Review

Orchestration of secretory protein folding by ER chaperones

Tali Gidalevitz a, Fred Stevens b, Yair Argon c,d,*

a Department of Biology, Drexel University, 418 Papadakis Integrated Science Bldg., 3245 Chestnut Street, Philadelphia, PA 19104, USA
b Argonne National Laboratory, Argonne, IL 60439, USA
c Division of Cell Pathology, Department of Pathology and Lab Medicine, The Children’s Hospital of Philadelphia, 3615 Civic Center Blvd., Philadelphia, PA 19104, USA
d The University of Pennsylvania, Philadelphia, PA 19104, USA

A R T I C L E   I N F O

Article history:
Received 11 December 2012
Received in revised form 27 February 2013
Accepted 1 March 2013
Available online 15 March 2013

Keywords:
Endoplasmic reticulum
Chaperones
Protein folding
Secretome

A B S T R A C T

The endoplasmic reticulum is a major compartment of protein biogenesis in the cell, dedicated to production of secretory, membrane and organelle proteins. The secretome has distinct structural and post-translational characteristics, since folding in the ER occurs in an environment that is distinct in terms of its ionic composition, dynamics and requirements for quality control. The folding machinery in the ER therefore includes chaperones and folding enzymes that introduce, monitor and react to disulfide bonds, glycans, and fluctuations of luminal calcium. We describe the major chaperone networks in the lumen and discuss how they have distinct modes of operation that enable cells to accomplish highly efficient production of the secretome. This article is part of a Special Issue entitled: Functional and structural diversity of endoplasmic reticulum.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

The great majority of extracellular, surface-expressed, secretory and endosomal eukaryotic proteins, totaling over 18,000 in the secreted protein database [1], fold and assemble in the endoplasmic reticulum (ER). Thus, in addition to being a major site for lipid biosynthesis and a major calcium store (reviewed elsewhere in this issue), folding of proteins is one of the main functions of the ER. Precise control over the rate of production and the “quality” of their structure is essential for every aspect of metazoan organisms, from morphogenesis during the embryonic development, to the complex endocrine regulation in aging. In this review, we will discuss how the complex processes of protein folding and assembly are supported by the specialized conditions in the ER.

2. Restricting an infinite folding landscape

Anfinsen’s axiom that the necessary information for folding of a protein is encoded within the amino acid sequence itself was derived from spontaneous refolding of pure protein samples after denaturation [2,3]. Though undoubtedly true, this principle does not suffice to explain protein folding in the cell, because the theoretical “folding space” is too vast [4], and therefore needs to be restricted to be compatible with biological constraints. Most importantly, the wide span of time-scales observed for protein folding suggests a paradigm distinct from random sampling of all possible conformations. Various mechanisms of limiting the conformational space have been proposed, including a bias towards native-like interactions [5], formation of intermediate states on the folding pathway [6], funneling into a particular folding pathway [7] with optional errors [8], and folding down the energy landscape through multiple pathways [9].

2.1. Chaperones limit the available conformational space

Molecular chaperones, by selectively interacting with certain sequence and structural elements, can both contribute to limiting the conformational space and favor particular folding pathways. For example, we traditionally consider that binding of hydrophobic sequences by HSP70 chaperones serves to protect nascent chains and folding intermediates from aggregation, by shielding these recognition regions from intermolecular hydrophobic interactions [10]. However, another consequence of HSP70 binding may be retarding the collapse of these hydrophobic regions into the core of the forming protein. Shielding of HSP70-target peptides from intramolecular hydrophobic interactions early in the folding process would favor a different final spatial arrangement of this sequence, compared to the arrangement that could result if it was allowed to participate in an early hydrophobic collapse. In fact, the hydrophobic residues that are involved in misfolding and aggregation are often the same residues that form chaperone recognition sites. For example, the ER HSP70 chaperone, BiP, binds to immunoglobulin (Ig) light chain...
(LC) mainly through two peptides on the LC variable domain [11, 12]. One of these two peptides is directly involved in the formation of aminophosphoryl fibrils by the variable domain, as shown in [13]. Interestingly, these two dominant BiP-binding sites are located in the center of each of the two β sheets, and the BiP-binding residues face the core of the β sandwich in the folded protein [12]. The release of BiP from the variable domain is coupled to the formation of the stabilizing disulfide bond [11]; BiP binding then serves to delay the closing of the sandwich. Thus, by retarding the incorporation of certain hydrophobic regions into the hydrophobic core, chaperone binding favors a different folding path leading to specific contacts. Finally, ATP-driven cycles of chaperone binding, and enzyme-like actions of protein disulfide isomerasers (PDI's) and protein prodlyn isomerasers (PPIs), also contribute to conformational remodeling during the folding process, by allowing proteins to escape unproductive intermediate states stabilized by non-native interactions. So far, we lack good understanding of how chaperones recognize such stabilized states.

2.2. Consequences of vectorial synthesis

The vectorial synthesis of polypeptides provides yet another means of restricting the conformational space, by controlling the order of folding steps. Because the emergent nascent chain starts folding concomitantly with synthesis, the folding opportunities for α helices and β sheets are vastly different: helix-forming amino acids are translocated sequentially and therefore are available immediately to form all the interactions that stabilize helices. Formation of β sheets, on the other hand, must be severely delayed by the translocation process, since each peptide that would assume an extended strand conformation must ‘wait’ until the other strands are synthesized before hydrogen bonds can be satisfied. This difference implies that if such secondary structure elements form the initial folding intermediates, then α helical proteins have a kinetic advantage in the ER. This difference also explains why the peptide binding chaperones of the HSP70 family evolved to bind extended β strand peptides. In addition to structural elements like β sheets, the organization of certain domains may require slowing down or pausing their folding until the entire domain has emerged from the translocon. For example, a C-type lectin fold, present in multiple metazoan extracellular proteins, has a loop-like form, connected at its base by a disulfide bond and an antiparallel β-sheet formed by N- and C-terminal β strands [β1, β5] coming close together [14]. In such arrangements, not only does the N-terminal β strand need to be prevented from inappropriately participating in the hydrophobic collapse, but the formation of the disulfide bond by the cysteine in the N-terminus has to be delayed as well. Other examples of such delay are the folding of influenza hemagglutinin [15–17] and chionic gonadotropin β chain [18], where binding of molecular chaperones is used to direct the folding pathways in vivo.

Vectorial synthesis has another global effect on the folding options available for large proteins. The organization of large proteins into a series of independently folding domains restricts the conformational space: as each domain folds, the number of options for the rest of the sequence are reduced. However, the complexity of the proteome and the oligomeric structure of many proteins mean that even this simple rule may not be sufficient, as discussed in the examples below.

3. Unique aspects of protein folding in the ER

The conditions for folding within the ER differ significantly from those in other cellular compartments like the cytosol, nucleus or mitochondria. The molecular crowding in the ER is 3–6× higher than in the cytosol [19], the redox potential is 1000× more oxidizing [20], free Ca ++ concentration oscillates and can reach 1 mM [21], and carbohydrates and a glycosylation machinery are unique to the ER. These conditions impact the energetics and kinetics of protein folding and provide the evolutionary pressure for the presence of specialized enzymes and chaperones to facilitate folding in the ER. Indeed, the most abundant luminal proteins (e.g. BiP, GRP94, PDI, HSP47 and calreticulin, see Table 1) all function in protein folding. We review the impact of these unique conditions first, and then discuss how the folding machinery responds to them. Our main focus is folding of proteins in the ER lumen and the folding of ER membrane proteins is considered elsewhere.

3.1. Proximity to the membranes

One consequence of the vectorial translocation of the nascent chain into the ER lumen is that it is positioned in proximity to the membrane, which can have detrimental effects on folding. Negatively charged membrane surfaces have been shown to influence conformational transitions and induce aggregation and fibrillogenesis of Ig LCs [22], lysozyme [23], and a proteolytic collagen fragment endostatin [24], as well as a variety of non-secretory proteins [25]. As there is no ER equivalent to the cytosolic ribosome-associated chaperones or to chaperonins, the protection of the emerging nascent chain from interaction with membrane surfaces and its channeling into folding pathways is mediated by the close spatial proximity of the BiP molecules which gate the translocon [26] and the oligosaccharide transferase [27].

3.2. Folding in the viscous luminal environment

One constraint on folding in the ER lumen is its high macromolecular density, which was measured to be 3–6 fold higher than in the cytosol and 9–18× higher than in typical aqueous buffers [19]. The viscosity of the ER lumen is high not only because of the presence of many proteins in various states of folding, including the high concentration of ER-resident chaperones, but also because of other macromolecules, estimated to occupy up to 30% of the volume [28]. Nonetheless, the diffusional mobility of GFP-tagged chaperones is only moderately slowed in the ER compared to the cytosol (E.L. Snapp, personal communication). The viscous environment is one evolutionary reason for molecular chaperones, whose activity imposes a timing sequence on the folding pathway that is not emulated in monodisperse in vitro systems. Molecular crowding clearly also affects the activity of the folding machinery, for example the enzymes that trim glycans [29].

