Proteomics of the endoplasmic reticulum to determine cellular organization and protein function

Michael D. Jain

Degree of Doctorate in Philosophy

Department of Anatomy and Cell Biology

McGill University
Montreal, Quebec, Canada
May 2009

This thesis is submitted to McGill University in partial fulfillment of the requirements for the degree of Doctorate in Philosophy.

©Copyright 2009 All rights reserved
ACKNOWLEDGMENTS

I would like to thank my supervisor Dr. John Bergeron for his enthusiasm and for enabling me to enjoy this work. I would like to thank the members of the Bergeron laboratory, particularly Hisao Nagaya, for their outstanding contributions to this project. I would like to thank the McGill M.D. – Ph.D. program for their confidence in me and continuous support. Most importantly, I would like to thank my family and friends for their affection and encouragement.
Thesis Objectives

This thesis has two overarching objectives:

1. To study the organization of the endoplasmic reticulum quality control pathway using a subcellular proteomics resource.

2. To identify novel proteins involved in endoplasmic reticulum quality control and to explore their function.
Original contributions

This body of work interrogates an organellar proteomics resource to make insights into the function of the endoplasmic reticulum in protein secretion. The following insights described in this thesis are original contributions to this scientific field:

1. Interrogation of the proteomics resource of Gilchrist et al. 2006, to determine the subdomain localization of over 100 proteins known to function in ER cargo processing.

2. Description of a spatiotemporal model of the ER: newly synthesized proteins traverse the ER from rough to smooth ER while undergoing synthesis, folding, and degradation. The resident proteins of the ER are segregated into physical and functional ER subdomains.

3. Identification of seven new targets of the mammalian unfolded protein response.

4. Identification of a new P-body protein (nudilin) and linking it to function at the ER.

5. Identification of a new constituent associated with the translocon (nicalin), the second protein suggested to regulate pre-emptive quality control, and a new constituent of the signal peptide peptidase complex.

6. Support for a novel process whereby the signal sequence of nicalin has a function separate from secretory pathway targeting, here involved in pre-emptive quality control.

7. Uncovering of four new proteins with chaperone-like characteristics (nicalin, stexin, erlin-1, and TMX2), with determination of the temporal ordering of their interaction with secretory cargo.

8. Domain studies on the short and long isoforms of TMX2 showing that the non-classical thioredoxin domain does not form mixed disulfides, yet is required for secretory cargo interaction.

9. Determining that the protein stexin interacts with diglucosylated forms of N-glycans.
Contributions of Authors

1. FLAG tagged constructs shown throughout this thesis were subcloned from vectors originally cloned by Hisao Nagaya.
2. Localization of nudilin (Fig. 3.1c) using immunofluorescence was performed by Hisao Nagaya.
3. The structure of nudilin (Fig. 3.2) was determined by Zheng Ye, Rangarajan Erumbi, Irene Mazzoni, Christine Munger, John Wagner, Michael Sacher, and Miroslaw Cygler.
4. Overexpression of nudilin in Fig. 3.3a was performed by Hisao Nagaya. The nudilin DNA construct for overexpression experiments shown throughout was prepared by Hisao Nagaya.
5. Nicalin localization in Fig. 4.1c was performed by Hisao Nagaya.
The nicalin YFP constructs used in Chapter 4 (including deletion mutants) were made by Hisao Nagaya.
6. Stexin, erlin, and TMX2 YFP localizations in Fig. 5.1b were performed by Hisao Nagaya.
7. FLAG-tagged TMX2 deletion mutant constructs used in Fig. 5.2c and the TMX2 SNDC to SNDA mutant construct used in Fig. 5.3c were made by Hisao Nagaya.
8. The abstract of this thesis was translated from English to French by Marie-Pascale Tremblay-Champagne.
9. All other attributions, including reagents obtained from other investigators, are indicated in the text.
# TABLE OF CONTENTS

Thesis Objectives .............................................................................................................. iii
Original contributions ........................................................................................................ iv
Contributions of Authors ................................................................................................... v
TABLE OF CONTENTS ......................................................................................................... vi
LIST OF FIGURES AND TABLES .................................................................................. viii
LIST OF ABBREVIATIONS ............................................................................................. ix

**INTRODUCTION** ............................................................................................................. 12
- Cells are defined by their organization ................................................................. 12
- Organization of the secretory pathway ................................................................. 13
- Exploring cells with a centrifuge .............................................................................. 14
- Organellar proteomics .............................................................................................. 15

**LITERATURE REVIEW** ................................................................................................ 17
- Overview of cellular proteins involved in cargo processing ................................. 17
- Post-transcriptional regulation of cargo messenger RNA: P-bodies and the ER........ 18
- Cargo synthesis and translocation (entry) into the endoplasmic reticulum ............. 19
- Removal of the signal sequence .............................................................................. 21
- The folding environment of the endoplasmic reticulum ........................................ 23
- Folding energetics in the endoplasmic reticulum ................................................... 24
- Chaperones of the endoplasmic reticulum ............................................................. 26
  - The glycoprotein recognition system ................................................................. 27
  - Disulfide bond formation and oxidative protein folding ....................................... 30
  - Hydrophobic domain recognition ........................................................................ 33
  - N-terminal proline recognition ............................................................................ 35
  - Cytosolic chaperones ............................................................................................ 36
- ER-associated degradation (ERAD) .......................................................................... 37
  - Selection of terminally misfolded/destabilized proteins for degradation ............ 37
  - Retrotranslocation into the cytoplasm ................................................................... 39
  - Proteosomal degradation ..................................................................................... 40
  - Non-proteosomal degradation ............................................................................. 41
- Secretion of cargo out of the ER: engaging the COPII coated vesicle ..................... 42
- ER retention signals .................................................................................................. 43
- Protein synthesis and folding generates a secretory “load” .................................... 44
- The unfolded protein response relieves secretory load .......................................... 44
- The mechanism of sensing protein folding in the ER .............................................. 46
- Increasing the capacity of protein folding through signal transduction ................. 47
  - Signalling downstream of Ire1 ............................................................................. 47
  - Signalling downstream of PERK ......................................................................... 48
  - Signalling downstream of ATF-6 ......................................................................... 49
- Interpretation of ER stress by the different transducers .......................................... 49
- Limiting protein entry into the ER during times of protein overload ...................... 50
  - Translational inhibition of secretory proteins ...................................................... 50
  - Co-translocational degradation .......................................................................... 52
  - Transcriptional repression of secretory proteins .................................................. 54
- ER stress-induced cell death ..................................................................................... 54
- Regulation of the size of the ER compartment ......................................................... 56
- Organellar proteomics ............................................................................................. 57
- Proteomics methods ................................................................................................. 58
  - Sample preparation ............................................................................................... 58
  - Mass spectrometry ................................................................................................ 59
  - Matching spectra to peptides and peptides to proteins ........................................ 60
  - Quantitation in Proteomics .................................................................................. 61
- Determining organization and function in the ER .................................................... 64

**CHAPTER 1: Interrogation of the proteomics resource reveals the spatial organization of ER cargo processing** ................................................................................. 67
LIST OF FIGURES AND TABLES

Fig. 1.1. Translation-associated proteins are found at the cytosolic face of the rough ER............. 71
Fig. 1.2: Translocon-associated proteins are found in the membrane of rough ER....................... 73
Fig. 1.3: Luminal ER folding occurs in the rough and smooth ER.............................................. 75
Fig. 1.4: Cytosolic chaperones associate with the smooth ER................................................. 76
Fig. 1.5: ER-associated degradation (ERAD) occurs in the smooth ER..................................... 77
Fig. 1.6: Distribution of ER secretory functions in the proteomics resource............................... 78
Fig. 1.7: Abundance and distribution of resident ER proteins involved in cargo processing
compared to cargo....................................................................................................................... 80
Fig. 1.8: Spatiotemporal model of the ER revealed by subcellular proteomics............................ 83
Table 1.1: Proteomics identification of proteins related to translation, translocation, folding,
ERAD, and ER-associated cytosolic protein folding................................................................. 90
Fig. 2.1: Sequence alignment identifies candidate cargo processing proteins of the ER............. 94
Fig. 2.2: HepG2 expression of selected transcripts corresponding to ER proteins of unknown
function......................................................................................................................................... 97
Fig. 2.3: Endpoint RT-PCR identification of novel targets of the unfolded protein response....... 98
Fig. 2.4: Real-time RT-PCR screen for unknown transcripts upregulated by tunicamycin in HepG2
cells............................................................................................................................................. 99
Fig. 2.5: Tyrosinase is a model cargo protein trafficking through the ER................................. 101
Fig. 2.6: Interaction between four resident ER proteins and the cargo protein tyrosinase ........ 102
Table 2.1: RT and Q-PCR primer sequences................................................................................ 107
Fig. 3.1: Localization of nudilin ................................................................................................. 114
Fig. 3.2: Nudilin is similar to the RNA decapping protein X29.................................................. 115
Fig. 3.3: Overexpression of nudilin increases the level of tyrosinase ....................................... 116
Fig. 3.4: siRNA knockdown of nudilin decreases tyrosinase levels.......................................... 117
Fig. 4.1: Localization of nicalin ................................................................................................. 129
Fig. 4.2: Time course of nicalin interaction with tyrosinase...................................................... 130
Fig. 4.3: Overexpression of nicalin diminishes tyrosinase levels............................................ 131
Fig. 4.4: Nicalin functions in proximity to translocon in pre-emptive quality control............. 132
Fig. 4.5: Nicalin interacts with signal peptide peptidase, found in proximity to the translocon... 133
Fig. 4.6: The signal sequence of nicalin is sufficient to diminish tyrosinase levels.................... 134
Fig. 5.1: Localization of stexin, erlin-1, and TMX2................................................................. 148
Fig. 5.2: The thioredoxin domain of TMX2 is required for cargo binding......................... 149
Fig. 5.3: TMX2 does not form mixed disulfides with substrate............................................ 150
Fig. 5.4: Stexin, erlin-1 and TMX2 show characteristics of chaperones................................. 151
Fig. 5.5: Stexin interacts with tyrosinase early in its folding..................................................... 152
LIST OF ABBREVIATIONS

A – aqueous (aqueous phase of Triton X-114)
D – detergent (Triton X-114 soluble)
DTT – dithiothreitol
ER – endoplasmic reticulum
ERAD – ER-associated degradation
ERAF – ER-associated folding
G – Golgi apparatus fraction
Glc – glucose
I – insoluble
IRES – internal ribosomal entry site
GFP – green fluorescent protein
MS – mass spectrometry
OST – oligosaccharyltransferase
PDI – protein disulfide isomerase
PPI – peptidyl-prolyl cis-trans isomerase
RM – rough microsomes
S – salt wash (2M KCl)
SDS-PAGE – sodium dodecyl sulfate polyacrilamide gel electrophoresis
SM – smooth microsomes
SPP – signal peptide peptidase
SRP – signal recognition particle
Tg – thapsigargin
Tm – tunicamycin
TRAM – translocating chain-associated membrane protein
TRAP – translocon-associated protein
UGGT – UDP:glucose glucosyltransferase
UPR – unfolded protein response
YFP – yellow fluorescent protein
ABSTRACT

This thesis describes the use of a proteomics resource to define the spatial organization of the endoplasmic reticulum (ER) and to predict the function of poorly characterized proteins. By proteomics, resident proteins involved in translation and translocation (eIFs, Sec61, TRAP, and the OST complex) are enriched in rough ER. Resident chaperones and folding enzymes (PDIs, calnexin, and BiP) are in equal abundance in rough and smooth ER. The ER-associated degradation (ERAD) machinery (derlin-1, VCP/p97, sel-1L, ubiquitin conjugating enzymes, and the proteasome), and also cytosolic chaperones (Hsp90, TRiC) all colocalize with their highest enrichment in smooth ER. This suggests a model that the ER is segregated into functional subdomains, which is useful to predict functional features of poorly characterized proteins localized in the proteomics resource. Nudilin is a cytosolic-located (P body) protein affecting translation which colocalizes with translation constituents at the cytosolic face of the ER membrane. Nicalin clusters with translocon constituents and is found to associate with signal peptide peptidase as well as negatively regulate the model secretory protein tyrosinase at the translocon. Stexin also clusters with translocon constituents and associates transiently with an early N-glycosylated form of tyrosinase. Erlin-1 and TMX2 co-distribute with chaperones found throughout the rough and smooth ER and are found to associate transiently with later N-glycosylated forms of tyrosinase. The ER is spatially organized to coordinate the temporal processing of cargo. Spatial localization in the ER predicts the function of resident uncharacterized proteins.
Cette thèse décrit l’usage d’une ressource protéomique afin de définir l’organisation spatiale du réticulum endoplasmique (RE) et de prévoir la fonction de protéines peu caractérisées à ce jour. Par la protéomique, des protéines résidentes impliquées dans la translocation et la traduction (eIFs, Sec61, TRAP, et le complexe OST) sont enrichies dans le RE granuleux. Des chaperones résidentes et des enzymes de pliage (PDIs, calnexin et BiP) sont d’abondance égale dans le RE granuleux et le RE lisse. Le système de dégradation associé au RE (ERAD) (derlin-1, VCP/p97, sel-1L, enzymes de conjugaison de l’ubiquitin, protéasome) ainsi que les chaperones cytosoliques (Hsp90, TriC) sont tous colocalisés dans le RE lisse, leur enrichissement y étant le plus important. Ceci suggère que le RE est ségrégué en sous-domaines fonctionnels, ce qui est utile pour prédire les caractéristiques fonctionnelles de protéines peu décrites, localisées dans la ressource protéomique. Nudilin est une protéine cytosolique (P body) affectant la traduction qui est colocalisée avec des constituants de la traduction à la surface cytosolique de la membrane du RE. Nicalin se regroupe avec des constituants du translocon et s’associe avec le signal peptide peptidase, en plus de réguler négativement la protéine de sécrétion tyrosinase au niveau du translocon. Stexin se regroupe aussi avec des constituants du translocon et s’associe de manière transitoire avec une forme N-glycosylée de tyrosinase trouvée immédiatement après sa synthèse. Erlin-1 et TMX2 se co-distribuent avec des chaperones trouvées à travers le RE lisse et granuleux et s’associent de manière transitoire avec des configurations N-glycosylée plus tardives de tyrosinase. Le RE est organisé dans l’espace afin de coordonner le traitement temporel de sa cargaison. La localisation spatiale des protéines résidentes non-caractérisées dans le RE prédit leur fonction.
INTRODUCTION

Cells are defined by their organization

The essence, or principle, that describes life is organization. There is no *animus vitae*, or life force that is a marker for a living thing. Life is instead thought to be an emergent property of organizing principles that coordinate otherwise inanimate materials (Ruse, 2006). This view has not always been held. For the ancient Greeks, life could not be described in purely physical or natural terms, and a force or substance was invoked to separate life from the inanimate. This persisted in places even in the 20th Century where European figures such as Bergson invoked a creative power or *elan vital*, which like other aspects of life could be transmitted by Darwinian selection (Ruse, 2006). These views lost favour when biological discoveries regarding the materials and organization of living things were collected and no such force could be located or shown to have causative power. Tellingly, discoveries such as the structure of DNA (Watson and Crick, 1953), showed ways of describing organized mechanisms for life-ascribed properties (i.e. self-replication), and found no need for an external substance or force.

Cells are thus ascribed life on the basis of their organization, or the coordination of otherwise inanimate complexes of proteins, ions, and other biochemical factors. This organization is coordinated in the physical space of the cell and maintained as ordered sequences of activity in time. The problem of spatial organization is rectified in eukaryotic cells by the use of intracellular membranes, which organize the intracellular environment into discrete membrane-bound compartments known as organelles. Across these membranes, differences in protein composition and chemical environment are maintained, such that activities in the lumen of an organelle can differ greatly from those at its outer (generally cytoplasmic) face. Such organization compartmentalizes cellular functioning in space. The problem of temporal organization is solved partially by the spatial
organization. When two enzymatic processes on the same substrate require that one occurs before the other, the two processes are often found in different subcellular compartments. The regulation of movement across membranes and between organelles in an organized fashion allows this temporal organization to be actualized.

Therefore, the essence of cells (what separates them as alive from inanimate or dead) depends upon maintaining a spatial and temporal organization of inanimate protein complexes, enzymes, and chemical environments. The evolutionary branching from prokaryotic cells with limited functions to more complicated eukaryotes required greater organization, and employed membrane bound organelles to divide up the intracellular space.

**Organization of the secretory pathway**

The process of protein secretion requires a dedicated pathway of organelles that spatially compartmentalizes specific environments and cellular machines. The specific temporal ordering of these events are largely governed by the transport of various messages and secretory products between these organelles. To start, DNA encoding a secretory protein is transcribed into mRNA in the nucleus, a double membrane-bound compartment. The introns are then spliced out and the mRNA is exported to the cytoplasmic space. In the cytoplasm, the mRNA is subject to post-transcriptional mechanisms, such as sequestration into structures such as P-bodies and degradation by cellular regulating factors. If the mRNA is correctly targeted, it can be translated by ribosomes attached to the cytoplasmic face of the rough endoplasmic reticulum (ER). The newly synthesized protein is co-translationally imported into the ER lumen, and then undergoes quality control for correct folding and enzymatic modification. A number of processes occur in the ER, which are spatiotemporally regulated. If the protein is correctly folded it is targeted to spatially confined ER exit sites and sent to the Golgi apparatus for further processing and secretion. If it is incorrectly folded (thereby representing a
conformational toxin), it is targeted to re-folding pathways found throughout the ER, or degradation pathways coordinating retrotranslocation sites with cytoplasmic proteasomes. Each of these functions must be spatiotemporally organized to maintain cellular viability. That many common diseases have been found to exhibit secretory pathway dysfunction, often resulting in cell death, reinforces the importance of such organization (Aridor and Hannan, 2000, 2002). Indeed, dysfunction of the secretory pathway may be causative in pathologies as common as type 2 diabetes, among others, making this organization important to human health (Marciniak and Ron, 2006). Moreover, compounds such as brefeldin A, which disperse organelles (in this example, the Golgi apparatus), are used as defensive toxins by some species of fungi, highlighting the necessity of organellar organization across kingdoms (Weber et al., 2004).

Exploring cells with a centrifuge

Initiated by the pioneering studies of Albert Claude, Christian de Duve, and George Palade, subcellular fractionation has been one method to explore the spatial organization of intracellular organelles as well as the functioning of the secretory pathway (De Duve, 1971; Palade, 1971). Other methods, such as electron and light microscopy, have been used in parallel to illuminate the morphology and organization of the secretory pathway (Orci, 1986; Presley, 2005).

Subcellular fractionation uses biophysical differences in organelle composition (such as density) to separate by centrifugation compartments of the cell, generating samples enriched for a particular organelle. Various protocols have been optimized to generate fractions highly enriched for organelles (such as those corresponding to rough ER, smooth ER, or the Golgi apparatus). All fractions have contamination from other organelles. In subcellular fractionation studies the extent of contamination has been quantified through enzyme assays, Western blotting, immunolabelling of organelles, and more recently quantitative
proteomics (Bell et al., 2001; Blondeau et al., 2004; Gagnon et al., 2002; Gilchrist et al., 2006).

**Organellar proteomics**

Protein sequencing by tandem mass spectrometry has been applied to isolated organelles (Au et al., 2007; Brunet et al., 2003; Yates et al., 2005). Typically this involves generating subcellular fractions enriched in a particular organelle, followed by tandem mass spectrometry and bioinformatics to determine the identity and quantity of the proteins found therein. Where classical studies used enzyme assays and marker proteins to define subcellular fractions, organellar proteomics is able to quantitate the abundance of hundreds of different proteins found in a particular fraction. Mapping the proteins found in each organelle-enriched fraction generates insight into intracellular organization by comprehensively describing the proteins found in a given fraction. If one compares the proteins found in two different organellar fractions, there are large differences in the abundance and presence of most proteins. This is due to cellular organization which localizes specific proteins and functions to specified organelles.

