Protein secondary structure determines the temporal relationship between folding and disulfide formation

Philip J Robinson[1], Shingo Kanemura[2], Xiaofei Cao[1], Neil J Bulleid[1*]

[1] Institute of Molecular, Cell and Systems Biology, College of Medical Veterinary and Life Sciences, Davidson Building, University of Glasgow, Glasgow, G12 8QQ, UK.

[2] Kwansei Gakuin University, 2-1 Gakuen, Sanda, Hyogo 669-1337 Japan.

ORCID: PJP (0000-0003-2576-5010), NJB (0000-0002-9839-5279)

*corresponding author: +44 141 330 3870, neil.bulleid@glasgow.ac.uk.

Classification: Biological Sciences, Biochemistry

Abstract

How and when disulfides form in proteins during their folding is a fundamental question in cell biology. Two models describe the relationship between disulfide formation and folding, the folded precursor model, in which formation of nascent structure occurs prior to the disulfides and the quasi-stochastic model where disulfides form prior to complete domain folding. Here we investigate oxidative folding within a cellular milieu of three structurally diverse substrates in order to understand the folding mechanisms required to achieve correct cysteine coupling. We use a eukaryotic translation system in which we can manipulate the redox conditions and produce stalled translation intermediates representative of different stages of translocation. We identify different disulfide bonded isomers by non-reducing SDS-PAGE. Using this approach, we determined whether each substrate followed a folding driven or disulfide driven mechanism. Our results demonstrate that the folding model is substrate-dependent with disulfides forming prior to complete domain folding in a cysteine-rich domain lacking secondary structure, whereas disulfide formation was absent in proteins with defined structural elements. In addition, we demonstrate the presence and rearrangement of non-native disulfides specifically in substrates following the quasi-stochastic model. These findings demonstrate why non-native disulfides are prevented from forming in proteins with well-defined secondary structure.

Key words

Protein folding, nascent chain, endoplasmic reticulum, disulfide formation

Significance statement

A third of human proteins contain structural elements called disulfide bonds that are often crucial for stability and function. Disulfides form between cysteines in the specialised environment of the endoplasmic reticulum (ER), during the complex process of protein folding. Many proteins contain multiple cysteines that can potentially form correct or incorrect cysteine pairings. To investigate how correct disulfide pairs are formed in a biological context, we developed an experimental approach to assess disulfide formation and rearrangement as proteins enter the ER. We found that a disulfide-dense protein domain with atypical secondary structure undergoes disulfide orchestrated folding as it enters the ER and is prone to incorrect disulfide formation. In contrast, proteins with defined secondary structure form folding dependent, native disulfides. These findings show how different mechanisms of disulfide formation can be rationalised from structural features of the folding domains.

\body

Introduction

A central question in protein folding is how the correct disulfide pattern is achieved for proteins that contain multiple cysteines. Intramolecular disulfide bonds consist of covalent cross-links between cysteines within a polypeptide (1), and their formation imposes conformational restraints that influence protein folding, stability and function. Disulfides form through an oxidative reaction between paired cysteines in the presence of an appropriate catalyst (2). In eukaryotic cells, the endoplasmic reticulum (ER) is the main site for disulfide formation, it provides a favourable redox environment and catalysts in the form of resident PDI (Protein Disulfide Isomerase) family members (3). When polypeptides contain more than two cysteines, correct pairing is required to form the native disulfide pattern; the spatial positioning of cysteine sidechains is required for this correct pairing and so disulfide formation is closely linked to the process of conformational folding.

Two models have been proposed that broadly explain the relationship between disulfide formation and conformational folding:(i) the folded precursor mechanism, in which nascent structure forms first, which then positions paired cysteines favourably to form a disulfide and (ii) the quasi-stochastic model, in which cysteines in an unfolded precursor pair more randomly, and the resulting disulfide influences further folding (4). Recent studies have identified substrates that favour the folded precursor mechanism (5-7), while other studies have identified contributions from both mechanisms at different stages of the folding process (8). Incorrect (non-native) disulfides can also form during folding, which are subsequently reduced and re-arranged (9, 10). In the ER, this reduction step is performed by specific PDI family members (11, 12) using reducing equivalents that originate in the cytosol (13). Evidence shows that non-native disulfides can act as important intermediates in the native folding of certain proteins (10).

Another important factor to consider when investigating folding and disulfide formation in the cell is the influence of the vectorial translation mechanism. During the early stages of translation, nascent polypeptides enter the ER through the Sec-translocon (14). The space constraints of the ribosome/Sec-complex limit structure formation until enough polypeptide has entered the ER-lumen (15). Every amino acid that is then added to the polypeptide expands the repertoire of intra-molecular interactions and the diversity of potential folds that might form. Both native and non-native disulfides can form at this co-translational stage (16, 17). The formation of native disulfides is dictated by the exposure of the nascent polypeptide to the ER, the folding of the nascent polypeptide and the accessibility of PDI family members (18). Non-native disulfides may also form owing to the partial exposure of nascent polypeptides to the ER, which prevents the completion of folding, leading to incorrect cysteine coupling.

The question of how proteins with multiple cysteines achieve the correct disulfide pattern is a difficult one to address due to the structural diversity of disulfide-containing proteins and the technical barriers that hinder folding assays performed in a biologically relevant context. In an earlier study, we devised an approach to investigate the folding of a protein containing a single disulfide as it emerges into the ER lumen (7). Here, we expand on this approach to study nascent polypeptides with diverse structures and with more -complex disulfide patterns to investigate the folding mechanisms by which correct disulfide bonding is achieved at the co-translational stage. To do so, we used a eukaryotic translation system, in which we control the ER-specific reductive pathway, in

order to characterise the rearrangement of non-native disulfides in translation intermediates. Both the timing of disulfide formation relative to domain exposure and the presence of non-native disulfides indicate which folding mechanism each substrate follows. These mechanisms can then be determined from the structures of the substrates studied to shed light on the process of disulfide bonding and folding in a biologically relevant context.

Results

Experimental design

The three proteins used in this study, β2-microglobulin (β2M), prolactin and the disintegrin domain of ADAM10, were chosen for their diversity of secondary structure and disulfide bonding (Fig. 1A). We added a C-terminal extension to each protein (Fig. 1B) to act as a tether to the ribosome and to enable the folding domain to exit from the ribosome before the completion of translation (Fig 1 Ci). DNA templates were synthesised that encode specific polypeptide lengths that lack stop codons. The templates were transcribed and the resulting mRNA was used to programme translation reactions that produce stalled nascent chains of differing lengths. Translations were performed using a rabbit reticulocyte lysate supplemented with semi-permeabilised (SP) cells to study folding in the ER (19) or without SP cells to represent folding in the cytosol. All samples were treated with NEM on completion to irreversibly modify thiols and freeze the disulfide status of the samples for downstream processing.

We adjusted the redox status of the translation reactions by adding specific components to the lysate before translation to influence correct disulfide formation. In all, we used three different lysates in our experiments. (1) An oxidising lysate that contained no additional components; this lysate allows disulfides to form in proteins synthesised in the absence or presence of SP cells. (2) A redox-balanced lysate supplemented with G6Pi to drive G6PD and TrxR1 activity (13) (Fig.1 Cii); this lysate is sufficiently reducing to prevent disulfide formation in proteins synthesised in the absence of SP cells but allows disulfide formation in translocated proteins when SP-cells are present (13). (3) And, a reducing lysate supplemented with DTT, in which proteins synthesised are unable to form disulfides in the absence or presence of SP-cells. The redox status of the lysate can also influence non-native disulfide formation in proteins translocated into the ER. This is because the cytosolic reducing pathway is the source of reducing equivalents for an ER-specific reductive pathway, which is required for disulfide rearrangement (Fig. 1 Ciii). Thus by comparing disulfide formation in an oxidising or redox-balanced lysate we can assay how the cytosolic reducing pathway influences the folding outcome of proteins translocated into the ER.

Stochastic disulfide formation is absent in β2M

β2M is a small, β-sheet rich protein that contains two cysteines and a single disulfide (Fig. 1Ai). Disulfide formation in β2M can be detected as a faster migrating species on non-reducing SDS-PAGE (7). The organisation of the β2M construct is illustrated in Fig. 2A. Previously, we used β2M to investigate folding events during ER entry and found that disulfide formation is a folding-driven process that requires the folding domain to be fully exposed to the ER (Fig. 2Bi) (7). We performed these initial experiments in a redox-balanced lysate; as such the ER-specific reductive pathway might have prevented stochastic disulfide formation prior to domain folding. To investigate this possibility we translated the 175 intermediate of β2M with SP cells in oxidising, redox-balanced and reducing

lysates and assayed disulfide formation by non-reducing SDS-PAGE (Fig. 2Bii). At this intermediate length, both cysteine residues will have exited the Sec-complex while part of the folding domain remains within. Each lane shows two products representing pre-protein (untargeted protein associated with SP cells) and mature protein (successfully targeted protein that has been translocated and undergone signal peptide cleavage). For the stalled samples, there was no change in gel-mobility when comparing oxidising, redox-balanced and reducing lysates, which reveals an absence of disulfide formation under each condition even though both cysteine residues were exposed to the ER lumen. The result in the oxidising lysate shows that the lack of disulfide formation is not dependent on the ER-specific reductive pathway. Only after the folding domain exited the Sec-complex, following nascent chain release did the mature protein shift in the oxidising and redox-balanced samples, showing that disulfide formation is absent until the complete folding domain is translocated.

