B. T. Nall and H. A. Scheraga, *Biochemistry*, 1985, **24**, 2680–2686.
75. J. M. Mason, M. J. Cliff, R. B. Sessions and A. R. Clarke, *J. Biol. Chem.*, 2005, **280**, 40494–40499.
76. N. Schonbrunner, G. Pappenberger, M. Scharf, J. Engels and T. Kiefhaber, *Biochemistry*, 1997, **36**, 9057–9065.
77. A. Yokota, K. Izutani, M. Takai, Y. Kubo, Y. Noda, Y. Koumoto, H. Tachibana and S. Segawa, *J. Mol. Biol.*, 2000, **295**, 1275–1288.
78. J. Clarke and A. R. Fersht, *Biochemistry*, 1993, **32**, 4322–4329.
79. V. I. Abkevich and E. I. Shakhnovich, *J. Mol. Biol.*, 2000, **300**, 975–985.
80. P. Kosuri, J. Alegre-Cebollada, J. Feng, A. Kaplan, A. Ingles-Prieto, C. L. Badilla, B. R. Stockwell, J. M. Sanchez-Ruiz, A. Holmgren and J. M. Fernandez, *Cell*, 2012, **151**, 794–806.
81. M. Qin, W. Wang and D. Thirumalai, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 11241–11246.
82. P. J. Robinson, M. A. Pringle, C. A. Woolhead and N. J. Bulleid, *J. Biol. Chem.*, 2017, **292**, 6978–6986.
83. E. Haber and C. B. Anfinsen, *J. Biol. Chem.*, 1961, **236**, 422–424.
84. C. B. Anfinsen, E. Haber, M. Sela and F. H. White Jr, *Proc. Natl. Acad. Sci. U. S. A.*, 1961, **47**, 1309–1314.
85. R. Geiger, M. Gautschi, F. Thor, A. Hayer and A. Helenius, *J. Biol. Chem.*, 2011, **286**, 5813–5822.
86. H. F. Lodish, *J. Biol. Chem.*, 1988, **263**, 2107–2110.
87. S. R. Pfeffer and J. E. Rothman, *Annu. Rev. Biochem.*, 1987, **56**, 829–852.
88. I. Braakman and D. N. Hebert, *Cold Spring Harbor Perspect. Biol.*, 2013, **5**, a013201.
89. A. Land, D. Zonneveld and I. Braakman, *FASEB J.*, 2003, **17**, 1058–1067.
90. R. F. Goldberger, C. J. Epstein and C. B. Anfinsen, *J. Biol. Chem.*, 1963, **238**, 628–635.
91. N. J. Bulleid and R. B. Freedman, *Nature*, 1988, **335**, 649–651.

92. N. Mesecke, N. Terziyska, C. Kozany, F. Baumann, W. Neupert, K. Hell and J. M. Herrmann, *Cell*, 2005, **121**, 1059–1069.
93. D. P. Sideris and K. Tokatlidis, *Antioxid. Redox Signaling*, 2010, **13**, 1189–1204.
94. J. M. Herrmann and J. Riemer, *J. Biol. Chem.*, 2012, **287**, 4426–4433.
95. A. Chacinska, S. Pfannschmidt, N. Wiedemann, V. Kozjak, L. K. Sanjuan Szklarz, A. Schulze-Specking, K. N. Truscott, B. Guiard, C. Meisinger and N. Pfanner, *EMBO J.*, 2004, **23**, 3735–3746.
96. R. C. Fahey, J. S. Hunt and G. C. Windham, *J. Mol. Evol.*, 1977, **10**, 155–160.
97. A. S. Lee, *Curr. Opin. Cell Biol.*, 1992, **4**, 267–273.
98. J. A. Carlson, B. B. Rogers, R. N. Sifers, M. J. Finegold, S. M. Clift, F. J. DeMayo, D. W. Bullock and S. L. Woo, *J. Clin. Invest.*, 1989, **83**, 1183–1190.
99. A. Jansens and I. Braakman, *Methods Mol. Biol.*, 2003, **232**, 133–145.
100. D. G. Bole, L. M. Hendershot and J. F. Kearney, *J. Cell Biol.*, 1986, **102**, 1558–1566.
101. M. Dong, J. P. Bridges, K. Apsley, Y. Xu and T. E. Weaver, *Mol. Biol. Cell*, 2008, **19**, 2620–2630.
102. G. K. Tofaris, A. Razzaq, B. Ghetti, K. S. Lilley and M. G. Spillantini, *J. Biol. Chem.*, 2003, **278**, 44405–44411.