An organellar proteomics approach has been used recently to map the proteins of the early secretory pathway: comprehensive proteomes have now been characterized for the rough and smooth endoplasmic reticulum (ER), Golgi, and COP I vesicle fractions of rat liver (Gilchrist et al., 2006). In addition, the rough and smooth ER were further mapped into biochemical fractions representing subspaces of the ER: a high salt wash removed proteins bound to the cytoplasmic face of the ER, while a subsequent detergent extraction separated integral membrane-associated proteins from non-membrane associated proteins. Clustering algorithms and bioinformatics predictive software were used to generate profiles showing which organelles carry the highest abundance for that protein (in this case, among the rough and smooth ER, Golgi and COP I vesicle
fractions). Furthermore, for ER proteins, profiles can be generated that show whether a protein is associated with the cytoplasmic face of the ER membrane, if it is an integral membrane protein or tightly associated with the membrane, or if it is non-membranous in the rough or smooth ER. This thesis uses such information to recapitulate the spatiotemporal organization of the ER and to predict protein function.

In addition to generating these profiles, the proteomics mapping of the early secretory pathway of rat liver uncovered 345 proteins of unknown function. In the course of this thesis work, some of the proteins were published as having functions important to protein secretion, for example the retrotranslocation protein Derlin and the chaperone Boca (Culi and Mann, 2003; Lilley and Ploegh, 2004; Ye et al., 2004). However, the vast majority of these proteins still have no known function. In systems terms, this means that we do not even have a complete “parts list” that annotates the function of resident organelle constituents, despite now knowing their abundance and intracellular localization. Furthermore, among these unknown proteins are likely to be important effectors of protein secretion, cellular biology, and human health. Determining the cellular function of unknown proteins is a high priority for current biomedical research.

This thesis is divided into several parts. Each examines the organization of function in the ER. In the Literature Review, the functions of the ER relevant to secretory cargo are described. These functions include protein synthesis, ER-associated protein folding (ERAF), ER-associated degradation (ERAD), cargo trafficking through the secretory pathway, and the maintenance of cellular homeostasis during protein secretion. In the Results chapters, organellar proteomics is used to understand the organization of these functions. In Chapter 1, each of the functional aspects of the ER (translation, translocation, folding, and degradation) are described to each have unique physical localizations within the ER itself. That is, biochemical subfractionations of the ER are found by
proteomics to separate translation proteins from translocation proteins, from folding enzymes, and from ER-associated degradation proteins.

In the remainder of the Results chapters, the consequence of this organization on the function of individual proteins is explored. Chapter 2 shows the identification of poorly studied ER proteins that function in enabling cargo secretion. Chapter 3 describes a poorly characterized protein (nudilin) that co-localizes with transcription and translation proteins in the proteomics resource. Indeed, evidence is presented that nudilin functions in regulating secretory cargo expression, as predicted by its localization. Chapter 4 describes nicalin, an ER membrane protein that co-localizes with the translocation machinery. Evidence is presented that it functions to affect the translocation process. Chapter 5 describes three additional proteins (stexin, TMX2, and erlin-1) that interact with secretory cargo. The order of interaction of these proteins with secretory cargo is found to agree with the prediction by the proteomics resource. The Discussion reviews these findings: the ER is organized into spatial domains corresponding to function. This allows the prediction of the function of poorly characterized proteins, because organization predicts function. Since the functional predictions are found to be supported by the data presented in this thesis, the model of the organization of the ER into spatial subdomains (generated by organellar proteomics) is here supported.

LITERATURE REVIEW

Overview of cellular proteins involved in cargo processing

Secretory cargo refers to proteins that traffic through the secretory pathway but do not reside or function within the ER or Golgi apparatus. Secretory cargo can be secreted to the extracellular space (e.g. albumin), targeted to the plasma membrane (e.g. the insulin receptor), or targeted to intracellular organelles such as endosomes and lysosomes (e.g. lysosomal hydrolases). A major function of the
ER and Golgi apparatus is to process cargo. This includes the synthesis, post-translational modification, quality control, and targeting of cargo. As this literature review will show, secretory cargo processing is fundamental to cell and organismal viability.

The ER is the gateway into the secretory pathway. Messenger RNAs corresponding to secretory cargo are targeted to the rough ER membrane where they code for newly synthesized proteins. Simultaneous to synthesis (co-translationally) the secretory cargo proteins traverse the ER membrane to enter the ER lumen (translocation), undergo post-translational modifications, begin to fold, and are packaged into carriers for ER-Golgi transport. Ensuring the productive protein folding of secretory cargo is a major function of the ER that occurs both during synthesis (co-translationally) and well after synthesis has finished. Despite the folding and refolding efforts of the ER, proteins can remain incorrectly folded and are then targeted for degradation in the cytosol. Degradation thereby requires a misfolded protein recognition machinery and reverse translocation into the cytosol. Global analysis of the proteins in the ER proteome from rat liver at steady state has shown that the much of the ER’s protein abundance is dedicated to protein synthesis and folding, followed afterwards by its detoxification function (Gilchrist et al., 2006). Therefore, the ER is an important secretory cargo processing compartment.

Post-transcriptional regulation of cargo messenger RNA: P-bodies and the ER

As for all mRNA, transcripts corresponding to secretory cargo are subject to post-transcriptional processing through the linked processes of mRNA degradation, RNA-mediated gene silencing and translational repression.

Post-transcriptional processing occurs in the cytoplasm as well as in discrete 100–300 nm cytoplasmic foci now known as P-bodies (Bashkirov et al., 1997; Sheth
and Parker, 2003). P-bodies are created as a consequence of RNA degradation since disruption of genes involved in RNA silencing and decay pathways also obliterate P-bodies (Eulalio et al., 2007b). P-body constituents include proteins involved in RNA decapping, nonsense mediated RNA decay, translational repression and other RNA metabolic functions (Eulalio et al., 2007a), establishing P-bodies as sites for post-transcriptional regulation.

P-body constituents not only affect secretory cargo transcript levels but are also found to affect ER morphology. Deletion of the Drosophila RNP granule protein Trailer Hitch decreases protein levels of the ER exit site constituent Sar1 and results in aberrant ER exit site morphology (Wilhelm et al., 2005). Deletion of the C. elegans P-body protein Car1 results in aberrant ER morphology during cytokinesis (Squirrell et al., 2006). The mechanism of action by which these P-body proteins are enacting ER morphological changes is not clear. In the case of trailer hitch, a complex between RNP granule constituents (a structure functionally related to the P-body) and ER exit site mRNAs has been observed (Wilhelm et al., 2005), suggesting a direct mechanism of a yet undetermined nature.

**Cargo synthesis and translocation (entry) into the endoplasmic reticulum**

Up to 1/4 of the proteome of a given cell is targeted to the secretory pathway (Wiseman et al., 2007a). Such proteins are synthesized by ribosomes bound to the cytosolic face of the ER membrane. It is the ribosome-studded appearance of this part of the ER membrane which gives it its morphological descriptor as rough ER. The decision to synthesize a protein while bound to the ER as opposed to synthesis in the cytosol is determined by the signal sequences of the newly synthesized protein (Blobel and Dobberstein, 1975a, b; Lingappa and Blobel, 1980). All newly synthesized proteins begin their translation on cytoplasmic ribosomes. Proteins that are to be synthesized on the ER and that will enter the secretory pathway contain signal sequences somewhere within their sequence.
Usually these sequences are found at the extreme N-terminus of the protein and do not end up as part of the final protein product, as they are cleaved once translation is completed. Signal sequences are not composed of a uniform recognition sequence; rather they all have similar biochemical properties among the amino acids used such that they can be recognized and targeted to the ER. These properties are so uniform, however, that they can be predicted with high reliability using bioinformatics. The properties of signal sequences generally include a positively charged segment on the extreme N-terminus and a short hydrophobic segment, followed by a polar region, with more detailed considerations refining the predictive power of computational algorithms (Emanuelsson and von Heijne, 2001). Once the signal sequence has been synthesized, a cytoplasmic receptor known as the signal recognition particle (SRP) binds to the newly synthesized signal sequence and causes a pause in translation. The ribosome in complex with the mRNA, the newly synthesized protein, and SRP then go to the ER membrane, where an ER resident transmembrane protein known as the SRP receptor interacts with SRP, allowing positioning of the ribosome on the rough ER membrane (Egea et al., 2005; Gilmore et al., 1982a; Gilmore et al., 1982b; Meyer et al., 1982). The ribosome interacts with a membrane channel known as Sec61 and translation resumes. The nascent chain of the newly synthesized protein now traverses the pore of the channel as it grows, thereby crossing the membrane separating the cytoplasm from ER lumen. This process is known as co-translational translocation because ER proteins translocate across the ER membrane as they are being synthesized. On account of its protein channel function in translocation, Sec61 is referred to as the translocon, and is assisted by translocon-associated binding partners such as TRAP and TRAM (Deshaies et al., 1991; Deshaies and Schekman, 1987; Gorlich et al., 1992; Hegde et al., 1998b).

A slightly different mechanism is used for proteins that enter the secretory pathway by using signal anchors instead of N-terminal signal sequences. These signal anchors are found anywhere in the protein and end up becoming
transmembrane domains in the resultant protein and are not cleaved. However, in
other respects the mechanism is the same, with recognition by SRP, targeting to
the ER membrane by the SRP receptor, and co-translational translocation into the
ER by Sec61 (Spiess, 1995).

A detail of the translocation process is that the proper orientation of
transmembrane proteins is achieved (Osborne et al., 2005). For example, the N-
terminus of such transmembrane proteins can either face the cytosol or the lumen
of the ER, and almost every molecule of that protein will face that one direction.
This correct orientation is achieved by plus and minus ends encoded into the
signal anchors, which determine the orientation of one of the transmembrane
domains. For TMD-containing proteins that use N-terminal signal sequences, the
N-terminus is cleaved in the lumen of the ER, establishing its orientation. It
should be noted that trans-acting factors associating with the translocon can
influence the orientation of a protein, although the amino acid sequence remains
the main driver of topology (Hegde et al., 1998a).

Determining the orientation of proteins in the membrane using bioinformatics of
the sequences encoded is useful, but not absolute. Errors can be made in
predictions of transmembrane domains, signal sequences, and whether a given
TMD traverses the membrane or merely interacts with it. For the same reasons as
for the difficulties with orientation, the correct assessment of an unknown protein
of the ER as luminal, integral membrane, or cytosolic-associated is also subject to
predictive error. The extent of the problem predicting topology depends on the
complexity of the protein and the program used; for example, even with diligent
methodologies, the topology of a 7 transmembrane spanning receptor is erroneous
20 - 25% of the time (Inoue et al., 2005).

**Removal of the signal sequence**
Following translocation, the N-terminal signal sequences of secretory proteins are specifically cleaved by signal peptidase on the luminal face of the ER membrane. This divides the secretory protein (with a new N-terminus in the lumen of the ER) from the signal sequence, which is left stranded in the membrane. To eliminate the cleaved signal sequence from the membrane, a presenilin-like aspartic protease called signal peptide peptidase (SPP) cleaves the signal sequence in the centre of the transmembrane region, turning the signal sequence into peptide fragments (Weihofen et al., 2002). In this respect, SPP functions in a manner similar to other proteases (such as presenilin) which catalyze intramembrane proteolysis of transmembrane proteins. A specific interest of intramembrane proteolysis is that it is a mechanism of liberating biologically active peptides from membranes, allowing for these peptides to have new functions in different compartments. For instance, intramembrane proteolysis of ATF6 by S1 and S2 proteases in the Golgi apparatus allows the newly cleaved cytosolic fragment of ATF6 to travel to the nucleus and act as a transcription factor (Ye et al., 2000). Indeed, homologues of signal peptide peptidase (SPP-like or SPPL) have been shown to coordinate intramembrane proteolysis of TNF-alpha at the cell surface, affecting TNF-alpha signalling (Fluhrer et al., 2006; Friedmann et al., 2006). Hence, intramembrane proteolysis can lead to protein degradation as well as can generate biologically active protein fragments. In Chapter 4, evidence will be presented that signal peptide peptidase cleavage of an ER resident protein (nicalin) generates an active fragment capable of co-translational secretory repression.

An additional function of signal peptide peptidase has recently been uncovered. Loureiro et al. have shown that host cell signal peptide peptidase is required by viruses to evade immune detection by down-regulating class I MHC (Loureiro et al., 2006). The virus-encoded immunoevasive molecules are found in a complex with signal peptide peptidase in order to accelerate retrotranslocation of MHC molecules from the ER lumen to the cytosol for proteasomal destruction. Retrotranslocation as a necessary step for ER-associated degradation will be
discussed in a further section, but this work points to functions for signal peptide peptidase outside of simply disposing of signal sequences stranded in the ER membrane.

The folding environment of the endoplasmic reticulum

Protein folding, or the formation of the three dimensional structure of proteins from linear chains of amino acids, occurs during protein synthesis. That is, folding occurs in the ER co-translationally, as domains capable of forming independent structures seek to bury their hydrophobic residues and achieve a stable (low) energy conformation as soon as they exit the ribosome. During this synthesis and folding process, a choreography of protein modifications and chaperone recruitment begins, with the end goal of achieving a stable conformational structure throughout the entire protein (Daniels et al., 2003).

The stability of a secretory protein is a determinant of its secretory efficiency. While the lowest energy conformation is dictated by the amino acid sequence of the secretory protein, finding this conformation does not always occur efficiently inside the cell. Therefore, the intensity of the cellular ER-associated folding (ERAF) process is also a determinant of secretory efficiency. The constituents of ERAF that affect secretory efficiency include resident protein machines of the ER such as chaperones, as well as the environment of the ER lumen, including oxidizing potential, protein crowding, and metabolite levels.

Once a newly synthesized secretory protein has entered the ER it has several possible fates, which all depend upon protein stability. Predominantly, newly synthesized proteins are properly folded and leave the ER by a secretory mechanism involving engagement of COPII vesicles for ER-Golgi transport (COPII vesicle engagement is discussed in more detail in subsequent sections). If the protein does not achieve a high enough stability (i.e. misfolded proteins), it is either retained in the ER and not secreted, or is retrotranslocated out of the ER.
and into the cytosol for degradation by a pathway known as ER-associated degradation (ERAD).

The process of ERAF and the pathways of ERAD and ER secretory exit compete for newly synthesized proteins to govern these three fates (Kincaid and Cooper, 2007). That is, every newly synthesized ER protein is either secreted, remains in the ER, or is degraded. If a protein enters the ERAD pathway it is degraded. While proteins engage ERAF they remain in the ER. If the protein has enough stability and has avoided the ERAD pathway and has left ERAF it can engage secretory exit. Retention in the ER is probably due to proteins that are unable to engage the ER secretory exit pathway, but are also not recognized for ERAD on account of reasonable stability.

**Folding energetics in the endoplasmic reticulum**

The specific features of a secretory protein that are recognized by the cell to assist folding will be discussed in detail in subsequent sections. However, from a global point of view, recent biophysical work has proposed that the eventual fate of a secretory protein is dependent first upon the biophysical stability of the protein in the ER, and then upon the inherent properties of ERAF and ERAD in a given cell type (Sekijima et al., 2005). That is, whether or not a protein is secreted can vary in different cell types based upon the constituents of ERAF and ERAD found in each cell and how they interact with a secretory protein of a given stability.

Secretory efficiency in a single cell type (that is, when ERAF and ERAD capabilities are held constant) is predicted when combining the inherent thermodynamic stability with the kinetic stability of a given secretory protein (Sekijima et al., 2005). Simplified, kinetics describe how fast a protein can fold, while thermodynamics describe how stable the protein is once it is folded (Wiseman et al., 2007a; Wiseman et al., 2007b). Individually, kinetic stability or thermodynamic stability do not predict secretion efficiency well. However, the
combination of these two stabilities is highly predictive of which cargo will be secreted and which will be degraded in a given cell type (Sekijima et al., 2005). Thus, secretion of cargo is dependent upon its folding energetics, which is generated from the primary amino acid sequence.

For any given cargo molecule which is expressed in different cell types, the intensity of ERAF and the threshold for ERAD determine secretory efficiency (Sekijima et al., 2005; Wiseman et al., 2007a; Wiseman et al., 2007b). That is, in some cell types a cargo molecule might be secreted, while in others it might be degraded. The determining factors are the ability of ERAF constituents to improve stability versus the threshold setting for ERAD to give up and degrade the cargo instead. This is not an all or nothing decision – the percentage of newly synthesized cargo that is eventually secreted in a given cell type is determined on this basis.

A useful example is transthyretin-mediated amyloidosis. Here, highly unstable disease mutations are less pathogenic than mildly unstable ones (Sekijima et al., 2005). This is because the disease occurs when unstable proteins are secreted into the extracellular space and over time are able to form amyloid fibrils (Hamilton and Benson, 2001). Highly unstable disease mutations predominantly undergo ERAD and as a result only tiny amounts are able to escape to form amyloid in the extracellular space. On the other hand, mildly unstable mutants are competent for secretory exit, are secreted, and cause more severe disease due to increased amyloid formation in the extracellular space. Additionally, certain disease mutations which have moderate stability only cause neurological disease and not liver disease. In this case, despite the fact that the choroid plexus and the liver both secrete transthyretin, differences in ERAF intensity and ERAD threshold cause the mutant proteins to undergo ERAD in the liver, while a certain amount of secretion is allowed in the choroid plexus. This secretion allows amyloid formation and a late-onset neurological disease while sparing the liver (Sekijima et al., 2005).
ERAf and ERAD efficiency depends upon the abundance and presence of a number of proteins and molecules in the ER. Chaperones and their folding assistants as well as metabolite levels appear to be the most important factors (Wiseman et al., 2007a; Wiseman et al., 2007b). A detailed discussion of ERAf and ERAD constituents will follow.

**Chaperones of the endoplasmic reticulum**

Hallmarks of molecular chaperones are that they bind and release proteins for the purpose of conformational stabilization and that by doing so they facilitate the correct fate of their clients (Hendrick and Hartl, 1993). There is a great variability in terms of the size and structure of secretory cargo proteins that need to be folded and traffic through the secretory pathway. While all of the information required to fold a protein is contained in the amino acid sequence, molecular chaperones increase the efficiency of folding and limit the formation of “off-pathway” proteotoxic and aggregating protein species (Ellis and Hartl, 1999). It has been postulated that some secretory proteins in certain cell types have specific chaperones to handle a high secretory load of a given protein (Martinek et al., 2007), however, in general each chaperone handles a great diversity of proteins. In one example, the cytosolic Hsp90 chaperone has at least 200 substrates in a single cell type (Zhao et al., 2005). This flexibility can also be seen in heterologous expression systems, whereby foreign proteins can be introduced into a cell and be properly folded, or in the case of viruses, which co-opt ERAf components for the proper folding and assembly of foreign viral proteins (Schelhaas et al., 2007).

In order to accommodate such a diversity of secretory traffic, each chaperone system has generalized biochemical determinants that distinguish properly folded proteins from incorrectly folded proteins. The glycosylation status of a protein, misfolded proteins with aberrant disulfide bonds, and the exposure of
hydrophobic domains are examples of determinants recognized by specific chaperone systems to aid in the re-folding of misfolded proteins. Indeed, these systems work in concert with a given misfolded protein being re-folded by multiple chaperone systems, and perhaps in multicomponent macromolecular complexes. The following gives a description of the main chaperone recognition systems in the ER, representing the most influential known constituents of the ERAF process.

The glycoprotein recognition system

Many newly synthesized proteins destined for the secretory pathway contain consensus sites (NXS/T) for N-linked glycosylation. The addition of an oligosaccharide chain at these sites occurs in the ER lumen and is catalyzed by the oligosaccharyltransferase (OST) complex. The OST complex directly interacts with the translocon and is positioned such that it can add N-linked glycans to secretory proteins co-translationally, as early as 65 amino acids into protein synthesis (Chavan and Lennarz, 2006; Whitley et al., 1996). Insight into a cellular function for glycosylation occurred when the protein calnexin was shown to be a lectin (glycan-binding) chaperone of the ER (Ou et al., 1993; Schrag et al., 2003). The specific interaction between calnexin and misfolded glycoproteins, but not non-glycoproteins, led to the positing of the “calnexin cycle”, in which the structure of the N-glycan is modified by the cell to signal the folding status of a protein in the ER (Hammond et al., 1994).