We also synthesised β2M in the oxidising lysate in the absence of SP-cells to test if substrates form disulfides independently of ER factors. For these experiments, we used a β2M sequence that did not contain an amino-terminal signal sequence in case this affected protein folding. Three intermediate lengths were chosen based on estimates assuming that ~40 amino acids span the length of the ribosome tunnel, 2/3 of the amino acids required to span the entire ribosome/Sec-complex (7) (20). The schematics in Fig. 2C illustrate the expected cysteine exposure in each case. Following translation, stalled ribosome-nascent chain complexes (RNCs) were either (i) isolated by ultracentrifugation through a sucrose cushion for direct analysis on SDS-PAGE or (ii) RNaseA treated to initiate release of the nascent polypeptide, for subsequent immunoisolation and SDS-PAGE analysis. At each intermediate length the stalled samples translated in the oxidising lysate have the same gel-mobility as those translated in the reducing or redox-balanced lysates, with no detectable disulfide-associated shift evident (Fig. 2C). Under these conditions both of the cysteines of the 141 and 165 intermediates were exposed to the oxidising lysate and yet the disulfide failed to form. It was only after the nascent chain was released from the ribosome that the shift associated with disulfide formation became visible in samples translated in oxidising lysate. The oxidising lysate is, therefore, sufficient to initiate β2M disulfide formation but only following the release of the protein from the ribosome.

These results with β2M thus favour a folding-driven mechanism for disulfide formation because disulfides failed to form until the entire folding domain exited the ribosomal tunnel. The prevention of premature disulfide formation does not appear to require protein disulfide isomerase as the requirement for tunnel exit was observed in the absence of an added source of ER (Fig.2C). This suggests that disulfides form only once specific structural parameters have been achieved.

Disulfide formation in prolactin favours a structured precursor-folding model

Next, we studied disulfide formation in bovine-prolactin, which has an α-helical structure and 3 native disulfide bonds (21). A schematic of the prolactin construct used in this study is illustrated in Fig. 3A. To determine the intermediate length required for the prolactin domain to be fully exposed to the ER lumen, we monitored N-linked glycosylation at position 230. Following translation of prolactin intermediates, cell pellets were isolated by centrifugation, resuspended in sample buffer and analysed by reducing SDS-PAGE. Glycosylation was detected from an intermediate length of 305 amino acids onwards (Fig. 3Bi). We calculated that 75 amino acids (305-230) are required to span

the distance between the P-site of the ribosome and the active site of the oligosaccharyl transferase (OST). Taking into account the ~12 residues it takes to reach from the exit of the Sec-complex to the OST, we estimate that 63 residues span the ribosome/Sec-complex (Fig. 3B ii), the same value observed with the β2M construct (7). Based on this result, we predicted the length of polypeptide (amino acids from the start codon) exposed to the ER for each intermediate (Fig 3. Biii). For the remaining experiments with the prolactin construct, we removed the glycosylation site to simplify SDS-PAGE analysis.

To assay for disulfide formation in prolactin nascent chains, we monitored the formation of a long-range disulfide (C88-C204), which increases the polypeptide chains' gel mobility relative to their reduced equivalent following non-reducing SDS-PAGE. This mobility shift was illustrated by comparing the mature form of wild type prolactin translated in the presence of SP cells in either a redox-balanced or a reducing lysate (Fig. 3C). The two short-range disulfides do not influence SDS-PAGE mobility (21) and so are not detected in these assays. There is also the potential for non-native disulfides to form, these non-native disulfides would be long range and likely to influence gel mobility. To determine if long-range disulfides do form in prolactin nascent chains we translated samples in a redox-balanced lysate and compared the resulting gel mobility by reducing (+DTT) or non-reducing (-DTT) SDS-PAGE (Fig. 3Di). Of the three intermediate lengths tested (275, 310 and 330), the migration of the signal-peptide cleaved polypeptide is the same under non-reducing or reducing gel conditions (dashed line), revealing a lack of disulfide formation in each case. We know that the first four cysteines and ~90 % of the prolactin sequence of the 275 intermediate is ER-exposed, whilst the entire prolactin sequence of the 310 and 330 intermediates is ER-exposed. Therefore, despite the presence of free thiols available for disulfide formation, the protein remains reduced. Following treatment with RNaseA, which releases the nascent chain from the ribosome, a change in gel mobility was detected for each intermediate when separated under non-reducing conditions (Fig. 3Dii). This result demonstrates that the formation of the long-range disulfide depends on release of the nascent chain from the ribosome; indicating that the tethering of the nascent chain prevents disulfide formation.

We next performed translations in an oxidising lysate to determine if the ER-specific reductive pathway is responsible for maintaining the prolactin intermediates in a reduced state. The translation of the 310 intermediate under these conditions produced the same gel-mobility as that produced on its translation in the redox-balanced or reducing lysate, with the disulfide associated shift present only after the nascent chain was released by RNaseA treatment (Fig 3Ei). This result demonstrates that the ER reductive pathway is not actively reducing the prolactin intermediates, providing further support for the theory that the polypeptide's physical attachment to the ribosome/Sec-complex prevents disulfides from forming. We next tested whether ribosome attachment prevents disulfides from forming when translation intermediates are exposed to an oxidising cytosol. To do so, we performed translations in the absence of SP cells (Fig 3Eii) and separated the translation products by non-reducing SDS-PAGE. For this assay, we translated a prolactin construct (residues 31-310) that did not contain the N-terminal signal sequence and was of sufficient length (280 aa) to ensure the folding domain exited the ribosomal tunnel. When this intermediate was translated in an oxidising, redox-balanced or reducing lysate, disulfide formation was absent while the nascent polypeptide was ribosome attached. Disulfides formed only in the oxidising lysate when the nascent chain was released from the ribosome. This demonstrates that the dependence of disulfide formation on ribosome release does not require ER-specific factors and is

caused by the physical attachment of the polypeptide to the ribosome, which prevents the completion of folding.

Mechanistically, the absence of both native and non-native disulfide formation at the co-translational stage, when cysteines are exposed to the oxidising milieu, shows that stochastic disulfide formation is absent at this point. This indicates that the folding of prolactin does not occur via the quasi-stochastic mechanism of folding, in which disulfides form first in unstructured, folding precursors before conformational folding takes place. Instead, the dependence of disulfide formation on ribosomal release is consistent with a folding-driven mechanism in which release provides the freedom that is necessary for conformational folding to occur, which then drives disulfide formation. These results thus favour the structured precursor model of oxidative folding.

Mixed disulfide populations in partially exposed disintegrin intermediates indicate a disulfide driven folding mechanism

Our results with β2M and prolactin both favour the structured precursor model of oxidative folding. However, both proteins have stable secondary structures in which the disulfide density is low. Could a different and contrasting protein structure fold via an alternative mechanism? To investigate this, we studied the disintegrin domain of ADAM10, which has a dense disulfide bonding pattern and little defined secondary structure (Fig.1 A iii) (22, 23). The dense disulfide content of this structure is predicted to favour the quasi-stochastic model of folding, in which disulfides form first and drive the later stages of folding (4). The organisation of the disintegrin construct and its pattern of disulfide bonding is illustrated in Fig 4A. To determine the intermediate length required for the disintegrin domain to be fully exposed to the ER-lumen, we monitored glycosylation at position 125, using the same procedure as described for the prolactin construct. Glycosylation occurred from a length of 200 amino acids onwards, which corresponds to the 63 amino acids of extension sequence required to span the length of the ribosome/Sec-channel. This is the same length observed for β2M (7) and prolactin (Fig. 3). The predicted ER exposure, in terms of polypeptide length and number of cysteines exposed, is shown in Fig. 4Bii. For the remaining experiments we removed the glycosylation site by mutagenesis to simplify SDS-PAGE analysis.

To determine whether disulfide bonding occurs in the disintegrin domain construct, we expressed a range (146-210 amino acids) of intermediates to achieve increasing levels of ER-exposure (Fig. 4C). When translated in a reducing lysate, we observed two products in each lane that corresponded to reduced pre-protein and reduced mature-protein (Fig 4.Ci). When translated in a redox-balanced lysate, the shorter intermediates ran as ~5 distinct species (Fig. 4Cii) (lanes 1-3). By comparing these to the reduced sample, we concluded that the slowest migrating species corresponds to un-translocated, reduced pre-protein and that the four faster migrating species represent translocated species with different disulfide bond patterns. The clear separation of the species indicates that the disulfides produce long-range constraints that contribute to changes in SDS-PAGE mobility. The formation of disulfides despite the disintegrin folding domain being only partially exposed to the ER contrasts with the folding of both β2M and prolactin. This early disulfide formation and the mixed population of species that form support a stochastic mechanism of cysteine coupling. When the length of intermediate was increased the abundance of each translocated species changed and once fully exposed, a faster migrating species became the major product observed (Fig 4Cii, lanes 4-6). This result indicates that the final disulfide pattern forms only once the entire folding domain exits

the Sec-complex and that disulfide rearrangements occur concomitant with nascent chain translocation.

