ORIGINAL RESEARCH COMMUNICATION

Association Between Foldability and Aggregation Propensity in Small Disulfide-Rich Proteins

Hugo Fraga,* Ricardo Graña-Montes,* Ricard Illa, Giovanni Covalada, and Salvador Ventura

Abstract

Aims: Disulfide-rich domains (DRDs) are small proteins whose native structure is stabilized by the presence of covalent disulfide bonds. These domains are versatile and can perform a wide range of functions. Many of these domains readily unfold on disulfide bond reduction, suggesting that in the absence of covalent bonding they might display significant disorder. **Results:** Here, we analyzed the degree of disorder in 97 domains representative of the different DRDs families and demonstrate that, in terms of sequence, many of them can be classified as intrinsically disordered proteins (IDPs) or contain predicted disordered regions. The analysis of the aggregation propensity of these domains indicates that, similar to IDPs, their sequences are more soluble and have less aggregating regions than those of other globular domains, suggesting that they might have evolved to avoid aggregation after protein synthesis and before they can attain its compact and covalently linked native structure. **Innovation and Conclusion:** DRDs, which resemble IDPs in the reduced state and become globular when their disulfide bonds are formed, illustrate the link between protein folding and aggregation propensities and how these two properties cannot be easily dissociated, determining the main traits of the folding routes followed by these small proteins to attain their native oxidized states. *Antioxid. Redox Signal.* 21, 368–383.

Introduction

LACKING HYDROPHOBIC CORES and stable secondary structure elements, the native structure of many small proteins is stabilized by the binding to metal ions or by the presence of disulfide bonds in close spatial vicinity (69). The covalent linkage of cysteine residues by disulfide bonds is an important and, in many cases, essential structural feature for numerous proteins (1, 3, 31, 58). Disulfide bonds usually cross-link distant regions of the protein sequence, decreasing the entropy of the unfolded state and making it less favorable relative to the folded conformation (7, 83). Sometimes, disulfide bonds also act enthalpically, by stabilizing local interactions (83). Although in most cases disulfides play a structural role and are, therefore, only indirectly essential for protein function, there are also examples in which they are involved in the regulation of protein activity (10, 48).

Small protein domains whose structural features are determined by their disulfide bonds are usually referred to as

disulfide-rich domains (DRDs) (24). The group of proteins that is covered by this definition is broad, including both secreted and intracellular proteins, and they are involved in a wide variety of functions, such as growth factors, toxins, enzyme inhibitors, and structural or ligand-binding domains within larger polypeptides (24). Since protein classification of small proteins is often unreliable using common sequence and structural comparison tools, Cheek *et al.* undertook a classification of 2945 DRDs found in the PDB on the basis of their structural and evolutionary relatedness as well as disulfide bonding patterns. Their classification leads to the recognition of 41-fold types (24). Domains in the same fold group share a structural core composed of secondary structure elements found in the same spatial arrangement. In this work, we exploit this database to investigate the folding and aggregation properties of DRDs, providing new insights on how the interplay between these properties shapes their structures and influences their folding pathways.

Departament de Bioquímica i Biologia Molecular, Institut de Biotecnologia i Biomedicina, Universitat Autònoma de Barcelona, Barcelona, Spain.

*These two authors contributed equally to this work.

Innovation

DRDs are small versatile proteins involved in a wide range of functions. The structuration of these kinds of proteins depends on the formation of a specific pattern of disulfide bonds through oxidative folding reactions, which are extremely slow when compared with the folding kinetics of globular proteins. This implies that DRDs populate unfolded or partially folded ensembles during relatively long periods of time, which might favor side aggregation reactions, resulting in an extremely dangerous situation in the cell. We demonstrate, however, that DRDs sequences have evolved to support a particularly low intrinsic aggregation load and, in fact, their solubility properties resemble those of IDPs, which constitute a paradigm of protein solubility. Moreover, we provide a rationale to forecast and explain the oxidative folding pathways of DRDs, a goal that remained elusive to date, on the basis of the interplay between folding and aggregative properties.

Results

In the absence of disulfide bonds, DRDs are predicted to be intrinsically unstructured

The structure of a protein results from the balance between the conformational entropy and the energy of residue interactions (39). During protein folding, the formation of a sufficient number of interactions is necessary to compensate the loss of conformational entropy. In this context, the influence of covalent disulfide bonds to protein stability is unique, as they mainly act by decreasing the entropy of the unfolded polypeptide (66). This, however, raises the question of what is the structure adopted by DRDs in their reduced state, that is, in the absence of the entropic strain imposed by the disulfide bonds. In order to address this issue, we undertook a series of *in silico* analysis using the Cheek's dataset (24). In this dataset, the different DRDs are classified according to their structural features, enabling the selection of proteins covering the complete spectrum of structures adopted by this group of proteins. Within the fold groups, the domains

are assembled into families of homologs. Accordingly, the complete dataset could be classified in 98 different families on the basis of structural and evolutionary relationships (24). We used a representative protein of each of these families for further analysis. The insulin-like family representative was, however, removed, as it consists of two different polypeptide chains, thus resulting in a dataset of 97 DRDs (Supplementary Table S1; Supplementary Data are available online at www.liebertpub.com/ars).

First, we analyzed the presence of disorder in the sequences of DRDs using the FoldUnfold server (42), assuming that the proteins are in the reduced state. This algorithm exploits the computed mean packing density in a globular state for each of the 20 natural amino-acid residues to generate packing density profiles and identifies disordered segments as those protein regions enriched in residues displaying low packing density values (42). We detail the results of such analysis in Supplementary Table S2. Interestingly, 41% of DRDs sequences are predicted to be unfolded in the reduced state by this algorithm (Fig. 1A). We mapped the predicted disordered regions over the secondary structure elements in the DRDs structure (Table 1). A majority (58%) of the predicted disordered residues map to unstructured regions in the native, disulfide-bonded conformation, another 28% to regular secondary structure elements, and the rest to turns.

Cys, Trp, Tyr, Ile, Phe, Val, and Leu are considered order-promoting residues, as they occur more frequently in ordered than in disordered structural regions (37, 72). In addition, in the FoldUnfold scale, Cys has a high mean packing density (42), reflecting that, in protein structures, Cys residues tend to exist in densely packed environments. However, in contrast to the rest of order-promoting residues, which are bulky and hydrophobic, in protein structures, Cys residues are closer to other Cys than to any other type of residue, including hydrophobic ones (59). This suggests that Cys are packed, because they can form disulfides and not due to any other intrinsic physicochemical property of this amino acid. Therefore, it is predicted that under reducing conditions, Cys would contribute less to protein order.

Despite there being no good substitution for Cys, it can tolerate substitutions with other small amino acids and in protein engineering studies, Cys is preferentially mutated to Ser or, alternatively, to Ala. These substitutions are considered

FIG. 1. Prediction of protein disorder in DRDs. Percentage of DRDs and their mutated variants predicted to be disordered according to (A) FoldUnfold, (B) FoldIndex, and (C) RONN algorithms. The fraction of DRDs sequences predicted to be unfolded is represented by black bars. In (C), the remaining domains with at least 10 consecutive residues predicted as disordered are represented by gray bars. Disorder predictions are shown for wt domains, DRDs mutants with cysteine mutated to alanine (Cys/Ala), and variants with cysteine mutated to serine (Cys/Ser). In (A), predictions for DRDs with cysteine mutated to a hypothetical amino acid X with the average packing density of the 20 natural amino acids (Cys/X) are also shown. wt, wild-type; DRD, disulfide-rich domain; RONN, regional order neural network.

TABLE 1. DISTRIBUTION OF DISORDERED RESIDUES IN DRDs, AS PREDICTED BY FOLDUNFOLD, AMONG THE DIFFERENT SECONDARY STRUCTURE ELEMENTS IN THEIR 3D STRUCTURES

SS type	Ratio of residues (%)
Helix	18.06
Strand	9.56
Turn	14.76
Unstructured	57.62

DRD, disulfide-rich domain; SS, secondary structure.

the less disruptive to study the role of covalent bonding in protein structures. We virtually mutated Cys to Ala and Ser in the 97 domains and re-analyzed the degree of predicted disorder (Supplementary Table S2). The 69% and 82% of the sequences in which Cys are mutated to Ala and Ser are now predicted to be unstructured, respectively (Fig. 1A). We also considered the mutation of Cys residues in DRDs to a hypothetical amino acid X displaying the average packing density of the 20 natural residues. In this case, 64% of the sequences are predicted to be unfolded (Fig. 1A).

In order to validate the results obtained with FoldUnfold we employed another two additional protein disorder detection tools: FoldIndex (70) and regional order neural network (RONN) (85), based on unrelated prediction models (Supplementary Tables S3 and S4).

The FoldIndex server developed by the Sussman lab (70) exploits an algorithm previously described by Uversky *et al.* (80) that takes into account the average hydrophobicity of natural amino acids in the Kyte-Doolittle scale and the average net charge of the polypeptide. FoldIndex provides a single score for the entire sequence, predicting whether it is folded (positive values) or not (negative values). Assuming that the proteins are reduced, it predicts 41% of DRDs sequences as intrinsically unstructured (Fig. 1B). When we used FoldIndex to analyze the effect of Cys to Ser mutations in the foldability of DRDs (Supplementary Table S3), 74% of the mutated sequences were predicted to be disordered (Fig. 1B). When we considered Cys to Ala mutations, 49% of the proteins still displayed negative values (Fig. 1B), despite Ala being considered a hydrophobic, order-promoting amino acid, in the FoldIndex scale (70).