103. I. Braakman, H. Hoover-Litty, K. R. Wagner and A. Helenius, *J. Cell Biol.*, 1991, **114**, 401–411.
104. S. Tiwari, J. A. Askari, M. J. Humphries and N. J. Bulleid, *J. Cell Sci.*, 2011, **124**, 1672–1680.
105. J. P. Schlebach, M. S. Kim, N. H. Joh, J. U. Bowie and C. Park, *J. Mol. Biol.*, 2011, **406**, 545–551.
106. C. Park and S. Marqusee, *Nat. Methods*, 2005, **2**, 207–212.
107. D. P. Goldenberg and T. E. Creighton, *Anal. Biochem.*, 1984, **138**, 1–18.
108. I. Braakman, L. Lamriben, G. van Zadelhoff and D. N. Hebert, *Curr. Protoc. Protein Sci.*, 2017, **90**, 14 11 11–14 11 21.
109. U. C. Krieg, A. E. Johnson and P. Walter, *J. Cell Biol.*, 1989, **109**, 2033–2043.
110. R. Daniels, B. Kurowski, A. E. Johnson and D. N. Hebert, *Mol. Cell*, 2003, **11**, 79–90.
111. W. Chen, J. Helenius, I. Braakman and A. Helenius, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 6229–6233.
112. W. W. Cleland, *Biochemistry*, 1964, **3**, 480–482.
113. I. Braakman, J. Helenius and A. Helenius, *EMBO J.*, 1992, **11**, 1717–1722.
114. I. Braakman, J. Helenius and A. Helenius, *Nature*, 1992, **356**, 260–262.
115. L. W. Bergman and W. M. Kuehl, *J. Supramol. Struct.*, 1979, **11**, 9–24.
116. L. W. Bergman and W. M. Kuehl, *J. Biol. Chem.*, 1979, **254**, 8869–8876.
117. L. W. Bergman and W. M. Kuehl, *J. Biol. Chem.*, 1979, **254**, 5690–5694.
118. R. A. Roth and S. B. Pierce, *Biochemistry*, 1987, **26**, 4179–4182.
119. H. Lilie, S. McLaughlin, R. Freedman and J. Buchner, *J. Biol. Chem.*, 1994, **269**, 14290–14296.
120. M. Mayer, U. Kies, R. Kammermeier and J. Buchner, *J. Biol. Chem.*, 2000, **275**, 29421–29425.

121. Y. Goto and K. Hamaguchi, *J. Mol. Biol.*, 1981, **146**, 321–340.
122. Y. Goto, T. Azuma and K. Hamaguchi, *J. Biochem.*, 1979, **85**, 1427–1438.
123. R. Hellman, M. Vanhove, A. Lejeune, F. J. Stevens and L. M. Hendershot, *J. Cell Biol.*, 1999, **144**, 21–30.
124. M. R. Knittler, S. Dirks and I. G. Haas, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 1764–1768.
125. M. Bertz, J. Buchner and M. Rief, *J. Am. Chem. Soc.*, 2013, **135**, 15085–15091.
126. M. J. Feige, S. Walter and J. Buchner, *J. Mol. Biol.*, 2004, **344**, 107–118.
127. Y. K. Lee, J. W. Brewer, R. Hellman and L. M. Hendershot, *Mol. Biol. Cell*, 1999, **10**, 2209–2219.
128. M. Vanhove, Y. K. Usherwood and L. M. Hendershot, *Immunity*, 2001, **15**, 105–114.
129. M. J. Feige, S. Groscurth, M. Marcinowski, Y. Shimizu, H. Kessler, L. M. Hendershot and J. Buchner, *Mol. Cell*, 2009, **34**, 569–579.
130. M. J. Feige, J. Behnke, T. Mittag and L. M. Hendershot, *J. Biol. Chem.*, 2015, **290**, 26821–26831.
131. E. van Anken, R. W. Sanders, I. M. Liscaljet, A. Land, I. Bontjer, S. Tillemans, A. A. Nabatov, W. A. Paxton, B. Berkhout and I. Braakman, *Mol. Biol. Cell*, 2008, **19**, 4298–4309.