3.3. Consequences of glycosylation

Carbohydrates serve several roles in protein folding. First, because of their hydrophilic nature, carbohydrates increase the solubility of the glycoprotein. Second, they generally mark the surface of folding modules and are not buried within them. Third, they make the process of translocation across the ER membrane less reversible by increasing the energy barrier to back-translocation. Apart from these general roles, however, carbohydrates also affect protein folding, albeit in individualized and idiosyncratic ways. Detailed studies with glycoproteins whose Asn-linked glycosylation sites were mutated systematically showed that in many cases glycosylation is needed for proper folding: under-glycosylated proteins form intracellular aggregates and are retained in the ER [30–32]. Somewhat paradoxically, however, almost no individual glycan is necessary for folding of VSV G protein or influenza hemagglutinin [33, 34]. Only one of seven hemagglutinin glycans, attached to Asn81, created a kinetic barrier to folding [32], and even this glycan effect could not be detected in another set of experiments [34]. Among the highly homologous MHC class I molecules, some are sensitive to the presence of glycans [35], whereas others fold equivalently in the presence or absence of the carbohydrates [36]. Furthermore, even proteins that are normally...
not glycosylated can sometimes benefit from the inclusion of ectopic glycans [37]. Thus, the ‘rules’ that govern the interplay between glycosylation and folding are at present still poorly understood. In part, the explanation lies in the observation that aromatic and hydrophilic amino acids are over-represented near those N-glycosylation sites that are important for proper folding [38]. This correlation was shown experimentally to be important because in such ‘aromatic sequons’ the aromatic side chain interacts with the first N-acetylgalcosamine of the glycan, and grafting such sequons onto non-glycosylated proteins increases their stability [39]. Table 2 provides comparative examples from several proteins where some N-glycosylation sites are known to impact, but others are dispensable for folding. This compilation further reinforces the potential relevance of neighboring hydrophobic residues. We suggest that for certain sites, an additional role for attachment of glycans [37]. Thus, the glycan, and grafting such sequons onto non-glycosylated proteins provides a parallel pathway for oxidative folding in the ER [46].

### 3.4. Folding in an oxidizing environment

A general property distinguishing proteins that fold in the ER from cytosolic proteins is the preponderance of disulfide bonds. Cys residues in newly synthesized secretory polypeptides tend to oxidize in the ER lumen because of its high oxidative redox potential. In the past decade there has been a major shift in the understanding of the redox buffer in the ER, from an emphasis on a glutathione-buffering system [20] to the realization that the eukaryotic ER uses a mechanism similar to the bacterial periplasm [40] to maintain its redox potential: a protein-based relay of oxidation/reduction reactions (Reviewed in [41,42]). The relay involves Ero1, a conserved FAD-dependent enzyme, which is oxidized by molecular oxygen and in turn acts as a specific oxidant of protein disulfide isomerase (PDI), which then directly oxidizes disulfide bonds in folding proteins. In addition to Ero1, PDIs can be oxidized by peroxiredoxin IV, which metabolizes the H2O2 formed by the Ero1 reaction, couples this oxidation to disulphide formation [43–45], and provides a parallel pathway for oxidative folding in the ER [46].

### Table 1
Relative of abundance of selected ER proteins.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Function</th>
<th>3T3a</th>
<th>HeLa b</th>
<th>Dendritic cells c</th>
<th>Caco2 d</th>
<th>Pancreatic REd e</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIP = GRP78</td>
<td>HSP70 family chaperone</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>GRP170 = Hpyu1</td>
<td>BIP exchange factor</td>
<td>11.14</td>
<td>12.22</td>
<td>20.02</td>
<td>18.18</td>
<td>12.00</td>
</tr>
<tr>
<td>SI1 = BAP</td>
<td>BIP exchange factor</td>
<td>0.96</td>
<td>0.34</td>
<td>0.13</td>
<td>0.55</td>
<td>0.10</td>
</tr>
<tr>
<td>EDrj3 = DnaJ11</td>
<td>J domain co-chaperone</td>
<td>4.21</td>
<td>7.65</td>
<td>5.38</td>
<td>6.09</td>
<td>5.80</td>
</tr>
<tr>
<td>EDrj5 = DnaJ10</td>
<td>J domain; reductase</td>
<td>1.17</td>
<td>0.89</td>
<td>1.01</td>
<td>0.95</td>
<td>4.00</td>
</tr>
<tr>
<td>PS36pk = DnaJ3</td>
<td>J domain co-chaperone</td>
<td>1.35</td>
<td>3.20</td>
<td>1.94</td>
<td>3.69</td>
<td></td>
</tr>
<tr>
<td>Sec63</td>
<td>Translocon J domain</td>
<td>0.37</td>
<td>9.98</td>
<td>0.61</td>
<td>6.88</td>
<td>39.60</td>
</tr>
<tr>
<td>EDrj1 = MTJ1 = DnaJ Cl</td>
<td>J domain co-chaperone</td>
<td>0.09</td>
<td>0.47</td>
<td>N.D.</td>
<td>1.15</td>
<td>7.20</td>
</tr>
<tr>
<td>HSP47 = Serpin H1</td>
<td>Collagen specific</td>
<td>204.29</td>
<td>62.09</td>
<td>N.D.</td>
<td>81.47</td>
<td></td>
</tr>
<tr>
<td>GRP94</td>
<td>HSP90 family chaperone</td>
<td>63.44</td>
<td>90.96</td>
<td>70.65</td>
<td>123.48</td>
<td></td>
</tr>
<tr>
<td>CNP73</td>
<td>Client-restricted GRP94 co-chaperone</td>
<td>0.94</td>
<td>2.91</td>
<td>2.85</td>
<td>4.44</td>
<td></td>
</tr>
</tbody>
</table>

### Glycoprotein quality control

<table>
<thead>
<tr>
<th>Protein</th>
<th>Function</th>
<th>3T3a</th>
<th>HeLa b</th>
<th>Dendritic cells c</th>
<th>Caco2 d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calnexin</td>
<td>Membrane lectin chaperone</td>
<td>19.76</td>
<td>56.75</td>
<td>17.84</td>
<td>97.65</td>
</tr>
<tr>
<td>Calreticulin</td>
<td>Lumenal lectin chaperone</td>
<td>39.45</td>
<td>129.61</td>
<td>35.40</td>
<td>164.94</td>
</tr>
<tr>
<td>UGTP</td>
<td>Glucosyl-transferase</td>
<td>2.91</td>
<td>3.98</td>
<td>65.11</td>
<td>3.72</td>
</tr>
<tr>
<td>Ribophorin I</td>
<td>Oligosaccharyl transferase complex</td>
<td>17.96</td>
<td>N.D.</td>
<td>14.87</td>
<td>42.70</td>
</tr>
<tr>
<td>Ribophorin II</td>
<td>Oligosaccharyl transferase complex</td>
<td>15.23</td>
<td>24.64</td>
<td>8.54</td>
<td>16.66</td>
</tr>
<tr>
<td>Glucosidase I α</td>
<td>Glycan trimming</td>
<td>14.33</td>
<td>33.94</td>
<td>N.D.</td>
<td>28.24</td>
</tr>
<tr>
<td>Glucosidase I β</td>
<td>Glycan trimming</td>
<td>9.50</td>
<td>32.74</td>
<td>4.76</td>
<td>33.61</td>
</tr>
<tr>
<td>OS-9</td>
<td>ERAD lectin</td>
<td>N.D.</td>
<td>0.31</td>
<td>N.D.</td>
<td>0.71</td>
</tr>
<tr>
<td>Erlentin = XTP3-B</td>
<td>ERAD lectin</td>
<td>0.66</td>
<td>0.00</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td>Mannosidase Man1b1</td>
<td>Glycan trimming</td>
<td>0.18</td>
<td>0.40</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td>EDEM2</td>
<td>ERAD lectin</td>
<td>N.D.</td>
<td>0.04</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>EDEM3</td>
<td>ERAD mannosidase</td>
<td>N.D.</td>
<td>0.28</td>
<td>0.11</td>
<td></td>
</tr>
</tbody>
</table>

### Redox proteins

<table>
<thead>
<tr>
<th>Protein</th>
<th>Function</th>
<th>3T3a</th>
<th>HeLa b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ero1</td>
<td>Redox generation</td>
<td>5.28</td>
<td>0.02</td>
</tr>
<tr>
<td>ERp44</td>
<td>5.30</td>
<td>11.06</td>
<td>5.10</td>
</tr>
<tr>
<td>ERp20</td>
<td>9.37</td>
<td>31.96</td>
<td>18.75</td>
</tr>
<tr>
<td>PDI1 = PDI</td>
<td>Isomerase and chaperone</td>
<td>78.66</td>
<td>109.76</td>
</tr>
<tr>
<td>PDI1 = ERp57</td>
<td>Glycoprotein preference</td>
<td>82.51</td>
<td>70.26</td>
</tr>
<tr>
<td>PDI4 = Erp71</td>
<td>16.09</td>
<td>28.58</td>
<td>15.01</td>
</tr>
<tr>
<td>PDI5 = Pdrr</td>
<td>1.16</td>
<td>1.34</td>
<td>0.13</td>
</tr>
<tr>
<td>PDIb = P5</td>
<td>30.95</td>
<td>21.35</td>
<td>30.34</td>
</tr>
<tr>
<td>Prdx4</td>
<td>Redox generation</td>
<td>18.74</td>
<td>31.90</td>
</tr>
</tbody>
</table>