The RONN algorithm developed by Yang *et al.* (85) aligns a query protein sequence to a set of prototype disordered/ordered regions and uses the alignment scores to classify the query sequence. RONN predicts 7% of reduced DRDs to have all of their residues unstructured, and an additional 29% are predicted to contain unstructured segments comprising more than 10 consecutive residues (Fig. 1C). The 21% and 45% of the sequences in which Cys are mutated to Ala and Ser are predicted as totally unstructured, respectively (Fig. 1C). An additional 61% and 45% of the sequences contain disordered regions in Ala and Ser mutants, respectively (Fig. 1C).

Venn diagrams analysis of the predictions for wild-type domains indicates a ~60% overlap between methods when compared in pairs (Supplementary Fig. S1).

In order to confirm that the predicted large increase in disorder on Cys to Ser mutation does not arise simply from a higher number of Ser residues in the mutated sequences, an effect that could be potentially important in short proteins, we

FIG. 2. Distribution of changes in DRDs stability upon mutation. Changes in free energy of unfolding ($\Delta\Delta G$) computed using the FoldX forcefield for Cys to Ala (yellow) and Cys to Ser (blue) mutants. Stability changes are represented as histograms that were calculated using $5 \text{ kcal} \cdot \text{mol}^{-1}$ bins, and histogram data were fitted to Gaussian distributions (black lines). To see this illustration in color, the reader is referred to the web version of this article at www.liebertpub.com/ars

analyzed whether there was any correlation between the length-normalized increase in the % of residues predicted to be disordered on mutation and the number of mutated Cys in a sequence. Pearson correlation coefficients of $r^2=0.008$, $r^2=0.004$, and $r^2=0.007$ were obtained for FoldUnfold, FoldIndex and RONN, respectively, enabling us to discard this effect.

The analysis of 100 randomly selected sequences shorter than 150 residues and devoid of disulfide bonds from the SCOP-derived ASTRAL40 dataset (15) with FoldUnfold and FoldIndex indicates that, as expected, reduced DRDs and especially their Ala and Ser analogues are predicted to display a higher degree of disorder than these domains (Supplementary Fig. S2).

Overall, independently of the considered algorithm and mutational scheme, these data converge to indicate that, in the absence of disulfide bonds, many DRDs would resemble intrinsically disordered proteins (IDPs) in structural terms.

In the absence of disulfide bonds, DRDs are predicted to be highly destabilized

A reduction of disulfide bonds usually leads to DRDs global unfolding (18, 30, 53). We calculated the theoretical differential stability of analogues of the 97 domains, in which

FIG. 3. Comparison of the aggregation properties of DRDs, small globular proteins, and IDPs. Value distribution of different parameters indicative of protein aggregation capability calculated using the AGGRESCAN algorithm: (A) average aggregation propensity (Na4vSS), (B) area of the aggregation profile above the threshold per residue (AATr), and (C) total area of the aggregation peaks – Hot Spots – per residue (THSAr). The distributions are shown in black for DRDs and in red or blue for the datasets of globular proteins or IDPs, respectively. IDP, intrinsically disordered protein. To see this illustration in color, the reader is referred to the web version of this article at www.liebertpub.com/ars

all the Cys residues involved in disulfide bonds (592 residues from a total of 613 Cys) were mutated to either Ala or Ser, using FoldX, a force field that enables a structure-based estimation of mutational free energy changes on the stability of a protein (44, 76). The histograms of $\Delta\Delta G$ values for the 194 resulting variants reveal a broad distribution with peaks around +5 and +8 kcal·mol⁻¹ for Ala and Ser mutants, respectively, with a tail toward more destabilizing $\Delta\Delta G$ values, especially in the case of Ser variants (Fig. 2) and a main contribution of entropic factors (Supplementary Fig. S3). $\Delta\Delta G$ values ≥ 5 kcal·mol⁻¹ can be considered strongly destabilizing (78) and might well account for the disruption of DRDs structure on reduction. No significant correlation ($r^2=0.025$) was found between the predicted increase in disorder in the sequence on Cys to Ser mutation according to FoldIndex and the calculated FoldX destabilization, as the energetic contribution of native disulfides depends on their topological connectivity. In addition, we did not observe any obvious relationship between the calculated destabilization and the specific fold or function of the domain.

DRDs sequences display low aggregation propensity and reduced binding sites to Hsp70 chaperones

The failure of proteins to fold, or to remain correctly folded, usually results in their aggregation, which is associated with the impairment of essential cellular processes. Protein aggregates might elicit stress responses, inhibit the proteasome, sequester chaperones and transcription factors, promote membrane pore formation and calcium overload, and cause oxidative stress and mitochondrial dysfunction, among other deleterious effects, being thus associated to pathological states (25). Recent results show that protein aggregation can be reliably correlated to a set of defined physicochemical parameters (26) and determined to a large extent by the primary sequence (49, 82). This has enabled the development of algorithms that are able to predict the presence of aggregation-prone regions (APRs) in proteins as well as their overall aggregation propensities (14).

As a general trend, evolution has endorsed globular proteins with solubility in their functional conformations. Nevertheless, when their stability is compromised and they become partially unfolded, the exposition of previously protected APRs might

trigger aberrant aggregation reactions (13, 38). Our results indicate that before they can form their disulfide bonds or in the eventual case that they become reduced, DRDs have both a high sequential and energetic propensity to occupy unfolded/disordered states, which implies that they might be at risk of aggregation. We used AGGRESCAN, an algorithm previously developed by our group, to predict protein aggregation in this protein set (11, 27, 34). We addressed how the aggregation properties of DRDs sequences can be compared with those of intrinsically disordered and globular proteins of a similar size. To this aim, we used the DISPROT (77) and the SCOP-derived ASTRAL40 datasets (15) and randomly selected 100 sequences shorter than 150 residues and devoid of disulfide bonds from each database. The following parameters were calculated for the three groups of proteins (all normalized by the protein length):

- (i) The average aggregation propensity of the sequence (Na4vSS)
- (ii) The number of aggregation-prone regions (NnHS)
- (iii) The average aggregating potency of the detected aggregation peaks (THSAr)
- (iv) The average aggregating potency of residues above the detection threshold (AATr), independently whether they are clustered in aggregating peaks or not.

DRDs display obviously lower average aggregation propensities than proteins from the ASTRAL40-derived dataset (Fig. 3A), with mean Na4vSS values of -16.4 and -6.3, respectively (the more negative the value, the more soluble the protein). A similar trend is observed for AATr, which is yet another measure of the overall aggregation propensity of the sequences (Fig. 3B). In addition, only 2% of the proteins from the ASTRAL40-derived dataset are devoid of aggregating stretches, whereas this value increases and includes 22% of the DRDs. In fact, 82% of the proteins devoid of APR were predicted as unfolded by FoldUnfold. In the same line, those DRDs presenting APRs have an average of 2.9 APRs for each 100 residues in contrast to the 3.9 APRs predicted for the ASTRAL40 group. Not only DRDs display fewer APRs, but also the average aggregating potency of the detected aggregation peaks in their sequences (THSAr) is lower (Fig. 3C). In fact, in terms of aggregation properties,

TABLE 2. AGGREGATION PROPERTIES OF THE IDPs, DRDs, AND GLOBULAR PROTEIN DATASETS AS PREDICTED BY AGGRESCAN

	IDPs	DRDs	Globular
Na4vSS	−21.2	−16.4	−6.3
AATr	0.106	0.109	0.154
NnHS	2.557	2.950	3.942
THSAr	0.085	0.076	0.121

IDP, intrinsically disordered protein.

DRDs sequences resemble much more those of IDPs than those of globular proteins (Fig. 3A–C and Table 2).

To ascertain whether the reduced intrinsic aggregation properties of DRDs might result from an amino-acid compositional bias, we compared their amino-acid content with the average composition of the proteins in the Swiss-Prot database (Fig. 4). Following the trend described earlier, DRDs are significantly depleted in Ile, Leu, and Val, three aliphatic residues that display high-aggregation propensity in most aggregation scales. In globular proteins, these residues are usually buried inside the tertiary structure and their depletion in DRDs would respond to the absence of hydrophobic cores in these disulfide-bonded domains. This might contribute to the lower aggregation propensity of DRDs. In contrast, they are enriched in Pro, a residue incompatible with the formation of β -sheet-enriched aggregates.

Overall, the analysis of the aggregating properties of DRDs suggests that as previously reported for IDPs (54), their sequences might have been under evolutionary control to minimize the presence of aggregation-prone residues as a strategy to prevent the aggregation of the solvent-exposed polypeptide chain before oxidative folding results in the attainment of a globular protective conformation.

FIG. 4. Comparison of the amino-acid composition of DRDs and small globular proteins. Differential composition of DRDs relative to the Swiss-Prot database 2013. Differential abundances are shown for each natural amino acid except cysteine (included in the computational analysis but not represented).

Disordered proteins have been shown to require, in general, less assistance from chaperones than ordered, globular proteins (47). To explore whether this could be the case of DRDs, we used LIMBO, a chaperone binding site predictor for the Hsp70 chaperones, trained from peptide binding data and structural modeling (81). We analyzed potential Hsp70 binding sites in the IDPs, DRDs, and ASTRAL40-derived datasets. As can be seen in Figure 5, both IDPs and DRDs display a significantly reduced number of Hsp70 binding sites when compared with globular proteins, reinforcing the structural similitude between reduced DRDs and disordered polypeptides. Whether DRDs exhibit also low binding to members of chaperone families other than Hsp70 should be further investigated.