Calnexin is a type I transmembrane protein restricted to the ER. It is similar in sequence to a soluble luminal protein of the ER called calreticulin, which is also a lectin chaperone (Peterson et al., 1995). Calnexin and calreticulin interact with glycoproteins containing a highly specific type of N-linked glycan: monoglucosylated N-linked glycans. In fact, these lectin chaperones will interact with proteins regardless of whether they are properly or improperly folded, so long as the glycan contains this monoglucosylated structure (Zapun et al., 1997).
In addition, the chaperone capability of calnexin and calreticulin appears to be mediated through interaction with the protein disulfide isomerase-like ERp57 (the function of PDIs in oxidative protein folding is discussed in detail in a subsequent section). Therefore it is through association with catalytic protein folding enzymes that monoglucosylated proteins are targeted for refolding by calnexin and calreticulin.

The calnexin/calreticulin cycle works as follows: newly synthesized proteins are glycosylated by the OST complex as they enter the ER lumen. This oligosaccharide has the structure of \( \text{Glc}^3\text{-Man}^9\text{(GlcNac)}_2 \) (where Glc is glucose, Man is mannose, and GlcNac is N-acetylglucosamine). Sequential trimming of the two terminal glucoses by glucosidase I and glucosidase II creates an N-oligosaccharide with the structure \( \text{Glc}^1\text{-Man}^9\text{(GlcNac)}_2 \). As noted, such a structure is bound by calnexin and calreticulin and the newly synthesized protein enters the cycle. Calnexin and calreticulin cease binding when glucosidase II re-enters and trims the final glucose to generate a \( \text{Man}^9\text{(GlcNac)}_2 \) oligosaccharide. If the protein remains misfolded, this is recognized by another member of the cycle known as UDP:glucose glucosyltransferase (UGGT). UGGT functions as a sensor of protein misfolding and responds to localized areas of misfolding in a protein by adding back the terminal glucose to the oligosaccharide, re-forming the \( \text{Glc}^1\text{-Man}^9\text{(GlcNac)}_2 \) oligosaccharide structure. Some controversy exists as to whether UGGT can modify glycans near or far from the misfolding lesion on the protein (Taylor et al., 2004), (Ritter et al., 2005), but the end result of UGGT activity is to generate a glycan that is re-recognized by calnexin and calreticulin. This allows another chance for ERp57-mediated oxidative folding. A protein that fails to become properly folded after the second round of calnexin binding can be trimmed again by glucosidase II, re-recognized and modified by UGGT, before returning to calnexin again. This defines the calnexin/calreticulin cycle, whereby modifications to the N-linked oligosaccharide marks a secretory protein for chaperone-mediated refolding.
If the protein becomes productively folded, UGGT does not add the terminal glucose and the protein is no longer recognized by calnexin or calreticulin and leaves the cycle. It may be that calnexin and calreticulin actively retain proteins in the ER and thus productively folded proteins are freed from this yoke and can now interact with the COPII vesicle machinery for secretion. However, the existence of multiple chaperone systems contributing to ERAF suggests that the calnexin/calreticulin cycle has a greater function on re-folding and folding efficiency than absolute protein retention, an assertion borne out by studies of fibroblasts deleted for components of the calnexin cycle. Here, knockouts of UGGT, calnexin, calreticulin, or ERp57 have been found to have only minor effects on the folding and secretion of most glycoproteins (Molinari et al., 2005; Pieren et al., 2005; Solda et al., 2006). That the whole animal knockouts of these proteins are embryonic or perinatally lethal (Denzel et al., 2002; Garbi et al., 2006; Luo et al., 2006; Mesaeli et al., 1999; Molinari et al., 2005) suggests an essential function to these proteins. However, it is unclear whether the toxicity is caused by the general folding inefficiencies such knockouts bring to the system, or if specific, largely unknown, secretory proteins have an absolute requirement for this cycle and are aberrant in these animals.

If a protein fails to become productively folded despite repeated cycles through calnexin/calreticulin and UGGT, it is considered to be “terminally” misfolded by the cell. There are two fates to such proteins, as uncovered by studies of disease-associated proteins that are misfolded. One fate is to be retained in the ER permanently, forming higher order protein structures such as aggregates, or being sequestered into autophagosome-like structures associated with the ER. These structures are thought to be protective, as experimental measures that inhibit their formation enhance ER stress-mediated cell death (Bernales et al., 2006; Granell et al., 2008). The second fate of terminally misfolded proteins is to be targeted for degradation. For glycoproteins, modification of the N-linked oligosaccharide chain is also the signal to send a protein down the degradative pathway; these
considerations will be discussed further in the section on ER-associated degradation.

**Disulfide bond formation and oxidative protein folding**

The relatively reducing environment of the cytosol disfavours the formation of disulfide bonds. Disulfide bond formation therefore occurs in the oxidizing environment of the endoplasmic reticulum. Disulfide bonding imparts structural stability from prokaryotes to eukaryotes (Heras et al., 2007). In addition, from the point of view of the cell, disulfide bonds represent a modification that can be monitored in the ER to detect aberrant protein folding.

Oxidative protein folding can be reconstituted *in vitro* using only three factors: FAD, the chaperone protein disulfide isomerase (PDI), and the FAD-binding protein Ero1 (Tu et al., 2000). Subsequent work showed that oxidative protein folding consists of a relay system of electrons between these molecules, with molecular oxygen (O₂) as the terminal electron acceptor (Tu and Weissman, 2002). The relay works as follows: FAD is freely transported into the ER lumen where it binds to Ero1. Ero1 in turn oxidizes PDI, becoming itself reduced. From there, PDI oxidizes cysteine pair-containing secretory proteins to catalyze disulfide bond formation, while Ero1 passes its electrons off on molecular oxygen.

There are several consequences to this relay system (Tu and Weissman, 2004). First, the relatively oxidizing environment of the ER is caused by FAD levels. Second, disulfide bond formation and oxidative protein folding reduces molecular oxygen. While it is unclear whether this directly leads to hydrogen peroxide (as is expected from such a reaction), it is known that disulfide bond formation creates oxidative free radical stress. For instance, overexpression of a misfolded secretory protein or the use of pharmacological protein misfolding agents can cause accumulation of endogenous peroxides and cell death, as cycles of oxidative re-
folding are attempted (Haynes et al., 2004); (Harding et al., 2003). Mutation of the cysteines in the misfolded secretory protein or deletion of Ero1 during pharmacological stress abrogates disulfide bond formation and greatly reduces toxicity (Harding et al., 2003; Haynes et al., 2004). Protein misfolding and oxidative re-folding may therefore represent an important source of toxic free radicals during pathological states.

Protein disulfide isomerase was the first chaperone to be discovered. Classic experiments by Christian Anfinsen in the 1960s showed that upon denaturation and disulfide bond reduction, the substrate protein RNase would spontaneously refold and recover its catalytic function in vitro, indicating that the amino acid sequence was alone sufficient to direct protein folding (Anfinsen and Haber, 1961). However, the inclusion of crude liver homogenates into the reaction greatly accelerated the rate of RNase reactivation (Goldberger et al., 1963) and an enzyme (protein disulfide isomerase) contributing to this acceleration was able to be purified (Goldberger et al., 1964). Subsequent work has shown that in addition to its ability to catalyze disulfide bond formation, PDI can also break or isomerize aberrantly disulfide bonded proteins and catalyze the reformation of new disulfide bonds. Therefore in complement to its function in catalyzing disulfide bond formation, PDI acts as a chaperone, breaking aberrant disulfide bonds and forming new ones in the hope that the new bonds will have greater stability.

Seventeen genes have now been annotated as belonging to the PDI gene family in the human genome (Ellgaard and Ruddock, 2005). This family has been constructed based upon sequence similarity to one or more of the four thioredoxin-like domains of PDI. These domains are named a, b, b’, and a’ on the basis of catalytic activity and their order of appearance in protein disulfide isomerase. The a and a’ domains are catalytically active and contain a characteristic CXXC motif. The b and b’ domains are redox inactive. The CXXC motif is posited to isomerize disulfide bonds through the formation of “mixed disulfides” with substrate proteins (i.e. with cargo) that can be visualized on non-
reducing gels as higher molecular weight complexes (Molinari and Helenius, 1999). Here, the first cysteine in the CXXC is thought to attack an existing disulfide bond in the secretory cargo protein (a bond that is likely to be aberrant since the protein is misfolded), and forms a new “mixed” disulfide bond between the PDI and the substrate protein. The second cysteine then attacks the mixed disulfide, freeing the PDI protein from the substrate, and allowing the substrate protein to attempt a new conformation of disulfide bonding through re-folding. In terms of the individual domain functions, the most complex organism from which the crystal structure of any PDI family member has been solved is that of the PDI gene in *Saccharomyces cerevisiae* (Tian et al., 2006). This structure revealed that the four thioredoxin domains form a twisted “U”, with the active a and a’ domains facing each other on at the ends of the U and the inactive b and b’ domains at the base. In addition, each of the four domains contains a surface-exposed hydrophobic patch, creating a hydrophobic surface in the cleft of the U that is likely involved in binding the misfolded proteins. In agreement with previous studies (Darby et al., 1998), truncations of the inactive b and b’ domains decreased the refolding activity of PDI, suggesting that the b domains play a role in substrate binding while also affecting the rate of enzyme refolding in *trans*.

It is not clear why so many different PDI family members are necessary. In some cases, protein complexes in the ER will include individual PDI members in order to obtain PDI activity for specialized functions. An example of this is the described interaction between the PDI member ERp57 and calnexin/calreticulin, which is used for the specific chaperoning of glycoproteins. Other PDI members are only found in certain tissues, such as testis, which may have individualized needs (van Lith et al., 2005). Certain secretory proteins may also require specific family members in order to achieve proper folding, however, there is as of yet no direct evidence to support this speculation. Interestingly, the catalytic activity and substrate binding abilities of the different PDI family members vary widely, and the family also includes non-catalytic members that continue to bind substrate (Alanen et al., 2006; Ellgaard and Ruddock, 2005). The relevance of this diversity
to cellular function is not known, however it is tempting to speculate that varying
affinities amongst PDI family members could contribute to a recognition system
for disulfide bonded proteins. This would be analogous to the glycoprotein
recognition system, which uses catalytic and non-catalytic glycan-binding
proteins to determine the fate of secretory glycoproteins.

For the purposes of this thesis, a more detailed examination of the sequence
features of one of the PDI family members, annotated as TMX2, is in order.
TMX2 contains a single thioredoxin domain that is homologous to the active a
domain of PDI. However, this domain is variant in TMX2, with a non-classical
SXXC motif at the putative catalytic site. Theoretically, the SXXC motif is
unlikely to be involved directly in disulfide bond formation (Ellgaard and
Ruddock 2005), although some weak evidence exists that another SXXC
thioredoxin protein, PDILT, is capable of forming mixed disulfides with
substrates (van Lith et al., 2005). It is also unclear how the lack of b domains and
a complementary a’ domain would affect the function of TMX2, although
oligomer formation or stabilization in a protein complex might offer functional
support. Indeed, other PDI family members which also contain single thioredoxin
domains have been proposed to have roles in oxidative folding (Anelli et al.,
2003; Haugstetter et al., 2005). A more detailed characterization of TMX2 is
presented in Chapter 5.

**Hydrophobic domain recognition**

Another determinant that can be recognized by chaperones of the ER are
hydrophobic-exposed residues of misfolded proteins. Energetically, hydrophobic
domains should be buried away from the aqueous environment of the cell,
therefore the exposure of hydrophobic domains represents an attractive marker of
protein misfolding. Recognition is mediated by the heat shock protein (Hsp)
families of chaperones. These have been best studied in the cytoplasm, however
both Hsp70 and Hsp90 family members are found in the lumen of the ER.
Additionally, both the Hsp70 and Hsp90 systems are recruited to the cytosolic face of the ER membrane and modulate those transmembrane cargo of the secretory pathway that have domains exposed to the cytosol.

The protein BiP (also known as GRP78) is a member of the Hsp70 family, is upregulated by ER stress and glucose deprivation, and is an abundant resident protein of the ER that stands as a prototypical chaperone that recognizes hydrophobic domains (Gething, 1999). More generally, the Hsp70s represent a versatile chaperone system with diversity and specificity generated by many different family members, as well as a large group of modulatory interacting factors (Bukau et al., 2006). Hsp70 family members contain a nucleotide binding domain that binds to ATP and a substrate binding domain that recognizes stretches of hydrophobic residues. ATP binding causes a conformational change in the substrate binding domain and converts Hsp70 proteins from an open state with high exchange rates to a closed state with low exchange rates (Minami et al., 1996). Such an activity is modulated by nucleotide exchange factors and DnaJ-domain containing proteins. DnaJ proteins include the Hsp40 family of proteins and stimulate ATP hydrolysis. Nucleotide exchange factors catalyze the opposite reaction, favouring ADP release and ATP formation. In this manner co-chaperones regulate the specificity and activity of the Hsp70 chaperone system.

Perhaps as a consequence of having chaperone activity that can activate conformational changes in other proteins, Hsp70 family members have diverse roles in the secretory pathway and elsewhere in the cell. For instance, roles for BiP in sealing the translocon (Hamman et al., 1998), functioning as a sensor in the unfolded protein response (Bertolotti et al., 2000), and other ER processes (Li and Lee, 2006) have been described.

The Hsp90 chaperones also comprise an important system operating at the ER. In the ER lumen, the Hsp90 protein GRP94 has been shown to interact with a variety of substrates as a chaperone (Argon and Simen, 1999). As with Hsp70 proteins,
Hsp90 chaperones possess ATPase activity, and have nucleotide and substrate binding domains. It is also regulated by a diversity of co-chaperones, with the tetratricopolypeptide repeat (TPR) proteins representing the most prominent family (Young et al., 2001). In the cytosol the Hsp90 protein has at least 200 substrates (Zhao et al., 2005), therefore a general role in protein folding is predicted in the ER for Hsp90 family members, with specificity imparted by ER-localized co-chaperones.

The Hsp70 and Hsp90 systems in the cytosol are both recruited to the cytosolic face of the ER. Here, the large cytosolic-exposed domains of proteins such as the cystic fibrosis transmembrane regulator (CFTR) require chaperone assistance from these systems. For such substrates, cellular decision making for ERAF, ERAD, or export can depend upon these systems and their co-chaperone regulation (Wang et al., 2006; Younger et al., 2006).

**N-terminal proline recognition**

A final type of chaperone found in the ER (as well as in other cellular compartments) isomerizes proline residues from their cis to trans form in order to sample different protein conformations and thereby assist protein folding. These are the peptidyl-prolyl isomerases (PPIases) and are characterized by two families named via their interactions with pharmacological reagents: cyclophilins bind the agent cyclosporin, while the FK-binding proteins (FKBPs) bind to FK506 (Zapun et al., 1999).

The determinants that specify which proteins receive cis/trans isomerization by a given family member are variable. In one well-known instance from the cytoplasmic compartment, the PPIase Pin1 isomerizes proline residues that follow a phosphorylated serine or threonine residue (Wulf et al., 2005). Here, specificity is imparted by sequence and feature recognition. As a consequence of this recognition motif, Pin1 does not so much increase protein stability, but instead
causes conformational changes in stable proteins, allowing modulation of phosphorylation signalling. In the ER, various PPIases have been found to be in a chaperone complex with BiP (Meunier et al., 2002), possibly allowing for the targeting of PPIases to misfolded proteins through recognition by BiP. In other cases, ER-localized PPIases are required for proper maturation of specific proteins. Here, sequence specificity may be important, as in certain instances PPIases have been required for maturation of only some members of a substrate family, while less conserved members are less dependent upon PPIase action (Stamnes et al., 1991). In general for the secretory pathway, the substrate specificities and recognition determinants of PPIases are less well characterized than for the other chaperone systems described. That said, PPIs clearly have important role in ER folding, particularly in the case of large and difficult to fold extracellular matrix molecules (Davis et al., 1998; Patterson et al., 2000). Recognition of misfolded proteins by the PPIases may depend on association with other protein misfolding sensors, or by the exposure of specific sequences containing proline residues available for isomerization.

**Cytosolic chaperones**

Some cargo proteins have large cytoplasmic domains that require folding assistance from cytoplasmic chaperones at the ER surface. These include important receptors such as the cystic fibrosis transmembrane regulator (CFTR), where the mutation causing the bulk of cystic fibrosis causes a protein folding defect at cytoplasmic sequences, leading to intracellular retention, and protein degradation (Cheng et al., 1990). The pathway of degradation for proteins with cytoplasmic folding “lesions” will be discussed in a subsequent section, however, folding factors equivalent to ERAF can assist in the proper folding of such proteins. Through a proteomics analysis, Balch and colleagues have found that cytoplasmic folding enzymes are most important in CFTR folding and that the modulation of these cytoplasmic factors can dictate proper trafficking to the plasma membrane (Wang et al., 2006). The determinants of cytosolic lesion
recognition largely use the Hsp70, Hsp90, and chaperonin systems, along with their regulators, as N-linked glycosylation and disulfide bonding only occur on the luminal side of the ER. The mechanism of action of Hsp70 and Hsp90 chaperones have been described. Chaperonins are large multi-subunit complexes (made up of CCT proteins) that are conserved from \textit{E. coli} to humans. They form nanocages that use ATP hydrolysis to catalyze conformational changes in misfolded proteins and then release them back into the cytoplasmic environment (Clarke, 2006). The determinants of chaperonin binding are largely hydrophobic, with eukaryotic chaperonins having slightly more specificity in that certain combinations of polar and non-polar residues are required (Clarke, 2006). That said, in general chaperonins are folding enzymes for a wide variety of substrates and can localize to the cytoplasmic face of the ER membrane to aid in the folding of diverse secretory cargo.

\textbf{ER-associated degradation (ERAD)}

\textbf{Selection of terminally misfolded/destabilized proteins for degradation}

Newly synthesized proteins that fail to achieve a certain threshold of stability are targeted for degradation by the ERAD pathway. The threshold set point for degradation is cell type dependent and is probably tuneable via signal transduction mechanisms (Nita-Lazar and Lennarz, 2005; Sekijima et al., 2005). This requires a certain level of recognition whereby highly unstable proteins are sequestered from ERAF into the ERAD pathway, while better folded proteins remain available additional folding or secretory exit.

Misfolded protein recognition by ERAD constituents use similar mechanisms to those by employed by chaperones. Particularly, the state of the N-glycan chain determines which proteins can enter ERAD. For glycoproteins, the resident ER protein Htm-1 in yeast and its mammalian orthologue EDEM recognize and target misfolded proteins for degradation (Hosokawa et al., 2001; Jakob et al., 2001; Molinari et al., 2003; Oda et al., 2003). These proteins act as \(\alpha_1,2\)-exo-
mannosidases to generate an N-glycoform recognized by the ER luminal protein Yos9p (Clerc et al., submitted). Yos9p has a mannose-6-phosphate homology domain that recognizes glycoproteins and targets their degradation (Bhamidipati et al., 2005; Kim et al., 2005; Szathmary et al., 2005). Thus, N-glycan modification by recognition proteins such as Htm-1/EDEM sequesters proteins into an ERAD pathway for degradation. It is not known whether nonglycoproteins are targeted for ERAD using recognition modules from other chaperone systems.