Data derived from published quantitative proteomic methods. The abundance of BIP/GRP78 in each dataset was set to 100 and the expression values of the other proteins were normalized to it.

a 3T3 proteomic data from [108].
b HeLa cell and Caco2 cell proteomics from [232], using peptide spectral counts. Similar results were reported by [233], based on SILAC quantitation.
c CD4-positive dendritic cells, peak intensity counts reported in [231].
d Values are from [107,234] and are based on purification of proteins.
The use of molecular oxygen as the terminal electron acceptor can lead to oxidative stress through the production of reactive oxygen species and oxidized glutathione. That cellular oxidative stress affects protein folding is shown, for example, by the slower maturation of LDL receptor under high H}_2O}_2 concentrations [47]. Disulfide bonds are very important in dictating folding pathways, because they form covalent interactions that stabilize folding intermediates [48] and severely restrict the landscape of available conformations. Formation of disulfide bonds begins very early in the life of the protein, sometimes as soon as the required Cys residues are available in the lumen [49,50]. On the other hand, some disulfide bonds do not form until much later in the folding sequence, even within the same domain [16,51]. Such dichotomy suggests that in vivo there are mechanisms that selectively delay some oxidation steps. As a further apparent contradiction, proteins can undergo post-translational oxidation and achieve the same native structure as with co-translational oxidation. Addition of reducing agents to the medium of live cells prevented disulfide bond formation in newly synthesized influenza hemagglutinin or asialoglycoprotein receptor and reduced the already oxidized glycoproteins inside the ER. When the reductant was washed out, the reduced proteins rapidly oxidized, folded correctly and assembled [52,53]. These examples show that apparently oxidation follows the same pathway when it occurs post- or co-translationally and underscores the concept that in most cases disulfide bonds serve to stabilize a local fold, achieved by multiple cooperative interactions, rather than to initiate the folding. In support of this view, mutagenesis experiments show that lysozyme can be reasonably well folded even without some disulfide bonds [54] and that if an Ig is made stable enough via genetic engineering, it can withstand the removal of its disulfide bonds [55].

All the above observations also indicate the need for precise control over protein oxidation in the ER. As discussed in Section 5.4, the PDIs are capable of shuffling disulfide bonds, stabilizing proper intermediates, and resolving aberrant disulfide bonds. One indication of the importance of the control over disulfide bonds is that the level of expression of PDIs in pancreatic β cells is proportional to insulin expression, consistent with oxidative folding being an important aspect of this hormone’s biogenesis [56].

### 3.5. Influence of calcium on folding

The ER is a major calcium store in the cell and therefore has to respond to the metabolic needs of the cell by rapid changes in free luminal Ca}_2^{+} levels. This must impact protein folding in the lumen, but the information available is surprisingly superficial. In principle, Ca}_2^{+} can influence folding directly through binding to amino acid’s side chains. Perturbation of Ca}_2^{+} levels also affects folding of luminal proteins indirectly, in a manner more pertinent for this review: several members of the folding machinery in the ER, including GRP94, BiP, calreticulin, ERP72, PIA1A, and PIA6, are Ca}_2^{+} binders (e.g. [57]) and their interactions can be modulated by Ca}_2^{+}, as suggested by interaction experiments [58–60]. The refolding activity of ERP72 and PIA6 [61] and the peptide-binding activity of GRP94 [62] are modulated by Ca}_2^{+}. Whether folding of a given protein is affected by Ca}_2^{+} directly or indirectly, is not easily predictable. Clearly, different secretory proteins respond differently to Ca}_2^{+} changes: depletion of Ca}_2^{+} has no effect on folding or secretion of albumin, but severely inhibits asialoglycoprotein receptor and α1-antitrypsin maturation [63,64]; The initial assembly step of large and small subunits of Heymann nephritis antigenic complex is calcium-dependent, but a later step is not [65]. Under low luminal Ca}_2^{+}, thyroglobulin is retained in the ER, but interestingly, it exits the calnexin/calreticulin cycle prematurely, while its interactions with the other chaperone axis of BiP and GRP94 is stabilized and prolonged [66,67]. Conversely, for other proteins, like TCRα, chelation of Ca}_2^{+} causes release from BiP [60]. Thus, the roles of chaperones can change depending on the Ca}_2^{+} concentration in the ER, but the resulting effects depend on the substrate protein.

### 3.6. A distinct spectrum of folds in secreted proteins

All the constraints on folding in the ER that are discussed here presumably provided evolutionary pressure that favored certain types of folds in secreted proteins and cell-surface receptors. There are several examples of types of domains that are over-represented in secretory proteins and under-represented in other types of proteins. One example is the immunoglobulin fold, which consists of a sandwich of two β-sheets with a Greek key topology and is found in hundreds of proteins of different functions, including integrins, adhesion molecules, interleukin receptors, receptor tyrosine kinases and histocompatibility proteins. The immunoglobulin fold is, however, under-represented among cytoplasmic proteins, and when it is found there, it is one subtype of the immunoglobulin fold (e.g. [68]). A second example is the ligand-binding domain of the low-density lipoprotein receptor. It is ~40 residue long with little recognizable secondary structure organized around a calcium ion [69], stabilized by Cys residues that form three disulfide bonds [69,70]. Variations of this ligand-binding domain are found in a many kinds of cell surface receptors and

### Table 2

<table>
<thead>
<tr>
<th>Protein</th>
<th>Sequence</th>
<th>Folding(\text{a})</th>
<th>Comment</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mouse tyrosinase</td>
<td>VYFN^{RT}</td>
<td>Necessary</td>
<td></td>
<td>[38]</td>
</tr>
<tr>
<td>Mouse tyrosinase</td>
<td>GDN^{TF}</td>
<td>Dispensable</td>
<td>Necessary in human tyrosinase</td>
<td>[235,236]</td>
</tr>
<tr>
<td>Mouse tyrosinase</td>
<td>RTAN^{ES}</td>
<td>Dispensable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mouse tyrosinase</td>
<td>IFNP^{GT}</td>
<td>Necessary</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human MHC class I(p)</td>
<td>CYN^{QOS}</td>
<td>Necessary</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human α3 integrin</td>
<td>AYPFVMN^{QOS}</td>
<td>Impedes</td>
<td>Gain-of-function mutant</td>
<td>[237]</td>
</tr>
<tr>
<td>Rat adenylate cyclase 6</td>
<td>RQIN^{SYS}</td>
<td>Dispensable</td>
<td>But needed for function</td>
<td>[239]</td>
</tr>
<tr>
<td>Rat adenylate cyclase 6</td>
<td>ASSN^{PET}</td>
<td>Dispensable</td>
<td>But needed for function</td>
<td>[240]</td>
</tr>
<tr>
<td>Human PDA2</td>
<td>TLFIRNKG{272}RT</td>
<td>Important</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human PDA2</td>
<td>LNITTELFVMN{246}QT</td>
<td>Important</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human PDA2</td>
<td>AAMPFERPA{137}ST</td>
<td>Dispensable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rat ICAM-5</td>
<td>RGGSLWNV{4}CS</td>
<td>Necessary</td>
<td></td>
<td>[241]</td>
</tr>
<tr>
<td>Rat ICAM-5</td>
<td>ETLKIR{57}GT</td>
<td>Dispensable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human chymotrypsin C</td>
<td>SLY省政府{57}GT</td>
<td>Necessary</td>
<td>Not for function</td>
<td>[242]</td>
</tr>
<tr>
<td>Human chymotrypsin C</td>
<td>RCLD{229}CGS</td>
<td>Dispensable</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(\text{a}\) Designation of the glycosylation site as necessary for proper folding, important but not essential, impeding or dispensible.

\(\text{b}\) The site in human HLA; rodent class I are functional without their N-glycosylation. Note that all sites are farther than 50 amino acids from the N-terminus, satisfying the prediction made in [166].
conversely are underrepresented in cytosolic proteins, presumably because of the requirement for oxidation of cysteines. Yet a third example is the EGF-like domains, present on a large number of membrane or extracellular proteins, but not in cytosolic proteins. These 30–45 residue-long domains consist of a small β sheet and a flexible loop, with three disulfide bonds stabilizing the structure [71].

4. Folding intermediates in vitro and in vivo

In vitro folding experiments often describe proteins with folding intermediates at equilibrium, which can be trapped by non-physiological conditions, like drastic pH shifts. In contrast, folding in vivo is often made non-reversible by post-translational modifications, such as proline or disulfide isomerization, which are driven by enzymes. Tyrosinase, a glycoprotein where the carbohydrates are essential for the native fold, exhibits several inactive intermediates, at least two of which are recognized by calnexin. If the association with calnexin is prevented, folding is actually more rapid, but the resulting protein fails to bind copper and is inactive [72]. In the large family of proteinase inhibitors, some intermediates are kinetically trapped and previously acquired disulfide-linked structures need to unfold in order to progress to the native state [73].

Whether through chaperone action or via the biophysical conditions that exist in the ER, restriction and simplification of the folding pathway is sometimes inherent in the protein: while exponential functions are required to describe the folding kinetics of phosphoglycerate kinase in vitro, when this cytosolic protein is engineered to fold in the ER, the measured folding kinetics are more two-state-like, reflecting the modifications imposed by the intra-luminal conditions [74].