132. P. L. Earl, B. Moss and R. W. Doms, *J. Virol.*, 1991, **65**, 2047–2055.
133. A. Otteken and B. Moss, *J. Biol. Chem.*, 1996, **271**, 97–103.
134. A. Land and I. Braakman, *Biochimie*, 2001, **83**, 783–790.
135. M. J. Papandreou, R. Barbouche, R. Guieu, S. Rivera, J. Fantini, M. Khrestchatisky, I. M. Jones and E. Fenouillet, *J. Biol. Chem.*, 2010, **285**, 13788–13796.
136. C. Yoon, S. C. Johnston, J. Tang, M. Stahl, J. F. Tobin and W. S. Somers, *EMBO J.*, 2000, **19**, 3530–3541.
137. S. Reitberger, P. Haimerl, I. Aschenbrenner, J. Esser-von Bieren and M. J. Feige, *J. Biol. Chem.*, 2017, **292**, 8073–8081.
138. I. Alloza, E. Martens, S. Hawthorne and K. Vandenbroeck, *Anal. Biochem.*, 2004, **324**, 137–142.
139. E. Martens, I. Alloza, C. J. Scott, A. Billiau and K. Vandenbroeck, *Eur. J. Biochem.*, 2000, **267**, 6679–6683.
140. T. Marquardt, D. N. Hebert and A. Helenius, *J. Biol. Chem.*, 1993, **268**, 19618–19625.
141. D. N. Hebert, B. Foellmer and A. Helenius, *EMBO J.*, 1996, **15**, 2961–2968.
142. D. N. Hebert, J. X. Zhang, W. Chen, B. Foellmer and A. Helenius, *J. Cell Biol.*, 1997, **139**, 613–623.
143. T. Solda, N. Garbi, G. J. Hammerling and M. Molinari, *J. Biol. Chem.*, 2006, **281**, 6219–6226.
144. F. F. Chiu, N. Venkatesan, C. R. Wu, A. H. Chou, H. W. Chen, S. P. Lian, S. J. Liu, C. C. Huang, W. C. Lian, P. Chong and C. H. Leng, *Biochem. Biophys. Res. Commun.*, 2009, **383**, 27–31.
145. N. L. Daly, J. T. Djordjevic, P. A. Kroon and R. Smith, *Biochemistry*, 1995, **34**, 14474–14481.

146. D. Fass, S. Blacklow, P. S. Kim and J. M. Berger, *Nature*, 1997, **388**, 691–693.
147. J. Gent and I. Braakman, *Cell. Mol. Life Sci.*, 2004, **61**, 2461–2470.
148. A. Jansens, E. van Duijn and I. Braakman, *Science*, 2002, **298**, 2401–2403.
149. M. M. Jorgensen, O. N. Jensen, H. U. Holst, J. J. Hansen, T. J. Corydon, P. Bross, L. Bolund and N. Gregersen, *J. Biol. Chem.*, 2000, **275**, 33861–33868.
150. O. B. Oka, M. A. Pringle, I. M. Schopp, I. Braakman and N. J. Bulleid, *Mol. Cell*, 2013, **50**, 793–804.
151. M. A. Lehrman, W. J. Schneider, T. C. Sudhof, M. S. Brown, J. L. Goldstein and D. W. Russell, *Science*, 1985, **227**, 140–146.
152. S. C. Blacklow and P. S. Kim, *Nat. Struct. Biol.*, 1996, **3**, 758–762.
153. M. Cemazar, S. Zahariev, J. J. Lopez, O. Carugo, J. A. Jones, P. J. Hore and S. Pongor, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 5754–5759.
154. S. S. Chng, M. Xue, R. A. Garner, H. Kadokura, D. Boyd, J. Beckwith and D. Kahne, *Science*, 2012, **337**, 1665–1668.