Amyloidogenic regions in DRDs overlap with secondary structure elements

Albeit aggregation-prone and amyloidogenic regions coincide in many cases (73), AGGRESCAN is not aimed at identifying short sequences that are able to nucleate the specific and ordered assembly of amyloids. This task is better attained using the WALTZ algorithm, as it employs a position-specific scoring matrix deduced from the analysis of the amyloid properties of a large set of hexapeptides (60). Using this program, we surveyed DRDs for the presence of potential amyloidogenic stretches. A total of 24 hits were obtained in 18 out of the 97 DRDs sequences. The same analysis in the ASTRAL40-derived dataset rendered 82 hits in 54 proteins, thus confirming the intrinsic low aggregation propensity displayed by DRDs.

Despite their low number, the detected sequences might potentially nucleate amyloid assembly if exposed to solvent. We explored the structures of the 18 DRDs and 54 globular proteins to map the detected amyloid stretches over their secondary structure elements. In both cases, the majority of residues in amyloid regions map to regular secondary structure elements (Table 3). This association between foldability and amyloidogenicity suggests a negative selection process acting to avoid the accumulation of aggregating sequences in unstructured regions. Interestingly enough, in DRDs, 44% of the amyloidogenic residues map to regions that adopt α -helical conformations in the native state and only 19% map to β -sheets, which is consistent with the view that one of the most effective strategies to neutralize amyloidogenic stretches in proteins is to hide them into α -helical structures (52, 79). This bias toward α -helices is not evident in the ASTRAL40 dataset, in agreement with previous data indicating that the aggregation propensity of all- α , all- β , and mixed α/β globular proteins is fairly similar (54).

Association between DRDs' predicted intrinsic properties and folding pathways

The folding and unfolding reactions of a large number of disulfide-containing proteins have been characterized (9, 58). These studies illustrated a high degree of diversity of oxidative folding pathways, especially for DRDs. The two extreme mechanisms are illustrated by (a) proteins that fold through a reduced number of intermediates containing mainly native disulfide bonds and (b) proteins that fold via a highly heterogeneous population of intermediates containing

FIG. 5. Predicted chaperone binding to DRDs, small globular proteins, and IDPs. Distribution of the number of chaperone binding sites per protein predicted with LIMBO for (A) the dataset of globular proteins, (B) DRDs, and (C) the dataset of IDPs.

mostly non-native disulfides, including fully oxidized scrambled isomers (19, 20). Leech-derived trypsin inhibitor (LDTI) and hirudin (Fig. 6A, B) represent two well-characterized models folding through the (a) and (b) mechanisms, respectively (4, 22, 23, 67).

To see whether there exists any relationship between the folding mechanism of a DRD and its intrinsic properties, we analyzed the sequences of LDTI and hirudin using FoldIndex (Fig. 6A, B). Eighty-nine percent of the LDTI sequence is predicted to be folded (unfoldability = 0.20). In contrast, 71% of the hirudin sequence is predicted to be unfolded (unfoldability = -0.14). The same trends are observed with FoldUnfold and RONN, which predict 13% and 0% of LDTI sequences to be disordered, respectively; however, for hirudin, these values increase to 32% and 45%, respectively. We used AGGRESCAN to compare the aggregation properties of the sequences of these two proteins. Reduced hirudin is predicted to be much more soluble than LDTI, with Na4vSS values of -33.3 and -3.2, respectively. No APRs are predicted for hirudin. In contrast, residues 23–27, overlapping the only α -helix in LDTI (residues 24–29), are predicted to be aggregation prone. In a previous work, we isolated and structurally characterized all the major intermediates in the LDTI folding pathway (67). Interestingly, all of them share a preformed α -helix, independently if it was covalently linked to rest of the structure or not, indicating that it forms early in the folding reaction.

The ligand binding module five (LA5) of the low-density lipoprotein (LDL) receptor is a DRD of 40 residues and three disulfide bonds with a calcium binding motif that is essential

for its structure. We have shown that in the presence of calcium, such as hirudin, its folding involves an initial phase of non-specific packing that gives rise to complex equilibrated populations of intermediates (2). This results in an inefficient folding reaction that requires 36 h to be completed. More importantly, in the absence of calcium, native LA5 is not formed and the oxidative folding reaction ends up with the formation of non-native scrambled forms through a stochastic process, indicating that non-covalent interactions, if they occur, are not sufficient to drive the folding of this domain. Accordingly, FoldIndex predicts 52% of the sequence to be unfolded (Fig. 6C), with an unfoldability value of -0.12. FoldUnfold and RONN predict 48% and 51% of LA5 to be disordered in the reduced form. AGGRESCAN predicts LA5 to be highly soluble with an Na4vSS value of -36.1, and WALTZ does not detect any amyloidogenic stretch. Thus, for LA5 also, disorder and solubility appear to be associated with an inefficient folding pathway. The LDL receptor contains seven LA repeats in tandem comprising residues 25–313. The average Na4vSS value for this large region is -32, with an unfoldability value of -0.10 and absence of amyloidogenic sequences. These values might explain why in intact cells the individual modules of the LDL receptor do not fold independently and first collapse into folding intermediates that are characterized by long-distance non-native disulfide bonding and absence of native structure and why the receptor does not aggregate in this misfolded state (51).

Fuller and co-workers have recently compared the oxidative folding of four conotoxins, GIHA, PIHA, SmIIA, and RIIK, sharing identical disulfide connectivity (41a). After 120 min of a reaction, more than 60% of initially reduced RIIK attained the native state. In contrast, less than 30% of the other proteins were folded at this time point, with a large accumulation of scrambled isomers. Again, we found a correlation between the efficiency of the reaction and the predicted aggregation and disorder properties of the proteins. In this way, the AGGRESCAN Na4vSS value of RIIK is 4.3; whereas it is -27.9, -49.0, and -52.7 for PIHA, SmIIA, and GIHA, respectively. Similarly, the FoldIndex unfoldability value of RIIK is 0.21; whereas for the other proteins, this value ranges between -0.37 and -0.43.

Even with DRDs that share sequence homology and 3D structure, it has been so far impossible to forecast whether they would fold by similar or different pathways (9). This is best illustrated by the cases of bovine pancreatic trypsin

TABLE 3. DISTRIBUTION OF PROTEINS AMYLOIDGENIC RESIDUES IN DRDs AND GLOBULAR PROTEINS DATASETS, AS PREDICTED BY WALTZ, AMONG THE DIFFERENT SECONDARY STRUCTURE ELEMENTS IN THEIR 3D STRUCTURES

SS type	Ratio of residues (%)	
	DRDs	Globular
Helix	43.52	35.37
Strand	18.65	37.69
Turn	10.88	4.63
Unstructured	26.94	22.31

FIG. 6. Relationship between DRDs aggregation properties and their foldability. APRs detected by employing the AGGRESCAN algorithm (in red), disordered segments predicted with FoldIndex (in blue), and domain regions where APRs and disordered fragments overlap (in magenta) are mapped over (A) LDTI (PDB: 2kmo), (B) hirudin (PDB: 2hir), (C) The ligand binding module five (LA5) of the low-density lipoprotein receptor (PDB: 1AJJ), (D) BPTI (PDB: 1d0d) and (E) TAP (PDB: 1d0d) structures. Unfoldability profiles computed with FoldIndex are shown below its corresponding structure for each domain; green and blue indicate folded and disordered regions, respectively. APRs, aggregation-prone regions; BPTI, bovine pancreatic trypsin inhibitor; LDTI, leech-derived trypsin inhibitor; TAP, tick anticoagulant peptide. To see this illustration in color, the reader is referred to the web version of this article at www.liebertpub.com/ars

inhibitor (BPTI) and tick anticoagulant peptide (TAP), two proteins that share a very similar structure and identical disulfide connectivity (Fig. 6D, E) but exhibit different oxidative folding pathways (17, 21, 84). Folding of BPTI proceeds *via* an orderly formation of native 1- and 2-disulfide bonds and native-like subdomains. Out of 74 possible 1-, 2-disulfide isomers that may potentially serve as folding intermediates of BPTI, only 5 intermediates were shown to populate the pathway and all of them adopt native disulfide bonds. Fully oxidized 3-disulfide scrambled isomers are absent. In contrast, folding intermediates of TAP consist of a highly heterogeneous population of 1- and 2-disulfide isomers, with at least 30 species identified. More importantly, 3-disulfide isomers serve as major folding intermediates in TAP, the reshuffling of which is required to attain the native conformation. Despite TAP and BPTI structures exhibiting an RMS deviation of only 1.01 Å at the backbone level, the two proteins differ in the degree of predicted disorder. FoldIndex predicts 80% and 49% of TAP and BPTI residues to be disordered, with unfoldability values of -0.10 and -0.01 , respectively (Fig. 6D, E). FoldUnfold does not predict any disordered region in the BPTI sequence, whereas 16% of the TAP sequence is disordered. RONN does not predict disordered regions in any of the two sequences. The two proteins are also expected to differ in their degree of flexibility in the oxidized state or extent of hydrogen–deuterium exchange, according to the H-protection server (56), which predicts 76% and 54% of the main chain hydrogen atoms to be exchangeable in TAP and BPTI, respectively; suggesting that, despite their structural similarity, the different degree of disorder encoded in their sequences might also impact the conformational properties of their folded states. To further analyze this issue, we used CABS-flex (50), an efficient procedure for the simulation of structure flexibility of folded

globular proteins. The analysis of TAP and BPTI structures with CABS-flex confirms that TAP experiences larger conformational fluctuations than BPTI, a property that can be more obviously appreciated in the residue mean-square-fluctuation profile of the respective simulations (Supplementary Fig. S4).