Stochastic cysteine coupling is likely to result in non-native disulfides that require subsequent rearrangement. Disulfide rearrangement mechanisms depend on a reductive pathway, which is active in the redox-balanced lysate. When we translated the same intermediates in an oxidising lysate (Fig. 4Ciii), we observed a distinctly different pattern. The absence of products that correspond to reduced protein shows that disulfide formation is ubiquitous in both the pre-protein and mature protein. For each sample, the fast-migrating species corresponded to intermediates with intra-chain disulfides, and the slower migrating species indicates the presence of inter-chain disulfides. The species that formed in the oxidising lysate showed more diffuse patterns on SDS-PAGE than did those synthesised in the redox-balanced lysate, indicating the presence of a heterogeneous mixture of disulfide-bonded polypeptides. As the intermediates increased in length, a less heterogeneous mix of species became evident (Fig.4 C iii lanes 5-6) with a similar migration pattern to the equivalent intermediates translated in the redox-balanced lysate. Together, these observations show that the reductive pathway has a direct impact on the disulfide-bonded species that form in partially translocated intermediates, indicating that disulfide rearrangements can occur while the nascent chain is attached to the ribosome.

ER-specific disulfide rearrangements occur in disintegrin nascent chains

In the disintegrin domain experiments, about 50 % of the protein is ER targeted, as assessed by the extent of signal-peptide cleavage in reducing samples (Fig. 4Ci). In addition, the products translated in the redox-balanced or oxidising lysate run as multiple species and it is therefore unclear which are untargeted and which are translocated. In order to use this system to investigate ER-specific rearrangements we ideally want to be able to isolate translation products that are specifically targeted to the ER. To do this we utilised the neo-epitope tag (NE) placed between the signal peptide and the disintegrin domain. This epitope is only recognised if the N-terminal signal peptide is cleaved, allowing the antibody-based isolation of mature protein without contamination from the pre-protein (24). Translation products were isolated in this way using a V5-antibody (which recognises the total translation product) or an antibody that recognises the neo-epitope (NE); products were then compared by non-reducing SDS-PAGE (Fig. 5A). We used intermediate lengths 146 and 210 to represent partial- and fully-translocated domains, synthesised in reducing, redox-balanced or oxidising lysates. Intermediates synthesised in the reducing lysate ran as two products when isolated with the V5 antibody (lanes 1 and 7), corresponding to pre-protein and mature protein, and as a single product when isolated with the NE antibody, corresponding to mature protein (lane 4 and 10). Results with the 146 samples showed that of the 5 species present following translation in the redox-balanced lysate (lane 2), the 4 species beneath the pre-protein are ER-specific (lane 5). For samples translated in the oxidising lysate, the fast-migrating smear visible in the V5-isolated sample (lane 3) was also visible in the NE-isolated sample (lane 6) confirming that this mix consists of ER-specific disulfide bonded species. For the 210 samples, the fast-migrating species share the same migration pattern (lanes 8-9 and 11-12) indicating that once the folding domain is exposed, disulfide rearrangements can occur.

The 146 intermediate shows some striking differences in its pattern of disulfide-bonded species, relative to the other intermediates, when synthesised in either an oxidising or a redox-balanced

lysate. This provides a useful tool to assay disulfide rearrangements. To determine whether disulfide rearrangement can occur following nascent chain synthesis in the oxidising lysate, we restored the cytosolic reductive pathway post-translationally. To do so G6Pi was added after nascent chain synthesis, and samples were then taken at specific time points and separated by non-reducing SDS-PAGE (Fig. 5B). Translation products synthesised in a reducing, redox-balanced or oxidising lysate (prior to the addition of G6Pi) were separated on the same gel to compare mobility (lanes 1-3). Following the addition of G6Pi, the mobility of the translation products changed from resembling the pattern prior to addition (lane 4, 1 min) to being equivalent to the pattern formed in the redox-balanced lysate (lanes 5,6 10-30 min). The NE-isolated samples show the same pattern as the V5-isolated samples but without the presence of reduced pre-protein. These results show that the reactivation of the reductive pathway following translation induces post-translational disulfide rearrangements in the nascent chain. Finally, we confirmed that thioredoxin reductase is the source of reducing equivalents for the disulfide rearrangements observed by performing G6Pi recovery experiments in the presence of the thioredoxin inhibitor auranofin. Treatment with auranofin prevents the change from the oxidised to the redox balanced gel migration pattern (Fig. 5.D compare lanes 5 and 7), which shows that cytosolic thioredoxin reductase is the source of reducing equivalents for these co-translational disulfide rearrangements.

Discussion

In this study, we used an *in vitro* translation system to assess nascent chain disulfide formation and rearrangement, in three proteins with diverse structures and disulfide bond patterns- β2M, prolactin and the disintegrin domain of ADAM10. Our results indicate that disulfide formation occurs via two mechanisms that depend on a protein's secondary structure. In substrates with a regular secondary structure, conformational folding drives disulfide formation. By contrast in substrates with an atypical secondary structure, the folding of disulfide-rich domains occurs through a disulfide driven process.

For β2M, disulfide formation depends on the protein's folding domain being fully exposed to the ER-lumen. For prolactin, the formation of the long-range disulfide requires that the protein is first released from the ribosome/Sec-complex. In both cases, there is a delay in disulfide formation, despite the exposure of multiple cysteines to the ER lumen, until specific conditions are reached. This shows an absence of stochastic cysteine coupling and instead favours the structured precursor mechanism of folding. For β2M, it is already established that disulfide formation follows folding (7) with an initial collapse of the nascent chain occurring during ER entry. In this case, the partial folding of early intermediates is likely to spatially separate cysteines and prevent stochastic disulfide formation, despite the favourable oxidation conditions. For prolactin, disulfide formation depends on the nascent chain's release from the ribosome/Sec-complex, which suggests that the tethering of the polypeptide to the ribosome prevents folding from reaching completion. Viable explanations for this dependence on release include a requirement for C- and N-terminal interactions to initiate the folding process (26, 27), and inhibitory interactions with cellular factors (28, 29) that are alleviated on release. Although ribosome tethering clearly prevents conformational folding from being completed, prolactin translation intermediates are unlikely to be completely unstructured because then stochastic disulfide formation would be expected. Instead, we propose a similar mechanism to that identified for β2M in which partially folded precursors form on the ribosome that spatially

separate cysteines and prevent premature disulfide formation. These findings fit with those of other studies that have demonstrated partial folding at the co-translational stage (30).

In contrast to β2M and prolactin, disulfide formation occurs in disintegrin intermediates prior to the folding domain's full exposure to the ER lumen and results in a mixed population of non-native disulfide-bonded species. The structure of the disintegrin domain is largely defined by disulfides and so its folding is likely to follow the quasi-stochastic model, where disulfides form first in an unstructured precursor before conformational folding (4). A combination of the structural restrictions on the polypeptide when it is partially ER-exposed, combined with a stochastic mechanism of cysteine coupling explains the array of species we observe. In a previous study we identified ADAM10 as a substrate of the reductase ERp57 (an association that implies non-native disulfides) along with other proteins that share densely disulfide-bonded domains and low levels of secondary structure (31). Given the role of folding in directing disulfide formation, it follows that proteins that lack a stable secondary structure, are more likely to follow a stochastic coupling process, and form non-native disulfides instead. Examples of other proteins that form non-native disulfides during folding include LDLR (10), and Bowman-Birk inhibitor (32), both of which contain disulfide-rich, irregularly structured domains. In this study, the non-native disulfides observed in the disintegrin domain form only during partial exposure; once the disintegrin domain is fully exposed, a single disulfide-bonded species forms independently of reductive pathway activity. Fitting this co-operative process with the quasi-stochastic model indicates that critical disulfides form first, and they then impose structural restrictions that favour the pairing of subsequent cysteines. This would explain why non-native disulfides are abundant under conditions where folding is restricted, while unrestricted folding allows correct, native disulfide pairings to be efficiently achieved.

In this study, we assayed disulfide formation in stalled translation intermediates to provide insight into the folding mechanisms that achieve correct cysteine coupling in native protein structures. Our results provide novel insights into the relationship that exists between disulfide formation and protein folding. Our findings also raise questions for future studies to address, including (i) whether these mechanisms apply to a broader number and diversity of substrates; (ii) the molecular interactions involved in these folding mechanisms ; and (iii) how folding events correlate with translation in real time. Such studies will help to further characterise the mechanisms described here and will advance our understanding of the fundamentally important process of oxidative protein folding in the cell.

Acknowledgements

This work was funded by the Wellcome Trust (grant number 103720) and the Royal Society (ref IE140492). XC is the recipient of a Chinese Scholarship Council award.

Declaration

The authors declare that they have no conflict of interest.

Materials and Methods

Molecular graphics

The ribbon diagrams (Fig.1) were drawn using the UCSF Chimera package, version 1.8, from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (33) using PDB files 1A1M (β2M), 1RW5 (prolactin) and 6BE6 (ADAM10 disintegrin).