155. D. Vitkup, C. Sander and G. M. Church, *Genome Biol.*, 2003, **4**, R72.
156. A. Fra, E. D. Yoboue and R. Sitia, *Front. Mol. Neurosci.*, 2017, **10**, 167.
157. J. Behnke, M. J. Feige and L. M. Hendershot, *J. Mol. Biol.*, 2015, **427**, 1589–1608.
158. A. Tannous, G. B. Pisoni, D. N. Hebert and M. Molinari, *Semin. Cell Dev. Biol.*, 2015, **41**, 79–89.
159. G. Lanzi, S. Ferrari, M. Vihinen, S. Caraffi, N. Kutukculer, L. Schiaffonati, A. Plebani, L. D. Notarangelo, A. M. Fra and S. Giliani, *Blood*, 2010, **116**, 5867–5874.
160. K. A. Patel, K. M. Bartoli, R. A. Fandino, A. N. Ngatchou, G. Woch, J. Carey and J. C. Tanaka, *Invest. Ophthalmol. Visual Sci.*, 2005, **46**, 2282–2290.
161. I. Bernascone, S. Vavassori, A. Di Pentima, S. Santambrogio, G. Lamorte, A. Amoroso, F. Scolari, G. M. Ghiggeri, G. Casari, R. Polishchuk and L. Rampoldi, *Traffic*, 2006, **7**, 1567–1579.
162. J. W. Wang, D. J. Groeneveld, G. Cosemans, R. J. Dirven, K. M. Valentijn, J. Voorberg, P. H. Reitsma and J. Eikenboom, *Haematologica*, 2012, **97**, 859–866.
163. S. Mitsuhashi, H. Mitsuhashi, M. S. Alexander, H. Sugimoto and P. B. Kang, *FEBS Lett.*, 2013, **587**, 2952–2957.
164. Q. H. Sun, C. Y. Liu, R. Wang, C. Paddock and P. J. Newman, *Blood*, 2002, **100**, 2094–2101.
165. N. Asai, T. Iwashita, M. Matsuyama and M. Takahashi, *Mol. Cell. Biol.*, 1995, **15**, 1613–1619.
166. S. Kjaer and C. F. Ibanez, *Hum. Mol. Genet.*, 2003, **12**, 2133–2144.
167. L. M. Mulligan, J. B. Kwok, C. S. Healey, M. J. Elsdon, C. Eng, E. Gardner, D. R. Love, S. E. Mole, J. K. Moore and L. Papi, *et al.*, *Nature*, 1993, **363**, 458–460.
168. M. Santoro, F. Carlomagno, A. Romano, D. P. Bottaro, N. A. Dathan, M. Grieco, A. Fusco, G. Vecchio, B. Matoskova and M. H. Kraus, *et al.*, *Science*, 1995, **267**, 381–383.

169. M. P. Cosma, M. Cardone, F. Carlomagno and V. Colantuoni, *Mol. Cell. Biol.*, 1998, **18**, 3321–3329.
170. A. A. Lobito, F. C. Kimberley, J. R. Muppidi, H. Komarow, A. J. Jackson, K. M. Hull, D. L. Kastner, G. R. Screaton and R. M. Siegel, *Blood*, 2006, **108**, 1320–1327.
171. H. H. Hobbs, M. S. Brown, J. L. Goldstein and D. W. Russell, *J. Biol. Chem.*, 1986, **261**, 13114–13120.
172. N. Lelli, R. Garuti, P. Pedrazzi, M. Ghisellini, M. L. Simone, R. Tiozzo, L. Cattin, M. Valenti, M. Rolleri and S. Bertolini, *et al.*, *Hum. Genet.*, 1994, **93**, 538–540.
173. N. Maksemous, R. A. Smith, L. M. Haupt and L. R. Griffiths, *Hum. Genomics*, 2016, **10**, 38.
174. H. Meng, X. Zhang, G. Yu, S. J. Lee, Y. E. Chen, I. Prudovsky and M. M. Wang, *PLoS One*, 2012, **7**, e44964.
175. S. Almeida, L. Zhou and F. B. Gao, *PLoS One*, 2011, **6**, e26454.
176. G. Kleinberger, A. Capell, N. Brouwers, K. Fellerer, K. Sleegers, M. Cruts, C. Van Broeckhoven and C. Haass, *Neurobiol. Aging*, 2016, **39**, 220 e217–220 e226.