AGGRESCAN predicts TAP sequence to be much more soluble than that of BPTI with Na4vSS values of -24.7 and -7.0 for TAP and BPTI, respectively. In BPTI, the regions encompassing residues 16–24 and 28–36, which overlap almost exactly with the two central β -sheets, are predicted to be aggregation prone (Fig. 6D). WALTZ also predicts the second stretch (residues 29–37) to be amyloidogenic. Interestingly, the two β -sheets in BPTI have been shown to constitute the nucleus from which conformational folding initiates (12). AGGRESCAN does not detect any APR in the first β -sheet of TAP. In addition, the aggregation-prone 35–39 region overlapping with the end of the second TAP β -sheet is predicted to display lower potency than the correspondent region in BPTI. WALTZ does not detect any amyloidogenic region in TAP. In order to analyze whether our predictions reflect experimental aggregation amyloid propensities, we synthesized two peptides corresponding to β -sheet 1 (18-IIRYFYN-24) (β 1-BPTI) and β -sheet 2 (22-LCQTFVY-35) (β 2-BPTI) in BPTI and two peptides corresponding to β -sheet 1 (22-ERAYFRN-28) (β 1-TAP) and β -sheet 2 (32-GCDSFWI-38) (β 2-TAP) in TAP and analyzed their *in vitro* aggregation properties.

β 1-BPTI and β 1-TAP peptides were incubated at $100 \mu\text{M}$ at pH 7.5 and 25°C for 48 h, and their aggregation was evaluated using synchronous light scattering (Fig. 7A). In excellent agreement with the predictions, β 1-BPTI exhibited two times higher scattering signals than β 1-TAP. We also monitored the binding of the two incubated peptides to the

FIG. 7. Light scattering and Th-T binding of BPTI and TAP β -sheet peptides.

Light scattering after incubation of (A) β 1 peptides from BPTI (red) and TAP (blue) and (B) β 2 peptides from BPTI (red) and TAP (blue) in the presence of TCEP. Th-T binding after incubation of (C) β 1 peptides from BPTI (red) and TAP (blue) and (D) β 2 peptides from BPTI (red) and TAP (blue) in the presence of TCEP. Black lines represent buffer scattering and free Th-T fluorescence, respectively. TCEP, *tris*(2-carboxyethyl)phosphine; Th-T, Thioflavin-T. To see this illustration in color, the reader is referred to the web version of this article at www.liebertpub.com/ars

FIG. 8. CR binding, structural and morphological properties of BPTI and TAP β 2 peptides incubated in the presence of TCEP. (A) Absorbance spectra of free CR (black) and after addition of β 2-BPTI (red) or β 2-TAP (blue) peptides after incubation. (B) Differential spectra of CR bound to β 2-BPTI (red) or β 2-TAP (blue) subtracted with free CR spectrum. (C) ATR-FTIR spectra of aggregated β 2-BPTI (red) and β 2-TAP (blue). IR spectra were deconvoluted to a maximum of five Gaussians. TEM micrographs of (D) β 2-TAP and (E) β 2-BPTI aggregates. CR, Congo red; ATR-FTIR, attenuated total reflectance–Fourier transform infrared spectroscopy; TEM, transmission electron microscopy. To see this illustration in color, the reader is referred to the web version of this article at www.liebertpub.com/ars

amyloid dye Thioflavin-T (Th-T) (Fig. 7C). Again, the increase in Th-T fluorescence emission elicited by the BPTI peptide was twofold that of the TAP peptide. However, none of the peptides promoted significant spectral changes in Congo red (CR), which is yet another amyloid dye (data not shown).

Since β 2-BPTI and β 2-TAP peptides contain a Cys residue, they were incubated as β -sheet 1 peptides, but 10 mM *tris*(2-carboxyethyl)phosphine (TCEP) was added in order to ensure their reduction. In agreement with the predictions, β -sheet 2 peptides exhibited significantly higher scattering signal and binding to Th-T than β -sheet 1 peptides (Fig. 7B, D). Again, the scattering and Th-T signals of β 2-BPTI were higher than those of β 2-TAP. Moreover, only the β 2-BPTI peptide was able to promote a red shift in the CR spectra; the difference between spectra of the dye in the presence and absence of the peptide exhibiting the characteristic amyloid band at 540 nm (Fig. 8A, B). Attenuated total reflectance–Fourier transform infrared spectroscopy (ATR-FTIR) enabled addressing the structural features of the aggregated β -sheet 2 peptides. Deconvolution of the absorbance spectra in the amide I region demonstrates significant differences in the secondary structure content of these aggregates (Fig. 8C). In this way, the IR spectrum of the β 2-TAP is dominated by a signal at 1641 cm^{-1} , accounting for 48% of the spectrum area, typically attributed to unordered structures; whereas the IR spectrum β 2-BPTI exhibits a mean band at 1628 cm^{-1} , which accounts for 45% of the total area, corresponding to a β -sheet secondary structure. The morphological features of incubated β -sheet 2 peptides were analyzed using transmission electron microscopy (TEM). As shown in Figure 8D and E, in both cases, we detected the presence of protein aggregates. Nevertheless, in good agreement with CR, IR data, and WALTZ predictions, the aggregates formed by β 2-TAP were amorphous whereas those of β 2-BPTI displayed an amyloid-like morphology. Overall, the experimental data confirm that the central elements of TAP structure are less aggregation prone than the topologically equivalent structures in BPTI.

Human proinsulin is a single polypeptide consisting of chains A and B which are connected by three disulfides and linked by a 35-residue-long connecting peptide that will be later cleaved off to yield mature insulin. The folding pathway of proinsulin involves a reduced number of intermediates and is, therefore, fast and efficient, being directed by the collapse of A and B chains to form a folding nucleus, with a negligible impact of the connecting peptide (46). This mechanism is correlated with the different predicted aggregation and disorder properties of the proinsulin domains. Chains A and B are predicted to be ordered by the three predictors and aggregation prone by AGGRESCAN, with unfoldability values of 0.34 and 0.21 and Na4vSS values of 15.9 and 10.9, respectively; whereas the connecting peptide is predicted to be disordered and highly soluble (Na4vSS = -34.7). Insulin-like growth factor-1 (IGF-1) shares proinsulin topology with two A and B chains that are linked by three conserved disulfide bonds with identical connectivity and a connecting peptide of 12 residues. In contrast to proinsulin, the folding of IGF-1 renders two different isomers, the native and a swapped form (46). Proinsulin and IGF-1 share most similarities structurally and functionally. Considering their high sequence homology and small size, their different folding behaviors are intriguing. While the predicted aggregation and

disorder properties of the IGF-1 A chain and the connecting peptide are similar to those of proinsulin, the B chain is predicted to be less ordered and specially more soluble, with unfoldability and Na4vSS values of 0.12 and -1.4 , respectively. These differences might well account for the different folding properties of these structurally similar proteins. A set of Proinsulin/IGF-1 hybrids has been constructed by exchanging their A and B chains, and their folding reactions are evaluated. According to our predictions, all molecules containing the less aggregation-prone B chain of IGF-1 folded into two isomers, whereas those containing the B chain of proinsulin folded into a unique stable structure (46).

Association between predicted intrinsic properties and the folding pathway of Nerita versicolor carboxypeptidase inhibitor

The data in the previous section suggest that DRDs, such as hirudin, LA5, or TAP, displaying sequences with significant disorder and low aggregation tendency might find it difficult to follow preferential folding pathways along which the protein folds with a high probability guided by the rapid formation of native-like local interactions. In these proteins, the folding reaction would occur through an initial stage of non-specific disulfide-bond bonding (packing), leading to the formation of scrambled species, which are further resolved into the native connectivity in an extremely slow reaction involving the establishment of non-covalent native-like contacts (3). In contrast, proteins such as LDTI and BPTI displaying lower disorder propensity and harboring sequences in which native secondary structure and aggregating regions overlap significantly would tend to fold through faster pathways that are dominated from the very beginning by non-covalent native-like interactions, resulting in the formation of a reduced number of folding intermediates in which native disulfides are predominant (3).

We explored whether this apparent association between predicted intrinsic protein properties and folding pathways would enable us to predict the way in which a previously uncharacterized DRD would fold. To this aim, we used *Nerita versicolor* inhibitor (NvCI), a novel 53-residue-long exogenous proteinaceous inhibitor of metalloproteinases from the marine snail *Nerita versicolor* (29). The NvCI structure defines a distinctive protein fold that is composed of a two-stranded antiparallel β -sheet connected by three loops (Fig. 9A). FoldIndex provides an unfoldability value of 0.10 for NvCI, FoldUnfold predicts the protein to be fully folded, and RONN predicts 85% of the sequence to be ordered. AGGRESCAN provides an Na4vSS value of -9.5 and predicts the regions encompassing residues 23–28 and 40–44, which overlap with the two central β -sheets, to be aggregation prone (Fig. 9A, B). WALTZ also predicts the second β -sheet (40–HVWTFE-45) to be amyloidogenic. Overall, the intrinsic properties of NvCI resemble much more those of LDTI/BPTI than those of hirudin/TAP proteins. To analyze whether, as predicted, this similitude results in an efficient folding reaction, we expressed, purified, and analyzed the conformational properties and oxidative folding of NvCI.