The in vivo conditions sometimes dictate a folding pathway that is not only simpler but actually different from that observed in vitro. For example, although each of the two domains of immunoglobulin light chains can fold autonomously when expressed alone, during immunoglobulin biosynthesis in vivo, the variable domain, which is N-terminal and emerges first, folds after the constant domain [75]. The reason is that BiP continues to engage its V domain peptide binding sites and, if this interaction persists, the light chain folding intermediate is targeted for ER-Associated Degradation (ERAD) [12,13]. Most light chain variants that form amyloids are mutated in the variable domains and do not complete folding in the cell. The partner heavy chain also has an important in vivo intermediate where the first constant domain is slow to fold [76]. Folding of this domain is coupled to assembly with the light chain, illustrating the importance of folding intermediates in vivo. Neither of the above folding intermediates would have been predicted by in vitro techniques and they illustrate how the complex luminal environment dictates the path to productive folding.

The insulin fold family provides an example for a related concept: intermediate structures necessary for attaining the native state, but not for activity of the protein. The N-terminal 8 amino acids of the B chain dictate the foldability of proinsulin, but once the native state is achieved, this peptide is dispensable for native structure, activity, or stability of mature insulin [77,78]. Thus, the folding process could, in itself, provide an evolutionary pressure for sequence conservation.

4.1. Oligomeric protein assembly

Once domains and polypeptides fold, many proteins require assembly of subunits before they are biologically active and are competent for export from the ER. Native viral glycoproteins often form non-covalent trimers that assemble in the ER [79] in an ATP-dependent fashion [80]. B and T cell receptors or MHC proteins assemble from 2 to 7 subunits into the biologically active entity before they exit the ER. In either case, there is at least one subunit whose folding is a pre-requisite for assembly: in the case of B cell receptor, the folding of the heavy chain is completed only upon subunit assembly [81,82] and in the case of the T cell receptor—only upon folding of the CD3 epsilon subunit [83]. A similar situation is observed for nicotinic acetylcholine receptor and likely many other receptors [84]: by stabilizing and sequestering subunits during assembly, chaperones like calnexin, BiP and ERP57 regulate the levels of assembled functional receptors. In general, unassembled subunits either aggregate or are targeted to degradation via the ERAD or autophagy pathways, but are not transported to the Golgi complex.

Intermediates do not only define the folding of protein domains or polypeptide chains, but are also characteristic of oligomeric assembly of many proteins. For example, the Heymann nephritis antigenic complex assembles in two discrete stages. First, a large glycoprotein (gp330) associates with a 44 kDa subunit early after synthesis, and then, >60 min after synthesis, a larger hetero-oligomer forms, still before acquisition of Endo H resistance [65].

What features enable the cell to discriminate between free subunits and assembled oligomers? No global answer can yet be provided, but in individual cases there are examples of features that may operate. As mentioned above, during Ig assembly there is a mechanism for coupling the folding of a ‘sentinel’ Cμ1 domain whose folding is retarded [85], to assembly of heavy and light chains by coordinating the formation an intra-domain disulfide bonds with a disulfide linking the two chains [51,81], thus releasing chaperones and folding enzymes [85,86], and see below). T cell receptor assembly depends heavily on hetero-pairing of subunits via transmembrane domains that contain charged amino acids [87–90]. If left unpaired, these subunits do not assemble and are dislocated from the ER membrane. The presence of unsatisfied intramembrane charge prevents stabilization of the unassembled subunits via homo-oligomerization and maintains them in a retrotranslocation-competent state [91]. The unassembled α chains can be retained in the ER by BiP, in Ca++-dependent fashion [60].

In glutamate transporters, an evolutionarily conserved arginine-based motif acts as an ER retention signal, while a luminal leucine motif is required for suppression of this signal and allows traffic to the Golgi complex [92]. This motif is likely a subject to conformation modulation that could occur upon subunit assembly. Another case of conformational control in the ER is the maturation of AMPA receptors to acquire ligand-binding activity [93]. The extracellular region that distinguishes isoforms of AMPA receptor can interact with binding proteins such as stargazin, or with other luminal isoform-specific proteins, to determine whether a transport-incompetent subunit has assembled sufficiently to traffic to the cell surface [94].

5. Chaperone networks in the ER

The complex demands of orchestrating the folding of secreted and membrane proteins require at least 10 ubiquitously expressed chaperones and some client-specific ones (Table 1, and other reviews in this issue). These chaperones do not generally work alone, but rather fulfill each task by forming transient networks of chaperones, cofactors and folding enzymes that operate in concert. The composition of these networks appears to adapt to the needs of the folding clients.

5.1. The BiP network

5.1.1. Client recognition

The most abundant chaperone network and the one involved with most client proteins is the BiP/GRP78 network (Table 1). BiP is a multifunctional chaperone, involved in nascent chain translocation, folding of polypeptides, resolving and removal of misfolded proteins and monitoring the ER stress transducers (entries in [95]). Consistent with these multiple functions and the variety of known client proteins, BiP is an essential protein. In the cell, it is usually not possible to reduce BiP expression below 40% of its normal level [96] (Argon, unpublished data), and in the whole organism, BiP deletion leads to
very early embryonic lethality at the pre-implantation stage [96]. The primary biochemical activity of BiP/GRP78, which unifies all these functions, is binding to 7–11 residue peptides that in the folded state form β strands and whose sequences tend to have alternate hydrophobic residues [97,98]. These residues are buried in the interior of the native fold but are exposed in early intermediates or in misfolded proteins [12,99]. The inherent affinity of BiP for peptides is in the range of 1 to 100 μM, as appropriate for a chaperone that had evolved for transient interactions. Higher affinity interactions would have inhibited folding altogether. It is estimated that a BiP-binding site would be created in a random sequence once every 36 amino acids [98]. This frequency would explain why BiP has a large number of client proteins. Yet, domains that have predicted BiP-binding sites do not always employ them to a significant degree in the cell, and this accounts for a chaperone-directed folding sequence in which BiP dictates the path of folding [13,100]. This can be exemplified by BiP binding to Ig light chain molecules: although light chain is predicted to have >5 binding sites in each of the variable (V\(\text{L}\)) or constant (V\(\text{C}\)) domains [12], BiP only binds unoxidized V\(\text{L}\) domain in cells, with no detectable binding to the constant domain [86,101] (Argov, unpublished data). Furthermore, two out of the potential five binding sites in V\(\text{L}\) are sufficient to explain the BiP-light chain interaction [12]. A combination of in vivo and in vitro folding studies revealed that the constant domain folds rapidly and stably even in the absence of an intradomain disulfide bond [75,86].

Thus, the simple presence of a BiP-binding site on a nascent chain does not ensure that BiP will bind and play a role in its folding. Instead, it appears that the rate and stability of protein folding determines whether or not a particular site is recognized, with BiP preferentially binding to proteins and domains that fold slowly or are less stable.

5.1.2. Regulation of ATPase cycle

The action of BiP in folding reactions, like its activity in other processes, requires nucleotides. Since there is no in situ ATP generator, the ER lumen depends on mitochondrial ATP that is imported into the ER via a 56 kDa permease [102,103]. In the lumen, ATP is needed for the action of various proteins. Binding of BiP to clients is very fast in the ATP-bound state, but hydrolysis must occur to stabilize the binding. Exchange of ADP to ATP is then needed to accelerate the release of the client protein. The peptide-binding and release activity of BiP thus requires its ATPase cycle [104], but, because BiP’s inherent ATPase activity is low and the spontaneous transition between the two states is extremely slow, its ATPase cycle depends on cofactors.

At least 7 proteins can modulate BiP’s ATPase cycle: Four HSP40-family proteins containing a J domain, two nucleotide exchange factors (Si11/BAP and GRP170/ROP150), and ERdj5, a PDI that also possesses a J domain. The J domain proteins stimulate BiP’s ATP hydrolysis and thus enhance BiP’s interaction with misfolded proteins, while Si11 and GRP170 exchange ADP for ATP so that BiP can dissociate from them more readily. The combined action of J proteins and nucleotide exchange factors creates a cycle of binding and release that is key to the mode of action of the BiP machine (reviewed in [105]).

GRP170 can form a stable complex with BiP in the absence of ATP, mediated through contacts between their nucleotide binding domains [106]. This association decreases the affinity of nucleotide binding to BiP but does not stimulate its ATPase activity [106]. It is noteworthy that there is a 9-fold excess of BiP over GRP170 and an even larger excess over Si11 or each of the J domain proteins [107,108] (Table 1), underscoring that these cofactors may be limiting components and must operate catalytically. Alternatively the cofactors may be only needed for a very limited set of substrates, but this seems unlikely, since the BiP cycle and not stable association seem to be the norm. Although GRP170 is an HSP70 family protein and has domains that can bind peptides [109,110], no direct role has yet been demonstrated for GRP170 in folding client proteins in vivo, though it is associated with immunoglobulins in B cells [111] and with misfolded mutants in other cells [112]. These interactions of GRP170 with both BiP and misfolded species may serve to target BiP to a specific pool of client proteins.

