Folding of small disulfide-rich proteins: clarifying the puzzle

Joan L. Arolas[1], Francesc X. Aviles[1], Jui-Yoa Chang[2] and Salvador Ventura[1]

[1]Institut de Biotecnologia i Biomedicina and Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona; 08193 Bellaterra, Barcelona, Spain.

[2]Research Center for Protein Chemistry, Institute of Molecular Medicine and Department of Biochemistry and Molecular Biology, The University of Texas; Houston, Texas 77030, USA.

Corresponding authors: Ventura, S. (salvador.ventura@uab.es), Chang, J.Y. (rowen.chang@uth.tmc.edu)

1

Abstract

The process by which small proteins fold to their native conformations has been intensively studied over the last few decades. In this field, the particular chemistry of disulfide bond formation has facilitated the characterization of the oxidative folding of numerous small, disulfide-rich proteins with results that illustrate a high diversity of folding mechanisms, differing in the heterogeneity and disulfide pairing nativeness of their intermediates. In this review, we combine information on the folding of different protein models together with the recent structural determinations of major intermediates to provide new molecular clues in oxidative folding. Also, we turn to analyze the role of disulfide bonds in misfolding and protein aggregation and their implications in amyloidosis and conformational diseases.

Introduction

In the absence of the cellular folding machinery, small proteins spontaneously fold to their native states under suitable conditions, implying that most of the information needed to specify a protein 3-D structure is contained within the amino acid sequence. To understand protein folding, over the past 30 years much experimental work has been focused on the identification and characterization of partially folded intermediates that occur along the folding reaction [1, 2]. However, this process is usually very fast therefore the commonly used spectroscopic techniques (e.g., fluorescence and circular dichroism) provide limited information about the nature of intermediates that accumulate transiently. The use of small, disulfide-rich proteins, displaying folding intermediates that can be chemically trapped in a time-course manner, purified and structurally characterized, has proved to be very valuable for folding studies [3-5]. A large number of protein models comprising divergent and physiologically relevant functions, i.e., protease inhibitors, proteases, nucleases, growth factors and venom toxins, have been analyzed in detail to date and show an unexpected scenario of folding diversity.

Intramolecular disulfide bonds (S-S) are important not only for the folding of secreted proteins, but also for their stability and function. In the last several years, there have been significant advances in the understanding of disulfide bond formation and rearrangement in the cell, for instance, concerning the role of protein disulfide isomerases (PDI) and their relatives (Dsb proteins) that catalyze both the formation and isomerization of disulfide bonds in the eukaryotic endoplasmic reticulum and prokaryotic periplasm, respectively (for reviews, see [6, 7]). Because of the complex environment of the cell, however, more basic research is necessary to fully understand the concomitant relationship between protein folding and disulfide bond formation. Failure of these processes is likely to cause protein degradation by proteases or misfolding and subsequent aggregation (e.g., prion disorders) [8, 9].

In this review, the folding behavior of the most-studied, small disulfide-rich proteins in the

literature is described. The determinants that account for the underlying diversity of folding landscapes and patterns of disulfide bond formation are addressed. Also, recent developments in the field, such as the structural determination of disulfide intermediates and studies of protein aggregation that may have important implications in protein folding and biomedicine are herein discussed.

Disulfide bonds and oxidative folding

The covalent link of cysteine residues by disulfide bonds is thought to serve distinct functions. First, they significantly influence the thermodynamics of protein folding: disulfide bonds stabilize the native conformation of a protein by lowering the entropy of the unfolded state, making it less favorable, as compared with the folded form. Second, they maintain protein integrity: oxidants and proteolytic enzymes in the extracellular environment can inactivate proteins. By stabilizing their structure, disulfide bonds protect proteins from damage, thus increasing their half-life [10]. Finally, although the disulfide bonds present in mature proteins have been considered chemically inert, now in some cases they appear to be cleaved or rearranged with significant consequences for their function (see review [11]).

The native set of disulfide bonds is the end-result of an often-complicated process involving covalent reactions such as oxidation (S-S formation), reduction (S-S breaking) and isomerization/reshuffling (S-S rearrangement) [12]. The term "oxidative folding" describes the composite process by which a reduced, unfolded protein gains both its native disulfide bonds (disulfide bond generation) and its native structure (conformational folding) (for experimental details of the oxidative folding technique, see Box 1). The course of oxidative folding is affected by three structural factors that have been identified from different *in vitro* folding studies, namely, the proximity, reactivity and accessibility of thiol groups and disulfide bonds [13]. The proximity of two reactive groups, defined as their effective intramolecular

concentration, is determined by the propensity of the protein backbone chain to juxtapose two groups; in unfolded species, this proximity is largely determined by loop entropy and enthalpic interactions. The reactivity of the groups relies on the disulfide reactions that occur through a thiol-disulfide exchange in which only the thiolate group is reactive to perform the nucleophilic attack; therefore, changes in the local electrostatic environment may affect this parameter. Nevertheless, the most critical factor seems to be the accessibility of the thiol groups and disulfide bonds. Because thiol-disulfide exchange reactions occur only when a thiol and a disulfide bond come into contact, their burial prevents contact, hence blocking the reaction.

Accordingly, the formation of a stable tertiary structure is a key event in the oxidative folding of proteins because it protects native disulfide bonds from reduction and reshuffling by making them inaccessible to protein thiols and redox agents [14]. The burial of both thiol groups and disulfide bonds in a disulfide intermediate may hinder any further progress in the folding reaction. In this regard, dead-end intermediates tend to be "disulfide-insecure" in that their structural fluctuations expose their disulfide bonds in concert with their thiol groups, leading to reshuffling rather than oxidation. These intermediates are normally long-lived "metastable" species that constitute rate-limiting steps in oxidative folding. In contrast, productive intermediates (leading to the native state) tend to be "disulfide-secure", meaning that their structural fluctuations preferentially expose their thiol groups while keeping their disulfide bonds buried. This distinction helps to understand the oxidative folding of many disulfide-rich proteins included in following sections.

Oxidative folding of BPTI and hirudin: two opposed models

Extensive studies on small, disulfide-rich proteins have shown divergent mechanisms of folding that are illustrated by: (a) the extent of heterogeneity of folding intermediates, (b) the

predominance of intermediates containing native disulfide bonds, and (c) the accumulation of scrambled isomers (fully oxidized species that contain at least two non-native disulfide bonds) as intermediates. On the basis of these points, bovine pancreatic trypsin inhibitor (BPTI) and hirudin represent two notable models with very different folding characteristics. The original studies on BPTI (58 residues; containing three disulfide bonds), conducted by Creighton *et al.* and later revised by Weissman and Kim, resulted in one of the most extensively studied models of oxidative folding [15-18]. They showed that the folding of BPTI is characterized by the predominance of a limited number of 1- and 2-disulfide intermediates (five of 75 possible) that adopt native disulfide pairings and native-like substructures (Figure 1). These intermediates seem to funnel protein conformations toward the native state and prevent the accumulation of 3-disulfide scrambled isomers. The rate-limiting step of the process is then the conversion of 2-disulfide intermediates into the intermediate processor that rapidly forms the third and final native disulfide. Venom neurotoxins like α62 (structured in a "three-finger fold"; four disulfide bonds), the squash trypsin inhibitor EETI-II and the cyclotide protein kalata B1 (28 and 29 residues; three disulfide bonds) share similar folding with that of BPTI [19-21]. Likewise, the folding of the 3-disulfide insulin-like growth factor (IGF-1; 70 residues), which folds into two isomers (native and swap) with different disulfide linkages but similar thermodynamic stability, bears resemblance to that of BPTI [22]. Interestingly, this kind of folding is in line with the "framework model", which stresses the importance of local interactions in reducing conformational search and in guiding efficient protein folding through the hierarchic condensation of native-like elements.