According to the crystal structure of NvCI (29), residues 41–VWT-43 in the amyloidogenic stretch are totally protected from the solvent in the native state, which would prevent the aggregation of the protein once it folds. We took profit of the buried Trp42 being the only tryptophan residue

FIG. 9. Conformational properties and oxidative folding of NvCI. (A) APRs detected employing the AGGRESCAN algorithm (in red), disordered segments predicted with FoldIndex (in blue) and domain regions where APRs and disordered fragments overlap (in magenta) are mapped over the NvCI structure (PDB: 4A94). (B) Unfoldability profile computed with FoldIndex; green and blue indicate folded and disordered regions, respectively. (C) NvCI reductive unfolding was followed after TCEP addition. The normalized area under each Trp fluorescence spectrum was plotted as a function of time. (D) TCEP 20 mM was added to oxidized NvCI, and Trp emission spectrum was recorded demonstrating a large structural shift on protein reduction (native and reduced fluorescence spectra are shown in blue and black, respectively). (E) In agreement with Trp intrinsic fluorescence, significant changes in NvCI secondary structure are observed when oxidized NvCI is reduced with TCEP (native and reduced CD spectra are shown in blue and black, respectively). (F) Time-dependent NvCI oxidation was followed by acid trapping and RP-HPLC. As predicted from its intrinsic properties, NvCI displays fast kinetics and its oxidative folding pathway involves the formation of only two major intermediates. R and N indicate the reduced and native states, respectively. NvCI, *Nerita versicolor* inhibitor. To see this illustration in color, the reader is referred to the web version of this article at www.liebertpub.com/ars

in NvCI to monitor whether this region becomes exposed and, therefore, at risk of aggregation in the reduced state. To this aim, we added 20 mM TCEP to reduce the protein and monitored the change in Trp fluorescence along time. As can be seen in Figure 9C, an exponential increase in fluorescence emission occurs on reduction. A comparison of the Trp42 spectra in the oxidized and reduced states confirms that the initially buried residue becomes fully exposed once the protein loses its disulfide bonds (Fig. 9D). This transition is accompanied by structural changes leading to unfolded conformations, as evidenced in the circular dichroism spectra of the oxidized and reduced NvCI forms (Fig. 9E).

The oxidative folding of NvCI was examined with acid trapping and RP-HPLC analysis of the disulfide intermediates that accumulate along the folding reaction. The reduced and unfolded protein was allowed to refold in Tris-HCl buffer at pH 8.4 (see Materials and Methods section). As shown in Figure 9F, only a few major intermediates populate the folding process of NvCI, resulting in a fast and efficient reaction that is almost completed after 4 h, even in the absence of any thiol catalyst or oxidizing agent. Mass spectrometry analysis of the conversion of the major folding intermediate, containing two disulphide bonds, into native NvCI indicates that this step of the oxidative folding reaction is very fast and efficient (Supplementary Fig. S5), enabling a significant population of scrambled isomers to be discarded, which would trap the reaction as in the cases of hirudin, TAP, or the LA5 module. Thus, as predicted, the folding pathway of this DRD resembles in kinetic terms that of BPTI/LDTI-like proteins, supporting an association

between DRDs intrinsic properties and oxidative folding pathways.

Discussion

Disulfide bond containing proteins are numerous and are involved in important biochemical processes. Extensive studies on the folding and conformational properties of selected DRDs have been performed *in vitro* under different experimental conditions (3). Here, we use a set of *in silico* approaches to systematically address the impact of disulfide bonds in these DRD features and experimentally validate theoretical predictions on selected DRD models. Oxidative folding catalysis occurs almost exclusively in the endoplasmic reticulum (ER), in the intermembrane space of mitochondria (41) in eukaryotes, and in the periplasm in bacteria (36). Systematic studies of DRDs folding rates *in vivo* are missing, mainly because disulfide bond formation occurs during translation and translocation, making the process complex to analyze. However, several data in the literature point out that under physiological conditions, DRDs oxidative folding is a relatively slow process. For mitochondrial DRDs, such as Cox19, oxidative folding and import are linked; the residence time of this protein in the cytosol in the reduced state has been shown to be 8 min, indicating that its *in vivo* oxidative folding rate is exceedingly slow (40). In the case of the LDL receptor, which contains several DRDs, it has been shown that, in the ER, several of these modules contain wrongly paired disulfides and require assisted re-shuffling, which makes their folding slow (51). For BPTI, it

has been shown that the effect of protein disulfide isomerase on the reduced protein and the 1-disulfide intermediates is negligible, suggesting that *in vivo* the first steps of oxidative folding are uncatalyzed and, therefore, slow (32). One of the best *in vitro* reconstitutions of a DRD *in vivo* folding reaction has been recently done by Purcell *et al.*, who evaluated the role of a number of ER-resident enzymes in the Conus venom glands in the folding of conotoxins produced by the same organism (74). The most favorable combination promoted folding of less than 40% of the total protein in 4 min. This rate and efficiency is very low when compared with those of globular proteins of the same size, which usually fold in the microseconds time scale (55). Thus, modeling DRDs properties in their reduced states is physiologically relevant, as it is likely that they populate this state significantly during the gap between ribosomal synthesis and the attainment of the native conformation at specific cellular compartments.

The comparative analysis of protein disorder and stability in wild-type and reduced DRDs analogues clearly indicates that, independently of the fold they adopt in the native state, most of these proteins would remain significantly unfolded before they can form the stabilizing covalent bonds. The persistence of such unfolded protein conformations is puzzling, as folding is known to be a primary strategy to prevent aggregation and, in globular proteins, even partial or transient unfolding might lead to pathological protein deposition (62). In DRDs, this danger is minimized, because they possess sequences with strongly reduced aggregation properties, relative to the remaining globular proteins. Experimental evidence supports the solubility of DRDs sequences. We had previously characterized the *in vitro* folding pathways of different DRDs (5, 6, 8, 67), with several of them involving the population of hundreds of transient, but long-lived, and unstructured intermediates. Despite refolding reactions from the initially reduced states being usually performed at high protein concentrations ($>1\text{ mg}\cdot\text{ml}^{-1}$) and some of them taking days to be completed in the absence of catalysts, we never observed side aggregation reactions. This behavior contrasts with the observation that accumulation of folding intermediates in globular proteins, even for small and fast folding polypeptides such as SH3 or WW domains, recurrently results in their aggregation (38, 45, 64).

In globular proteins, disulfide bonds strongly stabilize the structure, thus minimizing unfolding reactions that might expose previously hidden APRs (43). This is likely the underlying reason explaining why, as a trend, globular proteins with disulfide bonds tend to display more aggregation-prone sequences than proteins devoid of this bond (63). Our results indicate that this generic strategy is not enough, or is not acting, to ensure the solubility of DRDs, as their sequences seem to be shaped to exhibit low aggregation propensity and a reduced number of APRs despite being enriched in disulfides and highly stable in the oxidized state. This suggests that the low aggregation tendency detected in most of these domains is not intended to prevent aggregation on local unfolding of the native structure, but rather to maintain solubility during the process of oxidative folding. In fact, the presence of aggregating sequences in DRDs might have important consequences for their production. In this way, Hcpidin, a DRD hormone involved in iron homeostasis, displaying a strikingly high Na4vSS value of 17.1 is produced exclusively in inclusion bodies in bacteria and

is resistant to solubilization on chemical synthesis (57). In yet another example, the aggregation during the recombinant production of eight Kunitz proteins, all sharing the same cysteine pattern, could be rationalized just using the AGGRESCAN output (75).

Once conscious that, in the reduced state, DRDs structurally resemble IDPs, the fact that their sequences tend to be soluble is not surprising. IDPs are known to display a high intrinsic solubility and a reduced number of APRs as a strategy to prevent the aggregation of the fully solvent-exposed polypeptide chain in the absence of a protective secondary structure (35, 54). The depletion in hydrophobic residues also prevents IDPs from chaperone-mediated protein degradation (65). All these properties seem to be shared by DRDs in their reduced states.

It is now clear that the ability to assemble into highly organized and cytotoxic amyloid-like structures is a feature shared by an extensive number of polypeptides (25). This implies that, apart from the native structure, an alternative, ordered, and stable state exists that may be accessible to proteins. These two states compete in the cell, because the set of non-covalent interactions that stabilize the native fold are also responsible for the formation of aggregates and, therefore, the formation of highly structured globular proteins comes at the expense of an increased aggregation propensity (54). Despite the DRDs in our dataset displaying a low content of amyloidogenic sequences and not being known to be involved in depositional disorders, the analysis of mature insulin, amylin, and calcitonin, three small proteins containing disulfide bonds involved in injection-localized amyloidosis, type 2 diabetes, and medullary carcinoma of the thyroid, respectively, indicate that, on the average, they display a low degree of predicted disorder along with high aggregation and amyloidogenic propensities (Supplementary Table S5), which suggests that these properties might also result in the formation of toxic assemblies when they converge in DRDs. The association between foldability and aggregation is evident in DRDs, where $\sim 62\%$ of amyloid nucleating residues are located in sequences that adopt a regular secondary structure in the native, cross-linked, state. Interestingly, in DRDs, most of these regions reside in α -helices, likely because the presence of preformed β -sheets with high aggregation propensities is intrinsically dangerous, as shown for several disease-linked proteins (33) and here for the second β -sheet in BPTI. The bias toward the protection of amyloid regions in α -helices is apparently higher in DRDs than in globular proteins. Since the folding of α -helices essentially depends on local contacts, they are known to form faster than β -sheets; the adoption of helical structures early in the rather slow DRD oxidative folding reactions may provide a means to decrease the aggregation propensity of these amyloid stretches.