J domain proteins may also adapt the BiP complex to the demands of the specific task. Thus, ERdj3, can bind directly to several nascent, unfolded and mutant secretory proteins, even without BiP [113]. After ERdj3-client binding, BiP joins the complex, which leads to dissociation of ERdj3 and the client polypeptide and concomitant stimulation of BiP’s ATPase activity before folding is completed [113]. The higher ATPase activity converts BiP to its high-affinity state for clients, so ERdj3 seems to present the client for BiP-dependent folding [114]. Similarly, PS8(IPK)/DnaJC3 presents other clients for BiP-dependent folding [115]. On the other hand, binding of the J protein ERdj5 to misfolded α1-antitrypsin, with its subsequent recruitment of BiP and activation of its ATPase, leads to reduction of the client’s disulfide bond by the PDI activity of ERdj5 and its subsequent targeting by BiP and EDEM1 to disposal via ERAD [116]. How different J domain proteins target BiP is not understood. Although the various J proteins can bind alternatively to the same surface on BiP, the exact modes of binding are not identical, since a mutation in this binding region of BiP abolishes interactions with some J proteins but not with others [117].

5.1.3. Functional consequences of BiP binding

The usual activity of BiP involves ATP-dependent cycles of binding and release of the substrates. The cycling is mediated by the continual exposure of hydrophobic BiP binding sites on the immature folding intermediates. One of the exceptions to the ‘canonical’ binding and release action cycle is the BiP-substrate interaction that is sufficiently prolonged so as to be stable under physiological conditions. A prime example is BiP interaction with the C\(\text{L}\)\(\text{1}\) domain of Ig heavy chains [85].

This domain is intrinsically disordered in vitro, which sets it apart from other Ig domains. C\(\text{L}\)\(\text{1}\) folds only upon interaction with the C\(\text{I}\) domain of the Ig light chain. Structure formation proceeds via a trapped intermediate and can be accelerated in vitro by the ER-specific peptidyl-prolyl isomerase cyclophilin B [85]. BiP recognizes incompletely folded states of the C\(\text{L}\)\(\text{1}\) domain and competes for binding to the C\(\text{I}\) domain. Unlike BiP binding to many other sites in the Ig molecule [12,13], its binding to the C\(\text{L}\)\(\text{1}\) domain is persistent, ensuring that C\(\text{L}\)\(\text{1}\) remains unoxidized long after other domains have folded [81]. The ‘purpose’ of this built-in delay is to coordinate the formation of the intradomain disulfide with that of the inter-subunit disulfide bond [51,82], a key step that links completion of the heavy chain subunit folding to the assembly of heavy and light chains [51,85]. In vivo experiments demonstrate that these steps, including association with a folded C\(\text{I}\) domain and isomerization of a conserved proline residue, are essential for antibody assembly and secretion in the cell [85]. When chain assembly cannot be completed because the right partner is absent, or when a protein cannot fold due to mutation, the persistent BiP association leads to retention and/or ERAD [51,76,81,118]. The retention of incompletely folded proteins is therefore a consequence of the normal action of the BiP system. Nonetheless, what determines persistent BiP binding to the sites in C\(\text{L}\)\(\text{1}\) vs. short-lived interactions with other sites on the same molecule is not yet known.

In addition to the folding trajectories of their client proteins, chaperones like BiP can influence their aggregation pathways [119]. A mechanistic insight into this process is provided by considering the role of BiP in the folding and aggregation of the amyloidogenic mutant immunoglobulin light chain [120]. Of the two dominant BiP-binding sites important for the folding of V\(\text{L}\), one is directly involved in the formation of amyloid fibrils, as the synthetic peptide containing this site specifically inhibits V\(\text{L}\) amyloidogenesis [12,13]. Thus, binding to the same binding site mediates BiP’s function in promoting the efficient folding of V\(\text{L}\) in the cell and in the prevention of
amyloid formation by its mutant forms. This conclusion is echoed by the reactivity of the anti-oligomer antibody A11, which, in addition to recognizing aggregated forms of several dissimilar proteins, also recognizes a subset of molecular chaperones, including members of the HSP70 family [121]. A11 binding to these chaperones can interfere with the suppression of aggregation or the refolding of their substrates [121]. Therefore, increased competition for such a chaperone from other substrates, particularly if uncompensated for by stress response, may favor a shift from productive folding toward amyloidogenesis and the formation of toxic species by the disease-associated protein.

5.1.4. Organisinal roles of the BiP network

While BiP is essential in all cells, none of its associated J proteins seem to be essential in metazoa, and as far as we know none is a generalist J protein. Though some of them can bind misfolded poly-peptides directly (e.g. ERdj3 [113] or MTJ1 [122]), others function without client binding, and yet others, like Sec63, are specific not for protein folding, but rather for translocation across the ER membrane. Considering the important role of the J domain proteins in regulating BiP’s ATPase cycle, deletions of two J proteins in the mouse have surprisingly specific phenotypes. Deletion of ERdj5 leads to a pronounced salivary gland phenotype, consistent with failure of folding of amylose that induces persistent ER stress response [123]. This phenotype of ERdj5 −/− mice is very tissue-specific, since plasma cells or pancreatic β cells, which are also professional secretory cells, are not affected significantly. On the other hand, deletion of mouse PS8(IPK) has a β cell-specific phenotype [124]. The absence of phenotype in unaffected cells may be due to compensation for the loss of specific J domain protein by an unknown component, potentially pointing to a redundancy in the J domain proteins. Alternatively, the cell-specific dependence on a given J domain protein may be dictated by the specific protein clients expressed in these cells, pointing to much more specialized roles of these proteins.

Mutations in the SiIL nucleotide exchange factor [125] cause Marinesco–Sjögren syndrome (MSS) in humans [126,127] and ataxia and neurodegeneration in woozy mouse (wz) [128]. Since SiIL, like BiP and the J proteins, is expressed ubiquitously, this pathological presentation suggests selective vulnerability of certain cell types and tissues, most notably the cerebellum. In wz mice, loss of SiIL causes ER stress and accumulation of ubiquitinated proteins in cerebellar Purkinje cells, leading to degeneration and apoptosis [128]. Furthermore, even within Purkinje cells there is differential sensitivity to loss of this chaperone, as Purkinje cells in the vestibulocerebellum of wz mice are spared. Since SiIL functions to regulate the ADP to ATP exchange, and thus the chaperone cycle of BiP, the reasons for such selective vulnerability are again not immediately obvious. One potential explanation is that another BiP nucleotide exchange factor, GRP170/ORP150 (Lhs1) may be redundant with SiIL and thus compensates for its absence in protected cells [129]. However, ORP150 is upregulated in affected Purkinje cells of the wz mouse, so this upregulation is apparently insufficient to protect the cells [130]. On the other hand, early ectopic overexpression of ORP150 rescues ER stress, protein aggregation, and neurodegeneration, while its down-regulation exacerbates these phenotypes [130].

A fascinating result is that inactivation of one of the DnaJ proteins, PS8(IPK)/DnaJc3, which promotes the ATPase activity and substrate loading of BiP [115,131], also attenuated the phenotypes of SiIL −/− mice [130]. This result supports the view that disturbance in the BiP chaperone cycle is a cause of the neurodegeneration and other phenotypes in the wz mouse (and by extension, in MSS patients). The simultaneous absence of both the ATPase stimulating activity of PS8(IPK) and the nucleotide exchange activity of SiIL is predicted to decrease cycling of BiP, at least with respect to the specific set of substrates that the two co-chaperones may recognize. Thus, rescue of SiIL deficiency by deletion of PS8(IPK) points to the imbalance between the two opposing regulators of the BiP chaperone cycle as the potential etiological factor, rather than the deficiency of BiP chaperone activity per se. Since the ERdj5 knockout was not tested for genetic interaction with SiIL, it is not yet known whether the rescue property of the PS8(IPK)-deficient strain is unique to this J protein, or is shared by all.

Another non-obvious result is that deletion of PS8(IPK) by itself, unlike deletion of SiIL, does not cause brain pathology or ataxia, suggesting that sensitivity to the imbalance between J proteins and nucleotide exchange factors may be one-sided and cell-specific. The deletion of PS8(IPK) does cause pancreatic β cell failure, due to persistent UPR [124], but the ability of SiIL deficiency to mitigate the β cell phenotype was not reported [130]. Thus, the sensitivity to the imbalance between J proteins and exchange factors may be one-sided.

In addition to the differential sensitivity of Purkinje cells to SiIL depletion, there is evidence that they are also generally intolerant of protein misfolding in the ER. A missense mutation (Ser658 to Pro) in a ubiquitously expressed ERAD protein, SEL1L, in dogs leads to selective cerebellum-restricted neurodegeneration with marked loss of Purkinje cells [132]. As expected based on the role of SEL1L in disposal of misfolded proteins from ER, the affected Purkinje cells activated strongly the ER stress response [132]. Yet a third example of selective sensitivity of these cells can be seen in a mouse carrying the sticky (stf) mutation—a missense mutation in the editing domain of the alanyl-tRNA synthetase gene, which leads to mistranslation and possibly misfolding of many cellular proteins [133]. For unknown reasons, this mutation affected Purkinje cells selectively, causing accumulated ubiquitinated inclusions, increased expression of cytoplasmic chaperones and also activated the ER stress response. Persistent CHOP expression, which is thought to be associated with pathological levels of ER stress, was seen through the onset of degeneration, suggesting that perhaps the low but chronic levels of protein misfolding were not tolerated in these susceptible cells [133]. Therefore, the conspicuous sensitivity of Purkinje cells to slowing the BiP cycle may be indicative of their increased sensitivity to altered proteinostasis, as in the examples of mistranslation and defective ERAD.