Template generation

All protein constructs had the same C-terminal extension sequence with alternate N-terminal domains. This was either human β2M, bovine prolactin or human ADAM10 residues 456-550. The β2M and prolactin constructs contained their native N-terminal signal sequences while the ADAM10 construct had the human-β2M signal sequence and the Neo-epitope sequence (IIPVEEENP) added at the N-terminus. These constructs were synthesised as plasmid DNA (Table S1) using GeneArt™ gene synthesis (ThermoFisher Scientific), with the sequence of each provided in supplementary material (Fig. S1-S3). Mutations to remove glycosylation sites were introduced by site-directed mutagenesis using ACCUZYME™ (Bioline). Plasmid DNA was used as a template for PCR using an appropriate forward primer that adds a T7 promoter (Table S2) and various reverse primers that lack stop codons (Table S3). PCR products were ethanol precipitated, re-suspended in nuclease-free water, and transcribed into RNA templates using T7 RNA polymerase (Progmega) (at 37 °C, 2 h). The resulting RNA was ethanol precipitated and resuspended in nuclease-free water for use as template in translation reactions.

Cell free translation

Translations were performed using the Flexi® Rabbit Reticulocyte Lysate System (Promega) supplemented with 10 mM DTT (Melford) for reducing lysate, 5 mM G6Pi (Sigma) for redox-balanced lysate or ddH20 (oxidising lysate). Typical translation reactions contained amino acids except for methionine (20 µM), KCl (40 mM), EasyTag™ EXPRESS^{35}S Protein Labeling Mix (PerkinElmer) (1 µl/25-µl reaction) and RNA template. Semi-permeabilised HT1080 human fibrosarcoma cells were prepared as described previously (19) and added to a concentration of ~10^5 per 25 µl translation reaction if required. Following assembly of components, translation reactions were incubated at 30 °C for either 10 min or 30 min as indicated in the text. After this time, samples were either treated with RNase A (Sigma) (1 mg/ml, 5 min at 30 °C) or placed directly on ice. Samples were treated with *N*-ethylmaleimide (NEM) (Sigma) at a concentration of 20 mM before further processing. Samples for glycosylation assays or downstream processing through ultracentrifugation were also treated with cycloheximide (2 mM). Alternative translation conditions were optimised for prolactin 30-310 (Fig 3. Eii), with amino acids except for cysteine replacing amino acids except for methionine and the KCl concentration increased to 50 mM.

Sample processing

Samples containing SP cells were centrifuged (16,000 × *g*, 30 s) following translation to isolate cell pellets, which were washed with KHM buffer and re-suspended either in SDS-PAGE sample buffer

(20 µl) for direct SDS-PAGE analysis (crude samples) or in immunoisolation (IP) buffer (50 mM Tris, pH 7.5, 1% (v/v) Triton X-100, 150 mM NaCl, 2 mM EDTA, 0.5 mM PMSF, and 0.02% (w/v) sodium azide) for immunoisolation. Samples translated without SP cells, were processed by various procedures following NEM treatment. In each case, either (i) IP buffer (0.9 ml) was added for immunoisolation, (ii) crude translation product was run directly on non-reducing SDS-PAGE or (iii) ribosome-nascent chain complexes were isolated by ultracentrifugation. For the ultracentrifugation procedure, samples were loaded onto a sucrose cushion (2.5 M sucrose, 50 mM Tris pH 7.5, 10 mM $MgCl_2$, 25 mM KCl) for 90 min at 265,000 x g. Pellets were rinsed with nuclease free water and re-suspended in non-reducing sample buffer for SDS-PAGE analysis.

Immunoisolation procedure

Samples in IP buffer were incubated with 0.5% (v/v) Protein A-Sepharose (Generon) for 30 min (4 °C) and then centrifuged (2000 × *g*, 5 min) to clear samples of material that is bound non-specifically. The resulting supernatant was incubated with Protein A-Sepharose (0.5% v/v) and the indicated antibody at 4 °C overnight. The antibodies used were the polyclonal rabbit anti-human β2M antibody (Dako) (used at 1/1000), the anti-V5 antibody (Invitrogen) (used at 1/10,000) or amino-antibody to the mature form of PLAP153 (used at 1/500) (34), which we term the Neo-epitope (NE) antibody. For NE immunoisolation procedures, PAS was pre-treated with 1% Bovine serum albumin (BSA). Following overnight incubation with the antibody, samples were isolated by centrifugation (2000 × *g*, 5 min) and washed three times with 1 ml of IP buffer and once with 100 µl of 20 mM HEPES, pH 7.5. Protein was eluted by boiling the beads in non-reducing SDS-PAGE sample buffer. If samples were to be run under reducing conditions then 10 mM DTT was added to the sample before boiling and once cooled NEM was added (1 µl, 1M NEM) before loading. Samples were run on SDS-PAGE and the gels were fixed, dried and exposed to either BioMax MR film (KODAK) for analysis by autoradiography or to phosphoimager plates for analysis on a FLA-7000 bioimager (Fujifilm).

G6Pi recovery and auranofin treatment

Following translation of the 146 disintegrin intermediate (10 min at 30 °C) in oxidising lysate, G6Pi was added (5 mM) and samples incubated at 30°C for 10 min. Aliquots were removed at the time points specified and treated with NEM for subsequent processing using V5 and NE immunoisolation, as described above. For the auranofin treatment, 10 minute translations were performed before auranofin (sigma) was added at a concentration of 20 µM to relevant samples. Following a further 10 min incubation at room temperature, 5mM G6Pi was added for 30 min at 30°C. Samples were treated with NEM before V5 immunoisolation and SDS-PAGE analysis, as described above.

References

1. Fass D & Thorpe C (2018) Chemistry and Enzymology of Disulfide Cross-Linking in Proteins. *Chem. Rev.* 118(3):1169-1198.
2. Hudson DA, Gannon SA, & Thorpe C (2015) Oxidative protein folding: from thiol-disulfide exchange reactions to the redox poise of the endoplasmic reticulum. *Free Radic. Biol. Med.* 80:171-182.
3. Oka OB & Bulleid NJ (2013) Forming disulfides in the endoplasmic reticulum. *Biochim. Biophys. Acta* 1833(11):2425-2429.
4. Welker E, Wedemeyer WJ, Narayan M, & Scheraga HA (2001) Coupling of conformational folding and disulfide-bond reactions in oxidative folding of proteins. *Biochemistry* 40(31):9059-9064.
5. Qin M, Wang W, & Thirumalai D (2015) Protein folding guides disulfide bond formation. *Proc. Natl. Acad. Sci. USA* 112(36):11241-11246.
6. Kosuri P*, et al.* (2012) Protein folding drives disulfide formation. *Cell* 151(4):794-806.
7. Robinson PJ, Pringle MA, Woolhead CA, & Bulleid NJ (2017) Folding of a single domain protein entering the endoplasmic reticulum precedes disulfide formation. *J. Biol. Chem.* 292(17):6978-6986.
8. Lv JM*, et al.* (2018) Conformational folding and disulfide bonding drive distinct stages of protein structure formation. *Sci. Rep.* 8(1):1494.
9. Land A, Zonneveld D, & Braakman I (2003) Folding of HIV-1 envelope glycoprotein involves extensive isomerization of disulfide bonds and conformation-dependent leader peptide cleavage. *FASEB J.* 17(9):1058-1067.
10. Jansens A, van Duijn E, & Braakman I (2002) Coordinated nonvectorial folding in a newly synthesized multidomain protein. *Science* 298(5602):2401-2403.
11. Hirano N*, et al.* (1995) Molecular cloning of the human glucose-regulated protein ERp57/GRP58, a thiol-dependent reductase. Identification of its secretory form and inducible expression by the oncogenic transformation. *Eur. J. Biochem.* 234(1):336-342.
12. Oka OB, Pringle MA, Schopp IM, Braakman I, & Bulleid NJ (2013) ERdj5 is the ER reductase that catalyzes the removal of non-native disulfides and correct folding of the LDL receptor. *Mol. Cell* 50(6):793-804.
13. Poet GJ*, et al.* (2017) Cytosolic thioredoxin reductase 1 is required for correct disulfide formation in the ER. *EMBO J.* 36(5):693-702.
14. Rapoport TA, Li L, & Park E (2017) Structural and Mechanistic Insights into Protein Translocation. *Annu. Rev. Cell Dev. Biol.* 33:369-390.
15. Thommen M, Holtkamp W, & Rodnina MV (2017) Co-translational protein folding: progress and methods. *Curr. Opin. Struct. Biol.* 42:83-89.
16. Chen W, Helenius J, Braakman I, & Helenius A (1995) Cotranslational folding and calnexin binding during glycoprotein synthesis. *Proc. Natl. Acad. Sci. USA* 92(14):6229-6233.
17. Bergman LW & Kuehl WM (1979) Formation of an intrachain disulfide bond on nascent immunoglobulin light chains. *J. Biol. Chem.* 254(18):8869-8876.
18. Molinari M & Helenius A (1999) Glycoproteins form mixed disulphides with oxidoreductases during folding in living cells. *Nature* 402(6757):90-93.
19. Wilson R*, et al.* (1995) The translocation, folding, assembly and redox-dependent degradation of secretory and membrane proteins in semi-permeabilized mammalian cells. *Biochem J.* 307 (Pt 3):679-687.
20. Fedyukina DV & Cavagnero S (2011) Protein folding at the exit tunnel. *Annu Rev Biophys* 40:337-359.
21. Kaderbhai MA & Austen BM (1985) Studies on the formation of intrachain disulphide bonds in newly biosynthesised bovine prolactin. Role of protein-disulphide isomerase. *Eur. J. Biochem.* 153(1):167-178.