177. J. Wang, P. Van Damme, C. Cruchaga, M. A. Gitcho, J. M. Vidal, M. Seijo-Martinez, L. Wang, J. Y. Wu, W. Robberecht and A. Goate, *J. Neurochem.*, 2010, **112**, 1305–1315.
178. S. S. Shankaran, A. Capell, A. T. Hruscha, K. Fellerer, M. Neumann, B. Schmid and C. Haass, *J. Biol. Chem.*, 2008, **283**, 1744–1753.
179. J. Stoy, E. L. Edghill, S. E. Flanagan, H. Ye, V. P. Paz, A. Pluzhnikov, J. E. Below, M. G. Hayes, N. J. Cox, G. M. Lipkind, R. B. Lipton, S. A. Greeley, A. M. Patch, S. Ellard, D. F. Steiner, A. T. Hattersley, L. H. Philipson, G. I. Bell and G. Neonatal Diabetes International Collaborative, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 15040–15044.
180. K. He, C. N. Cunningham, N. Manickam, M. Liu, P. Arvan and B. Tsai, *Mol. Biol. Cell*, 2015, **26**, 3413–3423.
181. M. Liu, L. Haataja, J. Wright, N. P. Wickramasinghe, Q. X. Hua, N. F. Phillips, F. Barbetti, M. A. Weiss and P. Arvan, *PLoS One*, 2010, **5**, e13333.
182. E. Ikonen, N. Enomaa, I. Ulmanen and L. Peltonen, *Genomics*, 1991, **11**, 206–211.
183. K. Itoh, M. Shimmoto, K. Utsumi, N. Mizoguchi, N. Miharu, K. Ohama and H. Sakuraba, *Biochem. Biophys. Res. Commun.*, 1998, **247**, 12–17.
184. F. Malfait, S. Symoens, J. De Backer, T. Hermanns-Le, N. Sakalihasan, C. M. Lapiere, P. Coucke and A. De Paepe, *Hum. Mutat.*, 2007, **28**, 387–395.
185. A. Fertala, A. L. Sieron, E. Adachi and S. A. Jimenez, *Biochemistry*, 2001, **40**, 14422–14428.
186. K. P. Hoornaert, C. Dewinter, I. Vereecke, F. A. Beemer, W. Courtens, A. Fryer, H. Fryssira, M. Lees, A. Mullner-Eidenbock, D. L. Rimoin, L. Siderius, A. Superti-Furga, K. Temple, P. J. Willems, A. Zankl, C. Zweier, A. De Paepe, P. Coucke and G. R. Mortier, *J. Med. Genet.*, 2006, **43**, 406–413.
187. R. Ronzoni, R. Berardelli, D. Medicina, R. Sitia, B. Gooptu and A. M. Fra, *Hum. Mol. Genet.*, 2016, **25**, 642–650.

188. H. Okayama, M. Brantly, M. Holmes and R. G. Crystal, *Am. J. Hum. Genet.*, 1991, **48**, 1154–1158.
189. D. Comoletti, A. De Jaco, L. L. Jennings, R. E. Flynn, G. Gaietta, I. Tsigelny, M. H. Ellisman and P. Taylor, *J. Neurosci.*, 2004, **24**, 4889–4893.
190. L. Ulbrich, F. L. Favaloro, L. Trobiani, V. Marchetti, V. Patel, T. Pascucci, D. Comoletti, S. J. Marciniak and A. De Jaco, *Biochem. J.*, 2016, **473**, 423–434.
191. S. Jamain, H. Quach, C. Betancur, M. Rastam, C. Colineaux, I. C. Gillberg, H. Soderstrom, B. Giros, M. Leboyer, C. Gillberg, T. Bourgeron and S. Paris Autism Research International Sibpair, *Nat. Genet.*, 2003, **34**, 27–29.
192. A. De Jaco, M. Z. Lin, N. Dubi, D. Comoletti, M. T. Miller, S. Camp, M. Ellisman, M. T. Butko, R. Y. Tsien and P. Taylor, *J. Biol. Chem.*, 2010, **285**, 28674–28682.
193. H. Zhang, S. Means, A. R. Schultz, K. Watanabe-Smith, B. C. Medeiros, D. Bottomly, B. Wilmot, S. K. McWeeney, T. Kukenshoner, O. Hantschel and J. W. Tyner, *Cancer Res.*, 2017, **77**, 4258–4267.