The oxidative folding of hirudin (core domain, 49 residues; three disulfide bonds) differs from that of BPTI in the three folding features mentioned above: (a) folding intermediates are far more heterogeneous (at least 30 fractions of 1- and 2-disulfide intermediates have been identified), (b) predominant intermediates adopting native disulfides are absent, and (c)

3-disulfide scrambled isomers strongly accumulate [23, 24]. Taken together, the folding of hirudin can be dissected into two stages: an initial stage of non-specific disulfide bond formation (packing) leading to the formation of scrambled species, followed by a final stage of disulfide reshuffling (consolidation) of a heterogeneous scrambled population leading to the native structure (Figure 1). It is worth mentioning that the oxidative folding of many other small 3-disulfide proteins is similar to that of hirudin; among them, tick anticoagulant peptide (TAP; 60 residues), various venom neurotoxins and proteins stabilized by "cystine-knot" disulfide bonds such as potato carboxypeptidase inhibitor (PCI; 39 residues) and *Amaranthus* α-amylase inhibitor (AAI; 32 residues) [25-30]. The hirudin folding is consistent with the "collapse model" that depicts protein folding as an initial stage of rapid hydrophobic collapse, followed by slower annealing in which specific interactions refine the structure rather than dominate the folding code. Importantly, these studies have contradicted conventional wisdom, which considered scrambled isomers as abortive "off-pathway" folding intermediates. The presence of productive, "on-pathway" scrambled isomers seems to be frequent in the oxidative folding of small, disulfide-rich proteins.

Mixing folding mechanisms: the cases of EGF and LCI

The oxidative folding of epidermal growth factor (EGF) and leech carboxypeptidase inhibitor (LCI) serve to magnify the extent of diversity of disulfide folding landscapes because it displays both similarity and dissimilarity to the folding of BPTI and hirudin. EGF (53 residues; three disulfide bonds) folds through several 1-disulfide intermediates that rapidly form a single 2-disulfide intermediate with native disulfide bonds [31, 32]. This species acts as a major kinetic trap of the process since it can represent more than 85% of the protein. The pathway of this intermediate to the native structure proceeds through 3-disulfide scrambled isomers (Figure 2). Similar to EGF, the recently determined oxidative folding of LCI (67 residues; four

disulfide bonds) undergoes a sequential flow through 1- and 2-disulfide intermediates that rapidly accumulate as two predominant 3-disulfide intermediates with native disulfide pairings [33, 34]. These two species act as major kinetic traps of the process, which need structural rearrangements through the formation of 4-disulfide scrambled isomers to attain the native structure. Thus, EGF and LCI display an analogue folding that resembles that of BPTI by the presence of few predominant native-like intermediates. But, at the same time, it is also similar to that of hirudin due to the initial formation of heterogeneous populations of intermediates and the final rate-limiting step of conversion of scrambled isomers into the native protein.

Diversity in the oxidative folding of α-lactalbumin

Folding diversity of small, disulfide-rich proteins seems to depend crucially on the presence (or absence) of localized, stable domain structures. An outstanding example to illustrate this hypothesis is the mechanism of oxidative folding of α-lactalbumin (αLA) elucidated in the absence or presence of calcium (Figure 3) [35, 36]. αLA (122 residues; four disulfide bonds) comprises an α-helical domain and a β-sheet domain that is considerably stabilized upon binding to calcium [37]. In the absence of calcium, the oxidative folding of αLA resembles that of hirudin, proceeding through the formation of heterogeneous populations of 1-, 2- and 3-disulfide intermediates and with the final accumulation of 4-disulfide scrambled isomers. Binding of calcium changes the folding process of αLA significantly because it thermodynamically stabilizes the β-sheet domain. Consequently, the complexity of folding intermediates diminishes drastically and the process now involves the accumulation of two predominant intermediates that adopt native disulfide-bond pairings and native-like structures of the β-sheet domain. Thus, in the presence of calcium, the folding of αLA bears close resemblance to what is observed in BPTI folding.

To provide more insight into the role played by stable-structure elements in oxidative folding,

the reductive unfolding process of several model proteins has been examined (for details of this technique, see Box 1) [38, 39]. A striking correlation is observed between the mechanisms of oxidative folding and reductive unfolding. Those proteins with their native disulfide bonds reduced collectively in an "all-or-none" mechanism, without significant accumulation of partially reduced species, display both a high degree of heterogeneity of folding intermediates and the formation of scrambled isomers along their oxidative folding (e.g., hirudin, TAP, PCI, AAI and αLA). For these proteins, it is only at the final stage of folding (consolidation) where the attainment of the native disulfide bonds is guided by specific non-covalent interactions, which results in the cooperative and concerted stabilization of disulfide bonds [40]. In contrast, a sequential reduction of the native disulfide bonds is associated with the presence of predominant intermediates with native-like structures during folding (e.g., BPTI, EETI-II and calcium-bound αLA).

"Locking in" intermediates: the oxidative folding of RNase A

The oxidative folding of ribonuclease A (RNase A) has been extensively studied over the past several years by Scheraga *et al.* [13, 41-43]. In RNase A (124 residues), its four disulfide bonds are paired within very different secondary-structure elements: while 40-95 and 65-72 S-S occur in relatively flexible loop segments, 26-84 and 58-110 S-S join an α-helix to a β-sheet, stabilizing the surrounding local hydrophobic core. In the early stages of RNase A folding, 1-, 2-, 3- and 4-disulfide (scrambled) intermediates are formed by successive oxidations, until a pre-equilibrium is established among them (Figure 2). Interestingly, the distribution of disulfide intermediates within the 1S and 2S ensembles (collection of 1- and 2-disulfide species, respectively) is enthalpically biased toward native disulfide bonds, as concluded from the recent experiments performed with the compound *trans*-$[Pt(en)_2Cl_2]^{2+}$, which oxidizes thiol groups faster than the rate at which thiol-disulfide exchange reactions can take place [44].

The final and rate-limiting stage in the attainment of native RNase A is the formation of two intermediates, des[40-95] and des[65-72] (i.e., intermediates lacking the 40-95 and 65-72 S-S, respectively), with native disulfide-bond pairings and native-like structures. These species are formed mainly by the reshuffling of 3-disulfide intermediates, although a small fraction (up to 5%) may be formed by oxidation from the 2S ensemble. Upon formation of a stable tertiary structure, their three native disulfide bonds become protected from reduction and reshuffling ("locked in"), causing these "des species" to accumulate at high levels. However, their thiol groups remain solvent-accessible and, hence, these disulfide-secure intermediates oxidize relatively rapidly to the native protein. The protective structure of these two species is a critical factor in promoting the oxidative folding of RNase A [14]. The other two des species of the 3S ensemble, des[26-84] and des[58-110], are metastable dead-end intermediates that reshuffle preferentially to the 3S ensemble rather than oxidize to the native protein. Presumably, these two disulfide-insecure intermediates bury both their thiol groups and disulfide bonds in hydrophobic cores of a native-like structure, thus inhibiting oxidation as well as reduction and reshuffling. The specific reasons behind the accumulation of metastable intermediates have been determined recently from several structural characterizations of intermediates (see Box 2).