22. Janes PW*, et al.* (2005) Adam meets Eph: an ADAM substrate recognition module acts as a molecular switch for ephrin cleavage in trans. *Cell* 123(2):291-304.
23. Seegar TCM*, et al.* (2017) Structural Basis for Regulated Proteolysis by the alpha-Secretase ADAM10. *Cell* 171(7):1638-1648 e1637.
24. Bailey CA, Gerber L, Howard AD, & Udenfriend S (1989) Processing at the carboxyl terminus of nascent placental alkaline phosphatase in a cell-free system: evidence for specific cleavage of a signal peptide. *Proc. Natl. Acad. Sci. USA* 86(1):22-26.
25. Fariselli P & Casadio R (2001) Prediction of disulfide connectivity in proteins. *Bioinformatics* 17(10):957-964.
26. Krishna MM & Englander SW (2005) The N-terminal to C-terminal motif in protein folding and function. *Proc. Natl. Acad. Sci. USA* 102(4):1053-1058.
27. Haran G (2012) How, when and why proteins collapse: the relation to folding. *Curr. Opin. Struct. Biol.* 22(1):14-20.
28. Kramer G, Boehringer D, Ban N, & Bukau B (2009) The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nat. Struct. Mol. Biol.* 16(6):589-597.
29. Cabrita LD*, et al.* (2016) A structural ensemble of a ribosome-nascent chain complex during cotranslational protein folding. *Nat. Struct. Mol. Biol.* 23(4):278-285.
30. Holtkamp W*, et al.* (2015) Cotranslational protein folding on the ribosome monitored in real time. *Science* 350(6264):1104-1107.
31. Jessop CE*, et al.* (2007) ERp57 is essential for efficient folding of glycoproteins sharing common structural domains. *EMBO J.* 26(1):28-40.
32. Singh RR & Appu Rao AG (2002) Reductive unfolding and oxidative refolding of a Bowman-Birk inhibitor from horsegram seeds (Dolichos biflorus): evidence for "hyperreactive" disulfide bonds and rate-limiting nature of disulfide isomerization in folding. *Biochim. Biophys. Acta* 1597(2):280-291.
33. Pettersen EF*, et al.* (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25(13):1605-1612.
34. Spurway TD, Dalley JA, High S, & Bulleid NJ (2001) Early events in glycosylphosphatidylinositol anchor addition. substrate proteins associate with the transamidase subunit gpi8p. *J Biol Chem* 276(19):15975-15982.

Figure Legends

Figure 1. An experimental system to monitor disulfide formation and rearrangement in translation intermediates. (A) Ribbon diagrams representing the substrates used in this study: (i) human β2M (PDB 1A1M); (ii) human prolactin (PDB 1RW5); and (iii) the disintegrin domain of human ADAM10 (PDB 6BE6). Disulfide bonds are highlighted in yellow. (B) Schematic illustrating the C-terminal extension added to each folding domain to produce the extended constructs. The position of the signal peptide (SP) glycosylation site (NST), V5 epitope and 5 methionine residues (5xM) are indicated. (C) (i) Schematic of the ribosome-Sec complex expressing an ER-exposed nascent polypeptide. Phase 1 shows how the C-terminal extension retains the polypeptide attached to the ribosome to allow Phase 2. In Phase 2 the N-terminal folding domain of the polypeptide is fully exposed to the ER lumen. (ii) G6Pi drives the cytosolic reducing pathway, which (iii) is the source of reducing equivalents for disulfide rearrangements in the ER.

Figure. 2 Influence of redox status on the timing of β2M disulfide formation. (A) Organisation of the extended β2M construct with the single disulfide (C45-C100) displayed. (B) (i) Diagram illustrating the dependence of disulfide formation on domain exposure, arrows highlight the end of the β2M folding domain. (ii) Translation of the 175 intermediates with SP cells in redox-balanced (Bal), oxidising (Ox), and reducing (Red) lysate. RNaseA treatment following translation is indicated. (C) Non-reducing SDS-PAGE of isolated ribosome/nascent chain complexes representing β2M (i) 21-105, (ii) 21-141, and (iii) 21-165 translated in redox-balanced (Bal), oxidising (Ox) or reducing (Red) lysate without SP cells. Samples treated with RnaseA following translation are indicated. Schematics next to each panel indicate the predicted cysteine exposure at each intermediate length. All gels in this figure are representative of at least 3 repeats.

Figure 3. Conditions required for disulfide formation in prolactin (A) Organisation of the extended prolactin construct with disulfide bond positions indicated. (B) (i) Reducing SDS-PAGE of prolactin intermediates (285-310 aa in length), translated with SP cells. Glycosylated protein is highlighted for the 310 intermediate (*), (ii) a schematic of the extension length (63 aa) required for ER-exposure and (iii) a table of prolactin intermediate lengths containing the predicted exposure of amino acids and cysteine residues to the ER-lumen. (C) Non-reducing SDS-PAGE of wildtype prolactin translated with SP cells in reducing (Red) or redox-balanced lysate (Bal). The faster migrating oxidised mature protein is highlighted (arrow). (D) Prolactin intermediates (275-330) translated with SP cells in redox-balanced lysate and either (i) stalled or (ii) released through RNaseA treatment. Gels are run under reducing (+DTT) and non-reducing (-DTT) conditions. (E) Non-reducing SDS-PAGE of the 310 intermediate translated in reducing (Red), redox-balanced (Nat) or oxidising (Ox) lysate in (i) the presence or (ii) the absence of SP cells. Stalled samples are compared to the released (RNaseA-treated) samples. Representative data is shown in this figure from at least 3 experimental repeats in each case.

Figure 4. Domain exposure and redox conditions influence disulfide bonding in an extended disintegrin construct. (A) Organisation of the disintegrin construct with the sequence and native disulfide bonding pattern highlighted. (B) (i) Reducing SDS-PAGE to identify glycosylation in intermediates of lengths 180-225, with the table (ii) detailing expected exposure of amino acids and cysteine residues to the ER-lumen. (C) Non-reducing SDS-PAGE showing radiolabelled intermediates of increasing length (146-210 preprotein length). The mobility of these intermediate species varies depending on whether they are translated in reducing, redox-balanced or oxidising lysate. All gels in this figure represent translation reactions with SP cells immunoisolated to the V5 epitope, and are representative of 3 independent repeats.

Figure 5. ER-specific disulfide rearrangements in partially exposed nascent chain intermediates of the disintegrin domain. Non-reducing SDS-PAGE shows either total translation product (V5) or ER-specific species (NE) for the 146 and 210 intermediates of the disintegrin domain (A) when translated in reducing (Red), redox-balanced (Bal) and oxidising (Ox) lysate (B) when translated in

oxidising lysate, with G6Pi added after translation and samples taken at specific time points post G6Pi treatment (lanes 4-6). (C) As for B, 146 and 210 intermediates translated in oxidising lysate, to which G6Pi was added after translation in the presence or absence of auranofin (Aur) (lanes 4-7). Control samples for panels B and C were translated in reducing, redox-balanced and oxidising lysates (lanes 1-3 in both cases) for gel mobility comparison. Panel A and B were repeated 3 times and panel C twice, with representative data shown.

Supplementary information

Table S1: Plasmid list

Plasmid	**Description**
Extended-β2M	DNA Sequence in Fig.S1
Extended- β2M AST	Extended- β2M with N120A mutation
Extended-prolactin	DNA Sequence in Fig.S2
Extended-prolactin AST	Extended-prolactin with N230A mutation
Extended-disintegrin	DNA Sequence in Fig.S3
Extended-disintegrin AST	Extended-disintegrin with N125A mutation

Table S2: Forward primers for transcription/translation template generation

Extended-β2M Fwd*	GATGGCTAATACGACTCACTATAGGGTCAGGCCACCATGAGCAGATCTGTGGCCCTGG
Extended-β2M mature Fwd	GATGGCTAATACGACTCACTATAGGGTCAGGCCACCATGATCCAGCGGACCCCCAAGATCCAGGTGTA
Extended-prolactin Fwd	GATGGCTAATACGACTCACTATAGGGTCAGGCCACCATGGACAGCAAAGGTTCGTCGC
Extended-prolactin mature Fwd	GATGGCTAATACGACTCACTATAGGGTCAGGCCACCATGACCCCCGTCTGTCCCAATGGGC

*This primer is also used to generate disintegrin templates as the extended-disintegrin construct contains the β2M signal sequence.