The oxidative folding pathways of the analyzed DRDs illustrate an association between their foldability and aggregation propensity. DRDs displaying high disorder propensity in their reduced state are intrinsically highly soluble and devoid of regions with high aggregation propensity, but this comes at the cost of inefficient folding pathways. On the contrary, the presence of APRs overlapping with structural elements in DRDs seems to be associated with faster and more efficient folding pathways in which preferential intermediates are sequentially formed, likely because hydrophobic APRs would contribute to form the folding nucleus from which the reaction extends. Our analysis suggests that in

DRDs, folding and aggregation properties are finely tuned. On the one hand, proteins displaying slow folding reactions should be intrinsically soluble, as otherwise the persistence of unfolded or partially folded intermediates will result in a potentially harmful aggregation. On the other hand, since efficient DRDs folding reactions seem to depend on the presence of hydrophobic stretches, they have been associated with a higher aggregative risk; thus, the folding reaction should be inherently efficient, in order to avoid the accumulation of dangerous intermediates and APRs should be protected preferentially within regular structural elements.

Materials and Methods

DRD dataset

All the analyses described next were performed over the sequences and structures of a set of 97 domain representatives of the DRDs families in which the 41 DRD folds are divided, according to the classification scheme developed by Cheek *et al.* (24). In addition, protein disorder and aggregation properties were analyzed for LDTI, hirudin, the LDL receptor module LA5, BPTI, TAP, and NvCI wild-type sequences.

Analysis of protein disorder in DRD sequences

Intrinsic disorder in DRDs was predicted using the web-based algorithms FoldUnfold (42), FoldIndex (70), and RONN (85) with default settings for both wild-type and mutant sequences. An 11-residue sliding window was used in all cases. Predictions were performed with the wild-type sequences of DRDs and with mutant sequences where Cys were substituted by Ala or Ser. For the analysis made with FoldUnfold, mutant DRD sequences were also generated where Cys were substituted by a hypothetical amino acid X that possesses the average packing density of the 20 naturally occurring amino acids (42).

Analysis of the thermodynamic impact of disulfide deletion in DRDs

The effect of disulfide bond disruption on DRDs thermodynamic stability was evaluated using the FoldX algorithm (76). DRDs mutants, where disulfide-forming Cys were substituted by either Ala or Ser, were generated with the BuildModel command. This command computes energetic changes on mutation by comparing the mutant's rotamer energetics relative to that of the wild-type structure. FoldX calculations were obtained as the average of five independent runs performed with default physicochemical and energetic parameters. Disulfide deletion effect on energetic parameters was computed using isolated DRDs structures retrieved from the PDB; in those cases where DRDs were found as a part of a multi-domain protein structure, DRDs were extracted by employing the domain boundaries defined by Cheek *et al.* (24).

Analysis of the intrinsic aggregation properties of DRDs

Different parameters indicative of protein intrinsic aggregation properties were predicted for DRDs using the AGGRESCAN algorithm web server (27) with default settings. Directions about the use of AGGRESCAN and its outcome interpretation may be found in the AGGRESCAN

published tutorial (34). In order to compare DRDs aggregation properties with those of the globular and IDP ensembles, aggregation parameters were also calculated for two reference sets of 100 sequences with sizes below 150 residues and containing 1 or 0 Cys residues, which were retrieved among those proteins with less than 40% identity in the ASTRAL Compendium (ASTRAL40) (16) or among proteins in DISPROT database using pseudo-random numbers generated by the Mersenne Twister algorithm.

Identification of amyloidogenic regions in DRDs

The presence of amyloidogenic regions in DRDs and the ASTRAL40-derived dataset was evaluated using the web-based Waltz algorithm (60), employing default parameters.

Analysis of conformational flexibility in DRDs

The structural flexibility of DRDs was analyzed using the H-protection server (56) and the CABS-flex simulation algorithm (50), employing their default parameters.

Prediction of chaperone binding sites

Chaperone binding sites were predicted using the LIMBO algorithm webserver (81) employing default settings. Predictions were performed for the DRDs dataset and the sets of globular proteins and IDPs mentioned earlier.

Aggregation assay of BPTI and TAP derived peptides

Heptapeptides derived from BPTI and TAP corresponding to the sequences of strands $\beta 1$ ($\beta 1$ -BPTI: IIRYFYN and $\beta 1$ -TAP: ERAYFRN) and $\beta 2$ ($\beta 2$ -BPTI: LCQTFVY and $\beta 2$ -TAP: GCDSFWI) were purchased from EZBiolab, Inc. with a purity of >95%. Lyophilized peptides were dissolved in a 1:1 mixture of 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP) and trifluoroacetic acid (TFA) in order to obtain 5 mM peptide solutions. Dissolved peptides were filtered through 0.22 μ m filters to avoid the presence of pre-aggregated species and aliquoted to prepare peptide stocks. The solvent mixture was removed by centrifugal evaporation, and solid stocks were stored at -80°C . For aggregation assays, peptides were resuspended in 50 mM Tris pH 7.5, 150 mM NaCl to a 100 μ M concentration and sonicated on an ice bath for 10 min before incubation. Peptide concentration was verified by absorbance spectroscopy at 280 nm, and peptide samples were allowed to aggregate at 25°C and 500 rpm agitation. After 48 h of incubation, peptide aggregation properties were evaluated. For $\beta 2$ peptides, aggregation assays were performed in the presence of 10 mM TCEP to prevent the formation of intermolecular disulfide bonds.

Light scattering determination

Light scattering of incubated peptide samples was recorded from 300 to 700 nm at 90° using a FP-8200 spectrofluorimeter (Jasco) in synchronous mode. Scattered light spectra were obtained as the average of three consecutive scans.

Thioflavin-T binding

The binding of Th-T to incubated peptides was determined by diluting peptide samples in Th-T and recording the fluorescent emission spectra of the mixture between 460 and

600 nm, after equilibration at 298 K and employing an excitation wavelength of 445 nm in a FP-8200 spectrofluorimeter (Jasco). Final peptide and dye concentrations were 50 and 25 μM , respectively; and each spectrum was acquired as the accumulation of three consecutive scans.

Congo red binding

The binding of CR to aggregated peptides was evaluated by diluting fivefold the incubated peptides in CR, resulting in a final dye concentration of 10 μM , and recording the absorbance spectra of the mixture in the 380 to 670 nm range in a Cary 400 spectrophotometer (Varian).

Attenuated total reflectance–Fourier transform infrared spectroscopy

ATR-FTIR spectra were recorded for aggregated peptides samples as the accumulation of 16 scans in a Tensor 27 FTIR spectrophotometer (Bruker) using ATR mode default settings, and after evaporating water molecules under a $\text{N}_2(\text{g})$ stream. Infrared spectra were deconvoluted into overlapping Gaussian curves by employing the non-linear fitting program PeakFit (Systat Software).

Transmission electron microscopy

The morphology of the aggregated peptides was evaluated by TEM. Peptide samples were diluted 10-fold in milliQ water and then negatively stained by first allowing the deposition of 10 μl of the dilutions on carbon-coated copper grids for 5 min. After sample removal, the samples were stained for 1 min with 10 μl of 2% uranyl acetate, which was removed afterward. The grids were imaged at a 120 kV accelerating voltage in a JEM-1400 transmission electron microscope (JEOL).

Heterologous expression and purification of recombinant NvCI

The cDNA sequence of NvCI was fused in-frame to the *Saccharomyces cerevisiae* prepro- α -factor signal in the *Xho*I site of the pPICZ α A vector for secretion into the culture medium. Production of recombinant NvCI was carried out using a *Pichia pastoris* Zeocin hyper-resistant strain in an autoclavable bioreactor (Applikon Biotechnology). Purification of NvCI was performed using a combination of two ion exchange chromatographic methods: an initial weak cation exchange (AccellTM Plus CM; Waters) using 20 mM Tris-HCl (pH 7.0) and an ionic strength gradient (up to 1 M NaCl), followed by a second step of anion exchange (TSKgel[®] DEAE-5PW; Tosoh Bioscience LLC) using a linear gradient of 0%–100% 20 mM Tris-HCl (pH 8.5) containing 1 M NaCl. The purity of NvCI was determined by its molecular mass obtained by MALDI-TOF-MS, by Tris/Tricine/sodium dodecyl sulfate polyacrylamide gel electrophoresis, and by its functional activity against bovine carboxypeptidase 1.

Oxidative folding

NvCI (0.3 mg·ml⁻¹) was reduced in 50 mM Tris pH 8.4, NaCl 100, and 50 mM DL-dithiothreitol (DTT) for at least 2 h at 23°C. To initiate folding, samples were passed through a

PD-10 column (Sephadex-25; GE Healthcare) previously equilibrated with 50 mM Tris pH 8.4, NaCl 100 mM. To monitor the folding reaction, aliquots were removed at defined time points, quenched with 2% TFA, and analyzed by RP-HPLC as follows. Samples were injected in a Waters 2690 HPLC coupled to a UV detector set to 280 nm. A linear gradient of 20%–40% of 0.1% TFA in acetonitrile was applied for 70 min into a 250×4.6 (5 μM) C4 column (Phenomenex) at a flow rate of 0.75 ml·min⁻¹. To monitor conversion between species, 1 μl of sample extracted at selected times from the oxidation reaction were derivatized with 0.5 μl of 0.3 M vinylpyridine (Sigma) for 45 min, at room temperature, in darkness, then diluted to 15 μl with 0.1% TFA, and analyzed by mass spectrometry in a MALDI-TOF UltraFLEXtreme (Bruker Daltonics) in reflectron mode under 25 kV using external calibrators.

Intrinsic fluorescence

NvCI intrinsic tryptophan fluorescence was measured using NvCI (0.02 mg·ml⁻¹) after ON incubation with 20 mM TCEP (reduced state) and without any treatment (oxidized state). The spectra were measured in the 290–400 nm interval using a 280 nm excitation wavelength (10 nm excitation and emission slits, 0.1 s averaging time). NvCI reductive unfolding was analyzed by using tryptophan emission spectra every 30 s, as described earlier, after the addition of 20 mM TCEP.