The effects of "locking in" native disulfide bonds are also illustrated by the BPTI, LCI and lysozyme models. Despite its small size, BPTI has two hydrophobic cores that may fold semi-independently. Specifically, the formation of the 5-55 S-S induces global folding, while the formation of the 30-51 S-S causes only one core to fold. In intermediates containing either 5-55 or 30-51 S-S, the other native S-S (14-38) forms quickly, producing the des[30-51] and des[5-55] species, respectively [45]. However, these des species appear to be disulfide-insecure intermediates, preferentially reshuffling rather than oxidizing. Therefore, the productive precursor of native BPTI is the disulfide-secure des[14-38] intermediate and the

rate-limiting step of the folding process is either a reshuffling to form des[14-38] or the escape from the dead-end metastable des[30-51] and des[5-55] species. In the oxidative folding of LCI, the des[19-43] and des[22-58] species, lacking the native disulfide bond that connects the α-helix to the β-sheet and stabilizes the β-sheet core, respectively, also seem to behave as metastable, disulfide-insecure intermediates with a high content of a native-like structure [46]. Lysozyme (129 residues; four disulfide bonds) is a more complex protein that comprises two folding domains, called the α- and β-domains. A heterogeneous ensemble of relatively unstructured intermediates (containing 1S and 2S) is rapidly formed from the reduced protein after initiation of folding [47-49]. In these early stages, and similarly to BPTI and RNase A folding, the dominance of native-like interactions results in some conformational ordering and bias toward the attainment of native disulfide bonds. Three structured des species are subsequently formed: des[6-127], des[64-80] and des[76-94]. The majority of molecules slowly form the native 4-disulfide structure via rearrangement from des[76-94], which appears to be a native-like, metastable disulfide-insecure intermediate (the 76-94 S-S links the two domains of lysozyme). The rest of the population (~ 30%) folds more quickly via the other two des species.

Disulfide bonds and aggregation

The presence of insoluble protein deposits in human tissues correlates with the development of many debilitating disorders including amyloidosis and several neurodegenerative diseases. In a number of proteins related to conformational diseases, improper disulfide bond formation may result in structural rearrangements of the polypeptide chain leading to increased aggregation. One of the first pieces of evidence of that came from pioneering studies by Dobson *et al.* on lysozyme, a protein that later was shown to form amyloid fibrils in individuals suffering from non-neuropathic systemic amyloidosis [50, 51]. Lysozyme aggregation takes place during the

early stages of the folding reaction and is strongly prevented by the presence of PDI, which facilitates the attainment of the native state [52]. This reflects the importance of folding catalysts in physiological folding and suggests an important role avoiding aggregation of partially folded molecules in the intracellular environment.

The reduction of disulfide bonds has been shown to critically affect amyloid formation. For Bence Jones proteins, which are the major component of the amyloid fibrils in patients with systemic AL-amyloidosis, the reduction of native disulfide bonds leads to non-native protein association and formation of amyloid-like aggregates [53]. In the case of Amylin, found in islet amyloid deposits of Type II diabetes, disulfide bonds play a central role in the assembly mechanism and kinetics of fibril formation [54]. The disulfide linkage of β2-microglobulin, whose aggregation into amyloid deposits is common in patients with long-term hemodialysis, seems to protect against deposition by reducing conformational fluctuations and maintaining the global native-like topology [55]. Accordingly, disulfide bond reduction both destabilizes the native state and enhances the conformational flexibility of the polypeptide, resulting in increased formation of oligomeric structures. In addition, there is the special case of the prion protein, which causes transmissible spongiform encephalopathies. The oxidative folding of reduced, monomeric non-infective form PrP^C has been shown to result in the formation of an oligomeric, protease-resistant species with a high content of β-sheet structure joined by intermolecular disulfide bonding [56, 57]. More importantly, it has been demonstrated that the aberrant "scrapie" isoform (PrP^{Sc}) is able to convert native PrP^C into oligomers, thus proposing a mechanism for prion self-propagation (Figure 4).

Finally, the formation of new non-native intramolecular disulfide bonds may also result in aggregation. This is the case of superoxide dismutase (SOD), a protein that has been implicated in the familial form of the neurodegenerative disease amyotrophic lateral sclerosis [58]. SOD1 has four cysteines, two linked and two free in the native structure. Formation of a new disulfide

bond between the originally free cysteines serves to trap SOD1 in aggregation-prone conformations and produce accelerated protein deposition.

Concluding remarks and future perspectives

In the last few years, the results obtained from many studies on small, disulfide-rich proteins have substantially clarified the understanding of oxidative folding mechanisms. The folding of proteins like hirudin, PCI or AAI has shown the interdependence between conformational folding and assembly of native disulfide bonds. Because of the large heterogeneity of species at the start of the folding process, disulfide bonds are necessary to restrict the search in conformational space by cross-linking the protein in its unfolded state. In the cell this is most likely achieved with the help of folding catalysts, for instance, PDI, which may reduce and reshuffle non-native disulfide bonds at the early stages of folding to avoid misfolding.

Other works have highlighted the critical role of accessibility of both disulfide bonds and thiol groups, as observed in the oxidative folding of RNase A, BPTI or LCI. For these proteins, the critical step seems to be the formation of a stable tertiary structure that sequesters the native disulfide bonds (preventing subsequent rearrangements) but leaves the thiol groups of intermediates relatively exposed (or exposable) for subsequent oxidation. The "locking in" of native disulfide bonds by conformational folding could also assist the oxidative folding of larger disulfide-rich proteins. For such proteins, the absence of either structured intermediates or a strong native bias would hinder the folding reaction significantly. However, the independent folding of (sub)domains could help in overcoming this entropic barrier by successively locking in disulfide bonds and protecting native pairings from further rearrangements. This approach could also operate *in vivo*, where oxidative folding can occur co-translationally. Future experiments in this direction for regenerating large, multi-domain proteins would be very challenging.

Also, the assistance of computational algorithms able to analyze folding free energy landscapes of disulfide-rich polypeptides would be of much interest. Unfortunately and despite the exciting new insights on the folding of disulfide-free proteins provided by theoretical approaches [59], the apparent complexity and diversity of their folding pathways have kept bioinformatics research away form disulfide-rich proteins. As reported here, it is coming clear that general rules govern the disulfide folding pathways, thus providing a unique opportunity for the future development of computational methods to predict and design the folding of this set of proteins.

References

1.   Daggett, V., and Fersht, A.R. (2003) Is there a unifying mechanism for protein folding? *Trends Biochem Sci* 28, 18-25

2.   Onuchic, J.N., and Wolynes, P.G. (2004) Theory of protein folding. *Curr Opin Struct Biol* 14, 70-75

3.   Chang, J.Y. (2004) Evidence for the underlying cause of diversity of the disulfide folding pathway. *Biochemistry* 43, 4522-4529

4.   Creighton, T.E. (1997) Protein folding coupled to disulphide bond formation. *Biol Chem* 378, 731-744

5.   Wedemeyer, W.J.*, et al.* (2000) Disulfide bonds and protein folding. *Biochemistry* 39, 4207-4216

6.   Baneyx, F., and Mujacic, M. (2004) Recombinant protein folding and misfolding in Escherichia coli. *Nat Biotechnol* 22, 1399-1408

7.   Ellgaard, L., and Ruddock, L.W. (2005) The human protein disulphide isomerase family: substrate interactions and functional properties. *EMBO Rep* 6, 28-32

8.   Goldberg, A.L. (2003) Protein degradation and protection against misfolded or damaged proteins. *Nature* 426, 895-899