Table S3: Reverse PCR primers to generate templates for stalled intermediates

Length of intermediate from start codon (aa)			
β2M	Prolactin	Disintegrin	Primer sequence
105	-	-	CACGTGGTTCACTCTGCAG
141	-	146	CATCATCATCATCATGGTGCTGTC
-	-	155	AGCAGAGCTGGCGCT
-	-	165	AGAGCTGGCTGTGGCTC
165	275	-	GCCGCCAGCAGATGTAGAG
175	285	180	TGTAGATCCGCCTGTAGAGC
-	300	195	AGCTCCGCCAGTAGATCCT
-	305	200	TGTTCCAGTAGAGGCAGCTC
-	310	205	TCCGCCCCCAGCAGCT
-	-	210	AGAAGAAGCTCCGCCTCC

-	330	225	TGTTGTTCCGGTGCCTGTAG

Figure S1. DNA sequence of the β2M construct

ATGAGCAGATCTGTGGCCCTGGCTGTGCTGGCCCTGCTGTCTCTGTCTGGCCTGGAAGCCATCCAGCGGACCC
CCAAGATCCAGGTGTACAGCAGACACCCCGCCGAGAACGGCAAGAGCAACTTCCTGAACTGCTACGTGTCCG
GCTTCCACCCCAGCGACATCGAGGTGGACCTGCTGAAGAACGGCGAGCGGATCGAGAAGGTGGAACACAGC
GACCTGAGCTTCAGCAAGGACTGGTCCTTCTACCTGCTGTACTACACCGAGTTCACCCCCACCGAGAAGGACG
AGTACGCCTGCAGAGTGAACCACGTGACCCTGAGCCAGCCCAAGATCGTGAAGTGGGACCGGGACATGAAC
AGCACCGGCAAGCCCATCCCCAACCCTCTGCTGGGCCTGGACAGCACCATGATGATGATGATGTCCGGCACCG
CCAGCGCCAGCTCTGCTGGATCTGGCGGCGGAGCCACAGCCAGCTCTACATCTGCTGGCGGCACAAGCACCG
GCTCTACAGGCGGATCTACAGCAGGCGCTGCTGGCGCAACAGGCGGAGGATCTACTGGCGGAGCTGCCTCTA
CTGGAACAGCTGCTGGGGGCGGAGGCGGAGCTTCTTCTGGAACAGGCACAGGCGCCAGCGGCGCTACAGGC
ACCGGAACAACA

Figure S2. DNA sequence of the prolactin construct

ATGGACAGCAAAGGTTCGTCGCAGAAAGGGTCCCGCCTGCTCCTGCTGCTGGTGGTGTCAAATCTACTCTTGT
GCCAGGGTGTGGTCTCCACCCCCGTCTGTCCCAATGGGCCTGGCAACTGCCAGGTATCCCTTCGAGACCTGTT
TGACCGGGCAGTCATGGTGTCCCACTACATCCATGACCTCTCCTCGGAAATGTTCAACGAATTTGATAAACGGT
ATGCCCAGGGCAAAGGGTTCATTACCATGGCCCTCAACAGCTGCCATACCTCCTCCCTTCCTACCCCGGAAGAT
AAAGAACAAGCCCAACAGACCCATCATGAAGTCCTTATGAGCTTGATTCTTGGGTTGCTGCGCTCCTGGAATG
ACCCTCTGTATCACCTAGTCACCGAGGTACGGGGTATGAAAGGAGCCCCAGATGCTATCCTATCGAGGGCCAT
AGAGATTGAGGAAGAAAACAAACGACTTCTGGAAGGCATGGAGATGATATTTGGCCAGGTTATTCCTGGAGC
CAAAGAGACTGAGCCCTACCCTGTGTGGTCAGGACTCCCGTCCCTGCAAACTAAGGATGAAGATGCACGTTAT
TCTGCTTTTTATAACCTGCTCCACTGCCTGCGCAGGGATTCAAGCAAGATTGACACTTACCTTAAGCTCCTGAAT
TGCAGAATCATCTACAACAACAACTGCAACAGCACCGGCAAGCCCATCCCCAACCCTCTGCTGGGCCTGGACA
GCACCATGATGATGATGATGTCCGGCACCGCCAGCGCCAGCTCTGCTGGATCTGGCGGCGGAGCCACAGCCA
GCTCTACATCTGCTGGCGGCACAAGCACCGGCTCTACAGGCGGATCTACAGCAGGCGCTGCTGGCGCAACAG
GCGGAGGATCTACTGGCGGAGCTGCCTCTACTGGAACAGCTGCTGGGGGCGGAGGCGGAGCTTCTTCTGGA
ACAGGCACAGGCGCCAGCGGCGCTACAGGCACCGGAACAACA

Figure S3. DNA sequence of the disintegrin construct

ATGAGCAGATCTGTGGCCCTGGCTGTGCTGGCCCTGCTGTCTCTGTCTGGCCTGGAAGCCATTATTCCGGTGG
AAGAAGAAAACCCGGGGCAGCCTATCTGCGGCAACGGAATGGTGGAACAGGGCGAAGAGTGCGACTGTGG
CTACAGCGACCAGTGCAAGGACGAGTGCTGCTTCGATGCCAACCAGCCTGAGGGCAGAAAGTGCAAGCTGA
AGCCTGGCAAGCAGTGCAGCCCTAGCCAGGGACCTTGTTGCACAGCCCAGTGTGCCTTCAAGAGCAAGAGCG
AGAAGTGCCGCGACGACTCTGATTGTGCTAGAGAGGGCATCTGCAACGGCTTCACCGCTCTGTGTCCTGCCAG
CGATCCTAAACCTAACAGCACCGGCAAGCCCATCCCCAACCCTCTGCTGGGCCTGGACAGCACCATGATGATG
ATGATGTCCGGCACCGCCAGCGCCAGCTCTGCTGGATCTGGCGGCGGAGCCACAGCCAGCTCTACATCTGCTG
GCGGCACAAGCACCGGCTCTACAGGCGGATCTACAGCAGGCGCTGCTGGCGCAACAGGCGGAGGATCTACT
GGCGGAGCTGCCTCTACTGGAACAGCTGCTGGGGGCGGAGGCGGAGCTTCTTCTGGAACAGGCACAGGCGC
CAGCGGCGCTACAGGCACCGGAACAACA

Figure 1

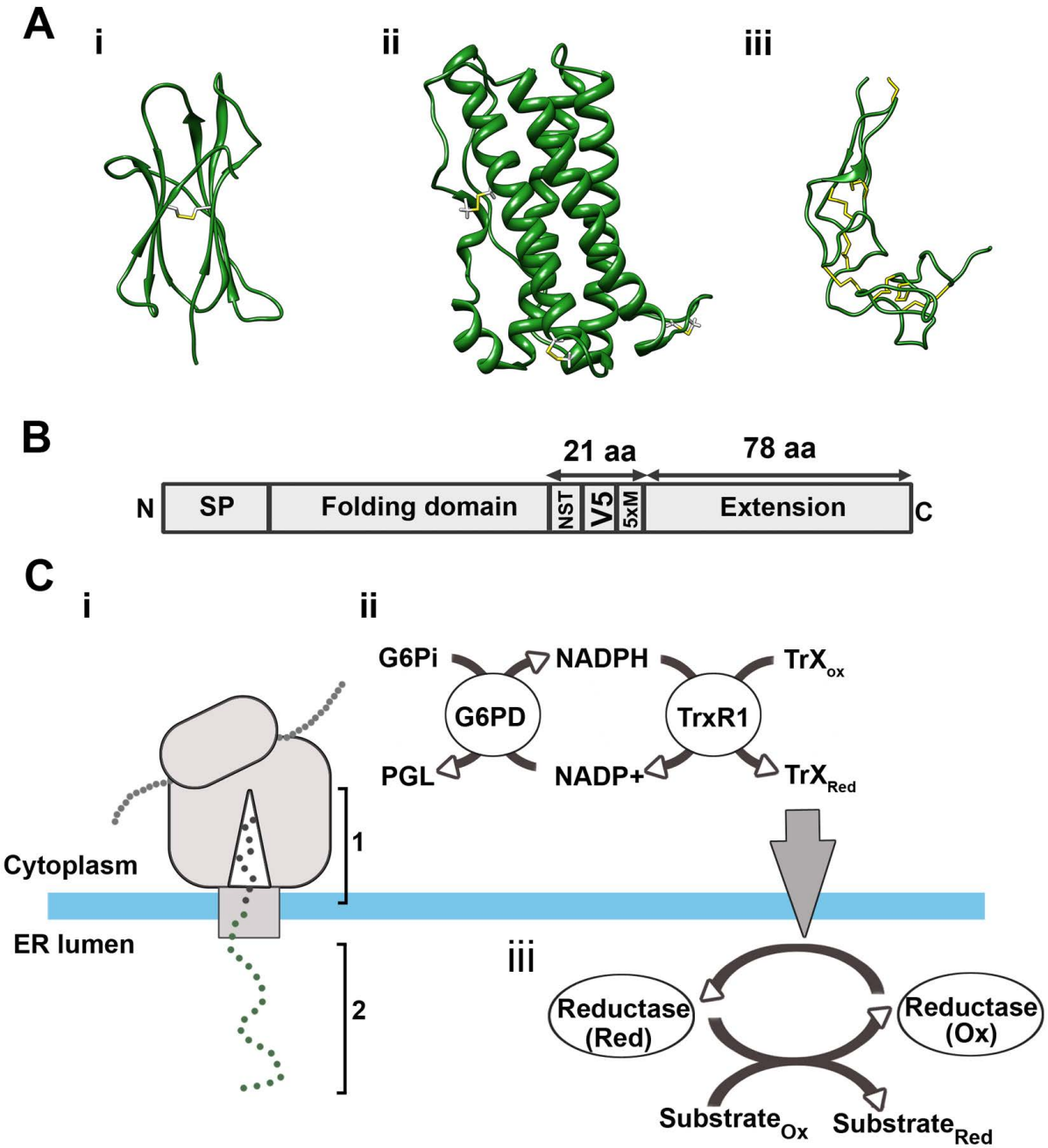

A
i
ii
iii
B
21 aa
78 aa
N
SP
Folding domain
NST
V5
5xM
Extension
C
C
i
ii
G6Pi
NADPH
TrX$_{ox}$
G6PD
TrxR1
PGL
NADP+
TrX$_{Red}$
1
Cytoplasm
ER lumen
2
iii
Reductase (Red)
Reductase (Ox)
Substrate$_{Ox}$
Substrate$_{Red}$