5.2. GRP94

5.2.1. Client recognition

A chaperone that is often associated with the BiP complex is GRP94, the single HSP90 family representative in the ER [134]. When working on a folding pathway together with BiP, GRP94 usually is the second chaperone in the sequence, thought to engage late folding intermediates (e.g., Ig [11,135,136], Tg mutant [137,138]), and this distinction seems to be due to different structural cues to which GRP94 is sensitive. Unlike extended peptides with specific sequence features that are recognized by BiP, no sequence or structural motives have emerged for GRP94. This may simply reflect the relatively restricted number of known GRP94 client proteins; alternatively, GRP94 may have a more complex recognition mode, similar to the HSP90’s ability to recognize metastable and near-native domains. The rules of engaging GRP94 with clients are also still unknown; the chaperone activity of GRP94 in vivo requires ATP [139], but unlike the BiP complex, no accelerator of ATPase nor an exchange factor has been discovered for GRP94. The only identified co-factor for GRP94 so far is CNYP3, which is widely expressed (Table 1), but is a client-specific co-chaperone that does not modulate the ATPase cycle of GRP94 [140].

The order of BiP-GRP94 interaction mirrors the HSP70–HSP90 axis in folding of cytoplasmic clients [141], and the position of GRP94 as the second in the sequence likely accounts for its restricted clientele. BiP’s sequence-dependent recognition of slow folding or unstable domains may account for most of the chaperone–dependent secretory proteins that are not engaged with the calnexin–calreticulin cycle; only a subset of these proteins is subsequently engaged by GRP94, guided by the yet unknown thermodynamic or structural features.
Such sequential action of chaperones provides a way to restrict the folding options of a client and direct it to a preferred pathway. While it is not clear whether GRP94 interacts physically with BiP, there is obviously a strong functional interaction between them, as demonstrated by the upregulation of expression of either chaperone when the other is silenced genetically or inhibited pharmacologically [142].

5.2.2. Organismal roles of GRP94

The GRP94−/− mouse provides a remarkable example of specific, client-restricted phenotype due to the lack of a chaperone. The embryonic lethality at E6.5−7.5 coincides with the developmental time when mesoderm induction occurs and when IGF-II, a client which is totally dependent on GRP94 activity is first expressed [143]. Another client, Ig, is not so strictly dependent on GRP94, and indeed targeted deletion of the chaperone in B lineage cells does not appreciably depress the circulating levels of antibodies [144]. The phenotypes of tissue-specific deletion of GRP94 also reflect its client interactions: deletion in skeletal muscle mostly affects the muscular growth pathway that depends on IGFs [145], and thrombocytopenia due to deletion in hematopoietic cells is caused by the effect on platelet glycoprotein Ib-IX-V complex [146]. Lack of GRP94 in a B cell line leads to a failure to respond to bacterial endotoxins, due to a selective defect in maturation of Toll-like receptors and integrins [144,147]. In all of these studies, GRP94 was not required for the cellular viability, which again is consistent with its selectivity.

Like GRP94, at least one other ER chaperone recognizes advanced folding intermediates. HSP47 preferentially recognizes Gly-X-Y repeats in the triple helices of various collagens, in a rather folded conformation [150]. Unlike GRP94, HSP47 does not require ATP and may dissociate from collagen in a pH-dependent manner [148], though it is not clear how this would be regulated in vivo. Ablation of the mouse gene for HSP47 leads to embryonic lethality at E11.5, accompanied by defective collagen biosynthesis [151], which is needed for mesenchymal tissues. These and other data are consistent with HSP47 being a collagen-specific chaperone.

5.3. The calnexin/calreticulin network

The ER has one unique type of chaperone—lectins that bind to glycoproteins. Two such chaperones are known: the luminal protein calreticulin (Crt) and the membrane-spanning chaperone calnexin (Cnx) [152]. Both prefer a glycoprotein whose glycan has a monoglycosylated intermediate, as a result of trimming after the initial glycosylation [153−155], although they also display distinct sensitivities to folding context of individual glycans [32,156−158]. There is also evidence that calnexin and calreticulin recognize features of proteins other than glycans. Dissecting intermediates in hemagglutinin, Hebert et al. observed that in contrast to Cnx, Crt bound primarily to early folding intermediates. Though the two chaperones share the same carbohydrate specificity, Crt binding depends on the oligosaccharides in the more rapidly folding top domain of HA, whereas calnexin is less discriminating [32]. Distinct epitopes also trigger calnexin and calreticulin binding in MHC class I heavy chains [159]. Another major difference between Crt and Cnx is the topology of their clients. Calreticulin selectively interacts with nascent luminal secretory proteins, such as transferring; however, when calreticulin was made artificially membrane-anchored, the spectrum of proteins it recognized became remarkably similar to that observed with calnexin [156]. Conversely, calnexin’s binding to its membrane-spanning client hemagglutinin was reduced when the latter was expressed as a soluble anchor-free protein [32].

5.3.1. Recruitment into the cycle

Calnexin and calreticulin participate in the dynamic sorting of immature proteins towards attainment of the native state vs. elimination by ERAD, in the so-called “calnexin/calreticulin cycle”. The sorting process consists of series of decision and commitment events. The first commitment event in the cycle is represented by the sequential action of oligosaccharyl-transferase complex (OST) and glucosidases I and II, which generates a mono-glucosylated protein intermediate [153,156,159−163]. Because OST and glucosidase I are localized to the translocon [164,165], the presence of a glycosylation site within N-terminal 50 residues ensures attachment of the glycan moiety early in the translocation process, thus directing the nascent protein into the Cnx/Crt cycle [166]. If the first glycosylation site is more C-terminal, the nascent protein may associate with BiP instead and proceed along a different chaperone-assisted folding path. The decision between the two chaperone complexes appears to be due to competition, since removing the glycosylation sites allows the protein that is normally a Cnx/Crt substrate to be recognized by BiP [166]. Cnx/Crt recruits into the complex disulfide isomerase ERp57 [167−170], which facilitates formation of the disulfide bonds in the immature glycoproteins.

5.3.2. Return to the cycle

Once the protein is released from the calnexin/calreticulin complex and further glucose-trimmed by glucosidase II to a high-mannose intermediate, it is subject to a second triage decision. If it is in a mature, secretion-competent conformation, it is targeted to the ER exit sites and proceeds to its destination. If, however, it is in a non-native conformation, it can be recognized by the folding sensor UDP-glucose: glycoprotein glucosyl transferase (UGT1) which re-glucosylates the high-mannose moiety, thus re-committing the incompletely folded polypeptide to the calnexin/calreticulin cycle for another round [153].

What proportion of glycoproteins is recommitted to repeated engagement with Cnx/Crt is unknown, and in three out of four cell types studied, UGT1 is quite sub-stoichiometric to both Cnx/Crt and glucosidase II (Table 1).

In addition to UGT1, the high-mannose intermediate can be recognized by α-mannosidase I, resulting in the trimming of the mannosyl moiety [171−173]. This step increases the chances of the polypeptide’s exit from the Cnx/Crt cycle, but does not commit it to exit. Further mannose trimming, for example by EDEM1 and EDEM3 [174,175], produces a protein that can no longer be re-glucosylated by UGT1, thus essentially removing it from the further refolding attempts, and instead committing it to the degradation process [173,176]. The relatively slow kinetics of α-mannosidase I enzymatic activity is thought to provide a timer [171] that allows for a certain residence time in the refolding part of the Cnx/Crt cycle, thus eliminating futile cycling of an intermediate that is unable to fold. Recognition of both—the trimmed mannose residues and the non-native polypeptide conformation by ERAD lectins such as OS9 and XTP3-B, is thought to be the final signal for degradation [177].

5.3.3. Client recognition

While we have a good idea of the sequence of events and identities of proteins in the calnexin cycle, there is much to be understood about both the recognition of the polypeptide by the different chaperones involved, and the consequences of their binding for the conformational maturation of the polypeptide. The first triage step—the attachment of the glycan to nascent protein emerging from the translocation channel—is unlikely to be driven by global conformational information. Local conformation, on the other hand, could affect the recognition of the glycosylation site by OST, since proline residues are not present at position +1 following the Asn, and are very rare in position +3 [178]. The binding by Cnx/Crt in vivo may be mediated by the specific glycan configuration, even though in vitro Cnx is able to bind unfolded, unglycosylated proteins [158]. For example, when glucose trimming is slowed down, the mature, folded forms of HA protein can be found bound to calnexin [163], suggesting that only the sugar residues were mediating the binding in this case. On
the other hand, complete inhibition of glucose trimming has been shown to induce a prolonged association of a normally Cnx/Crt substrate (p62 viral glycoprotein) with BiP [166]. Given that BiP preferentially binds immature folding intermediates, selection into Cnx/Crt cycle is not likely to be based on the folding state of the polypeptide.