9.   Ross, C.A., and Poirier, M.A. (2004) Protein aggregation and neurodegenerative disease. *Nat Med* 10 Suppl, S10-17

10.   Zavodszky, M.*, et al.* (2001) Disulfide bond effects on protein stability: designed variants of Cucurbita maxima trypsin inhibitor-V. *Protein Sci* 10, 149-160

11.   Hogg, P.J. (2003) Disulfide bonds as switches for protein function. *Trends Biochem Sci* 28, 210-214

12.   Creighton, T.E.*, et al.* (1995) Mechanisms and catalysts of disulfide bond formation in proteins. *Trends Biotechnol* 13, 18-23

13.   Narayan, M., *et al.* (2000) Oxidative folding of proteins. *Acc Chem Res* 33, 805-812

14.   Welker, E., *et al.* (2001) Structural determinants of oxidative folding in proteins. *Proc Natl Acad Sci U S A* 98, 2312-2316

15.   Creighton, T.E. (1990) Protein folding. *Biochem J* 270, 1-16

16.   Darby, N.J., *et al.* (1995) Refolding of bovine pancreatic trypsin inhibitor via non-native disulphide intermediates. *J Mol Biol* 249, 463-477

17.   Goldenberg, D.P. (1992) Native and non-native intermediates in the BPTI folding pathway. *Trends Biochem Sci* 17, 257-261

18.   Weissman, J.S., and Kim, P.S. (1991) Reexamination of the folding of BPTI: predominance of native intermediates. *Science* 253, 1386-1393

19.   Daly, N.L., *et al.* (2003) Disulfide folding pathways of cystine knot proteins. Tying the knot within the circular backbone of the cyclotides. *J Biol Chem* 278, 6314-6322

20.   Heitz, A., *et al.* (1995) Folding of the squash trypsin inhibitor EETI II. Evidence of native and non-native local structural preferences in a linear analogue. *Eur J Biochem* 233, 837-846

21.   Ruoppolo, M., *et al.* (2001) Slow folding of three-fingered toxins is associated with the accumulation of native disulfide-bonded intermediates. *Biochemistry* 40, 15257-15266

22.   Yang, Y., *et al.* (1999) Probing the folding pathways of long R(3) insulin-like growth factor-I (LR(3)IGF-I) and IGF-I via capture and identification of disulfide intermediates by cyanylation methodology and mass spectrometry. *J Biol Chem* 274, 37598-37604

23.   Chatrenet, B., and Chang, J.Y. (1993) The disulfide folding pathway of hirudin elucidated by stop/go folding experiments. *J Biol Chem* 268, 20988-20996

24.   Lu, B.Y., and Chang, J.Y. (2005) Assay of disulfide oxidase and isomerase based on the model of hirudin folding. *Anal Biochem* 339, 94-103

25.	Cemazar, M.*, et al.* (2004) Oxidative folding of Amaranthus alpha-amylase inhibitor: disulfide bond formation and conformational folding. *J Biol Chem* 279, 16697-16705

26.	Chang, J.Y. (1996) The disulfide folding pathway of tick anticoagulant peptide (TAP), a Kunitz-type inhibitor structurally homologous to BPTI. *Biochemistry* 35, 11702-11709

27.	Chang, J.Y., and Ballatore, A. (2000) Structure and heterogeneity of the one- and two-disulfide folding intermediates of tick anticoagulant peptide. *J Protein Chem* 19, 299-310

28.	Chang, J.Y.*, et al.* (2006) Fully oxidized scrambled isomers are essential and predominant folding intermediates of cardiotoxin-III. *FEBS Lett*

29.	Fuller, E.*, et al.* (2005) Oxidative folding of conotoxins sharing an identical disulfide bridging framework. *Febs J* 272, 1727-1738

30.	Venhudova, G.*, et al.* (2001) Mutations in the N- and C-terminal tails of potato carboxypeptidase inhibitor influence its oxidative refolding process at the reshuffling stage. *J Biol Chem* 276, 11683-11690

31.	Chang, J.Y.*, et al.* (2001) A major kinetic trap for the oxidative folding of human epidermal growth factor. *J Biol Chem* 276, 4845-4852

32.	Wu, J.*, et al.* (1998) Trapping of intermediates during the refolding of recombinant human epidermal growth factor (hEGF) by cyanylation, and subsequent structural elucidation by mass spectrometry. *Protein Sci* 7, 1017-1028

33.	Arolas, J.L.*, et al.* (2004) Role of kinetic intermediates in the folding of leech carboxypeptidase inhibitor. *J Biol Chem* 279, 37261-37270

34.	Salamanca, S.*, et al.* (2003) Major kinetic traps for the oxidative folding of leech carboxypeptidase inhibitor. *Biochemistry* 42, 6754-6761

35.	Chang, J.Y. (2002) The folding pathway of alpha-lactalbumin elucidated by the technique of disulfide scrambling. Isolation of on-pathway and off-pathway intermediates. *J Biol Chem* 277, 120-126

36.   Chang, J.Y., and Li, L. (2002) Pathway of oxidative folding of alpha-lactalbumin: a model for illustrating the diversity of disulfide folding pathways. *Biochemistry* 41, 8405-8413

37.   Permyakov, E.A., and Berliner, L.J. (2000) alpha-Lactalbumin: structure and function. *FEBS Lett* 473, 269-274

38.   Chang, J.Y. (1997) A two-stage mechanism for the reductive unfolding of disulfide-containing proteins. *J Biol Chem* 272, 69-75

39.   Chang, J.Y., *et al.* (2000) The underlying mechanism for the diversity of disulfide folding pathways. *J Biol Chem* 275, 8287-8289

40.   Arolas, J.L., *et al.* (2004) Secondary binding site of the potato carboxypeptidase inhibitor. Contribution to its structure, folding, and biological properties. *Biochemistry* 43, 7973-7982

41.   Narayan, M., *et al.* (2003) Characterizing the unstructured intermediates in oxidative folding. *Biochemistry* 42, 6947-6955

42.   Ruoppolo, M., *et al.* (2000) Contribution of individual disulfide bonds to the oxidative folding of ribonuclease A. *Biochemistry* 39, 12033-12042

43.   Shin, H.C., and Scheraga, H.A. (2000) Catalysis of the oxidative folding of bovine pancreatic ribonuclease A by protein disulfide isomerase. *J Mol Biol* 300, 995-1003

44.   Narayan, M., *et al.* (2003) Native conformational tendencies in unfolded polypeptides: development of a novel method to assess native conformational tendencies in the reduced forms of multiple disulfide-bonded proteins. *J Am Chem Soc* 125, 2036-2037

45.   Creighton, T.E., *et al.* (1996) The roles of partly folded intermediates in protein folding. *Faseb J* 10, 110-118

46.   Arolas, J.L., *et al.* (2005) Study of a major intermediate in the oxidative folding of leech carboxypeptidase inhibitor: contribution of the fourth disulfide bond. *J Mol Biol* 352, 961-975

47.   Dobson, C.M., *et al.* (1994) Understanding how proteins fold: the lysozyme story so far. *Trends Biochem Sci* 19, 31-37

48.   Matagne, A., and Dobson, C.M. (1998) The folding process of hen lysozyme: a perspective from the 'new view'. *Cell Mol Life Sci* 54, 363-371

49.   Klein-Seetharaman, J., *et al.* (2002) Long-range interactions within a nonnative protein. *Science* 295, 1719-1722

50.   Canet, D., *et al.* (2002) Local cooperativity in the unfolding of an amyloidogenic variant of human lysozyme. *Nat Struct Biol* 9, 308-315