Another recognition mechanism, utilizing a series of checkpoints involving analysis by E3 ligases and subsequent selection for degradation by ubiquitination has been posited for secretory proteins with misfolded cytosolic or membrane domains (Carvalho et al., 2006; Vashist and Ng, 2004). Secretory proteins engineered to have misfolding “lesions” at selected sites on the protein have defined separate ERAD pathways for luminal versus membrane versus cytosolic misfolding (ERAD-L, luminal; ERAD-M, membrane; and ERAD-C, cytosolic). In all cases, the ERAD recognition modules are coupled to retrotranslocation and targeting to the ubiquitin-proteasome system through the formation of larger complexes consisting of misfolded protein recognition modules and ubiquitin ligases. For instance, a complex has been isolated in yeast consisting of Yos9p and Hrd3p (oriented to the ER lumen for misfolded protein recognition) along with the putative retrotranslocon channel Der1p, as well as with cytosolic-oriented ubiquitin ligases and adaptors of the p97/VCP orthologous complex (Carvalho et al., 2006; Denic et al., 2006). In total, this multi-constituent complex is important for the degradation of luminal substrates in the ERAD-L pathway and couples substrate recognition to retrotranslocation and proteasomal degradation (discussed in the next section). On the other hand, a complex consisting of different ubiquitin ligases along with orthologous p97/VCP adaptors was isolated that is important for the ERAD-M and ERAD-C pathways (Carvalho et al., 2006). Similar complexes may exist in mammalian cells; for instance, a complex of
Derlin-1 with E2 and E3 ubiquitin ligases cooperates with the Hsp70 system to regulate ERAD of cytosolic domain mutants of CFTR (Younger et al., 2006).

Such complexes are likely to define the pathway of ERAD: misfolded glycoproteins are marked as such by EDEM/Htm-1 modification of the N-glycan. This leads to recognition by Yos9p, which is in a complex with the retrotranslocation machinery (see next section) and the ubiquitination machinery required for presentation to the proteasome.

**Retrotranslocation into the cytoplasm**

Unstable luminal and transmembrane cargo are not degraded in the ER, but by the cytosolic proteasome. This requires that proteins targeted for the ERAD pathway are extracted from the ER (lumen or membrane) into the cytosol in a process known as retrotranslocation (i.e. the opposite of translocation into the ER). Initially, studies suggested that misfolded proteins were retrotranslocated for degradation using the same translocon channel (Sec61) used by secretory proteins to enter the ER (Wiertz et al., 1996b). However, not all proteins had impaired degradation in Sec61 mutants, suggesting the presence of additional pathways for retrotranslocation (Jarosch et al., 2002).

Insight into the identification of new members of the retrotranslocation machinery has largely depended on the analysis of viral proteins which block MHC I expression in order to prevent detection by the immune system. Specifically, the human cytomegalovirus proteins US2 and US11 have been found to promote retrotranslocation of MHC I into the cytosol, causing its subsequent downregulation through proteasomal degradation (Wiertz et al., 1996a; Wiertz et al., 1996b). Analysis of the human interacting partners of US2 and US11 in transfected cells has identified several ER proteins necessary for retrotranslocation. These include Derlin-1, a postulated channel for retrotranslocation, signal peptide peptidase and Sel1L, a member of a ubiquitin
ligase complex (Lilley and Ploegh, 2004; Loureiro et al., 2006; Mueller et al., 2006; Ye et al., 2004). How these constituents fit together to functionally enable retrotranslocation is not known, and additional constituents are also likely to exist. It now appears that there are multiple pathways for retrotranslocation, possibly requiring multiple channels, or even a lipid droplet-based system (Ploegh, 2007). Speculatively, this complexity may be required for the cell to maintain degradative capacity in the face of a high misfolded protein load. As described, the process of recognition of misfolded proteins is coupled to the retrotranslocation and degradative pathways (Carvalho et al., 2006), therefore further definition of the misfolded protein surveillance complexes of the ER should help to define the molecular logic of ERAD.

Proteasomal degradation

Retrotranslocation of proteins targeted for degradation is coupled to the cytoplasmic proteasome by the ubiquitin system. For many ERAD substrates, this is achieved using a complex consisting of the cytoplasmic AAAtpase p97/VCP and various adaptors that bind ubiquitin and ubiquitin ligases as well as recruit p97/VCP to the ER membrane. Here, the p97/VCP provides the energy to extract proteins from membranes and act as an “unfoldase” in order to allow misfolded protein transit across the (likely narrow) retrotranslocation channel, although the exact structural mechanism contains several open questions (Bukau et al., 2006). The adaptors provide specificity to p97/VCP, since this protein also functions in a variety of other cellular processes, including degradation of cytoplasmic proteins, organellar membrane fusion, cell cycle regulation, and DNA repair (Halawani and Latterich, 2006). As an example of adaptor function, in yeast the adaptor UBX2 recruits its p97/VCP orthologue to the ER membrane and deletion of this factor causes impaired ERAD (Neuber et al., 2005; Schuberth and Buchberger, 2005). In mammalian cells the putative retrotranslocation channel Derlin-1 interacts with p97/VCP, providing a physical linkage between retrotranslocation and protein extraction (Ye et al., 2004), although additional factors may be necessary for the
recruitment step of p97/VCP to the ER. The prototypical ERAD-specific p97/VCP adaptors are Npl4 and Ufd1, a complex which is specific to degradation of proteins from the ER (Tsai et al., 2002). Ufd1 binds to polyubiquitinated proteins and thereby links p97/VCP to the ubiquitin machinery (Ye et al., 2003). Further work in yeast has now defined the pathway from the orthologue of p97/VCP to the proteasome. Oligoubiquitinated substrates collected by the orthologous p97/VCP complex are multiubiquitinated by the yeast E4 ubiquitin ligase Ufd2, before being bound by a Rad23/Dsk2 complex that binds ubiquitin conjugates for presentation to the proteasome (Richly et al., 2005). In this manner, misfolded proteins extracted from the ER are able to interact with the ubiquitin-proteasome system for degradation.

In addition to being unfolded prior to proteasomal degradation, glycosylated proteins must be deglycosylated. This is achieved in the cytoplasm using a peptide N-glycanase enzyme that has been found to interact with both Derlin-1 at the ER membrane as well as Rad23 in proximity to the proteasome (Katiyar et al., 2005; Katiyar et al., 2004). These locations would allow for flexibility in the timing of deglycosylation, as misfolded ER proteins could be deglycosylated at any step along the way from retrotranslocation to proteasomal degradation.

**Non-proteasomal degradation**

A number of ER proteins that undergo degradation appear to not use the proteasomal system (Schmitz and Herzog, 2004). Although proteolysis occurring inside of the ER has been postulated, mechanistic detail about such a process remains undescribed. On the other hand, the proteasome-independent degradation mechanism of autophagy has recently been shown to be important for several substrates and protein misfolding diseases (Bukau et al., 2006). Autophagy is the collection of protein aggregates into autophagosomal vesicles and subsequent fusion with the lysosome for degradation. Such a pathway may be necessary for the clearance of aggregated proteins, since protein aggregates are prone to causing
proteasome blockage (Bence et al., 2001). The autophagy process is regulated by a large number of Atg constituents. Interestingly, several of these constituents are upregulated by ER stress and regulate ER expansion during acute protein misfolding conditions (Bernales et al., 2006). Determining the autophagic pathways of ER protein degradation and the function of autophagosomes during ER stress represent an open area of future investigation for the ERAD field.

**Secretion of cargo out of the ER: engaging the COPII coated vesicle**

COPII vesicles mediate transport of secretory cargo from the ER to the Golgi. These are coated vesicles that form at specialized sites known as ER exit sites (Murshid and Presley, 2004). In order to engage the COPII vesicle and leave the ER, secretory cargo must avoid ERAD and escape ER retention by ERAF. Also essential is that secretory cargo are competent to engage the COPII vesicle machinery. Much effort has been put into determining the mechanism of cargo selection by COPII vesicles, however it remains difficult to accurately predict, based upon sequence alone, those proteins that are resident to the ER versus those that are secreted. This may be for several reasons. First, many ER resident proteins are themselves competent to engage the COPII machinery and can be secreted from the ER to the Golgi. These proteins maintain their ER localization through the use of ER retention sequences (i.e. dilycine motifs, KDEL) that cause the recycling of ER resident proteins from the Golgi back to the ER (described in the next section). Second, cargo are not automatically competent for secretion simply because they have stable folds. For instance, it has been found that stable proteins lacking certain ER export sequences (i.e. DXE) are retained in the ER because they are unable to engage the COPII vesicle (Nishimura and Balch, 1997). Similarly, COPII vesicles can be engaged even by misfolded proteins if their concentrations rise within the ER. For instance, in experiments where aspects of the ERAD machinery are mutated, misfolded proteins can be secreted so long as DXE-type ER export signals for COPII engagement are preserved (Kincaid and Cooper, 2007).
The COPII coated vesicle is conserved in all eukaryotes, with human homologues named after proteins initially discovered in yeast. COPII vesicles are structurally composed of a cytosolic heterodimeric adaptor complex Sec23 and Sec24, which directs cargo selection (Gurkan et al., 2006). The Sec13 and Sec31 proteins form the vesicle cage, which has a unique cube-like (cuboctahedral) structure (Stagg et al., 2006). The mechanism of COPII selectivity is thought to involve direct interactions between Sec24 and the ER exit signals (such as DXE) of the secretory cargo. To account for the diversity of secretory proteins that must undergo ER-Golgi transport, several cargo binding sites have been found on Sec24, each of which seems to recognize different types of exit signals (Miller et al., 2003). Moreover, there are several isoforms of Sec24 found in each species, many of which are tissue-specific (Gurkan et al., 2005) and could widen the spectrum of secretory proteins able to engage COPII vesicles. In total, the efficiency of COPII vesicles to engage a specific secretory protein, the vigilance of the ERAD recognition machinery, and the abundance of ERAF all work together to determine the secretion level of a cargo protein of a given stability.

**ER retention signals**

As noted, resident proteins of the ER are often competent to engage COPII vesicles and become transported to the Golgi apparatus. To counteract this “leakage”, the cell uses a Golgi-ER recycling system that recognizes short sequences on the C-terminus of ER resident proteins. For ER resident membrane proteins a cytoplasmic-oriented dilycine motif of KKXX interacts directly with COPI coatomer, a constituent of the Golgi-ER transport machinery, in order to be recycled back to the ER (Cosson and Letourneur, 1994; Letourneur et al., 1994). For ER resident luminal proteins, a C-terminal K(or H)DEL sequence is recognized by a KDEL receptor in the Golgi apparatus, and this facilitates retrieval back to the ER (Pelham, 1991). Some variability around these sequences is found among resident ER proteins, however 60-70% of the known ER-localized
proteins lack a retention signal of any sort (Scott et al., 2004). Several other mechanisms by which ER proteins are segregated from secretory proteins have been proposed (Teasdale and Jackson, 1996), but exactly how the majority of ER resident proteins avoid accumulation in the Golgi apparatus remains unclear.

**Protein synthesis and folding generates a secretory “load”**

Protein secretion imparts stresses on the cell that must be buffered by cellular homeostasis; if secretory demand exceeds the cellular capacity, secretion falters and/or disease results. First, the supply of free amino acids available to a given cell is finite. Metabolic pathways such as gluconeogenesis regulate the extracellular availability of amino acids, while the proteasome and other mechanisms regulate intracellular availability. Protein secretion by definition causes a net loss of freely available amino acids to the cell; however, the trapping of free amino acids into proteins occurs immediately upon new protein synthesis and is independent of whether or not the proteins are properly folded or secreted (for instance, protein aggregates can sequester amino acids, causing amino acid deprivation). Secondly, synthesis rates and the kinetic and thermodynamic rates of folding can be limited by the availability of cellular assistance (ERAF), which may need upregulation (i.e. tyrosinase requires a melanocyte-specific factor for rapid folding) (Francis et al., 2003). Third, as described above, oxidative protein folding and disulfide bond formation generate free radicals which cause oxidative stress and need to be buffered. Fourth, the size of the ER compartment is highly plastic, with secretory cells having far more extensive ER membranes, suggesting a structural demand for membrane building when secretion increases. All of these demands fluctuate during times of low versus high secretory load, and must be more prominently accommodated in secretory cell types.

**The unfolded protein response relieves secretory load**
The ER is able to sense the accumulation of misfolded proteins and respond in a manner that mitigates the oxidative, metabolic, and structural stresses to the cell. Misfolded proteins can accumulate under several conditions, including genetic mutations (which manifest as protein conformational changes), high secretory demand (due to the relative increase of proteins that can fall “off pathway” in their folding), heat shock, and high expression of viral proteins after infection. Such cell biological conditions can occur during normal and pathological states. An example of a normal activation of ER stress is during B-cell development where the high secretory load of antibody secretion activates ER stress sensors to drive differentiation from immature to mature B-cells (Reimold et al., 2001). Pathologically, ER stress is activated in a diverse range of diseases from type II diabetes to cancer to atherosclerosis (Marciniak and Ron, 2006).

A transcriptional program known as the “unfolded protein response” couples ER stress to the cellular response. Transciptome profiling by microarray studies has revealed that the UPR activates a diversity of targets both inside and outside the ER, with several hundred genes upregulated after pharmacological ER stress (Harding et al., 2003; Travers et al., 2000). Annotation of these targets has suggested a certain logic to UPR-upregulated genes. For ER-localized genes, these include constituents of the translocation, glycosylation, folding, degradation, and lipid biosynthesis machineries, whereas outside of the ER proteins that maintain amino acid availability, redox regulation, and other metabolic regulators are part of the UPR response. Together these genes fit a profile whereby ER stress is accommodated through upregulation of ERAF and ERAD in the ER, as well as by proliferation of the ER compartment, and through metabolic buffers elsewhere in the cell. Here, the UPR attempts to re-fold and degrade misfolded proteins while supporting general cellular protection against their metabolic consequences. That is, the UPR is a transcriptional program that acts as a buffer against a high secretory load. In Chapter 2, 7 new targets of the mammalian UPR are identified; these ER-localized proteins are hypothesized to help accommodate ER stress.
The mechanism of sensing protein folding in the ER

The accumulation of misfolded proteins is recognized by stress transducers in the ER that activate the movement of transcription factors from the ER to the nucleus upon stimulation. Such transducers of the UPR are Ire1, Perk, and ATF-6 (Marciniak and Ron, 2006). Additional proteins suggested to also function as transducers have been identified, including the ER-localized transcription factors Oasis and Bbf2h7 (Kondo et al., 2005; Kondo et al., 2006). These two proteins seem to display tissue-specific expression.

The mechanism by which the ER recognizes misfolded proteins and triggers the UPR is best understood for Perk and Ire1. Having structurally similar ER luminal domains, Perk and Ire1 both interact directly with the ER chaperone BiP and it has been noted that activation of Perk and Ire1 by acute protein misfolding stress coincides with the loss of BiP interaction (Bertolotti et al., 2000). As such, the following model has been proposed: under normal folding conditions, BiP is in excess abundance compared to its requirements as a chaperone and is freely available to bind to the luminal domains of PERK and Ire1, maintaining their inactivity. However, during conditions of high secretory load or protein misfolding, BiP preferentially binds to its misfolded client substrates, leaving Perk and Ire1 exposed and available for dimerization and activation. The problem with this model is that in yeast a mutant form of Ire1 that does not bind BiP has been generated. This mutant Ire1 does not exhibit hyperactivation under normal or protein misfolding conditions, in direct contrast to the model (Kimata et al., 2004). Crystal structures are now available for the luminal domain of Ire1 in yeast (Credle et al., 2005) and humans (Zhou et al., 2006). Both structures suggest that the dimerization event creates an MHC-like peptide binding groove. While in yeast it is suggested that this groove could bind misfolded protein directly (and thereby sense protein misfolding), the human groove is thought to be too narrow to accommodate protein-protein interactions. Moreover, human Ire1 can dimerize in the absence of peptide binding and the residues thought to be
necessary for binding peptides in yeast are buried in the human structure. There may be differences in the mechanism by which yeast and humans sense protein misfolding, although high structural and sequence conservation exists. At present, the specific roles of the MHC-like groove and BiP binding in the cellular detection of protein misfolding are not mechanistically clear.

Chaperone binding to the other ER stress transducer, ATF-6, is also thought to regulate its activation. ATF-6 is not activated by dimerization, but by trafficking to the Golgi, where it is cleaved by Golgi-resident site-1 and site-2 proteases (S1P, S2P). Upon Golgi cleavage, ATF6 trafficks to the nucleus and functions as a transcription factor (Marciniak and Ron, 2006). The binding of ATF6 to BiP or to the calnexin/calreticulin chaperone cycle acts negatively on ATF6, by retaining it in the ER. For BiP, the interaction is thought to cover over ER exit signals in the sequence of ATF-6 (Shen et al., 2002). For calreticulin, binding is thought to prevent Golgi trafficking of underglycosylated forms of ATF-6 (Hong et al., 2004). Therefore, in all known cases of UPR transduction, chaperone binding is found to be a negative regulator of activation. The ER therefore monitors its own chaperone capacity (that is, ERAF capacity) as a surrogate measure of the secretory load it experiences.

**Increasing the capacity of protein folding through signal transduction**

**Signalling downstream of Ire1**

In contrast to the other transducers, Ire1 is present in all eukaryotes and represents an evolutionarily conserved ER stress transduction mechanism. Activation of Ire1 through dimerization leads to an autophosphorylation event using its cytosolic kinase domain (Shamu and Walter, 1996). In addition to the kinase domain, the cytosolic domain of Ire1 contains an atypical endoribonuclease (RNase) domain (Sidrauski and Walter, 1997). Autophosphorylation leads to activation of the RNase domain, which results in the cleavage of highly specific RNAs. In fact, up
until recently (see the section on transcriptional repression), Ire1 was only known to cleave a single mRNA, that of the transcription factor Xbp-1 (Hac1p in yeast). Cleavage of Xbp-1 mRNA results in the use of a different C-terminus upon translation, which has increased activity as a transcription factor. As a CREB-like transcription factor, Xbp-1 recognizes specific promoter elements to drive transcription of UPR targets (Cox and Walter, 1996). These include specific DNA sequences known as ER-stress elements (ERSE), however, additional unknown sequences likely also exist since Xbp-1 targets cannot be predicted \textit{in silico} purely on the basis of their promoter sequences.

In addition to the activation of the transcription factor Xbp-1, phosphorylated Ire1 is able to bind adaptors coupling ER stress the JNK signalling pathway (Urano et al., 2000). Here, the protein folding state of a cell can be integrated with information from the extracellular environment. Pathologically, ER stress signalling through the Ire1-JNK pathway has been observed to activate insulin receptor substrate-1 and thereby cause insulin resistance in mouse models of type II diabetes (Ozcan et al., 2004). Whether this mechanism of insulin resistance is adaptive is unclear, however, chemical chaperones that ameliorate ER stress have been observed to relieve diabetic symptoms in genetically obese mice (Ozcan et al., 2006).

Additionally, Ire1 signalling can be regulated by other signalling pathways in the cell. Molecular modulators of Ire1 signalling include Nck and PTP-1B (Gu et al., 2004; Kebache et al., 2004). The purpose of such modulation is not known, however, it does provide an opportunity for fine tuning of the UPR response during different cellular circumstances.

\textbf{Signalling downstream of PERK}

Activation of the kinase PERK leads to two consequences: phosphorylation of eIF2alpha and activation of the transcription factor Nrf2 (Cullinan et al., 2003;
Harding et al., 1999). Nrf2 phosphorylation results in the activation of genes involved in resisting oxidative stress insults, consistent with the general role of the unfolded protein response (Cullinan and Diehl, 2004). On the other hand, phosphorylation of eIF2alpha has effects on both transcription and translation. In general, eIF2alpha phosphorylation results in translational control, the mechanism of which is discussed in the next section. Here, Perk signalling leads to transient inhibition of translation in order to prevent additional secretory loading of an already overloaded ER. Perk phosphorylation of eIF2alpha has a second consequence, which is to paradoxically promote the translation of the transcription factor ATF-4, through a mechanism also described in the next section. ATF-4 controls genes involved in amino acid metabolism as well as oxidative stress protection. Therefore, PERK signalling initiates an “integrated stress response” to protein malfolding, characterized by cellular reprogramming to refold and degrade secretory proteins, whilst upregulating the cellular factors that buffer the toxic byproducts of such a state (Harding et al., 2003).