The conformational maturation of polypeptides released from Cnx/Crt is accessed by UGT1. The structural basis for this triage step is still not well understood. Different modes of recognition have been reported for UGT1 in vitro, from near-native conformation of chymotrypsin inhibitor-2 [179], to non-native conformations of several proteins with exposed hydrophobic residues on the surface [180], to small, local deviations from the native fold [181]. In plants, inactivating mutation of UGT1 results in less stringent ER retention of a biochemically active brassinosteroid receptor mutant harboring localized structural distortion [182]. It is possible that the degree of non-nativeness recognized by this folding sensor is dependent on, or influenced by, other, perhaps protein-specific, structural or conformational information. Clearly, a better characterization of the substrate recognition by UGT1 will be necessary for understanding the principles of triage of folding intermediates in the ER.

When the above triage mechanism is not sufficient, for example under ER stress conditions, an additional lectin-based mechanism exists. The ER membrane protein malectin specifically binds the Glc3Man9GlcNAc2 intermediate glycan, presumably formed during the glucosylation process, through a novel carbohydrate-binding site in its luminal domain [183]. Glycoprotein recognition by malectin does not affect their entry to the Cnx/Crt chaperone system and malectin binds after Cnx [184]. Malectin is induced by ER stress and associates preferentially with unfolded conformers of glycoproteins, which is consistent with a role as a backup to the lectin chaperone system that would be activated when misfolded forms accumulate in the ER [184].

5.3.4. Functional consequences of lectin chaperone binding

What is the contribution of Cnx/Crt binding to the folding of their substrates? One straightforward explanation, common to all chaperones interacting with early intermediates, is preventing aggregation of immature proteins [185]. The charged nature of glycans may in itself serve an anti-aggregation function. Second, Cnx/Crt binding mediates the recruitment of ERP57, which monitors the proper disulfide bond structure of the glycoprotein (reviewed in [186]).

Another possibility is that attachment of the glycan residue and binding of the lectin chaperones anchor the surrounding amino acids to the surface of the protein. In addition to reducing the conformational freedom of peptide backbone around the glycosylation site, such anchoring may allow conformations where amino acids that are normally buried in the core remain on the surface, without compromising the protein stability. Examination of 506 glycoprotein crystals showed enrichment for surface-exposed aromatic and hydrophobic residues in close spatial proximity to the N-glycosylation sites [38]. Petrescu et al. [187] propose that without attached glycans, these residues will seek the hydrophobic core of the molecule, thus favoring a different conformation. For example, among four used glycosylation sites of mouse tyrosinase, two are dispensable and two are necessary for efficient folding/function. The dispensable sites are GDEN203FT and RTAN237FS (see Table 2), while the two whose occupancy is important for folding efficiency are VYFN208RT and IFMN227GT, with neighboring aromatic side chains. In fact, other studies suggest that glycans at certain positions can directly protect hydrophobic residues. In recombinant erythropoietin, the inner regions of highly branched glycans appear to stabilize the mature protein conformation by clamping to the hydrophobic protein surface areas [188], while in an unassembled alpha subunit of human chorionic gonadotropin, a glycan residue directly shields the hydrophobic core region that is normally protected by interaction with the beta subunit [189].

5.3.5. Organisal roles of the chaperones of the calnexin/calreticulin cycle

The fundamental role of monitoring the quality of glycoproteins would predict that the calnexin/calreticulin cycle is essential, but perhaps the overlapping specificity of calnexin and calreticulin would render each one of them dispensable. This expectation is born out in tissue culture cells [190] and, surprisingly, also in knock-out animals. Two different Cnx−/− mouse models had normal embryonic development, but exhibited defective postnatal growth and neurological and behavioral deficits, accompanied by myelination-related defects [176,191]. On the other hand, calreticulin−/− mice [192] showed variably early embryonic lethality, with notable defects in heart development and function and a failure to absorb the umbilical hernia, they had no other gross morphological changes. At least in part, the heart function defects could be attributed to the calcium-buffering function of calreticulin rather than to its chaperone activity, since Crt−/− embryos had inhibited bradykinin-induced Ca2+ release by the InsP3-dependent pathway [192].

The restricted phenotypes and normal development of most tissues in these knock-out animals suggest, despite the distinctions in client recognition between Crt and Cnx, as discussed above, that for the majority of clients, the lectin chaperone requirement for their efficient folding is relatively relaxed. Alternatively, there may be sufficient ability in the non-lectin chaperone machinery to support the folding of many, but not all clients. The specific neurological phenotypes, myelination defects, and heart defects in knock-out animals suggest specific, non-redundant requirements for either calnexin or calreticulin function. At present, we don’t know whether this requirement reflects a heightened chaperone dependency of the secretory proteome in the susceptible cells (for example, the presence of substrates that are highly dependent on calnexin or calreticulin, or the absence of compensating chaperones), or perhaps an increased sensitivity of these cells to protein misfolding (or decreased calcium buffering) in the ER.

5.4. Protein disulfide and prolyl isomerases

In addition to chaperones, which do not directly alter the structure of itinerant proteins in the ER, the lumen is rich with enzymes that can change protein structure. Two of the important types of enzymes are protein disulfide isomerases (PDI) and prolyl isomerases (PPI), and each type is represented by multiple family members (Table 1). As discussed above, the importance of dealing with the constraint of an oxidizing luminal environment provides obvious roles to several PDIs. However, what is the significance of such a large number of related proteins with thioredoxin domains? Do they have unique properties, or are many of them redundant? Similarly, among the PPIs, do all of them function as prolyl isomerases and do they have distinct substrates and functions?

5.4.1. The ER PDIs

At last count, there are 21 ER proteins with one or more thioredoxin-like domains, the defining feature of the PDI family [193]. Each thioredoxin domain is a platform for the CXXC active site that mediates the electron transfer involved in forming disulfide bonds (see chapter 6 and [42]). The PDIs are oxidoreductases, which can oxidize, reduce or isomerise disulfides, depending on the redox potential. Some PDIs, like Erdj5, are reductases in vivo and some, like Erp29, are catalytically inactive. In addition, PDIs are also involved in other activities: the highly abundant PDIa1 is the β subunit of prolyl 4-hydroxylase [194] and Erp44 binds to IP3 receptor [195]. The most abundant PDIs in a number of distinct cell types are the 60 kDa protein disulfide isomerase PDIa1 and two structurally-related proteins, P5 and Erp57 (Table 1).

The relevance of these PDIs to the fate of secretory proteins has been established using pulse-chase experiments in the presence or
absence of individual PDIs. For example α-fetoprotein can still form disulfide bonds in the absence of PDI or Erp57, but its traffic to the Golgi is delayed, consistent with the presence of non-native folding intermediates that may require a disulfide isomerization reaction [196]. Yet, given the large number of ER PDIs, an important question is to what extent are they functionally redundant.

Genetic insight into the redundancy question is provided by the phenotypes of null alleles of the three PDIs in Caenorhabditis elegans: two of them are not essential, but the third is non-redundant and is required for prolyl 4-hydroxylase activity in collagen biosynthesis [197]. Williams et al. used an RNAi knockdown approach to address the same question in mammalian cells [198]. They showed that some PDIs clearly have defined tasks. Pdia1 depletion impacted oxidative folding of each one of several well-characterized secretory liver proteins. However, the phenotype was surprisingly modest, suggesting that other PDIs can compensate for Pdia1 depletion, albeit with lower efficacy. In contrast, depletion of Erp72 or P5, either alone or in combination with PDI or Erp57, had minimal impact on oxidative folding [196]. The RNAi approach also showed that one PDI, Erp57, also has broad specificity, but with a clear preference for glycoproteins. This specificity is explained because Erp57 must be physically associated with the calnexin cycle components to catalyze isomerization reactions with most of its substrates [198]. Yet, while for some glycoproteins, like influenza virus hemagglutinin, the action of Erp57 is important for post-translational, later phases of oxidative folding, for many other glycoproteins association with Erp72 could replace Erp57 and maintain folding competence [199]. In perhaps similar fashion, P5 is a co-factor associated with BiP and interacts with (at least some) BiP substrates [200].

If the calnexin cycle has evolved with a specialized oxireductase to facilitate native disulfide formation in complex glycoproteins, the process of ER-associated degradation involves one PDI with a unique activity: Erdj5 acts as a reductase, not an oxidase, on misfolded proteins and it is the only PDI with a J domain that enables it to bind to BiP and thus couple the peptide binding activity to the reductase activity in order to triage proteins for ERAD [116]. Erp72 has also been shown to interact with mutant thromboglobulin [201] and to participate with Erp29, Erp57 and Pdia1 in the unfolding of polypeptide virus proteins during viral infection [202]. This unfolding is initiated by Erp29, a PDI-family member without oxireductase activity, which coordinates Erp57 and Pdia1 to unfold the C-terminus of the capsid protein [202]. A final example of a unique role for a PDI is pErp1, which is specialized for supporting antibody secretion by plasma cells [203,204]. All these studies now enable the conclusion that although there is redundancy in supporting disulfide bond formation, a number of PDIs perform unique functions due to specific protein interactions mediated by their non-thioredoxin domains.