Figure 2

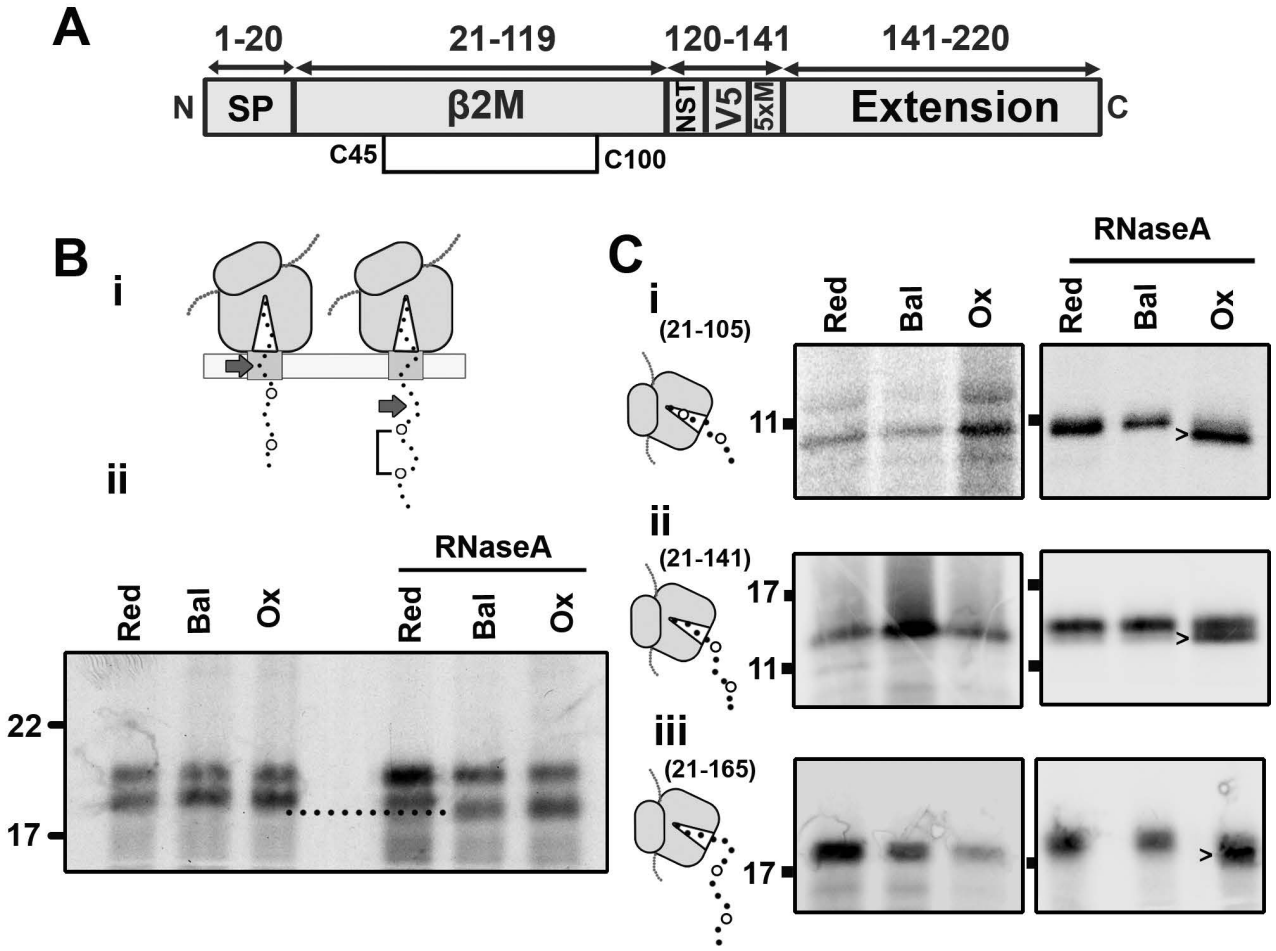

A
1-20
21-119
120-141
141-220
N
SP
β2M
NST
V5
5xM
Extension
C
C45
C100
B
i
ii
RNaseA
Red
Bal
Ox
Red
Bal
Ox
22
17
C
i
(21-105)
Red
Bal
Ox
RNaseA
Red
Bal
Ox
11
ii
(21-141)
17
11
iii
(21-165)
17