Circular dichroism spectroscopy

Far UV CD spectra were acquired on a Jasco-710 spectropolarimeter that was continuously purged with nitrogen and thermostated at 25°C. Since both TCEP and DTT interfered with far UV CD spectra, protein was reduced and before CD analysis, the reducing agent was removed using the protocol described earlier.

Acknowledgments

This work was supported by grants BFU2010-14901 from Ministerio de Ciencia e Innovación (Spain) and 2009-SGR-760 from AGAUR (Generalitat de Catalunya). S.V. has been granted an ICREA Academia award (ICREA).

Author Disclosure Statement

No competing financial interests exist.

References

1. Anfinsen CB, Haber E, Sela M, and White FHJ. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain *Proc Natl Acad Sci U S A* 47: 1309–1314, 1961.
2. Arias-Moreno X, Arolas JL, Aviles FX, Sancho J, and Ventura S. Scrambled isomers as key intermediates in the oxidative folding of ligand binding module 5 of the low density lipoprotein receptor. *J Biol Chem* 283: 13627–13637, 2008.
3. Arolas JL, Aviles FX, Chang JY, and Ventura S. Folding of small disulfide-rich proteins: clarifying the puzzle. *Trends Biochem Sci* 31: 292–301, 2006.
4. Arolas JL, Bronsoms S, Aviles FX, Ventura S, and Sommerhoff CP. Oxidative folding of leech-derived tryptase

- inhibitor via native disulfide-bonded intermediates. *Antioxid Redox Signal* 10: 77–85, 2008.
5. Arolas JL, Bronsoms S, Lorenzo J, Aviles FX, Chang JY, and Ventura S. Role of kinetic intermediates in the folding of leech carboxypeptidase inhibitor. *J Biol Chem* 279: 37261–37270, 2004.
 6. Arolas JL, Bronsoms S, Ventura S, Aviles FX, and Calvete JJ. Characterizing the tick carboxypeptidase inhibitor: molecular basis for its two-domain nature. *J Biol Chem* 281: 22906–22916, 2006.
 7. Arolas JL, Castillo V, Bronsoms S, Aviles FX, and Ventura S. Designing out disulfide bonds of leech carboxypeptidase inhibitor: implications for its folding, stability and function. *J Mol Biol* 392: 529–546, 2009.
 8. Arolas JL, Lorenzo J, Rovira A, Vendrell J, Aviles FX, and Ventura S. Secondary binding site of the potato carboxypeptidase inhibitor. Contribution to its structure, folding, and biological properties. *Biochemistry* 43: 7973–7982, 2004.
 9. Arolas JL and Ventura S. Protease inhibitors as models for the study of oxidative folding. *Antioxid Redox Signal* 14: 97–112, 2011.
 10. Aslund F and Beckwith J. Bridge over troubled waters: sensing stress by disulfide bond formation. *Cell* 96: 751–753, 1999.
 11. Belli M, Ramazzotti M, and Chiti F. Prediction of amyloid aggregation *in vivo*. *EMBO Rep* 12: 657–663, 2011.
 12. Bulaj G and Goldenberg DP. Phi-values for BPTI folding intermediates and implications for transition state analysis. *Nat Struct Biol* 8: 326–330, 2001.
 13. Calloni G, Zoffoli S, Stefani M, Dobson CM, and Chiti F. Investigating the effects of mutations on protein aggregation in the cell. *J Biol Chem* 280: 10607–10613, 2005.
 14. Castillo V, Grana-Montes R, Sabate R, and Ventura S. Prediction of the aggregation propensity of proteins from the primary sequence: aggregation properties of proteomes. *Biotechnol J* 6: 674–685, 2011.
 15. Chandonia JM, Hon G, Walker NS, Lo Conte L, Koehl P, Levitt M, and Brenner SE. The ASTRAL Compendium in 2004. *Nucleic Acids Res* 32: D189–D192, 2004.
 16. Chandonia JM, Walker NS, Lo Conte L, Koehl P, Levitt M, and Brenner SE. ASTRAL compendium enhancements. *Nucleic Acids Res* 30: 260–263, 2002.
 17. Chang JY. The disulfide folding pathway of tick anticoagulant peptide (TAP), a Kunitz-type inhibitor structurally homologous to BPTI. *Biochemistry* 35: 11702–11709, 1996.
 18. Chang JY. A two-stage mechanism for the reductive unfolding of disulfide-containing proteins. *J Biol Chem* 272: 69–75, 1997.
 19. Chang JY. Evidence for the underlying cause of diversity of the disulfide folding pathway. *Biochemistry* 43: 4522–4529, 2004.
 20. Chang JY. Diversity of folding pathways and folding models of disulfide proteins. *Antioxid Redox Signal* 10: 171–177, 2008.
 21. Chang JY and Li L. Divergent folding pathways of two homologous proteins, BPTI and tick anticoagulant peptide: compartmentalization of folding intermediates and identification of kinetic traps. *Arch Biochem Biophys* 437: 85–95, 2005.
 22. Chang JY, Schindler P, and Chatrenet B. The disulfide structures of scrambled hirudins. *J Biol Chem* 270: 11992–11997, 1995.
 23. Chatrenet B and Chang JY. The disulfide folding pathway of hirudin elucidated by stop/go folding experiments. *J Biol Chem* 268: 20988–20996, 1993.
 24. Cheek S, Krishna SS, and Grishin NV. Structural classification of small, disulfide-rich protein domains. *J Mol Biol* 359: 215–237, 2006.
 25. Chiti F and Dobson CM. Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* 75: 333–366, 2006.
 26. Chiti F, Stefani M, Taddei N, Ramponi G, and Dobson CM. Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* 424: 805–808, 2003.
 27. Conchillo-Solé O, de Groot NS, Avilés FX, Vendrell J, Daura X, and Ventura S. AGGRESCAN: a server for the prediction and evaluation of “hot spots” of aggregation in polypeptides. *BMC Bioinformatics* 8: 65, 2007.
 28. This reference has been deleted.
 29. Covalada G, del Rivero MA, Chavez MA, Aviles FX, and Reverter D. Crystal structure of novel metallo-carboxypeptidase inhibitor from marine mollusk *Nerita versicolor* in complex with human carboxypeptidase A4. *J Biol Chem* 287: 9250–9258, 2012.
 30. Creighton TE. Intermediates in the refolding of reduced ribonuclease A. *J Mol Biol* 129: 411–431, 1979.
 31. Creighton TE. Protein folding coupled to disulphide bond formation. *Biol Chem* 378: 731–744, 1997.
 32. Creighton TE, Hillson DA, and Freedman RB. Catalysis by protein-disulphide isomerase of the unfolding and refolding of proteins with disulphide bonds. *J Mol Biol* 142: 43–62, 1980.
 33. de Groot N, Pallares I, Aviles F, Vendrell J, and Ventura S. Prediction of “hot spots” of aggregation in disease-linked polypeptides. *BMC Struct Biol* 5: 18, 2005.
 34. de Groot NS, Castillo V, Grana-Montes R, and Ventura S. AGGRESCAN: method, application, and perspectives for drug design. *Methods Mol Biol* 819: 199–220, 2012.
 35. de Groot NS and Ventura S. Protein aggregation profile of the bacterial cytosol. *PLoS One* 5: e9383, 2010.
 36. Denoncin K and Collet JF. Disulfide bond formation in the bacterial periplasm: major achievements and challenges ahead. *Antioxid Redox Signal* 19: 63–71, 2013.
 37. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, Ausio J, Nissen MS, Reeves R, Kang C, Kissinger CR, Bailey RW, Griswold MD, Chiu W, Garner EC, and Obradovic Z. Intrinsically disordered protein. *J Mol Graph Model* 19: 26–59, 2001.
 38. Espargaro A, Castillo V, de Groot NS, and Ventura S. The *in vivo* and *in vitro* aggregation properties of globular proteins correlate with their conformational stability: the SH3 case. *J Mol Biol* 378: 1116–1131, 2008.
 39. Fersht AR. The sixth Datta Lecture. Protein folding and stability: the pathway of folding of barnase. *FEBS Lett* 325: 5–16, 1993.
 40. Fischer M, Horn S, Belkacemi A, Kojer K, Petrunaro C, Habich M, Ali M, Kuttner V, Bien M, Kauff F, Dengjel J, Herrmann JM, and Riemer J. Protein import and oxidative folding in the mitochondrial intermembrane space of intact mammalian cells. *Mol Biol Cell* 24: 2160–2170, 2013.
 41. Frand AR, Cuozzo JW, and Kaiser CA. Pathways for protein disulphide bond formation. *Trends Cell Biol* 10: 203–210, 2000.
 - 41a. Fuller E, Green BR, Catlin P, Buczek O, Nielsen JS, Olivera BM, Bulaj G. Oxidative folding of conotoxins sharing an identical disulfide bridging framework. *FEBS J* 272: 1727–1738, 2005.
 42. Galzitskaya OV, Garbuzynskiy SO, and Lobanov MY. FoldUnfold: web server for the prediction of disordered