51.   De Felice, F.G., *et al.* (2004) Formation of amyloid aggregates from human lysozyme and its disease-associated variants using hydrostatic pressure. *Faseb J* 18, 1099-1101

52.   van den Berg, B., *et al.* (1999) The oxidative refolding of hen lysozyme and its catalysis by protein disulfide isomerase. *Embo J* 18, 4794-4803

53.   Klafki, H.W., *et al.* (1993) Reduction of disulfide bonds in an amyloidogenic Bence Jones protein leads to formation of "amyloid-like" fibrils in vitro. *Biol Chem Hoppe Seyler* 374, 1117-1122

54.   Koo, B.W., and Miranker, A.D. (2005) Contribution of the intrinsic disulfide to the assembly mechanism of islet amyloid. *Protein Sci* 14, 231-239

55.   Chen, Y., and Dokholyan, N.V. (2005) A Single Disulfide Bond Differentiates Aggregation Pathways of ss2-Microglobulin. *J Mol Biol*

56.   Lee, S., and Eisenberg, D. (2003) Seeded conversion of recombinant prion protein to a disulfide-bonded oligomer by a reduction-oxidation process. *Nat Struct Biol* 10, 725-730

57.   Welker, E., *et al.* (2001) A role for intermolecular disulfide bonds in prion diseases? *Proc Natl Acad Sci U S A* 98, 4334-4336

58.   Khare, S.D., *et al.* (2005) Sequence and structural determinants of Cu, Zn superoxide dismutase aggregation. *Proteins* 61, 617-632

59.   Kuhlman, B., and Baker, D. (2004) Exploring folding free energy landscapes using computational protein design. *Curr Opin Struct Biol* 14, 89-95

60.   Creighton, T.E. (1986) Disulfide bonds as probes of protein folding pathways. *Methods Enzymol* 131, 83-106

61.   Li, Y.J*., et al.* (1995) Mechanism of reductive protein unfolding. *Nat Struct Biol* 2, 489-494

62.   van Mierlo, C.P*., et al.* (1993) Partially folded conformation of the (30-51) intermediate in the disulphide folding pathway of bovine pancreatic trypsin inhibitor. 1H and 15N resonance assignments and determination of backbone dynamics from 15N relaxation measurements. *J Mol Biol* 229, 1125-1146

63.   van Mierlo, C.P*., et al.* (1991) Two-dimensional 1H nuclear magnetic resonance study of the (5-55) single-disulphide folding intermediate of bovine pancreatic trypsin inhibitor. *J Mol Biol* 222, 373-390

64.   Kortemme, T*., et al.* (1996) Comparison of the (30-51, 14-38) two-disulphide folding intermediates of the homologous proteins dendrotoxin K and bovine pancreatic trypsin inhibitor by two-dimensional 1H nuclear magnetic resonance. *J Mol Biol* 257, 188-198

65.   Laity, J.H*., et al.* (1997) Structural characterization of an analog of the major rate-determining disulfide folding intermediate of bovine pancreatic ribonuclease A. *Biochemistry* 36, 12683-12699

66.   Shimotakahara, S*., et al.* (1997) NMR structural analysis of an analog of an intermediate formed in the rate-determining step of one pathway in the oxidative folding of bovine pancreatic ribonuclease A: automated analysis of 1H, 13C, and 15N resonance assignments for wild-type and [C65S, C72S] mutant forms. *Biochemistry* 36, 6915-6929

67.   Cemazar, M*., et al.* (2003) Oxidative folding intermediates with nonnative disulfide bridges between adjacent cysteine residues. *Proc Natl Acad Sci U S A* 100, 5754-5759

68.   Gehrmann, J.*, et al.* (1998) Structure determination of the three disulfide bond isomers of alpha-conotoxin GI: a model for the role of disulfide bonds in structural stability. *J Mol Biol* 278, 401-415

69.   van den Berg, B.*, et al.* (1999) Characterisation of the dominant oxidative folding intermediate of hen lysozyme. *J Mol Biol* 290, 781-796

70.   Arolas, J.L.*, et al.* (2005) NMR structural characterization and computational predictions of the major intermediate in oxidative folding of leech carboxypeptidase inhibitor. *Structure (Camb)* 13, 1193-1202

Figure legends

Figure 1. Schematic overview of opposite landscapes in oxidative folding. (A) For proteins like BPTI, a limited number of 1- and 2-disulfide intermediates, comprising native disulfide pairings and native-like structures, funnel protein conformations toward the native state. (B) Hirudin-like proteins fold through an initial stage of disulfide bond formation (packing) followed by the reshuffling of scrambled isomers (consolidation) to form the native protein. The heterogeneous ensembles of 1-, 2- and 3-disulfide intermediates are indicated by 1S, 2S and 3S, respectively.

Figure 2. Comparative folding of small disulfide-rich proteins. R and N indicate the reduced and native forms, respectively. 1S, 2S, 3S, and 4S are ensembles of molecules with the corresponding number of disulfide bonds. The disulfide pairings of folding intermediates are shown between parentheses and the missing disulfides between square brackets. The native structure of each protein is depicted as a ribbon plot, including native disulfide pairings (in yellow). The calcium atom bound to αLa is represented by a red sphere. Interestingly, BPTI and TAP, both containing three disulfide bonds, have an almost identical size, structure and disulfide pairing but differ completely in their mechanism of oxidative folding.

Figure 3. Oxidative folding of αLA in the absence or presence of calcium. Folding of αLA (0.5 mg/mL) was carried out in Tris-HCl buffer (pH 8.4) containing 0.2 mM 2-mercaptoethanol, with or without 5 mM $CaCl_2$. Folding intermediates were trapped by acidification (4% trifluoroacetic acid) in a time-course manner and further analyzed by RP-HPLC. N and R indicate the elution position of the native and fully reduced protein, respectively. IIIA and IIA stand for two predominant 3- and 2-disulfide intermediates, respectively, comprising native disulfide bond pairings. Adapted, with permission, from Ref.

[36].