Figure 3

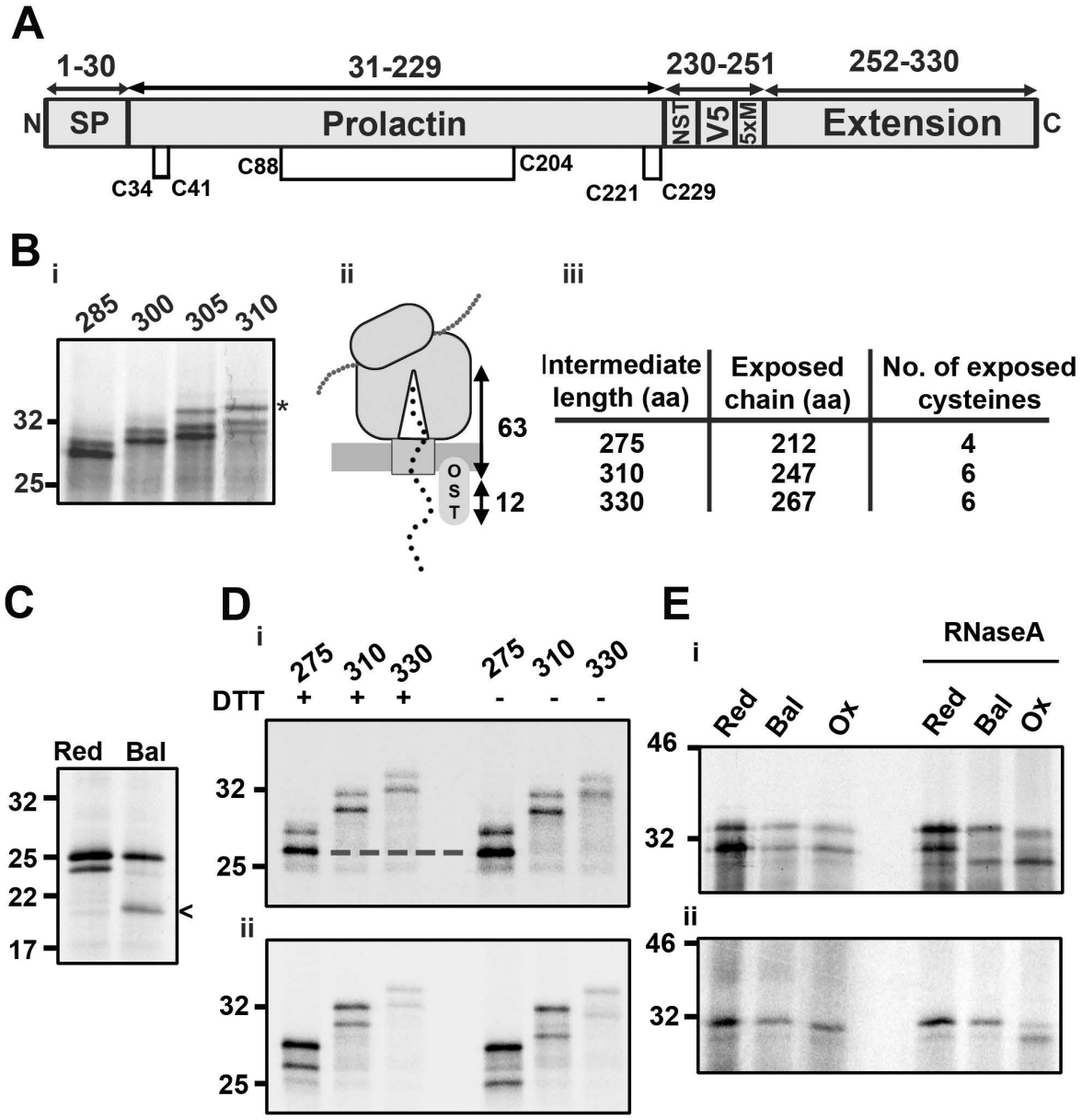
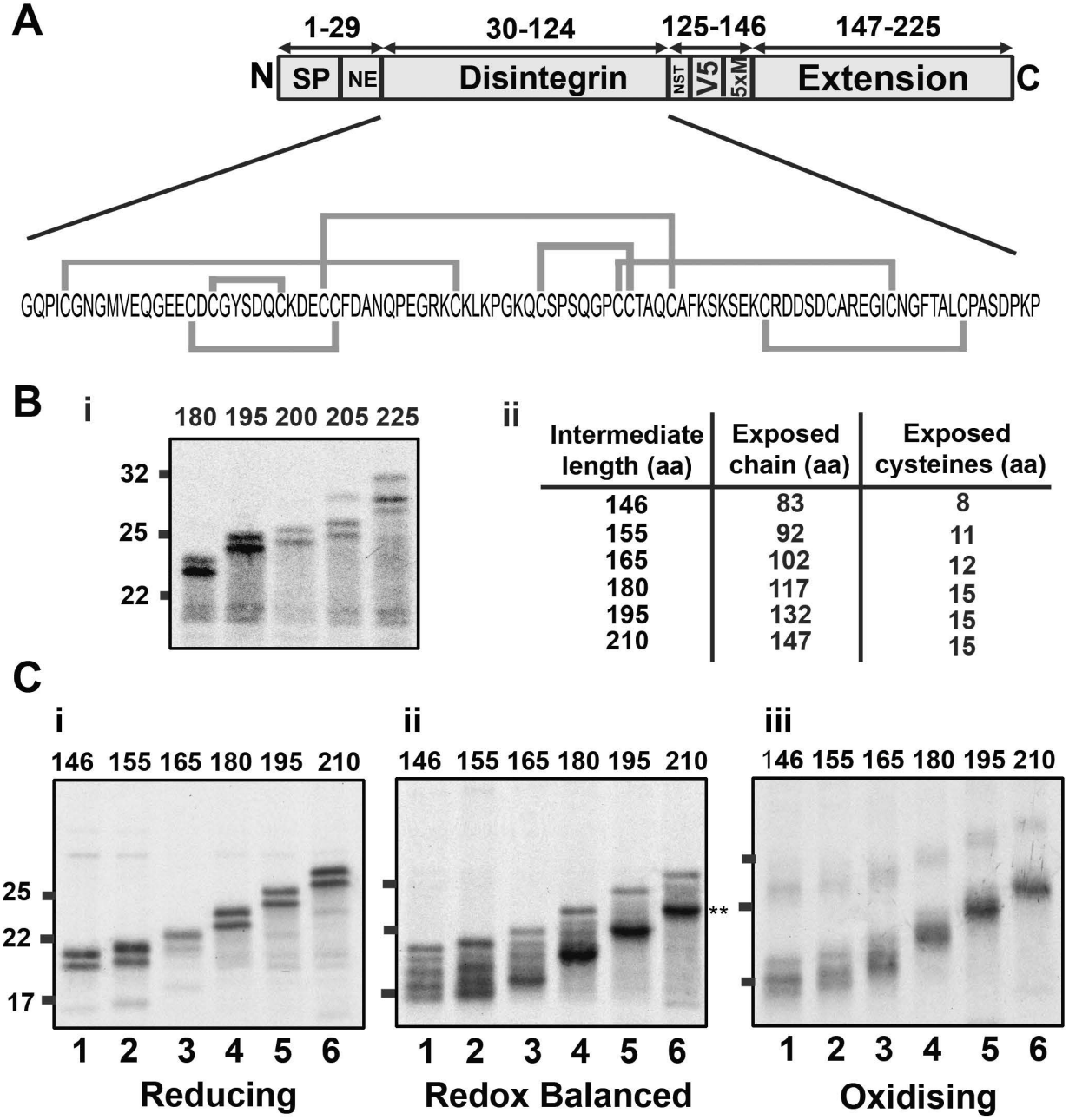
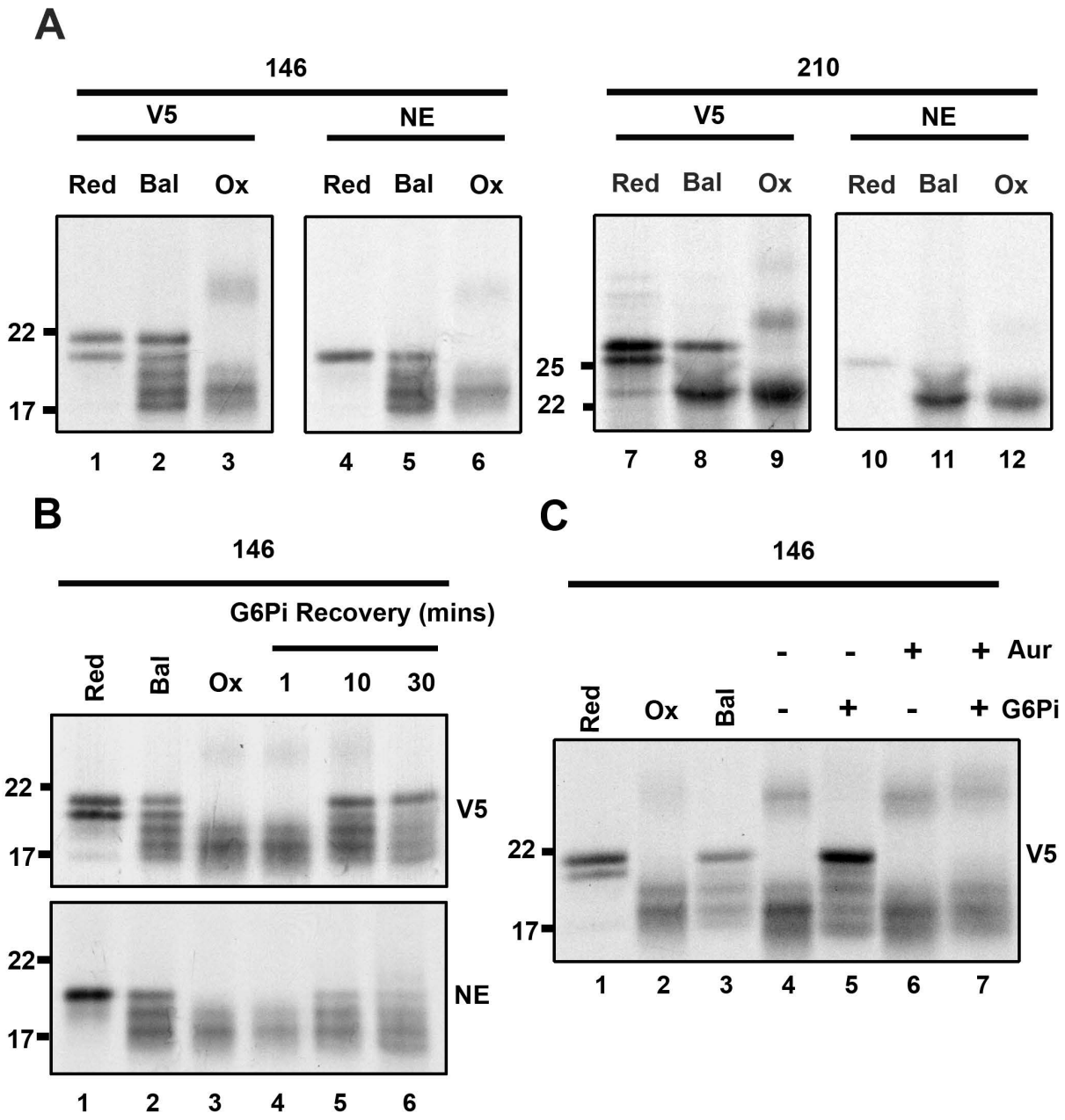

A
1-30
31-229
230-251
252-330
N
SP
Prolactin
NST
V5
5xM
Extension
C
C34
C41
C88
C204
C221
C229
B
i
285
300
305
310
32
25
*
ii
63
12
OST
iii
Intermediate length (aa)
Exposed chain (aa)
No. of exposed cysteines
275 212 4
310 247 6
330 267 6
C
Red
Bal
32
25
22
17
<
D
i
275
310
330
275
310
330
DTT
+
+
+
-
-
-
32
25
ii
32
25
E
i
RNaseA
Red
Bal
Ox
Red
Bal
Ox
46
32
ii
46
32

Figure 4



Intermediate length (aa)	Exposed chain (aa)	Exposed cysteines (aa)
146	83	8
155	92	11
165	102	12
180	117	15
195	132	15
210	147	15

Figure 5


A
146
210
V5
NE
V5
NE
Red Bal Ox
Red Bal Ox
Red Bal Ox
Red Bal Ox
22
17
25
22
1 2 3 4 5 6
7 8 9 10 11 12
B
146
G6Pi Recovery (mins)
Red Bal Ox 1 10 30
22
17
V5
22
17
NE
1 2 3 4 5 6
C
146
- - + + Aur
Red Ox Bal - + - + G6Pi
22
17
V5
1 2 3 4 5 6 7