Signalling downstream of ATF-6

Following protease cleavage in the Golgi apparatus, ATF-6 migrates to the nucleus where it interacts with various ER stress elements upstream of target genes (Yamamoto et al., 2004). The transcriptional targets of ATF-6 are constituents of the ERAF and ERAD machineries, as well as lipid biosynthetic pathways (Okada et al., 2002).

Interpretation of ER stress by the different transducers

Recently, the activation of the Perk, Ire1, and ATF-6 pathways were compared after treatment with the pharmacological ER stress agents tunicamycin, dithiothreitol (DTT) and thapsigargin (DuRose et al., 2006). Each of these agents has a different mechanism of action: tunicamycin inhibits glycosylation, dithiothreitol reduces disulfide bonds, while thapsigargin causes a reduction in ER
luminal calcium levels. Each causes protein misfolding, however, the Perk, Ire1, and ATF-6 pathways were found to have different types of activation, depending upon the drug used. For instance, the Ire1 pathway was most rapidly activated by DTT, the PERK pathway (eIF2alpha phosphorylation) was most responsive to thapsigargin, and the ATF-6 pathway, while sensitive to tunicamycin and DTT, was relatively unchanged after thapsigargin treatment. These results were interpreted as showing the ability of the cell to optimize responses to protein misfolding under physiological conditions, depending upon the type of stress encountered. In addition, the three pathways have also been shown to be mutually dependent on each other and can engage in signalling cross-talk. For instance, ATF-6 cleavage is required for activation of Ire1 signalling (Lee et al., 2002). In general, however, how these pathways are integrated to upregulate the appropriate cellular responses to a specific type of stress is not well understood. Examples of two genes of unknown function that respond to thapsigargin and DTT but not tunicamycin are shown in Chapter 2.

Limiting protein entry into the ER during times of protein overload

As described above, failure to accommodate the load of secretion can lead to amino acid starvation, oxidative free radical generation, and compromise of cellular functions. To avoid these circumstances, the UPR can activate mechanisms that shut off cellular secretion, thereby allowing the ER time to resolve its problem of overload. Here, limiting toxicity takes priority over continued secretory function. That said, the decision to shut off secretion can only be a temporary measure since there are essential cellular proteins (such as those targeted to intracellular compartments) that require an intact secretory pathway. If the ER cannot resolve its problems of protein overload, apoptosis often results. Three potential mechanisms that limit the influx of new secretory proteins during times of ER overload have been described.

Translational inhibition of secretory proteins
Translational control, or the shutting down of protein synthesis, is a general stress-response mechanism conserved in eukaryotes. Viral infection, amino acid starvation, iron deficiency, and ER protein misfolding are among the conditions sensed by the cell that result in translational control (Kaufman, 2004). Each of these conditions activate kinases (as described above for PERK) that phosphorylate the translation initiation factor eIF2alpha on a single conserved serine residue. Translation initiation begins when eIF2 (a heterotrimer consisting of eIF2alpha, beta, and gamma), forms a ternary complex by binding to the small ribosomal subunit and a methionyl initiator tRNA. With the help of other initiation factors, the ternary complex is localized to mRNA to search for a start codon. The GEF eIF2B is required to release the ternary complex in order to allow another cycle of translation initiation. Phosphorylation of eIF2alpha by the various stress sensors causes eIF2alpha to bind and sequester eIF2B. This diminishes the amount of available ternary complex, and translation initiation (and protein synthesis) is thereby stalled (Rowlands et al., 1988).

As described, sensation of misfolded proteins by PERK causes the phosphorylation of eIF2alpha and in this manner represses translation. Thus, the UPR prevents the translation of new proteins into an already overloaded ER. Such a translational repression is transient, lasting in culture about 30 minutes after the start of pharmacological misfolding stress (Rutkowski and Kaufman, 2004). Reversal of the translational repression is mediated by protein phosphatases such as PP1 and CReP (Connor et al., 2001; Jousse et al., 2003; Novoa et al., 2001), allowing protein synthesis to be restored at later time points. In addition, specific proteins continue translation during eIF2alpha-mediated translational control, via the use of internal ribosomal entry sites (IRES) to bypass the requirement for eIF2 in translation initiation. In particular, the use of an internal ribosomal entry site has been shown for the abundant chaperone BiP (Sarnow, 1989), which functions to re-fold proteins during ER stress and would be helpful for the cell to upregulate during the early part of the UPR. Furthermore, the previously mentioned
transcription factor ATF-4 continues to be translated under conditions favouring eIF2alpha phosphorylation, independent of an IRES. Since ATF-4 can act as a transcription factor to upregulate stress-responsive genes during the UPR it is useful during protein misfolding states favouring eIF2alpha phosphorylation. The mechanism of ATF-4 translation during ER stress relies upon two open reading frames available at the 5’ end of the ATF-4 mRNA (Lu et al., 2004a). Under low stress conditions, the second ORF is preferentially translated, which inhibits translation of the rest of the ATF-4 message and results in low ATF-4 protein levels. During eIF2alpha phosphorylation, it is thought that some translation initiation is spared, and as a consequence of diminished translation initiation, only the first ORF is translated (Lu et al., 2004a). This allows translation of the rest of the ATF-4 message and increased ATF-4 levels. The paradox of translation inhibition leading to increased ATF-4 protein is resolved by this mechanism as well as by the concept that ATF-4 is a low abundance protein. Here, small increases in protein level can have effects relevant to its function as a transcription factor. In general, however, abundant proteins (except those that use an IRES) have decreased synthesis during translational control, allowing relief for the ER from additional secretory load. The importance of translational control under ER stress conditions is underscored by the finding that eIF2alpha mutant mice are highly susceptible to high fat diet-induced type II diabetes and have dysfunctional insulin secretion in the pancreas (Scheuner et al., 2005).

Co-translocational degradation

While protein synthesis and folding occurs co-translationally to secretory proteins as they are translocated into the ER, a new mechanism whereby proteins are also selected for co-translocational degradation has recently been proposed (Oyadomari et al., 2006). Implicated in this process is the UPR-inducible protein p58-IPK, a protein known to inhibit the ER stress sensor PERK as part of a feedback loop to restore translation after prolonged ER stress (Yan et al., 2002). As part of a PERK-independent function, however, p58-IPK associates with the
Sec61 translocon and has a role in the proteasomal targeting of secretory proteins as they are translocated into the ER (Oyadomari et al., 2006). In this case, co-translocational degradation might represent an additional line of defense against protein entry into an overloaded ER. In one model, protein malformation stress leads to activation of PERK and translation inhibition as described above. If after a period of time this intervention fails to mitigate the ER stress, translational recovery is mediated by p58-IPK via reversal of PERK phosphorylation. While this restoration of translation is likely necessary to maintain cellular function, it would not help in eliminating the ER stress; as noted, continued synthesis of secretory proteins into an overloaded ER is likely toxic. With co-translocational degradation, difficult to fold proteins stalled in the translocon (the substrates most likely to be causing the ER stress) are preferentially selected for degradation through an unknown mechanism, but using the same protein, p58-IPK. Since p58-IPK contains a DNA J domain involved in modulating the function of Hsp70 family members, one might speculate a role for cytosolic E3 ubiquitin ligase/Hsp70 complexes in this process. Indeed, the E3 ligases CHIP and RMA1, recently shown to cooperate to cause degradation of misfolded chloride channels in cystic fibrosis (Younger et al., 2006), represent interesting candidates. That p58-IPK and this process has an important physiological role is suggested by the finding that deletion of p58-IPK in mice causes exacerbated type II diabetes in an ER stress model (the Akita mouse) of the disease (Oyadomari et al., 2006).

In another model, co-translational degradation is posited to occur during acute ER stress. Here, non-glycosylated forms of the prion protein were observed to accumulate after acute ER stress, whereas glycosylated forms were diminished (Kang et al., 2006). The non-glycosylated form of prion protein was found in fact to be unglycosylated, on account of having never been allowed to enter the ER and therefore was prevented from accessing the oligosaccharyltransferases that mediate glycosylation. Thus, ER stress activates cellular mechanisms that prevent entry of newly synthesized proteins into the ER, possibly related to the co-translocational degradation mechanism described above. A selection process for
pQC of only certain proteins entering the ER was also discovered, as this effect was dependent upon the signal sequence that was appended to the prion protein. Apparently, the cell is able to “read” the signal sequences of proteins and mediate differential translocation under various cellular conditions. Evidence for co-translocational inhibition of secretory proteins mediated by the signal sequence of the novel chaperone nicalin is presented in Chapter 4.

**Transcriptional repression of secretory proteins**

Another mechanism by which the UPR limits the influx of secretory proteins during times of cellular stress is recently described (Hollien and Weissman, 2006). Here, activation of the UPR leads to cleavage and degradation of mRNAs specific to secretory proteins, thereby preventing their synthesis and entry into the overloaded ER. This transcriptional repression appears to be mediated by the endonuclease function of the UPR stress sensor Ire1. Where the previously described Ire1-mediated splicing of Xbp-1 mRNA leads to the production of a transcription factor form of Xbp-1, it appears that Ire1 is also able to splice mRNAs specific to secretory proteins for the purpose of RNA degradation. Interestingly, substrate specificity for Ire1-mediated degradation is dependent upon the RNA bases corresponding to the substrate’s ER signal sequence. While the exact mechanism whereby Ire1 selects secretory proteins for degradation is not clear, transcriptional repression of secretory proteins during the UPR could be an important process in the cellular regulation of secretion as well as for cytoprotection during stress states. As well, as for co-translocational degradation, it is becoming apparent that the signal sequence of a protein encodes important information for the ER stress pathway, apart from their canonical function in localizing proteins to the secretory pathway.

**ER stress-induced cell death**
From a physiological perspective, the accumulation of misfolded proteins inside a secretory cell represents a dangerous situation. Unstable proteins may be secreted, resulting in amyloid-type formations in the extracellular environment. Furthermore, unstable proteins inside the cell can cause the misfolding of other partly stable (metastable) proteins that exist in every proteome (Gidalevitz et al., 2006). This is thought to involve depletion of the cellular folding capacity including saturation of ERAF and ERAD, and also increases the chance for secretion of proteotoxic species. Therefore from an organismal perspective death of individual cells can be an appropriate response to protein misfolding.

The apoptotic decision is also mediated by the unfolded protein response. In one model, the proapoptotic BAK and BAX proteins form a direct interaction with the UPR sensor Ire1 and modulate Ire1 signalling (Hetz et al., 2006). Although it is unclear how the various signals integrate to activate apoptosis, the physical proximity of the core ER stress and apoptosis machinery suggests a tightly coupled mechanism. Importantly, mice knocked out for both BAK and BAX show exaggerated toxicity after pharmacological ER stress (tunicamycin), lending credence to an important protective role for ER stress-induced apoptosis at the organismal level (Hetz et al., 2006). Other constituents of the core apoptosis machinery (i.e. the proapoptotic BCl-2 family) have also been found to be downstream of ER stress signalling. These primarily involve phosphorylation signalling pathways leading from upstream UPR activators such as CHOP and JNK to downstream BCl-2 activation (Szegezdi et al., 2006).

While the UPR is able to send apoptotic signals, mild ER stress can be adaptive against cell death in a mechanism that also uses the UPR (Lu et al., 2004b). This results in the paradox that the UPR can upregulate both pro-apoptotic and anti-apoptotic genes and yet make a commitment to one outcome versus the other under specific stress circumstances. Such a decision has recently been found to involve the differential stability of pro-apoptotic versus anti-apoptotic gene mRNAs during different stress conditions (Rutkowski et al., 2006). In this model,
mild stress is able to activate the UPR, but unknown negative regulators of apoptosis cause downregulation of pro-apoptotic genes. If ER stress is acute and persistent at a high level, pro-apoptotic genes are highly activated and are therefore unable to be downregulated sufficiently to prevent cell death. Identification of the unknown transcriptional regulators of pro-apoptotic UPR genes would be useful to advance this model. Nevertheless, it is clear that the UPR controls the apoptotic versus adaptive cellular response to ER stress, and that this decision is important to organismal physiology.

**Regulation of the size of the ER compartment**

The amount of ER membrane found in a cell varies widely depending upon the type of cell as well as its intracellular and extracellular environments (Federovitch et al., 2005). For instance, professional secretory cells (i.e. B-cells, pancreatic endocrine and exocrine cells, etc.) that receive a high secretory load have high ER membrane content. Hepatocytes in the liver also have prominent ER membranes on account of their secretory function as well as to support the many detoxification reactions that occur in the lumen or across the ER membrane. These membranes further expand during certain drug treatments which upregulate the cytochrome p450 detoxification enzymes. Various neuronal cell types also have well developed ER compartments in order to store high levels of calcium ions. Generally, the rough ER expands during secretory-type stresses whereas the smooth ER expands during lipid biosynthetic or detoxification stresses, consistent with their partly divergent functions (Federovitch et al., 2005).

Some of the molecular stimulators of ER membrane proliferation are now known. That the UPR directly activates programs involved in membrane proliferation has been shown in cell culture models where overexpression of the active form of Xbp-1 can alone cause expansion of the ER compartment (Sriburi et al., 2004). As previously described, Xbp-1 is spliced into a transcript corresponding to an active transcription factor after ER stress, via the direct activity of the ER stress sensor.
Ire1. Xbp-1 is also necessary for differentiation from immature B to mature B-cells, two cell types which differ greatly in terms of antibody secretory capacity and ER compartment size (Reimold et al., 2001). From another direction, it has been shown that a protein as generic as GFP, if it is overexpressed, targeted to the ER as a transmembrane protein, and improperly allowed to dimerize, can cause smooth ER proliferation (Snapp et al., 2003). Since dimers but not monomers caused proliferation in this paradigm, weak homotypic interactions between the cytosolic domains of proteins on apposing ER membranes may represent a physically sufficient state to activate the requisite signalling pathways for phospholipid upregulation. These results may also explain the finding that individual ER residents such as the p450 enzymes can cause ER compartment proliferation upon upregulation or overexpression (Profant et al., 1999). Reversal of this proliferation, requiring membrane destruction, is also poorly understood but may use lysosomal degradation pathways in the form of autophagy (Bernales et al., 2006).

**Organellar proteomics**

Mass spectrometry-based approaches to protein identification have recently been applied to cell biology (Yates et al., 2005). Organellar proteomics typically involves the generation of a subcellular fraction (by centrifugation and other biochemical means) enriched for a specific organelle or cellular structure. Mass spectrometry alongside of bioinformatics analysis can then be used to quantitatively determine the complement of proteins found within the organelle or cellular structure. Near-complete proteomes have been mapped from enriched fractions of smooth microsomes (smooth ER), rough microsomes (rough ER), Golgi, and Golgi-derived COPI vesicles in rat liver (Gilchrist et al., 2006). In this study, a quantitative and comprehensive mapping uniquely identified more than 1400 proteins in these fractions. For any individual protein, a profile can be generated that shows the abundance of that protein across the different fractions. For example, the translocon channel protein Sec61 is found almost entirely in
rough ER, with very little detected in smooth ER or Golgi fractions (Gilchrist et al., 2006). A separate biochemical fractionation of the ER shows Sec61 to be a membrane protein. In this way, the abundance and localization of all the proteins identified can be determined (Bell et al., 2007b). In addition, 345 proteins of unknown function were uncovered, representing an impressive accomplishment of protein discovery. Finally, insights into biological function were made. For instance, the finding that the Golgi-derived COPI vesicles are enriched for resident enzymes over cargo suggested strong support for a particular model of Golgi function, that of cisternal maturation (Gilchrist et al., 2006). Similarly fruitful approaches have been applied to other organelles and cellular structures, including the phagosome (Gagnon et al., 2002; Garin et al., 2001; Houde et al., 2003; Stuart et al., 2007), the clathrin-coated vesicle (Blondeau et al., 2004; Wasiak et al., 2002), and other subcellular structures (Au et al., 2007; Yates et al., 2005).

Proteomics methods

Here, I will give a brief and broad overview of the methods of proteomics used by groups such as Gilchrist et al. 2006, since this type of data is used in this thesis. Proteomics relies upon mass spectrometry-based identification of proteins enzymatically cleaved into peptide fragments. The general methods used can be broken up into three main steps. 1. Sample preparation. 2. Mass spectrometry. 3. Bioinformatics matching of spectra to peptides. The methods used by Gilchrist et al. 2006 can be found in more detail elsewhere (Au et al., 2007). It should be noted that the technical aspects of each step are highly involved and efforts at technical improvements of each step are ongoing (Au et al., 2007; Bantscheff et al., 2007; Van Oostrum, 2008).

Sample preparation
This includes the biological methods used to generate the sample of interest. For instance, by the methods of Gilchrist et al. 2006, rat liver is made into subcellular and biochemical fractions by biochemical fractionation and centrifugation. The fractionation is described in more detail in Chapter 1. The samples (fractions) are run on 1D SDS-PAGE and stained with Coomassie blue. The entire gel lane, corresponding to a sample, is then cut into slices (approximately 1 mm). Each gel slice is digested with trypsin for in-gel protein digestion. This generates small peptide fragments (tryptic peptides) suitable for mass spectrometric analysis.

**Mass spectrometry**

In this case, the basic utility of mass spectrometry is that it is able to generate mass spectra corresponding to the tryptic peptides. These mass spectra can then be used with high confidence to identify the peptide sequence. Finally, these peptide sequences are matched to proteins, thereby enabling identification of the proteins found within complex biological mixtures (Aebersold and Mann, 2003).

Mass spectrometry is more accurate when the samples entering the mass spectrometer are less complex. Complex samples include biological samples (i.e. containing many different proteins of variable abundance). Sample preparation by subcellular fractionation is one way to decrease sample complexity and improve proteomics identifications (Brunet et al., 2003). SDS-PAGE followed by cutting the gel lane into 1 mm slices as described under sample preparation also reduces sample complexity. The next step to reduce sample complexity is to introduce the sample into a liquid chromatography column, which causes peptides to elute off the column in an ordered fashion based on their hydrophobicity. This further simplifies the sample as it is introduced into the mass spectrometer.

Only ions can be manipulated into and within the electric fields used by the mass spectrometer. Therefore, peptides are ionized using an ionization source. Ionization is commonly performed by electrospray ionization or MALDI (matrix-
assisted laser desorption ionization) depending on the instrument used (Lin et al., 2003). Both methods achieve ionization of peptides. In electrospray ionization, peptides eluting off of the liquid chromatography column are subjected to a strong electric field. This causes the peptides to desolvate and ionize. In addition to causing ionization, the electric field directs peptides into the mass spectrometer.

The mass spectrometer detects the mass-to-charge (m/z) ratio of ionized peptides and records a mass spectrum. Different types of mass spectrometers exist. The data set used in this thesis used a tandem mass spectrometry (tandem MS) approach (Gilchrist et al., 2006). Here, ionized peptides fly through a quadrupole selector that selects certain types of peptides for further analysis. Selection in the quadrupole is based on predetermined peptide characteristics such as m/z, charge, and/or intensity. Selected peptides are then accelerated into a collision cell containing an inert gas such as argon or nitrogen. The collisions of the ions with the inert gas results in the breaking of peptide bonds, or “fragmentation” of the peptide into smaller constituent peptides of shorter sequence. These fragments then fly through a time-of-flight (ToF) analyzer into a detector that records the m/z and mass spectrum of the fragmented peptide. This fragmentation (tandem MS) spectrum is then used to determine the sequence of the original peptide, via a computational matching process.