Figure 4. Disulfide-dependent aggregation of the prion protein. (A) Model of oligomerization and conversion of prion protein PrP from its normal cellular form (PrP^C) to an infectious "scrapie" isoform (PrP^{Sc}). A terminal thiolate of the disulfide-linked oligomer attacks the intramolecular disulfide bond of PrP^C forming an intermolecular disulfide bond. This builds up the polymer while maintaining a reactive thiolate in terminal position ready to attack the next monomer. A conformational transition to a more rigid, β-enriched scrapie isoform would protect the intermolecular disulfide bonds, inhibiting the reaction of depolymerization. (B, C) Electron micrographs showing the oligomeric form of PrP[90-231] formed after reduction of the conserved disulfide bond. (B) Oligomers at completion of the process. (C) Oligomers at completion of the process in which the initial solution has been seeded with the oligomers from B (ratio of PrP^{Sc} seeds to PrP^C is 1:10). Adapted, with permission, from Ref. [56].
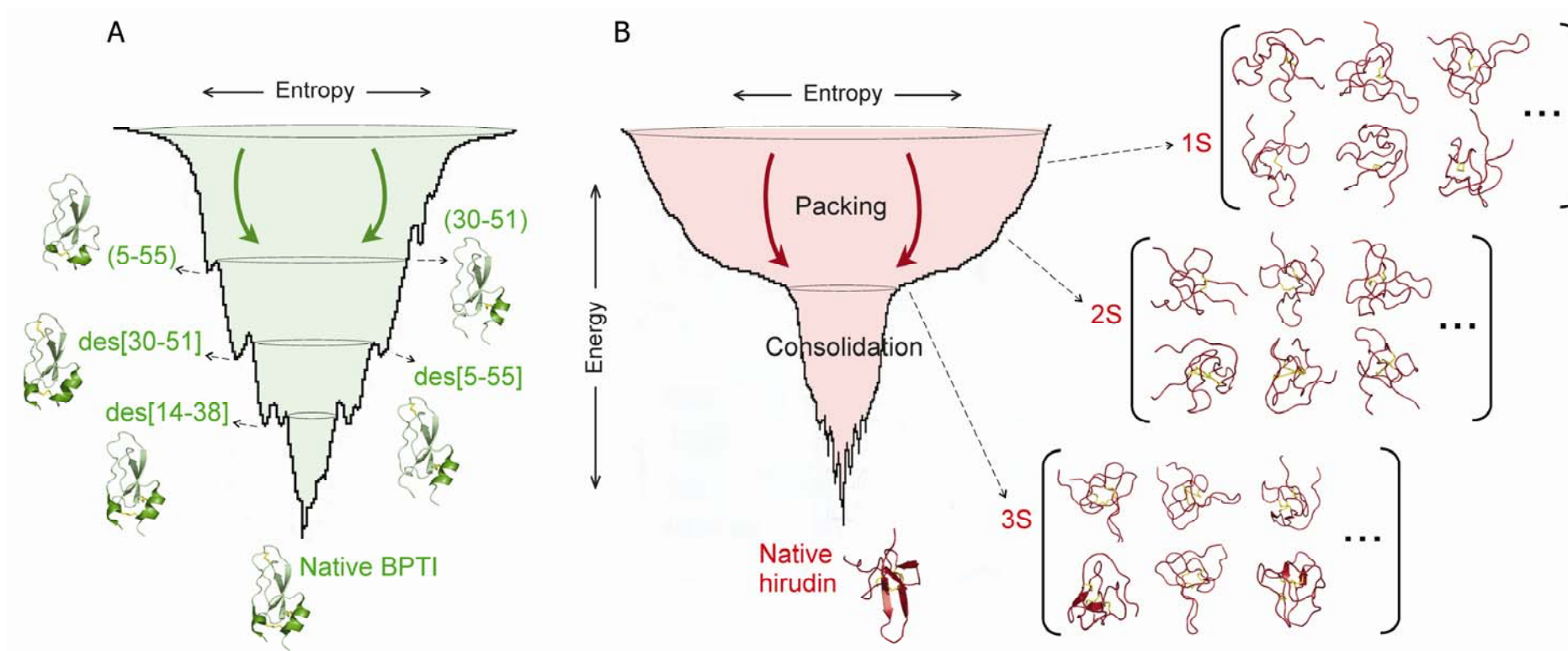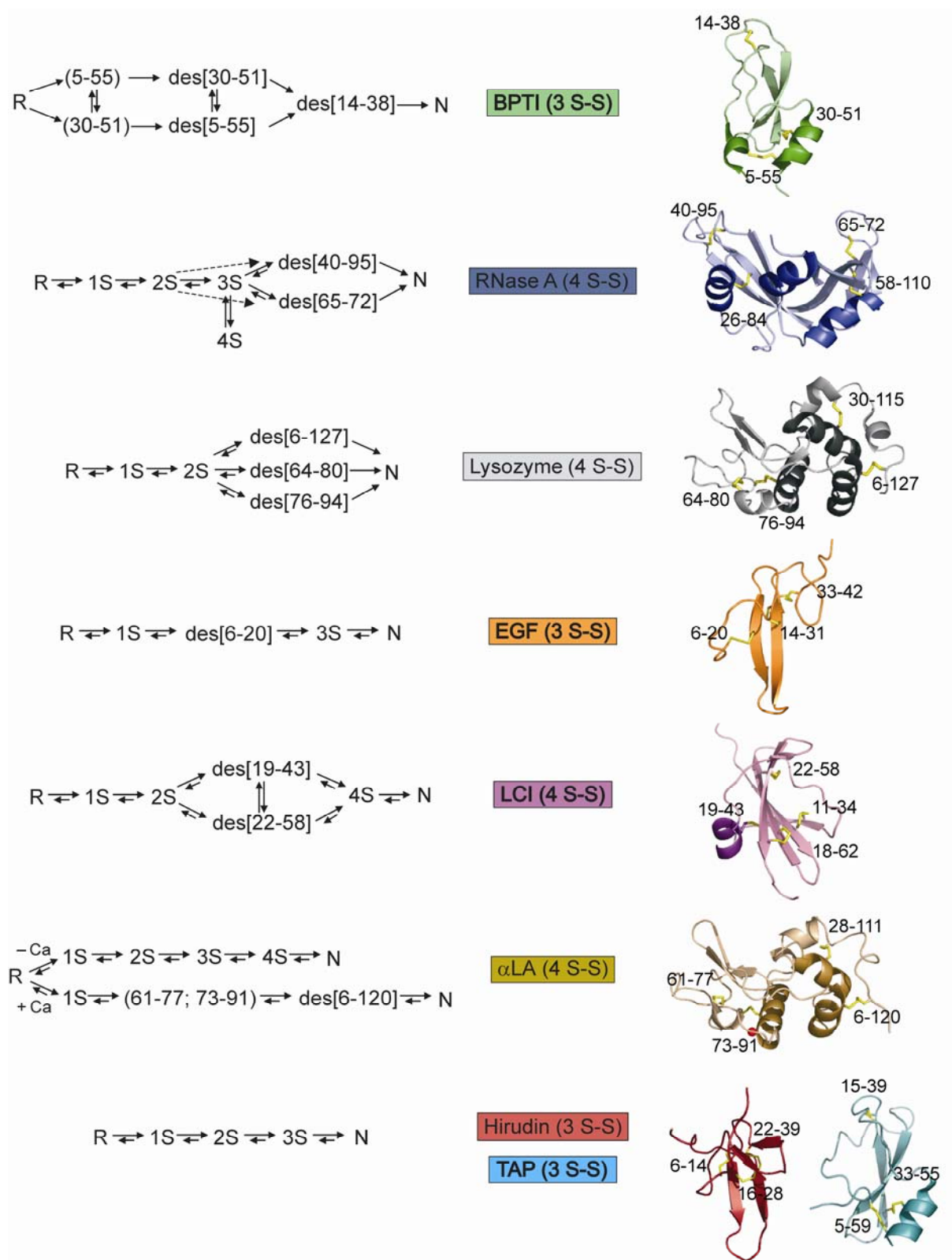
Figure 1

Figure 2

R ⟶ (5-55) ⟶ des[30-51]
R ⟶ (30-51) ⟶ des[5-55] ⟶ des[14-38] ⟶ N

BPTI (3 S-S)

14-38
30-51
5-55

R ⇌ 1S ⇌ 2S ⇌ 3S ⇌ des[40-95] ⟶ N
3S ⇌ des[65-72]
4S

RNase A (4 S-S)

40-95
65-72
58-110
26-84

R ⇌ 1S ⇌ 2S ⇌ des[6-127]
des[64-80] ⟶ N
des[76-94]

Lysozyme (4 S-S)

30-115
6-127
64-80
76-94

R ⇌ 1S ⇌ des[6-20] ⇌ 3S ⇌ N

EGF (3 S-S)

33-42
6-20   14-31

R ⇌ 1S ⇌ 2S ⇌ des[19-43] ⇌ 4S ⇌ N
des[22-58]