5.4.2. The ER immunophilins

The immunophilins/PPIs family is characterized by peptidyl-prolyl cis/trans isomerase activity (at least in vitro) and sensitivity to one or more of the immunosuppressive compounds cyclosporine A (CsA), FK506 and rapamycin. Of the three branches of the family, two are represented in the ER—the cyclosporine A (CsA)-binding cyclophilin B (CypB), and at least five FK506-binding proteins, FKBP13, FKBP19, FKBP22, FKBP23, and FKBP65. All of these immunophilins can catalyze isomerization of prolines, often a rate-limiting step in folding [205], but their mode of substrate recognition, and indeed the importance of their PPI activity in vivo is still poorly understood. Some of the ER immunophilins are induced when misfolded proteins accumulate, and even associate with misfolded substrates [206]. Yet, there is little detailed understanding for how the ER immunophilins participate in folding in the secretory pathway. Transferrin biosynthesis is sensitive to cyclosporine A [207], so since CypB is the main target of CsA in the ER [208], this immunophilin is implicated as a folding catalyst. CypB also accelerates the resolution of a kinetically trapped, on-pathway intermediate in immunoglobulin heavy chain folding, explaining the requirement for isomerization of Pro32 in cells [85].

Most of the data about the other immunophilins documents association and genetic necessity, but with relatively little mechanistic understanding. The large FKBp65 associates with the extracellular matrix proteins tropoelastin and collagen during its residence in the ER [209]. FKBp19 associates with CD81 in osteoblasts, which interacts with CD9 and a prostaglandin receptor regulator, and these associations cumulatively result in expression of interferon-inducible genes [210]. No specific substrate proteins for FKBp13 have been documented.

Human mutations in the FKBp14 gene are a cause of a type of Ehlers–Danlos syndrome (EDS) with progressive kyphoscoliosis, myopathy, and hearing loss [211]. Consistent with this phenotype, FKBp14-deficient fibroblasts exhibit altered assembly of the extracellular matrix in culture [211]. Both humans and mice deficient for CypB have Osteogenesis imperfecta and other bone abnormalities during development, as well as molecular defects in collagen biosynthesis [212,213]. Nonetheless, the null phenotypes of these immunophilins do not exhibit growth, development, fertility, immunity or other defects aside from bone and connective tissues, as would have been inferred from cell biological and proteomic studies. One would have expected, for example, a defect in humoral immunity, given the interactors of FKBp19 and CypB [85,210,214]. Similarly, FKBp13 knockout mice do not have an obvious phenotype that would be consistent with effects on folding of clients (Argon, unpublished data).

C. elegans, an important model genetic organism, has three secretory pathway FKBP s, expressed in hypodermal cells [215]. Neither one is an essential protein, but as a functional group, all three are essential for normal nematode development, collagen biogenesis, and the formation of an intact exoskeleton under adverse physiological conditions [215]. Thus, the genetic evidence indicates extensive functional redundancy among the ER immunophilins in relation to protein biosynthesis.

While studies implicating immunophilins in folding of specific substrates are few, there are studies about the association of immunophilins with the ER folding machinery. One functional interaction is between FKBp23 and BiP. The Neurospora FKBp23 homologue binds physically to BiP, which enhances the chaperone activity of the FKBP [216]. Mouse FKBp23 also binds BiP [217]. This binding is Ca++-dependent, can suppress the ATPase activity of BiP through the PPIase activity of FKBp23 and is mediated by catalyzing the cis/trans isomerization of Pro117 in the ATPase domain of BiP [218].

An intriguing recent finding has been that ER immunophilins interact with six of the protein disulfide isomerases [214], suggesting that PPIs and PDIs can modulate each other’s enzymatic activity in vivo, as demonstrated in some in vitro cases [219,220]. CypB, in particular, interacts with three PDIs (Pdia1, Pdia4 and Pdia6) as well as with three major chaperone networks in the ER (BiP, Cnx/Crt and Gpr94). The interaction of CypB with the lectin chaperones is mediated by their proline-rich P domain, through the same surface that binds Erp57, and is not sensitive to cyclosporine A [221]. Therefore, the protein interaction module is distinct from the enzymatic activity of CypB [221]. While the dynamics of these interactions in vivo are yet to be explored, a number of cells in fact contain a sufficient copy number of CypB to accommodate all these interactions (Table 1).

Recently, a role for CypB has been reported in ERAD. Either pharmacological inhibition of CypB with cyclosporine A, or its depletion with siRNA inhibits the degradation of a subset of misfolded soluble luminal proteins, those that contain cis proline residues [208]. CypB
is apparently the only ER-resident target of cycloporsine A and its catalytic activity likely enhances disposal from the ER by resolving local structures that are retro-translocated inefficiently. Functional roles in vivo for the other PPIs are yet to be elucidated.

5.5. Coordination of chaperone actions

Though we present each of the major chaperone systems separately, their various modes of recognition and activities are all needed, in parallel or in sequence, to fold proteins. In addition to the aforementioned sequential interaction of BiP and GRP94 with some clients, there are also other examples of coordinated chaperone action. Many of the folding factors that process non-glycoproteins interact sufficiently frequently to be isolated as a multi-chaperone complex from antibody-producing cells [222]. Intermediates of apolipoprotein B-100 were associated with GRP94, ERp72, BiP, calreticulin, and cyclophilin B [223]. Remarkably, as judged by subcellular fractionation, this array of chaperones remains associated with apolipoprotein B-100 during subsequent processing in the Golgi complex [223]. Presumably, the association of chaperones and lectins and/or folding enzymes creates a task-specific functional complex that works ad hoc to fold a particular client.

When a large glycoprotein like thyroglobulin is bound by BiP, GRP94, calreticulin, ERp29 and others [67,224], the binding is not necessarily simultaneous; different domains of thyroglobulin likely require different chaperones [225], and the cycles of binding and release may be temporally programmed [226]. Likewise, when a disulfide-bonded glycoprotein is targeted to ERAD, it is released from the lectin chaperones, and its disulfides are reduced by Erdj5, which uses its J domain to couple the activity to BiP [116], in preparation for escort by OS-9 to the SEL1L/Hrd1 retrotranslocon [116]. These examples illustrate the dynamic nature of the chaperone networks, which are transient, come together based on individual protein's 'needs', as well as on the stage of folding that is being chaperoned. The dynamic association of chaperones, enzymes and their substrates and clients is also evident by the use of live cell imaging of their fusions with fluorescent protein [227–229].

6. Conclusions

The folding machinery in the lumen of the ER has evolved to accommodate the unique features of the ER environment as well as common signatures of secretory proteins, such as disulfide bonds and glycans. Thus, while some of the ER chaperones and enzymes work in similar fashion to homologous cytosolic chaperones (e.g. BiP and HSP70, FKBP13 and FKBP12), other components (lectin chaperones and adaptors, PDI) are unique to this folding compartment. Even GRP94, which is structurally highly similar to the two cytosolic HSP90s, has evolved a distinct mode of chaperone action. To recognize a particular substrate, the various ER chaperone systems use distinct structural features, and for the most part engage in cycles of binding and release, during which distinct folding reactions are executed. The dynamics is even more complicated, since different chaperones may work on different domains of a given protein molecule, and may collaborate with an isomerase or a glycosylation enzyme, as required ad hoc by the substrate. The identity of the substrate dictates which chaperone acts as a hub, that in turn forms transient complexes with one or more other folding factors, in characteristic order.

A common biophysical role that the ER folding machinery fulfills is restriction of the folding options early on, at a time that many pathways can be followed. Perhaps the unifying mechanism for this action is “pinning” a patch of residues to either the core or the surface as a means of favoring a particular folding pathway (e.g. [230]). However, the BiP and the lectin chaperone networks achieve this goal by recognizing distinctly different structures (peptides vs. glycans). The BiP network marks peptides that are destined to be shielded from solvent, and thereby allowing other segments of the polypeptide to fold around these peptides. The attachment of glycans and the subsequent association with the lectin chaperone network overcomes the thermodynamic tendency of nearby patches enriched in hydrophobic and aromatic residues to avoid display on the protein surface, thus initiating particular folding pathways. The action of PDI and PPIs obviously restricts folding options by introducing irreversible steps. The GRP94 network may also restrict conformational space, but in a way that is currently not understood because the structural cues of GRP94 engagement with folding intermediates are yet to be deciphered.

Despite the impressive progress of our knowledge, many questions are still not resolved, including some that are common to several chaperone systems. How is the ‘productive’ folding intermediate distinguished from the ‘unproductive’ one? Are the rules intrinsic to the intermediate states, or are they different for different chaperones? When a substrate binds cyclically to a chaperone, how is the number of repetitive interactions determined? If a substrate presents several chaperone binding sites, when does its folding depend on just one dominant site (e.g. one glycan), and when are multiple sites required? When does the abundance of limiting co-chaperones, like GRP170, or sensors, like UGT1, become an impediment for efficient folding?

Future studies should also explain how unassembled subunits stay occupied with the ER chaperone networks, rather than sent for degradation. Our knowledge of substrate recognition by components like OS-9, which participate in such triage decisions, is still incomplete and it will be important to understand when and how such components engage with the folding machinery, and when they are committed to the substrate disposal via ERAD or autophagy.

Acknowledgements

We thank Drs. J. Brodsky (Univ. Pitt.), E. Snapp (Albert Einstein College of Medicine), M. Molinari (Institute for Research in Biomedicine, Bellinzona, Switzerland), P. Cresswell (Yale Univ.) and D. Hebert (Univ. Mass.) for helpful comments and personal communications. Work in the authors’ lab was supported by NIH grants AG18001 and GM07748.

References