**Matching spectra to peptides and peptides to proteins**

What a tandem MS spectrum should look like for any tryptic peptide of known sequence can be theoretically predicted. One can therefore construct a computational database containing the theoretical mass spectra for all peptides in a defined database (i.e. the database of non-redundant human sequences). Probabalistic software is then used to match the experimentally-obtained spectra to the theoretical spectra. One example of such software is called “mascot” (Perkins et al., 1999). This software finds the best match between an experimentally determined spectrum and a theoretical spectrum of known
sequence. By matching the combined information of the peptide mass (which is matched first) and then the fragmentation (tandem MS) spectra, the sequence of a peptide is determined. Since the matching is probabilistic, a score greater than 95% confidence is often used to qualify spectra for peptide sequence matching. In practice, this approach leads to greater than 50% of the total tandem mass spectra assigned to peptides, with less than 1.5% false positive assignments (Au et al., 2007).

Once the sequence of a peptide is determined, the cognate protein is identified. Peptides that hit only one sequence with perfect match are deemed “unique” peptides, and this identifies the protein unambiguously. Peptides that align perfectly with multiple proteins (sharing homologous sequences) are called “shared” peptides. One method of determining protein abundance (see below) utilizes the number of tandem MS found to identify a given protein. For shared peptides, these are apportioned to proteins based on the ratio of unique peptides seen. To illustrate, imagine two proteins, X and Y that share 10 peptides (10 peptides that align equally well with both X and Y). If X has 2 unique peptides and Y has 3 unique peptides, then X will be assigned 4/10 shared peptides while Y will be assigned 6/10 peptides, leading to a total abundance of 6 peptides for X and 9 peptides for Y in that sample.

**Quantitation in Proteomics**

Proteins can be identified within a sample by proteomics using the methods described above. Mass spectrometry can also be used to quantitate the abundance of proteins within complex mixtures (i.e. biological samples). Different methods exist to quantitate protein abundance. These include methods using protein labelling (i.e. SILAC, and iTRAQ) as well as label free methods (Bantscheff et al., 2007).
In the data used in this thesis, that of Gilchrist et al. 2006, a label free method called redundant peptide counting was used. The theory behind this method is that in a complex mixture of proteins, abundant proteins are more likely to be “seen” by the mass spectrometer (“seen” meaning selected for fragmentation, followed by identification at 95% confidence as described) than less abundant proteins. For example, an abundant protein like albumin may be identified by the mass spectrometer hundreds of times. That is, the mass spectrometer selects peptide ions for fragmentation corresponding albumin sequences hundreds of times in a comprehensive proteomics study from a rat liver sample. In contrast, a low abundance protein such as a transcription factor may have corresponding tandem mass spectra identified only once or twice within the same mixture. Hence, the number of tandem mass spectra (i.e. the number of peptides) “seen” by the mass spectrometer may be a surrogate for abundance, particularly when taking into account proteins of similar molecular weight and number of tryptic peptides. This intuitive idea has been supported by several studies. For instance, McPherson and colleagues (Blondeau et al., 2004) were able to recapitulate the stoichiometry of the clathrin coat complex by this method. That is, the relative abundance of the clathrin heavy chain, light chain, and adaptor proteins measured by redundant peptide counting were found to be at ratios as expected from previous literature. Gilchrist et al. 2006 further showed that redundant peptide counting to quantitate individual protein amounts (i.e. albumin) was similarly quantitative to the use of radiolabelled antibodies used in Western blotting the same protein. Similarly, enzymatic activities and quantitation of marker proteins across different cellular fractions correlated well with the representation of abundance provided by redundant peptide counting (Gilchrist et al., 2006).

A recent critical review suggests that such label free quantification is superior to other methods in terms of dynamic range and cost, but at the present time has more uncertainty in accuracy compared to labelling methods (Bantscheff et al., 2007). More recently, the label free approach is extended through the use of computational learning algorithms to obtain absolute levels of protein
quantification from peptide counts (Bergeron and Hallett, 2007; Lu et al., 2007; Mallick et al., 2007). Further refinement of label free methods may yield additional improvements in accuracy and use.

Other methods employ labelling, which add a mass tag to the peptides entering the mass spectrometer. One method, named SILAC (stable isotope labeling by amino acids in cell culture), introduces stable isotope-labelled arginine and lysine into cell culture so that the resultant tryptic peptides are all labelled (Ong et al., 2002). Two samples can be compared quantitatively by having one sample to be labelled with the other unlabelled. When the samples are mixed, the mass spectra of pairs of peptides (labelled and unlabelled) are compared for intensities of specific peaks, thereby quantitating the difference between the samples. This method theoretically contains the least opportunity for error to be introduced into the quantitation process (Bantscheff et al., 2007). Recently, this method has been extended in vivo with a SILAC diet being fed to mice, allowing proteome comparisons between wild type and knockout animals (Kruger et al., 2008).

Another approach is to label protein mixtures after they have been processed biologically, but prior to mass spectrometry. An example is the ICAT approach whereby cysteines from tryptic peptides are chemically modified with a tag that once again allows comparison between peptide pairs by peak intensities on mass spectra (Gygi et al., 1999). This method has been refined such that a commercial iTRAQ (a variant of ICAT) reagent purports to multiplex up to 8 samples at a time for comparison (Ross et al., 2004). Post-processing labelling methods are less expensive than SILAC, but have some additional technical difficulties in data analysis and are generally not thought to be as quantitative as SILAC (Bantscheff et al., 2007).

In summary, mass spectrometry is used on complex mixtures developed from cell biological experiments. This allows a quantitative representation of the protein constituents found in biological samples. In their current forms, metabolic
labelling (i.e. SILAC), peptide labelling (i.e. ICAT), and label free methods (redundant peptide counting) have variable advantages and disadvantages as proteomics methods.

**Determining organization and function in the ER**

The preceding literature review has described the functions of the endoplasmic reticulum with respect to its central role in protein secretion. There is a great degree of complexity required for efficient protein secretion, with many elegant cellular solutions to the biochemical problems posed by newly synthesized proteins. Despite many years of effort in cell biology, many questions remain. Global approaches to look at all the constituents of the secretory pathway have recently led to a prominent set of questions about this complexity, which is, with at least 1400 resident proteins of this pathway, what are all these constituents doing? Furthermore, how are these able to be organized to prevent disarray?

This thesis begins with the proteomics resource of Gilchrist et al. 2006, which makes the unexpected finding that proteins can be reliably localized to multiple sublocations. Some proteins were found to be exclusively rough ER, some proteins both in the rough and smooth ER, while others are found exclusively in the smooth ER. Hence, the work of Gilchrist et al. 2006 suggest that resident proteins of the ER are highly organized through restricted localizations.

The simplest explanation for such an order would be that the ER is made up of subdomains (defined by a group of proteins localizing together in the proteomics resource), where proteins which localize together function together. Chapter 1 uses the secretory pathway function of the ER to study this question. Indeed, the sublocalization of known proteins in the proteomics resource corresponds with high reliability to specific functional steps in secretory cargo synthesis, maturation, and degradation; that is, each of these functional processes have their own spatial location within the ER. Specifically, synthesis and translocation occur
in the rough ER (cytosolic face and integral to the membrane respectively), while the degradation machinery is segregated away from these machines into a subdomain of the smooth ER. This subdomain (an ERAD compartment) contains at the cytosolic surface of the ER molecular chaperones, the ubiquitin-conjugating machinery, p97, and the proteasome.

This suggests a model of ER function whereby newly synthesized cargo traverse functional groups that are restricted in space within the ER. This coincides with the work of Balch and Kelly, which find on a protein energetics basis that cargo maturation can be modeled by separating the same functions to which is here ascribed separated compartments (Wiseman et al., 2007a; Wiseman et al., 2007b). It also agrees with all of a limited number of previous studies making observations on diffusion of ER resident proteins (Nikonov et al., 2002; Snapp et al., 2006).

To test this spatial model of the ER, further chapters have focused on its predictive power. If known proteins are localized into specific locations, then unknown proteins that have the same restricted location are predicted to function in the same process. Chapter 2 identifies candidate unknown proteins for a function in the secretory machinery of the ER. Chapters 3, 4, and 5 focus on individual proteins and uses the proteomics localization to predict protein function. In Chapter 3 the cytosolic rough ER protein nudilin is found to localize to the P-body and affect the secretory cargo protein tyrosinase at a step upstream of ER entry. In Chapter 4, the transmembrane rough ER protein nicalin interacts with signal peptide peptidase and affects tyrosinase translocation into the ER. Chapter 5 focuses on three additional proteins which interact transiently with tyrosinase. Stexin is localized to rough ER, TMX2 to rough and smooth ER, and erlin-1 is localized to smooth ER. Consistent with these localizations, stexin interacts with an early N-glycosylated form of tyrosinase while TMX2 and erlin-1 interact with later forms of tyrosinase. The features of nudilin, nicalin, stexin, TMX2, and erlin-1 examined in this thesis are all congruent with the localization
predicted by the proteomics resource. Hence, the model of the ER as segregated into functional subdomains has predictive power into determining the function of individual unknown proteins.
CHAPTER 1: Interrogation of the proteomics resource reveals the spatial organization of ER cargo processing

Abstract

The proteomics resource (Gilchrist et al., 2006) gives a quantitative mapping of over 1400 proteins found in isolated fractions enriched for rough ER, smooth ER, and Golgi. The localization of each protein is described by its abundance in each of the three fractions. In this chapter, the proteomics resource is interrogated for over 100 proteins known to function in protein synthesis, entry into the ER (translocation), folding, and degradation. Remarkably, the proteins in each of these functional categories showed distinct localizations across the rough and smooth ER. Proteins involved in translation and translocation were found predominantly in rough ER. Proteins involved in ER folding were found equally in rough and smooth ER. Proteins involved in proteasome-mediated degradation and cytosolic folding were found in the smooth ER. Greater resolution of these functions were found when interrogating the Triton X-114 extracted ER fractions in the proteomics resource. Translation proteins were found in the salt wash fraction of rough ER, indicating a cytosolic association with the ER. Translocation was found to occur in the detergent-extracted (membrane) portion of the rough ER. Luminal ER folding was found in the aqueous phase of both rough and smooth ER fractions, while degradation and cytosolic folding were found in the aqueous phase of smooth ER. These results suggest that the ER is organized into spatially restricted subdomains corresponding to function. These domains are restricted with respect to rough and smooth ER and by the membrane. A new model of cargo trafficking is posited, whereby newly synthesized cargo traverses functional complexes from rough to smooth ER in order to attain their correct fate in secretion or degradation.

Introduction
The ER is a single continuous compartment with free diffusion of small molecules and proteins throughout (Lippincott-Schwartz et al., 2000). At the same time, visualization of ER morphology shows that it has a high degree of heterogeneity. For instance, the nuclear envelope is visualized as continuous with the rough ER but looks completely different due to the localization of nucleus-specific protein complexes (such as nuclear pore complexes) anchored to lamins inside the nucleus (Aitchison and Rout, 2002). Another example is the classical morphological distinction between rough ER and smooth ER on electron microscopy, generated by the presence of ribosomes on the rough ER membrane only. In yeast cells, the ER has at least two morphological types. The ER surrounding the nuclear envelope (“cortical ER”) consists of flattened sheets. The peripheral ER, which is sparsely connected (though continuous) with the cortical ER, consists of tubules that extend through the cytoplasm (Shibata et al., 2006). An intermediate compartment between the ER and Golgi (ERGIC) that does not contain ribosomes can also be distinguished on morphological grounds in many cell types (Appenzeller-Herzog and Hauri, 2006).

Functionally, the ribosome-containing ER is involved in protein synthesis. The function of the smooth ER is thought to be involved in detoxification reactions. Cells containing large amounts of smooth ER, such as hepatocytes, have high concentrations of detoxification and lipid modification enzymes. Furthermore, induction of the detoxifying p450 enzymes by drugs such as barbiturates cause proliferation of the smooth ER (Federovitch et al., 2005). At times, other subdomains of the ER have been imparted with functional significance. Although it is not directly continuous with the ER itself, ERGIC is well established to function in the forward transport of cargo from the ER to the Golgi as well as in the retrograde retrieval of ER resident enzymes from the Golgi back to the ER (Appenzeller-Herzog and Hauri, 2006). Quality control compartments of various composition continuous with the ER have also been proposed (Kamhi-Nesher et al., 2001), although studies on chaperone diffusion suggest free movement of quality control constituents (Lippincott-Schwartz et al., 2000). The function of the
sheets and tubules in yeast cells are also currently unclear, as deletion of reticulon family members (causing loss of the tubules) has no appreciable effect on ER function and cell viability (Voeltz et al., 2006). Hence, while the morphological and functional heterogeneity of ER-related functions are clear, the relationship between morphology, organization, and function in the ER remains elusive in many cases.

A proteomics mapping of the ER, which reports the protein composition and abundance of organellar fractions, has been pursued in our lab (Gilchrist et al., 2006). Here, rat liver was separated into various fractions by centrifugation and the protein composition was quantitatively mapped by mass spectrometry. Fractions enriched for rough ER, smooth ER and Golgi apparatus were collected and mapped. In addition, the rough and smooth ER were subfractionated into biochemical fractions. First, the ER microsomes were washed with 2M KCl (salt wash) to remove peripheral (cytoplasmic) proteins associated with the ER membrane. Next, the washed membranes were extracted with the detergent Triton X-114, yielding a detergent insoluble fraction (cytoskeleton and large complexes), a detergent soluble fraction (ER integral membrane proteins), and an aqueous fraction (ER luminal and tightly associated peripheral ER proteins). Hence, the structural features of the ER with regards to rough and smooth ER and with respect to the membrane were separated and mapped. It has been established that the proteomics resource describes a quantitative measure of protein abundance (Gilchrist et al., 2006). Localization of a specific protein is determined by finding the fraction(s) with the highest relative abundance for that protein (Gilchrist et al., 2006).

In this chapter, the proteomics resource is interrogated for resident proteins involved in the processing of cargo. Known proteins involved in translation, translocation, ER folding, cytoplasmic folding, and ER-associated degradation are localized in the proteomics resource. Remarkably, proteins with similar function had similar localization in the proteomics resource. Translation-related proteins
were found in the cytoplasm of rough ER. Translocation was found in the membrane of rough ER. ER folding was found equally in rough and smooth ER. Surprisingly, cytoplasmic folding and ERAD constituents were found almost exclusively in smooth ER. Hence, the ER is organized with a spatial structure corresponding to specialized functions.

Results

The proteomics resource of Gilchrist et al. 2006 was interrogated for resident proteins involved in one of 5 functional categories in cargo protein secretion: Translation-related proteins, translocation-related proteins, ER folding, cytoplasmic folding, and ubiquitin-proteasome degradation. Of the proteins found in the proteomics resource, 106 have previous evidence documenting a function in one of these categories (Table 1.1).

Next, the localization profile of each individual protein was obtained as a function of its abundance. The measure of a protein’s abundance in a particular fraction is found by counting the number of mass spectra (number of peptides) corresponding to that protein. This is because the number of times a protein is “seen” by the mass spectrometer increases with more abundant proteins (Gilchrist et al., 2006). In the proteomics resource, the abundance of a particular protein can be compared across different organelle fractions by normalizing the abundance as a percentage of the total protein found in that organelle. For instance, if a given protein comprises 1% of all peptides found in the rough ER, 0.5% of all peptides found in the smooth ER, and is not identified in the Golgi, the protein is thought to be twice as abundant in rough ER as compared to smooth ER, and below detection in the Golgi.
Fig. 1.1. Translation-associated proteins are found at the cytosolic face of the rough ER. 
a) Translation-associated proteins in the total fractions. Proteins known to function in protein translation and RNA metabolism found in the organellar proteomics resource (Gilchrist et al. 2006) are plotted as a heat map to indicate their abundance in the R (rough ER), S (smooth ER) and G (Golgi) total fractions. 
b) Cumulative distribution of the translation associated proteins. The abundances shown in panel A were summated and expressed as a proportion out of 1.0. Here, over 80% of the identifications of the proteins shown were found in the R (rough ER) fraction. 
d) Cumulative distribution of the proteins indicated in the R (rough) and S (smooth) biochemical fractions, expressed as a proportion out of 1.0. Error bar: +/- SD for n = 3 fractions.
Proteins known to function at the level of RNA metabolism or protein synthesis are plotted in Figure 1.1 as a heat map. Here, the intensity of colour is related to the abundance of each protein found in each fraction. The topology of the ER proteome is such that the vast majority of proteins are of low abundance, while a small number of proteins are highly abundant (Lesimple et al. in preparation). Accordingly, the highly abundant proteins are indicated in green heat maps, with lesser abundance proteins indicated in red heat maps in this chapter. When the translation-related proteins were plotted a striking pattern emerged whereby these proteins were largely restricted to the rough ER and had relatively low abundance in the smooth ER and Golgi (Figure 1.1a). This indicates a primary localization to rough ER. In order to find the sum localization for this functional category, these proteins were summated and then expressed as a proportion out of 1 for each of the fractions. Here, the majority of the protein abundance for translation-related proteins was found to be in the rough ER (Figure 1.1b). Since rough ER is defined by the presence of ribosomes, that these synthesis-related proteins are found predominantly associated with rough ER is as expected; that proteins involved in RNA metabolism are also found at the rough but not smooth ER is a novel finding.

Independently, the biochemical fractions showed the same pattern of rough ER localization for translation-associated proteins (Figure 1.1c and d). In addition, these proteins were largely removed by a high salt wash and ended up in the salt wash fraction (Figure 1.1c and d). This is as expected for translation-associated proteins, which are known to be associated with the ER at its cytoplasmic face and can be expected to be washed off the membrane using a high salt treatment.

Proteins involved in entry into the ER and the closely associated process of N-linked glycosylation [which occurs immediately upon exit from the Sec61 channel (Yan and Lennarz, 2005)] displayed a different distribution. As for translation-associated proteins (and as expected given the co-translational nature of ER
Fig. 1.2: Translocon-associated proteins are found in the membrane of rough ER.

a) Translocon-associated proteins in the total fractions. Proteins known to function in ER entry (translocation) and N-linked glycosylation were found in the organellar proteomics resource (Gilchrist et al. 2006) and plotted as a heat map to indicate their abundance in the R (rough ER), S (smooth ER) and G (Golgi) total fractions.

b) Cumulative distribution of the translocon-associated proteins. The abundances shown in panel A were summated and expressed as a proportion out of 1.0. Here, translocon-associated proteins are found predominantly in the membrane (detergent phase) of rough ER. Error bars +/- SD for n = 3 fractions.


d) Cumulative distribution of the proteins indicated in the R (rough) and S (smooth) biochemical fractions, expressed as a proportion out of 1.0.
entry), these translocation-associated proteins were found predominantly in rough ER (Figure 1.2a-d). However, the translocon-associated proteins partitioned into the detergent phase of Triton X-114, as befitting the integral membrane status of the Sec61 translocon (Figure 1.2c and d). The SRP receptor, the majority of the subunits of the translocon (Sec61, TRAP and TRAM) known to be in a complex, signal sequence cleavage (signal peptidase and signal peptide peptidase) and N-linked glycosylation (ribophorins, oligosaccharyltransferase) were all identified and as an aggregate, displayed characteristics of integral membrane proteins of the rough ER (Figure 1.2c and d).

Highly abundant are chaperones and protein folding enzymes in the ER lumen. These proteins partitioned equally into rough and smooth ER, with little abundance in the Golgi (Figure 1.3). An exception is ERp44, a protein previously thought to be in the ER (Higo et al., 2005), shown by proteomics to be a Golgi protein (Gilchrist et al., 2006), and now thought to be an important stimulator of IgM polymerization in the ERGIC and cis-Golgi (Anelli et al., 2007). The chaperones and protein folding enzymes of the ER largely partitioned into the aqueous phase of Triton X-114, as is expected of luminal proteins (Figure 1.3c and d). An exception is the transmembrane chaperone calnexin, which partitioned into the Triton X-114 detergent phase of both rough and smooth ER, an apparently unique localization among ER folding constituents.