LCI (4 S-S)

22-58
19-43   11-34
18-62

−Ca 1S ⇌ 2S ⇌ 3S ⇌ 4S ⇌ N
R
+Ca 1S ⇌ (61-77; 73-91) ⇌ des[6-120] ⇌ N

αLA (4 S-S)

28-111
61-77
6-120
73-91

R ⇌ 1S ⇌ 2S ⇌ 3S ⇌ N

Hirudin (3 S-S)
TAP (3 S-S)

15-39
22-39
6-14
16-28
33-55
5-59

25

Figure 3
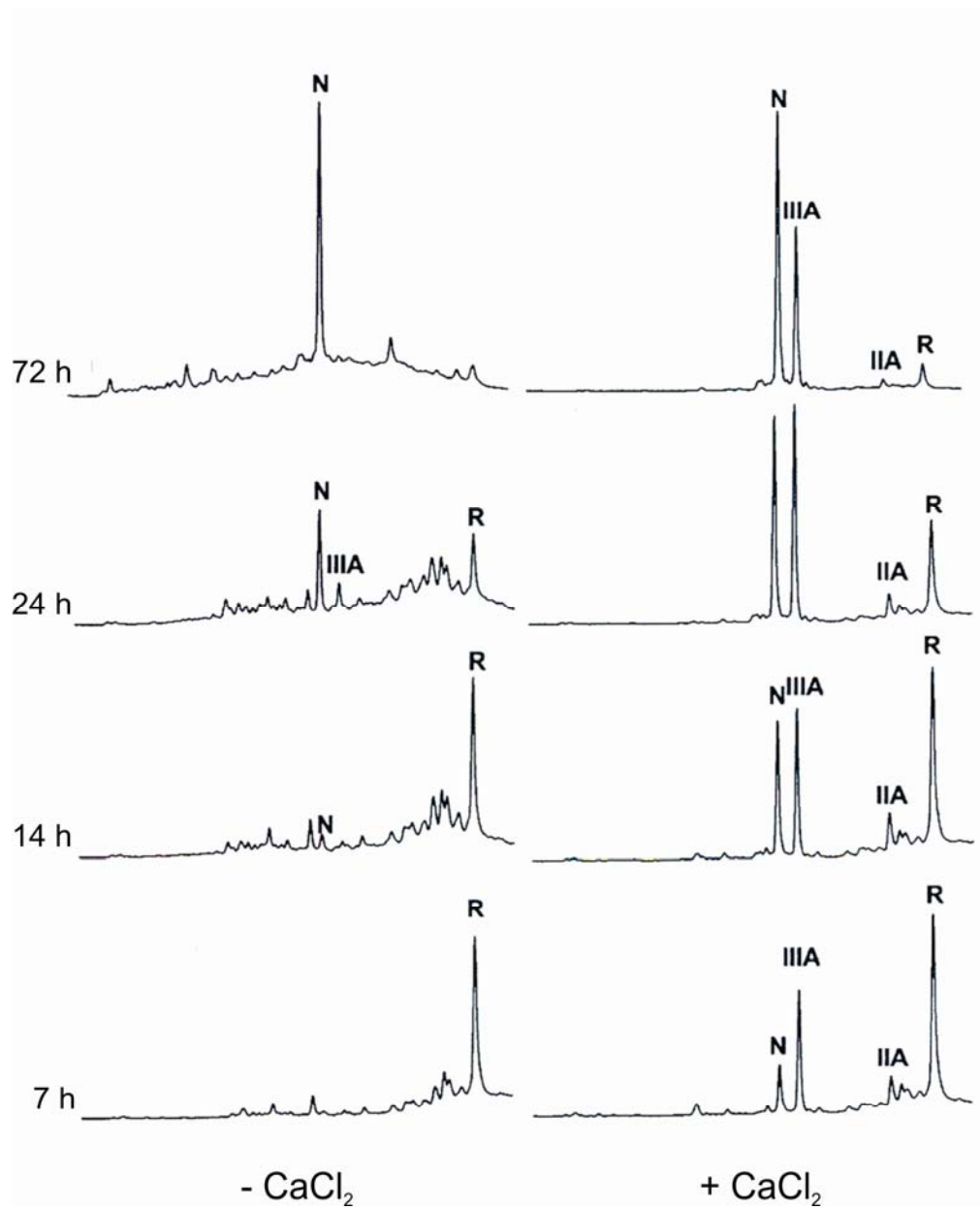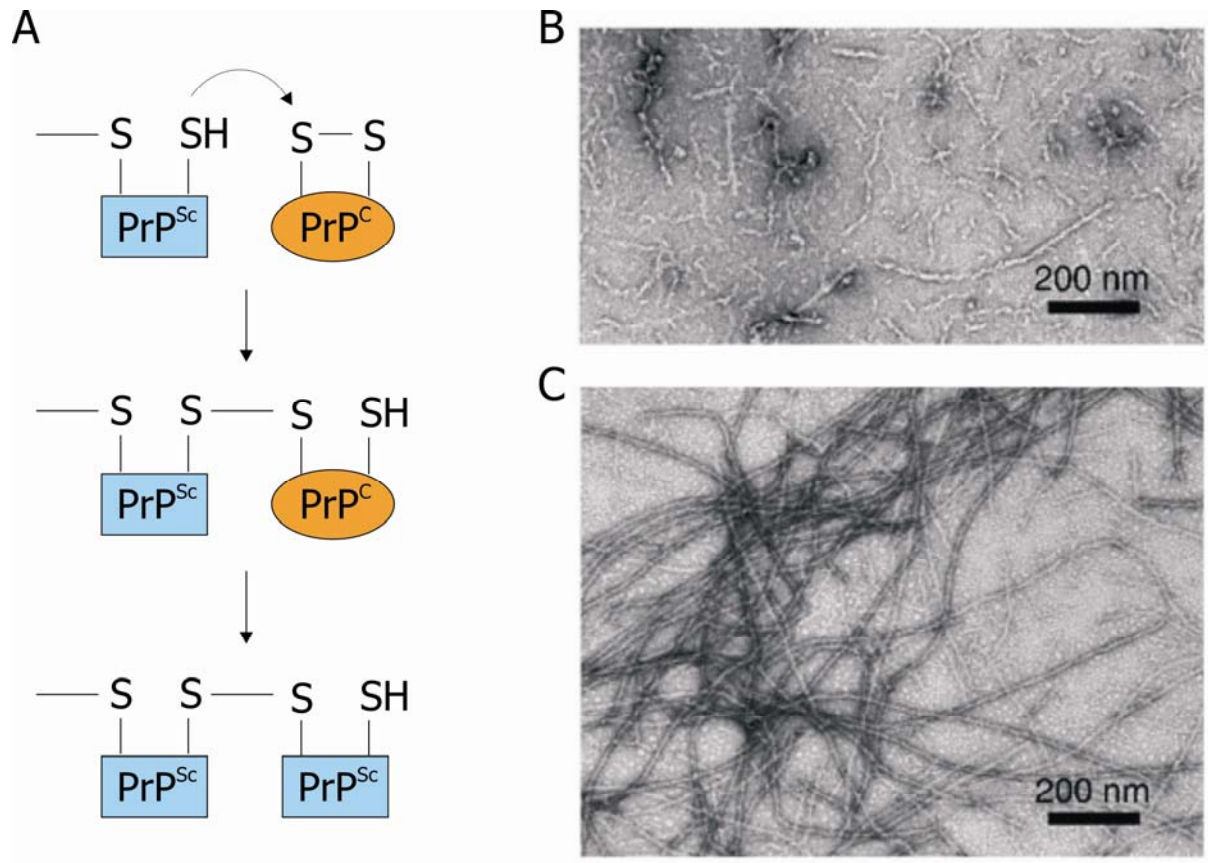