- regions in protein chain. *Bioinformatics* 22: 2948–2949, 2006.
43. Grana-Montes R, de Groot NS, Castillo V, Sancho J, Velazquez-Campoy A, and Ventura S. Contribution of disulfide bonds to stability, folding and amyloid fibril formation: the PI3-SH3 domain case. *Antioxid Redox Signal* 16: 1–15, 2012.
 44. Guerois R, Nielsen JE, and Serrano L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol* 320: 369–387, 2002.
 45. Guijarro JI, Sunde M, Jones JA, Campbell ID, and Dobson CM. Amyloid fibril formation by an SH3 domain. *Proc Natl Acad Sci U S A* 95: 4224–4228, 1998.
 46. Guo ZY, Qiao ZS, and Feng YM. The *in vitro* oxidative folding of the insulin superfamily. *Antioxid Redox Signal* 10: 127–139, 2008.
 47. Hegyi H and Tompa P. Intrinsically disordered proteins display no preference for chaperone binding *in vivo*. *PLoS Comput Biol* 4: e1000017, 2008.
 48. Hogg PJ. Disulfide bonds as switches for protein function. *Trends Biochem Sci* 28: 210–214, 2003.
 49. Ivanova MI, Thompson MJ, and Eisenberg D. A systematic screen of beta(2)-microglobulin and insulin for amyloid-like segments. *Proc Natl Acad Sci U S A* 103: 4079–4082, 2006.
 50. Jamroz M, Kolinski A, and Kmiecik S. CABS-flex: server for fast simulation of protein structure fluctuations. *Nucleic Acids Res* 41: W427–W431, 2013.
 51. Jansens A, van Duijn E, and Braakman I. Coordinated nonvectorial folding in a newly synthesized multidomain protein. *Science* 298: 2401–2403, 2002.
 52. Kallberg Y, Gustafsson M, Persson B, Thyberg J, and Johansson J. Prediction of amyloid fibril-forming proteins. *J Biol Chem* 276: 12945–12950, 2001.
 53. Li YJ, Rothwarf DM, and Scheraga HA. Mechanism of reductive protein unfolding. *Nat Struct Biol* 2: 489–494, 1995.
 54. Linding R, Schymkowitz J, Rousseau F, Diella F, and Serrano L. A comparative study of the relationship between protein structure and beta-aggregation in globular and intrinsically disordered proteins. *J Mol Biol* 342: 345–353, 2004.
 55. Lindorff-Larsen K, Piana S, Dror RO, and Shaw DE. How fast-folding proteins fold. *Science* 334: 517–520, 2011.
 56. Lobanov MY, Suvorina MY, Dovidchenko NV, Sokolovskiy IV, Surin AK, and Galzitskaya OV. A novel web server predicts amino acid residue protection against hydrogen-deuterium exchange. *Bioinformatics* 29: 1375–1381, 2013.
 57. Luo X, Jiang Q, Song G, Liu YL, Xu ZG, and Guo ZY. Efficient oxidative folding and site-specific labeling of human hepcidin to study its interaction with receptor ferroportin. *FEBS J* 279: 3166–3175, 2012.
 58. Mamathambika BS and Bardwell JC. Disulfide-linked protein folding pathways. *Annu Rev Cell Dev Biol* 24: 211–235, 2008.
 59. Marino SM and Gladyshev VN. Cysteine function governs its conservation and degeneration and restricts its utilization on protein surfaces. *J Mol Biol* 404: 902–916, 2010.
 60. Maurer-Stroh S, Debulpaep M, Kuemmerer N, Lopez de la Paz M, Martins IC, Reumers J, Morris KL, Copland A, Serpell L, Serrano L, Schymkowitz JW, and Rousseau F. Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nat Methods* 7: 237–242.
 61. This reference has been deleted.
 62. Monsellier E and Chiti F. Prevention of amyloid-like aggregation as a driving force of protein evolution. *EMBO Rep* 8: 737–742, 2007.
 63. Mossuto MF, Bolognesi B, Guixer B, Dhulesia A, Agostini F, Kumita JR, Tartaglia GG, Dumoulin M, Dobson CM, and Salvatella X. Disulfide bonds reduce the toxicity of the amyloid fibrils formed by an extracellular protein. *Angew Chem Int Ed Engl* 50: 7048–7051, 2011.
 64. Mu Y, Nordenskiöld L, and Tam JP. Folding, misfolding, and amyloid protofibril formation of WW domain FBP28. *Biophys J* 90: 3983–3992, 2006.
 65. Neklesa TK and Crews CM. Chemical biology: greasy tags for protein removal. *Nature* 487: 308–309, 2012.
 66. Pace CN, Grimsley GR, Thomson JA, and Barnett BJ. Conformational stability and activity of ribonuclease T1 with zero, one, and two intact disulfide bonds. *J Biol Chem* 263: 11820–11825, 1988.
 67. Pantoja-Uceda D, Arolas JL, Aviles FX, Santoro J, Ventura S, and Sommerhoff CP. Deciphering the structural basis that guides the oxidative folding of leech-derived tryptase inhibitor. *J Biol Chem* 284: 35612–35620, 2009.
 68. This reference has been deleted.
 69. Polticelli F, Raybaudi-Massilia G, and Ascenzi P. Structural determinants of mini-protein stability. *Biochem Mol Biol Educ* 29: 16–20, 2001.
 70. Prilusky J, Felder CE, Zeev-Ben-Mordehai T, Rydberg EH, Man O, Beckmann JS, Silman I, and Sussman JL. FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* 21: 3435–3438, 2005.
 71. This reference has been deleted.
 72. Rezaei-Ghaleh N, Blackledge M, and Zweckstetter M. Intrinsically disordered proteins: from sequence and conformational properties toward drug discovery. *Chembiochem* 13: 930–950, 2012.
 73. Rousseau F, Schymkowitz J, and Serrano L. Protein aggregation and amyloidosis: confusion of the kinds? *Curr Opin Struct Biol* 16: 118–126, 2006.
 74. Safavi-Hemami H, Gorasia DG, Steiner AM, Williamson NA, Karas JA, Gajewiak J, Olivera BM, Bulaj G, and Purcell AW. Modulation of conotoxin structure and function is achieved through a multienzyme complex in the venom glands of cone snails. *J Biol Chem* 287: 34288–34303, 2012.
 75. Salinas G, Pellizza L, Margenat M, Flo M, and Fernandez C. Tuned *Escherichia coli* as a host for the expression of disulfide-rich proteins. *Biotechnol J* 6: 686–699, 2011.
 76. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, and Serrano L. The FoldX web server: an online force field. *Nucleic Acids Res* 33: W382–W388, 2005.
 77. Sickmeier M, Hamilton JA, LeGall T, Vacic V, Cortese MS, Santos A, Szabo B, Tompa P, Chen J, Uversky VN, Obradovic Z, and Dunker AK. DisProt: the Database of Disordered Proteins. *Nucleic Acids Res* 35: D786–D793, 2007.
 78. Siekierska A, De Baets G, Reumers J, Gallardo R, Rudyak S, Broersen K, Couceiro J, Van Durme J, Schymkowitz J, and Rousseau F. α -Galactosidase aggregation is a determinant of pharmacological chaperone efficacy on Fabry disease mutants. *J Biol Chem* 287: 28386–28397, 2012.
 79. Tzotzos S and Doig AJ. Amyloidogenic sequences in native protein structures. *Protein Sci* 19: 327–348, 2010.
 80. Uversky VN, Gillespie JR, and Fink AL. Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins* 41: 415–427, 2000.

81. Van Durme J, Maurer-Stroh S, Gallardo R, Wilkinson H, Rousseau F, and Schymkowitz J. Accurate prediction of DnaK-peptide binding *via* homology modelling and experimental data. *PLoS Comput Biol* 5: e1000475, 2009.
82. Ventura S, Zurdo J, Narayanan S, Parreno M, Mangues R, Reif B, Chiti F, Giannoni E, Dobson CM, Aviles FX, and Serrano L. Short amino acid stretches can mediate amyloid formation in globular proteins: the Src homology 3 (SH3) case. *Proc Natl Acad Sci U S A* 101: 7258–7263, 2004.
83. Wedemeyer WJ, Welker E, Narayan M, and Scheraga HA. Disulfide bonds and protein folding. *Biochemistry* 39: 4207–4216, 2000.
84. Weissman JS and Kim PS. Reexamination of the folding of BPTI: predominance of native intermediates. *Science* 253: 1386–1393, 1991.
85. Yang ZR, Thomson R, McNeil P, and Esnouf RM. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics* 21: 3369–3376, 2005.

Address correspondence to:

Dr. Salvador Ventura
 Departament de Bioquímica i Biologia Molecular
 Institut de Biotecnologia i Biomedicina
 Universitat Autònoma de Barcelona
 Barcelona 08193
 Spain

E-mail: salvador.ventura@uab.es

Date of first submission to ARS Central, July 17, 2013; date of final revised submission, March 6, 2014; date of acceptance, March 16, 2014.

Abbreviations Used

APRs = aggregation-prone regions
 ATR-FTIR = attenuated total reflectance–Fourier transform infrared spectroscopy
 BPTI = bovine pancreatic trypsin inhibitor
 CR = Congo red
 DRDs = disulfide-rich domains
 DTT = DL-dithiothreitol
 ER = endoplasmic reticulum
 HFIP = 1,1,1,3,3,3-hexafluoro-2-propanol
 IDPs = intrinsically disordered proteins
 IGF-1 = insulin-like growth factor-1
 LDL = low density lipoprotein
 LDTI = leech-derived trypsin inhibitor
 NvCI = *Nerita versicolor* carboxypeptidase inhibitor
 RONN = regional order neural network
 TAP = tick anticoagulant peptide
 TCEP = tris(2-carboxyethyl)phosphine
 TEM = transmission electron microscopy
 TFA = trifluoroacetic acid
 Th-T = Thioflavin-T
 wt = wild-type