Cytosolic chaperones, which help to fold transmembrane cargo containing domains exposed to the cytoplasm, were also detected, including the major cytosolic chaperones of the Hsp90 system and the TriC chaperonin complex (CCT) (Figure 1.4). Surprisingly, these chaperones almost exclusively partitioned into smooth ER. In the biochemical fractions cytosolic chaperones were found in the salt wash as well as aqueous and insoluble phases, suggesting a rather tight association with the ER membrane compared to the more easily removable translation constituents (Figure 1.4c and d). This localization to smooth ER
Fig. 1.3: Luminal ER folding occurs in the rough and smooth ER.

a) Resident proteins of the ER folding machinery were found in the total fractions of the organellar proteomics resource (Gilchrist et al. 2006) and plotted as a heat map to indicate their abundance in the R (rough ER), S (smooth ER) and G (Golgi) total fractions. b) Cumulative distribution of the translocon-associated proteins. The abundances shown in panel A were summated and expressed as a proportion out of 1.0, showing a roughly equal distribution in rough and smooth ER. c) Resident ER folding proteins identified in the biochemical fractions of rough (RM) and smooth (SM) microsomes. S, salt wash. I, detergent insoluble. D, detergent phase of Triton X-114. A, aqueous phase of Triton X-114. d) Cumulative distribution of the proteins indicated in the R (rough) and S (smooth) biochemical fractions, expressed as a proportion out of 1.0. Error bars +/- SD for n = 3 fractions.
Fig. 1.4: Cytosolic chaperones associate with the smooth ER.

a) Cytosolic chaperones of the Hsp90 class and TriC were found in the total fractions of the organellar proteomics resource (Gilchrist et al. 2006) and plotted as a heat map to indicate their abundance in the R (rough ER), S (smooth ER) and G (Golgi) total fractions. b) Cumulative distribution of the cytosolic chaperones. The abundances shown in panel A were summated and expressed as a proportion out of 1.0, showing preferential association with the smooth ER. c) Cytosolic chaperones identified in the biochemical fractions of rough (RM) and smooth (SM) microsomes. S, salt wash. I, detergent insoluble. D, detergent phase of Triton X-114. A, aqueous phase of Triton X-114. d) Cumulative distribution of the cytosolic chaperone constituents in the R (rough) and S (smooth) biochemical fractions, expressed as a proportion out of 1.0. These chaperones are partially removed by salt wash, indicating their peripheral association with the membrane. Cytosolic chaperones are also found in the insoluble and detergent phases, but predominantly in the smooth ER fractions. Error bars +/- SD for n = 3 fractions.
Fig. 1.5: ER-associated degradation (ERAD) occurs in the smooth ER. ERAD constituents including p97/VCP, derlin-1, ubiquitin ligases, and the proteasome are predominantly found in the smooth ER fractions of the organellar proteomics resource. a) Heat map of known ERAD constituents identified in the total fractions. b) Cumulative distribution of ERAD constituents in the total fractions. c) ERAD constituents found in the biochemical fractions. The proteasome and additional ubiquitin ligases are uncovered in the biochemical fractions because enrichment by subfractionation extends the identifiable proteasome for mass spectrometry. d) Cumulative distribution of ERAD constituents, expressed as a proportion out of 1.0. RM, rough ER; SM, smooth ER; S, salt wash fraction; I, insoluble fraction; D, detergent (Triton X-114) fraction; A, aqueous fraction of Triton X-114. Error bars +/- SD for n = 3 fractions.
Fig. 1.6: Distribution of ER secretory functions in the proteomics resource

The cumulative distribution of the known proteins found in translation, translocation, ER folding, ERAD, and cytosolic folding were calculated (see Fig. 1.1) and plotted as line graphs here. These functional categories represent the resident proteins involved in cargo processing; cargo undergo each of these steps in an ordered manner during maturation. In the total fractions (right), these processes are seen to be ordered as occurring from rough ER to smooth ER. In the biochemical fractions (left), a similar directionality is seen, with resident functional complexes also being segregated with respect to the membrane. Cargo can be imagined to travel as a wave through maturation, originating in the cytoplasm of the rough ER and proceeding through the membrane towards the smooth ER for degradation to occur.
suggests a late role in protein folding, or perhaps a greater role for protein unfolding for cytoplasmic degradation (see discussion).

The proteins of degradation, including the putative retrotranslocation channel derlin, the AAAtpase p97, the retrotranslocation protein Sel1, a variety of ubiquitin ligases (E1, E2, and E3), and the subunits of the proteasome were also identified. These proteins were found nearly exclusively in smooth ER (Figure 1.5). While derlin mainly partitioned into the detergent phase, the proteasome was found to be tightly associated with the ER membrane by virtue of its resistance to salt wash and partitioning into the aqueous phase (Figure 1.5c and d).

When the distribution of each functional category is compared (Figure 1.6), distinct localizations are observed. Protein synthesis occurs distinctly in the cytoplasm of rough ER, entry into the ER in the membrane of rough ER, folding in the rough and smooth ER lumen, with degradation occurring in the membrane and cytoplasm of the smooth ER.

The functional groups described are not of the same abundance. If the total abundance of all the proteins involved in cargo processing is measured, 1/10 is for protein synthesis, 2/10 in protein translocation and glycosylation, 6/10 in protein folding, and the remainder involved in cytoplasmic folding and degradation (~3% each) (Figure 1.7a). Moreover, the sum abundance of cargo in the ER greatly exceeds that of any other category and is on par with the total abundance of the cargo processing machinery (Figure 1.7a). Similarly, the distribution of the cargo processing machinery in total (all 106 cargo handling proteins considered together) shows a roughly equal distribution between rough and smooth ER, with little in the Golgi (Figure 1.7b). In contrast, cargo is enriched in the Golgi, but while it is in the ER it makes little preference for rough vs. smooth ER (Figure 1.7c). Hence, cargo in its various forms of maturation is spread evenly throughout the ER, while the low abundance degradation system makes up only a small portion of cargo processing and is sequestered to smooth ER. Further, the
Fig. 1.7: Abundance and distribution of resident ER proteins involved in cargo processing compared to cargo.

a) The abundance in rough (RER) and smooth (SER) of the known proteins involved in protein synthesis (translation), entry into the ER (translocation), ER folding, cytoplasmic folding and degradation and cargo. “All cargo processing” represents the sum total of the other categories (excluding cargo). Each bar indicates the % of the total peptides identified corresponding to the function indicated, giving an idea of the abundance of a particular functional category. b) Normalized distribution of resident cargo processing proteins across the total fractions. c) Normalized distribution of cargo across the rough and smooth ER and Golgi fractions in the proteomics resource.
abundance of cargo is greater than that of the resident cargo processing proteins of the ER.

Discussion

Balch and Kelly have described a framework to understand the tissue specificity of protein folding (Sekijima et al., 2005; Wiseman et al., 2007a; Wiseman et al., 2007b). In their FoldEx model, cargo fate in terms of secretion or degradation is dependent upon the inherent stability of the cargo protein. This stability is modifiable by the complement of ER-associated folding (ERAF) factors that exist in any given cell. For cargo of a given stability, ER-associated degradation (ERAD), ERAF and secretory exit all compete to choose cargo fate. The more stable a cargo protein is, the less it is recognized by ERAD and it becomes secreted. The more active the ERAD machinery, the greater is susceptibility to degradation and the more stable a protein must be in order to be secreted. Conversely, ERAF is important for large and slowly folding cargo where folding rates are rate limiting compared to export. The specific composition of ERAF, ERAD, and the secretory machinery in a given cell type determines the capacity for secretion. Here, the quality control processes of ERAF and ERAD are found to have separate spatial localizations within the ER. Hence, the necessary steps in quality control determined by biophysical studies are found here to have a spatial correlate. Furthermore, it is shown here that there is a high concentration of folding enzymes relative to a low concentration of ERAD proteins (Fig. 1.7a). This gives support to a high cellular commitment to preserving protein secretion. Packing the ER with folding enzymes aimed at increasing cargo stability is apparently preferred to allowing the ERAD system to recycle a larger proportion of slow-folding cargo.

The Ron laboratory has described the flux of chaperone capacity against cargo protein load in the ER (Marciniak and Ron, 2006). Here, newly synthesized proteins and in particular misfolded proteins cause a metabolic and oxidative load on the ER. This is counteracted by the unfolded protein response, which halts
protein synthesis early on, and then upregulates ERAF and ERAD constituents to decrease ER load. Thus, the UPR modifies the fate of cargo by affecting ERAF and ERAD composition.

In this chapter, the abundance and spatial segregation of the synthesis, ERAF, and ERAD machineries have been described. Notably, synthesis is restricted to rough ER and represents the influx of protein load in the ER. ERAF is highly abundant and is not restricted, with diffusion between rough and smooth ER. ERAD constituents are low abundance and are spatially restricted to smooth ER (Figure 1.8). This suggests several novel aspects of ER function that are compatible and give insight into the work proposed by Ron, Balch and Kelly. In the model supported by the data shown here, cargo are restricted to rough ER upon immediate synthesis, but then are found in both rough and smooth ER in an unrestricted manner over time. The resident machinery of the ER however, is spatially restricted. Newly synthesized cargo enter through the rough ER, are folded at the rough and smooth ER, but then undergo degradation in the smooth ER, suggesting a directionality of cargo trafficking away from the ribosome (Figure 1.8).

Since the degradation machinery is of low abundance and is sequestered in the smooth ER, this predicts that degradation is rate limiting. If highly unstable proteins are degraded more rapidly than moderately unstable proteins, then in a degradation-limiting environment (shown to occur in this chapter), the degradation versus secretion of moderately unstable proteins is partially dependent upon the capacity of ERAD. Indeed, Morimoto has recently reported that the introduction of highly unstable proteins prevents the degradation of moderately unstable proteins which are otherwise degraded in disease models of neurodegeneration (Gidalevitz et al., 2006). Further, Balch and Kelly indicate that moderately unstable proteins are not secreted in cell types that have high capacities of ERAD. Upregulation of ERAD constituents by the UPR, or variations in ERAD levels therefore could account for the tissue specificity
Figure 1.8. Spatiotemporal organization of the endoplasmic reticulum revealed by subcellular proteomics. Above, extension of the ER proteome by subfractionation into biochemical fractions reveals the membrane topology of the ER functional processes indicated. Below, comparison of the proteome of the rough ER, smooth ER, and Golgi apparatus reveals the directionality of cargo processing in the ER, proceeding from the rough ER-associated ribosome towards the smooth ER-associated proteasome. The figure at right is adapted from Kaufman et al. *Nat. Rev. Mol. Cell Biol.* (2002).
proposed by Balch and Kelly. These findings are in agreement with an ER with ERAD constituents in limiting quantities.

The spatial sequestration of degradation shown here has theoretical advantages for handling protein flux in the ER. For instance, high local concentrations of degradation factors would allow these overall low abundance proteins to form the necessary complexes for ERAD. This concentration of ERAD factors would also allow the dwarfing of ERAF abundance locally so that once targeted to the local degradation space in the smooth ER, unstable cargo could be passed along the various steps (glycan modification, retrotranslocation, unfolding, deglycosylation, ubiquitination) in correct sequence for organized degradation. This is to say that degradation is normally a stochastic process once targeting occurs. This is supported by the lack of degradation intermediates during steady-state conditions. For instance, ubiquitinated intermediates that do not complete degradation have been observed to accumulate with proteasome inhibition and in pathological states (Bence et al., 2001; Bennett et al., 2007). Glycosylated intermediates can be accumulated by PNGase F inhibition (Hirsch et al., 2003). Intermediates failing to retrotranslocate can also be accumulated using inhibition of derlin (Lilley and Ploegh, 2004). Under normal conditions, degradation occurs to completion despite its low abundance and these species cannot be observed. In pathological states intermediates are generated, further attesting to the limiting concentration of ERAD constituents.

This is in contrast to conformational forms interacting with ERAF, which have variability in disulfide bonding, cis/trans conformational states, and repetitive cycles of folding under normal circumstances; folding is not a stochastic process proceeding directly from instability to stability. Instead, ERAF permeates the entire ER and is composed of many highly abundant constituents with a diversity of determinants enacting folding. Thus, the level of any one chaperone may or may not be of particular importance for the folding of most cargo. In fact, cargo
(and their particular misfolding lesions) likely attract many chaperones in the tight folding environment of the ER. This is substantiated by chaperone knockout studies, where the deletion of chaperones such as the abundant lectin calnexin causes increased secretion of some proteins but decreased secretion of others (Pieren et al., 2005). The level of stability attained by any particular cargo is predicted by Balch and Kelly to be dependent upon the difficulty of folding predicted by the primary sequence and the particular milieu of ERAF factors that can optimize their folding (Wiseman et al., 2007a). That ERAF is found here to be highly abundant with a vast number of folding constituents available for any given cargo protein generally supports this idea. The abundance suggests also a mass effect whereby ERAF is the default fate for a secretory protein, where ERAD and secretory exit must compete with ERAF to clear secretory proteins from the ER.

Work on the UPR also is also compatible with a view of segregated ERAF and ERAD. The major activator of the UPR is the load of proteins in the ER (Ron and Walter, 2007), a load that permeates both the rough and smooth ER where ERAF is present. The halt on synthesis enacted by the UPR allows the lesser abundance secretion and degradation pathways to clear the ER. Upregulation of ERAF and ERAD by the late arms of the UPR are attempts to preserve the cell by clearing the ER on one hand while still allowing the secretion of cargo necessary for cell survival, through attempts to optimize folding of moderately unstable proteins for engaging the exit machinery.

The model of spatially restricted synthesis, translocation, and degradation machinery with spatially unrestricted cargo and ERAF in the ER is reminiscent of functioning in the Golgi apparatus. In the Golgi apparatus secretory cargo is further processed, notably through carbohydrate modifications, via a stepwise series of interactions with Golgi modification enzymes. To enact this function the Golgi is organized into a series of stacks, with early modification enzymes located at the cis face and later modification enzymes located at the trans face (Glick,
2000). The ER might function in a similar fashion, just not organized into the characteristic stacked shape of the Golgi apparatus. The function of the stacked shape of the Golgi is presently controversial; it may enable a cisternal maturation model whereby cargo are transported through the en bloc movement of entire cisternae through the stack from cis to trans (Pelham, 2006). A similar process of maturation could occur in the ER whereby rough ER proteins are “retrieved” to ribosome organizing sites in a retrograde fashion, while cargo carry on directionally from rough to smooth ER. That membranous continuities between Golgi stacks have been observed (Trucco et al., 2004), similar in magnitude to the continuities between yeast cortical and peripheral ER, suggests some similarities in membrane organization between the two organelles.

Studies on the diffusion of resident ER proteins supports the view of the ER as having spatial restrictions in the manner described here. For instance, translocons, as marked by the GFP-tagged protein Dad1 have been shown to have restricted diffusion in the membrane compared to that predicted if diffusion were completely free for these substrates (Nikonov et al., 2002). Conversely, chaperones such as calreticulin-GFP as well as cargo-GFP have completely free diffusion (Nikonov et al., 2002; Snapp et al., 2006). Even under conditions of high cargo misfolding load, which might be expected to create retentive cargo:chaperone complexes, mobility of calreticulin-GFP remained far above that of the translocon (Snapp et al., 2006). This is in agreement with the model proposed here which finds restriction of translocation to specific sites in rough ER, but a much broader and unrestricted localization of ERAF constituents.

The fractions used to generate the proteomics resource are from rat liver, which feature highly abundant smooth ER membranes as part of the detoxification function of liver ER. The spatial segregation in liver ER may be exaggerated by the distinct ER types found in this organ, amplifying the segregation phenomenon. That said, ER subdomains of varying function, including a “quality control compartment” have been proposed to exist (Borgese et al., 2006; Kamhi-
Nesher et al., 2001). Specifically, the work by Rapoport and colleagues, who have recently found that the reticulon family are segregated to smooth ER and help to create the specified shape of tubular ER in yeast cells, suggests that spatial subdomain restriction is at least important for morphological determination of organellar structure (Voeltz et al., 2006).

In summary, this chapter finds that resident proteins of the ER exhibit a spatial segregation with respect to function and that the degradation system is of low abundance. The structural and morphological features of the ER A model of the ER is suggested whereby secretory cargo traverse spatially restricted functional complexes to determine their fate. The ERAF system is abundant, spread across the ER, and is diverse while the ERAD system is of low abundance, is locally specified, and may operate in a stochastic manner. This is in agreement with studies of secretion of variably stable cargo and with the known functioning of the UPR to increase ER load. This discovery of the ER as spatially organized was made possible only by the comprehensiveness of organellar proteomics, which allow a view beyond that displayed by the few marker proteins amenable to specific antibody labelling.

Materials and Methods

The organellar proteomics resource contains supplementary tables detailing the abundance of individual proteins across biochemical and total fractions (Gilchrist et al., 2006). After manual annotation and curation, proteins known to function in translation, translocation, ER folding, ERAD, or cytosolic folding were selected. These were plotted as heatmaps using Treeview and line graphs using Excel. Standard deviations for categories (i.e. Fig. 1.1b) were derived by taking the sum of peptides for all the proteins of that category and comparing this value across the n = 3 biological replicates for each fraction analyzed by Gilchrist et al. 2006.
<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Name</th>
<th>Link to Human Gene</th>
<th>Comment</th>
<th>Reference Link</th>
<th>RM</th>
<th>SM</th>
<th>Chignyl</th>
</tr>
</thead>
<tbody>
<tr>
<td>EEF1A</td>
<td>EF-1 alpha</td>
<td></td>
<td></td>
<td></td>
<td>0.17</td>
<td>0.06</td>
<td>0.00</td>
</tr>
<tr>
<td>EEF1G</td>
<td>EF-1 gamma</td>
<td></td>
<td></td>
<td></td>
<td>0.02</td>
<td>0.03</td>
<td>0.00</td>
</tr>
<tr>
<td>EEF1F</td>
<td>EF-1 gamma</td>
<td></td>
<td></td>
<td></td>
<td>0.02</td>
<td>0.03</td>
<td>0.00</td>
</tr>
<tr>
<td>EEF2</td>
<td>EF-2 alpha</td>
<td></td>
<td></td>
<td></td>
<td>0.04</td>
<td>0.05</td>
<td>0.00</td>
</tr>
<tr>
<td>EEF1</td>
<td>EF-2 alpha</td>
<td></td>
<td></td>
<td></td>
<td>0.04</td>
<td>0.05</td>
<td>0.00</td>
</tr>
<tr>
<td>EEF1</td>
<td>EF-2 alpha</td>
<td></td>
<td></td>
<td></td>
<td>0.04</td>
<td>0.05</td>
<td>0.00</td>
</tr>
<tr>
<td>EEF1</td>
<td>EF-2 alpha</td>
<td></td>
<td></td>
<td></td>
<td>0.04</td>
<td>0.05</td>
<td>0.00</td>
</tr>
<tr>
<td>EEF1</td>
<td>EF-2 alpha</td>
<td></td>
<td></td>
<td></td>
<td>0.04</td>
<td>0.05</td>
<td>0.00</td>
</tr>
<tr>
<td>EEF1</td>
<td>EF-2 alpha</td>
<td></td>
<td></td>
<td></td>
<td>0.04</td>
<td>0.05</td>
<td>0.00</td>
</tr>
<tr>
<td>EEF1</td>
<td>EF-2 alpha</td>
<td></td>
<td></td>
<td></td>
<td>0.04</td>
<td>0.05</td>
<td>0.00</td>
</tr>
</tbody>
</table>