72 h

24 h

14 h

7 h

- CaCl₂

+ CaCl₂

Figure 4

Box 1. Oxidative folding and reductive unfolding techniques

In the technique of oxidative folding, pioneered by Creighton *et al.* [60], disulfide-containing proteins are initially fully reduced and denatured. After removal of both reagents, the reduced and denatured proteins are allowed to refold (at ~ pH 8.5) in the presence of selected buffers containing different redox agents. The folding intermediates are then trapped in a time-course manner and subsequently analyzed using several techniques. Thus, the disulfide-folding landscape is elucidated from the mechanism of formation of disulfide bonds, and characterized by the heterogeneity and structures of disulfide species that accumulate along the process of oxidative folding. The original studies of oxidative folding were based on the irreversible trapping of intermediates by alkylation (with iodoacetate) and further analysis by ion-exchange chromatography. Nevertheless, rearrangements of intermediates during the trapping procedure were observed for both BPTI and RNase A [18]. The most effective method seems to be the reversible quenching of the solution by acidification (usually with trifluoroacetic acid at pH ≤ 2) and analysis of the trapped intermediates by reversed-phase high performance liquid chromatography (RP-HPLC) [32]; see the example of αLA in Figure 3. This strategy permits not only the structural characterization of the intermediates, but also stop/go studies of them after readjusting the pH. In oxidative folding, the influence of redox agents on the folding process and the efficiency of recovery of the native protein can be achieved by using different redox agents, for instance, cystine and oxidized glutathione (GSSG), which promote disulfide formation, and cysteine, reduced glutathione (GSH) and 2-mercaptoethanol, which promote disulfide bond reshuffling.

In the reductive unfolding technique [61], native disulfide-containing proteins are treated with different concentrations of reducing agent (e.g., dithiothreitol; DTT) in the absence of denaturant. To monitor the unfolding reaction (performed at ~ pH 8.5), aliquots of the sample are removed at various time intervals, quenched by acidification (or alternatively by alkylation)

and analyzed by RP-HPLC. The mode of reduction will have crucial implications in interpreting the stabilization of disulfide bonds in the native state by protein structure (see main text).
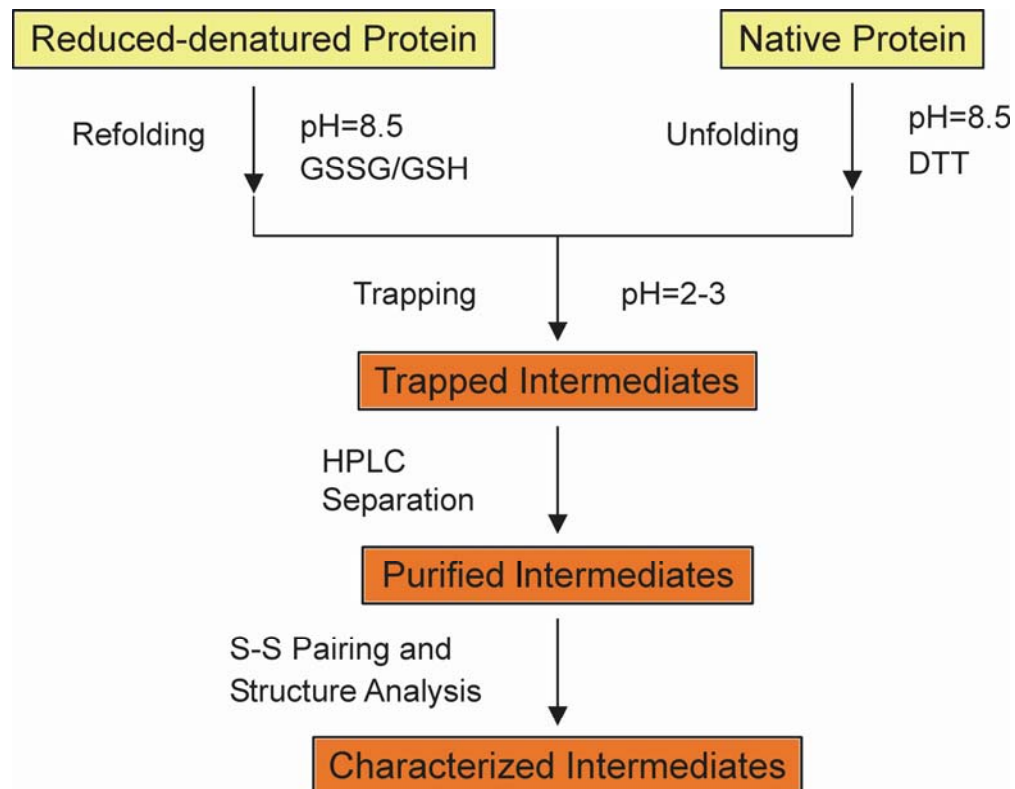
Box 2. Structures of intermediates: shedding light on protein folding

Although some intermediates have been directly purified from the folding reaction (see below), most studies have taken advantage of the construction of analogues in which the free cysteine residues are replaced with alanine or serine residues. The first intermediate analogues to be analyzed were those of BPTI. NMR studies on analogues [5-55] and [30-51] showed a partially folded structure that was structurally similar to that of native BPTI [62, 63]. Additionally, an analogue of the des[5-55] species displayed a structure almost identical to native BPTI, despite some differences in the N- and C-terminal regions [64]. For RNase A, NMR analysis of the des[40-95] and des[65-72] intermediate analogues also revealed native-like structures [65, 66]. However, small structural differences spatially adjacent to the mutation sites accounted for lower stability compared to the native protein (Figure I). An analogue of the MFI scrambled isomer of AAI was recently characterized by NMR, providing clues on the role of structural constraints in directing the folding process: the compact fold brings the cysteine residues into close proximity, thus facilitating reshuffling to native disulfide bonds [67]. Also, scrambled isomers of the peptide α-conotoxin GI were synthesized and characterized with implications for its structure and stability [68].

"Real" intermediates of other small, disulfide-rich proteins (with reactive cysteines, unlike the analogues) have been trapped, isolated by RP-HPLC and further characterized by NMR. The highly native-like structure of the des[76-94] intermediate of lysozyme clearly depicted the reasons behind its accumulation during folding: its Cys94 thiol is not solvent-accessible and therefore direct oxidation to the fully native protein is restricted [69]. The recently determined structure of the major intermediate of LCI, des[22-58] (named III-A), correlates with the finding on lysozyme [70]; III-A displays a native-like structure and its two free cysteine residues are buried (confirmed by NMR amide-proton exchange experiments) (Figure I). Besides the accessibility of thiol groups, other factors like proximity may affect the formation

of disulfide bonds. Thus, for kalata B1, the recently determined structure of its major intermediate, des[1-15], revealed a native-like structure that maintains the two cysteine residues distant from each other, thereby preventing direct oxidation to the final disulfide bond [19]. The overall results presented here emphasize the necessity of high-resolution structure determinations of real intermediates to fully understand disulfide folding.

Figure I. Ribbon structures of native RNase A and LCI together with their predominant folding intermediates des[40-95] and des[22-58].

Native RNase A   des[40-95] RNase A

Native LCI   des[22-58] LCI