**Translocation-related proteins**

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Name</th>
<th>Link to Human Gene</th>
<th>Comment</th>
<th>Reference Link</th>
<th>RM</th>
<th>SM</th>
<th>Chignyl</th>
</tr>
</thead>
<tbody>
<tr>
<td>339911</td>
<td>EEF1</td>
<td></td>
<td></td>
<td></td>
<td>0.33</td>
<td>0.18</td>
<td>0.04</td>
</tr>
<tr>
<td>166583</td>
<td>EEF2</td>
<td></td>
<td></td>
<td></td>
<td>0.45</td>
<td>0.26</td>
<td>0.07</td>
</tr>
<tr>
<td>169356</td>
<td>EEF2</td>
<td></td>
<td></td>
<td></td>
<td>0.45</td>
<td>0.26</td>
<td>0.07</td>
</tr>
<tr>
<td>114363</td>
<td>EEF2</td>
<td></td>
<td></td>
<td></td>
<td>0.45</td>
<td>0.26</td>
<td>0.07</td>
</tr>
<tr>
<td>114363</td>
<td>EEF2</td>
<td></td>
<td></td>
<td></td>
<td>0.45</td>
<td>0.26</td>
<td>0.07</td>
</tr>
<tr>
<td>114363</td>
<td>EEF2</td>
<td></td>
<td></td>
<td></td>
<td>0.45</td>
<td>0.26</td>
<td>0.07</td>
</tr>
<tr>
<td>114363</td>
<td>EEF2</td>
<td></td>
<td></td>
<td></td>
<td>0.45</td>
<td>0.26</td>
<td>0.07</td>
</tr>
<tr>
<td>114363</td>
<td>EEF2</td>
<td></td>
<td></td>
<td></td>
<td>0.45</td>
<td>0.26</td>
<td>0.07</td>
</tr>
<tr>
<td>114363</td>
<td>EEF2</td>
<td></td>
<td></td>
<td></td>
<td>0.45</td>
<td>0.26</td>
<td>0.07</td>
</tr>
</tbody>
</table>

**Translation-Related Proteins**

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Name</th>
<th>Link to Human Gene</th>
<th>Comment</th>
<th>Reference Link</th>
<th>RM</th>
<th>SM</th>
<th>Chignyl</th>
</tr>
</thead>
<tbody>
<tr>
<td>141092</td>
<td>EIF-1A</td>
<td></td>
<td></td>
<td></td>
<td>0.21</td>
<td>0.11</td>
<td>0.06</td>
</tr>
<tr>
<td>21967</td>
<td>EIF-2</td>
<td></td>
<td></td>
<td></td>
<td>0.35</td>
<td>0.19</td>
<td>0.10</td>
</tr>
<tr>
<td>51432</td>
<td>EIF-3</td>
<td></td>
<td></td>
<td></td>
<td>0.18</td>
<td>0.09</td>
<td>0.05</td>
</tr>
<tr>
<td>18626</td>
<td>EIF-4</td>
<td></td>
<td></td>
<td></td>
<td>0.14</td>
<td>0.07</td>
<td>0.04</td>
</tr>
<tr>
<td>18626</td>
<td>EIF-4</td>
<td></td>
<td></td>
<td></td>
<td>0.14</td>
<td>0.07</td>
<td>0.04</td>
</tr>
<tr>
<td>18626</td>
<td>EIF-4</td>
<td></td>
<td></td>
<td></td>
<td>0.14</td>
<td>0.07</td>
<td>0.04</td>
</tr>
<tr>
<td>18626</td>
<td>EIF-4</td>
<td></td>
<td></td>
<td></td>
<td>0.14</td>
<td>0.07</td>
<td>0.04</td>
</tr>
<tr>
<td>18626</td>
<td>EIF-4</td>
<td></td>
<td></td>
<td></td>
<td>0.14</td>
<td>0.07</td>
<td>0.04</td>
</tr>
<tr>
<td>18626</td>
<td>EIF-4</td>
<td></td>
<td></td>
<td></td>
<td>0.14</td>
<td>0.07</td>
<td>0.04</td>
</tr>
<tr>
<td>18626</td>
<td>EIF-4</td>
<td></td>
<td></td>
<td></td>
<td>0.14</td>
<td>0.07</td>
<td>0.04</td>
</tr>
<tr>
<td>18626</td>
<td>EIF-4</td>
<td></td>
<td></td>
<td></td>
<td>0.14</td>
<td>0.07</td>
<td>0.04</td>
</tr>
</tbody>
</table>

**Translation-Related Proteins**

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Name</th>
<th>Link to Human Gene</th>
<th>Comment</th>
<th>Reference Link</th>
<th>RM</th>
<th>SM</th>
<th>Chignyl</th>
</tr>
</thead>
<tbody>
<tr>
<td>21967</td>
<td>EIF-2</td>
<td></td>
<td></td>
<td></td>
<td>0.35</td>
<td>0.19</td>
<td>0.10</td>
</tr>
<tr>
<td>51432</td>
<td>EIF-3</td>
<td></td>
<td></td>
<td></td>
<td>0.18</td>
<td>0.09</td>
<td>0.05</td>
</tr>
<tr>
<td>18626</td>
<td>EIF-4</td>
<td></td>
<td></td>
<td></td>
<td>0.14</td>
<td>0.07</td>
<td>0.04</td>
</tr>
<tr>
<td>18626</td>
<td>EIF-4</td>
<td></td>
<td></td>
<td></td>
<td>0.14</td>
<td>0.07</td>
<td>0.04</td>
</tr>
<tr>
<td>18626</td>
<td>EIF-4</td>
<td></td>
<td></td>
<td></td>
<td>0.14</td>
<td>0.07</td>
<td>0.04</td>
</tr>
<tr>
<td>18626</td>
<td>EIF-4</td>
<td></td>
<td></td>
<td></td>
<td>0.14</td>
<td>0.07</td>
<td>0.04</td>
</tr>
<tr>
<td>18626</td>
<td>EIF-4</td>
<td></td>
<td></td>
<td></td>
<td>0.14</td>
<td>0.07</td>
<td>0.04</td>
</tr>
<tr>
<td>18626</td>
<td>EIF-4</td>
<td></td>
<td></td>
<td></td>
<td>0.14</td>
<td>0.07</td>
<td>0.04</td>
</tr>
<tr>
<td>18626</td>
<td>EIF-4</td>
<td></td>
<td></td>
<td></td>
<td>0.14</td>
<td>0.07</td>
<td>0.04</td>
</tr>
</tbody>
</table>

**Translation-Related Proteins**
Table 1.1: Proteomics identification of proteins related to translation, translocation, folding, ERAD, and ER-associated cytosolic protein folding.

The proteins characterized are from Gilchrist et al. 2006. Prominent missing proteins in each of the categories (e.g. TRAP-beta for the translocon) were identified when our mass spectra were matched to the mammalian database but not when matched against the rat database; the rat protein database contains incomplete and incorrectly annotated genome data as explained in Gilchrist et al. 2006. RM, rough microsomes (correspond to rough ER). SM, smooth microsomes (corresponds to smooth ER). Golgi, Golgi apparatus. SLT, salt wash. INS, insoluble in Triton X-114. DET, detergent phase of Triton X-114. AQU, aqueous phase of Triton X-114. Abundance is given as the percentage of the total peptides in the listed fraction (i.e. RM DET) that corresponds to the protein name shown. Protein name is shown as its usual name, as well as the name linked in NCBI Gene (blue). The comment is based on a review of the literature for the protein described, with reference links occasionally shown.

<table>
<thead>
<tr>
<th>Cytosolic Protein Folding</th>
<th>RM</th>
<th>SM</th>
<th>Golgi</th>
<th>RM SLT</th>
<th>RM INS</th>
<th>RM DET</th>
<th>RM AQU</th>
<th>SM SLT</th>
<th>SM INS</th>
<th>SM DET</th>
<th>SM AQU</th>
<th>AQU SLT</th>
<th>AQU INS</th>
<th>AQU DET</th>
<th>AQU AQU</th>
</tr>
</thead>
<tbody>
<tr>
<td>139645 HIPPO</td>
<td>HIPPO</td>
<td>0.011</td>
<td>0.002</td>
<td>0.005</td>
<td>0.003</td>
<td>0.012</td>
<td>0.000</td>
<td>0.007</td>
<td>0.000</td>
<td>0.003</td>
<td>0.003</td>
<td>0.000</td>
<td>0.006</td>
<td>0.000</td>
<td>0.031</td>
</tr>
<tr>
<td>329568 HIPRO</td>
<td>HIPRO</td>
<td>0.003</td>
<td>0.010</td>
<td>0.000</td>
<td>0.003</td>
<td>0.029</td>
<td>0.000</td>
<td>0.046</td>
<td>0.000</td>
<td>0.146</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>160049 HIPYOS</td>
<td>HIPYOS</td>
<td>0.000</td>
<td>0.000</td>
<td>0.003</td>
<td>0.000</td>
<td>0.002</td>
<td>0.000</td>
<td>0.003</td>
<td>0.000</td>
<td>0.007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>72115 OCT ALPH</td>
<td>OCT</td>
<td>0.002</td>
<td>0.007</td>
<td>0.005</td>
<td>0.002</td>
<td>0.000</td>
<td>0.002</td>
<td>0.017</td>
<td>0.000</td>
<td>0.012</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>202700 OCT BETA</td>
<td>OCT2</td>
<td>0.000</td>
<td>0.016</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90653 OCT GAMMA</td>
<td>OCT3</td>
<td>0.000</td>
<td>0.029</td>
<td>0.002</td>
<td>0.000</td>
<td>0.014</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>198155 OCT DELTA</td>
<td>OCT4</td>
<td>0.000</td>
<td>0.015</td>
<td>0.004</td>
<td>0.000</td>
<td>0.016</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>206941 OCT ZETA</td>
<td>OCT6</td>
<td>0.000</td>
<td>0.039</td>
<td>0.002</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>101403 OCT ETA</td>
<td>OCT7</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.010</td>
<td>0.001</td>
<td>0.000</td>
<td>0.013</td>
<td>0.000</td>
<td>0.018</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>52335 OCT DELTA</td>
<td>OCT8</td>
<td>0.000</td>
<td>0.000</td>
<td>0.005</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 2: Identification of candidate unknown proteins functioning in cargo processing

Abstract

The proteomics resource of Gilchrist et al. 2006 describes over 300 proteins with poorly characterized functions. In the previous chapter it was found that the selection of known proteins involved in cargo processing showed specific spatial localizations in the ER corresponding to functional categories. However, the ER has a diversity of additional functions (detoxification, lipid metabolism, etc.) not directly related to cargo processing. This chapter describes efforts to identify unknown proteins in the proteomics resource that are involved generally in cargo processing and in subsequent chapters their sublocalization will be explored. The majority of the unknown proteins did not contain domain or sequence features indicative of function. However, informatics efforts identified a nudix domain protein of interest (nudilin) that shows sequence similarity to RNA binding proteins, as well as a thioredoxin family member (TMX2) with similarity to protein folding enzymes. In a separate approach to find unknown proteins involved in quality control, cognate genes were screened for upregulation by the unfolded protein response. Seven genes were found to be upregulated by the UPR. Finally, of these candidates, 4 proteins were found to interact with the model secretory cargo tyrosinase (TMX2, nicalin, stexin, erlin-1). Thus, 5 proteins in total (4 interacting with tyrosinase plus nudilin) are candidates for functioning in cargo processing.

Introduction

Synthesis, folding, and targeting (processing) of cargo destined for intracellular organelles, the plasma membrane, or the extracellular space is a major function of the ER. The ER also functions in detoxification, lipid metabolism, calcium storage, and other important cellular and physiological processes.
The 176 proteins of unknown function found in the ER by proteomics may function in one of these ER processes, or may be involved in an undiscovered process. The particular interest of this thesis are those proteins involved in the processing of secretory cargo from message to transit into the Golgi. Identifying unknown proteins broadly involved in cargo processing is the interest of this chapter. This was attempted in two ways.

First, bioinformatics can give domain and feature information about unknown proteins based upon sequence similarity to known families. This was used for the 176 proteins of unknown function and of these, two proteins were selected as having a specific interest for cargo processing (nudilin, TMX2).

Second, the unfolded protein response (UPR) is a transcriptional program that upregulates genes involved in accommodating an increased secretory protein load. The UPR can be elicited by pharmacological means and the transcripts corresponding to proteins of unknown function can thereby be assessed for upregulation. From microarray studies, genes upregulated by the UPR are found to be enriched for functioning in cargo processing (Harding et al., 2003; Travers et al., 2000).

Finally, candidates for involvement in cargo processing at the protein level were tested for interaction with cargo using the model substrate tyrosinase. In this chapter, 5 proteins in total were found to be involved in cargo processing: 1 protein with sequence similarity to RNA binding proteins (nudilin), and 4 proteins interacting with tyrosinase (nicalin, stexin, TMX2, erlin-1).

**Results**

Bioinformatics efforts on the 176 proteins yielded domain information, as well as predicted transmembrane domains, and signal sequence features [(Gilchrist et al., 2006) and data not shown]. Of these proteins 2 were selected. The protein we have named nudilin has a nudix domain found in RNA decapping proteins;
nudilin has sequence homology to syndesmos (a plasma membrane protein) and the Xenopus laevis protein X29, which has been shown to bind RNA (Peculis et al., 2007) (Figure 2.1a). TMX2 had two forms in the database, which differ by an exon and are presumed splice variants. The 296 a.a. form of TMX2 is denoted here as TMX2L (long) and the 258 a.a. form of TMX2 is denoted here as TMX2s. Both forms contain a conserved thioredoxin domain with homology to the PDI family of protein folding enzymes (Figure 2.1b, c). The atypical SXXC motif found in the thioredoxin domain of TMX2 is discussed further in Chapter 5.

Proteins of unknown function with transmembrane domains were prioritized for further analysis. Here, 33 genes corresponding to proteins of unknown function were selected for analysis by RT-PCR. These genes represented 32 transmembrane proteins plus one protein (UBXD2) that contains a UBX domain predicted to bind the AAAtpase VCP/p97; this protein is predicted to associate with the ER at its cytosolic face. In the selection of these genes, the use of a microarray project analyzing stress-responsive human genes (including ER stress) was attempted (Murray et al., 2004). We were unable to accept this set as entirely reliable on account of a great deal of noise, however, it should be noted that these genes were not selected on a completely random basis; some were chosen on the basis of possible positive results on these previous high-throughput microarray studies.

First, primer sequences corresponding to the human orthologues of the 33 genes were designed. These orthologues were found using Homologene on NCBI as well as BLAST verification to ensure that the closest match to the rat sequence identified by proteomics was used. PCR primers were selected using the program Primer 3 (Rozen and Skaletsky, 2000) according to their best score. To ensure cross-species compatibility, high-scoring primers were aligned with the rat and mouse sequences, and primers with high nucleotide identity between the three species were preferred. The sequences used are found in Table 2.1.
Fig. 2.1: Sequence alignment identifies candidate cargo processing proteins of the ER

a) Nudilin shares homology with the proteins syndesmos and X29. Using Clustal W, X29 and nudilin are seen here to share the core region essential to the nudix domain. These residues are not found in syndesmos. Homologous amino acids are shaded in grey. b) TMX2 shares homology with protein disulfide isomerase. Two forms of TMX2 identified in the database, a short version one containing 259 aa (TMX2s) and a longer version containing 296 aa (TMX2L) are shown here.
To determine if these genes were expressed in HepG2 cells and to find their linear range of amplifications, cDNA sets from untreated HepG2 cells were generated and amplified with varying rounds of PCR. Most genes amplified linearly between 23 and 27 rounds of PCR, although some more abundant genes such as the controls GAPDH and actin required only 18 – 20 rounds to reach the linear range, while others required 27 – 30 rounds. Genes requiring greater than 30 rounds of amplification were not used, and the lower number of rounds in the linear range (i.e. 23 for those in the 23-27 rounds range) was subsequently used to rule out genomic amplification. Examples of the amplification testing used is shown in Figure 2.2.

HepG2 cells were then grown to 70% confluence and treated with pharmacological activators of ER stress for 12 hours. That the DTT, tunicamycin, and thapsigargin treatments effectively stressed the cells was validated by upregulation of the UPR markers CHOP and Herp (Figure 2.3a). These markers are routinely found to be upregulated 20-fold and greater on microarray studies (Harding et al., 2003), and were highly upregulated here. Next, genes corresponding to proteins of unknown function were assessed for transcript upregulation. A total of 7 genes were found to be upregulated by ER stress, as assessed by endpoint PCR (Figure 2.3b and c).

To determine quantitatively this gene upregulation, Q-PCR using tunicamycin was then used on the 33 genes. CHOP and Herp were found to be upregulated 20.7 and 26.5 fold, respectively, consistent with microarray studies. In addition, 5 genes of unknown function were found to be upregulated with statistical significance, at a level greater than 1.5-fold (Figure 2.4). Indeed, the level of upregulation ranged from 1.8 – 3.3-fold for the 5 genes.

Two genes, stexin and nicalin, were upregulated by endpoint RT-PCR but not Q-PCR. This is explained by the finding that stexin and nicalin were upregulated by thapsigargin and DTT on endpoint PCR, but not tunicamycin, which was the drug
used in the Q-PCR screen (Figure 2.3 and 2.4). Therefore, 7 genes were identified in total by the RT-PCR screening.

Stexin, nicalin, and erlin-1 will be discussed further in this thesis. Naming of the remaining ER stress upregulated proteins is proposed as follows: We named sterlin (stress upregulated ER protein) and sterubin (stress upregulated ER ubx domain protein). Two proteins were previously described as lyric and CVAK104, respectively (Duwel and Ungewickell, 2006; Sutherland et al., 2004). Bioinformatics interrogation of these proteins predicted a UBX domain for sterubin, a kinase domain for CVAK104, and no further motifs in the other proteins. Sterlin and lyric are predicted transmembrane proteins based upon TMHMM and their biochemical solubility in the detergent phase of Triton X-114 (Gilchrist et al., 2006).

Tyrosinase was tested for suitability as a model substrate of protein secretion. Consistent with previous reports (Halaban et al., 2002), tyrosinase overexpression induced a colour change visible upon cell lysis and 14000g centrifugation (data not shown). Next, a pulse-chase protocol demonstrated several forms of tyrosinase: First, glycosylated tyrosinase forms previously demonstrated to represent 7, 6, and 5 occupied glycosylation sites were observed after 15 min. pulse with radiolabelled methionine (no chase) (Ujvari et al., 2001) (Figure 2.5). Second, these forms of tyrosinase shifted upward after chase, representing glycosylation modification in the Golgi apparatus; these forms of tyrosinase are mature and have left the ER (Halaban et al., 1997). Third, introducing protein misfolding through incorporation of a bulky proline analogue, azetidine-2 carboxylic acid (Azc), prevents tyrosinase maturation from the ER form to the Golgi form. Fourth, a non-glycosylated form of tyrosinase is visible immediately after pulse and is rapidly degraded (CHO-). This recapitulates the known features of tyrosinase as a model of protein secretion in heterologous cell systems (Wang and Hebert, 2006).
Fig. 2.2: HepG2 expression of selected transcripts corresponding to ER proteins of unknown function

RNA was extracted from HepG2 cells at steady state and converted to cDNA using reverse transcriptase. cDNA sets were PCR amplified using primers specific to the unknown proteins indicated for the number of rounds indicated. The linear range of amplification (for future experiments) was determined by inspection of increased band intensity.
Fig. 2.3: Endpoint RT-PCR identification of novel targets of the unfolded protein response

HepG2 cells were treated with the conditions indicated prior to RNA extraction and generation of cDNA. Primers indicated were amplified into the linear range by PCR and visualized. GAPDH is used as a nonresponsive control.  

a) Upregulation of the unfolded protein response markers CHOP and Herp after pharmacological ER stress. DTT, dithiothreitol (2.5 mM); Tm, tunicamycin (1 μg/ml); Tg, thapsigargin (1 μM); Glc(-), incubation in glucose free medium.

b) Upregulation of unknown ER proteins by 12 h thapsigargin.

c) Upregulation of nicalin by DTT and Tg but not Tm.
Fig. 2.4: Real-time RT-PCR screen for unknown transcripts upregulated by tunicamycin in HepG2 cells.

HepG2 cells were treated with 1 μg/ml tunicamycin for 12 h or vehicle control (n = 3) and RNA was isolated. cDNA sets were generated using reverse transcriptase followed by analysis using real-time RT-PCR. a) Screen of 33 genes of unknown function. Those genes revealing increased expression (greater than 1.5 fold, n = 3 with 3 technical replicates each) and that are significantly upregulated (p > 0.05) are marked with an asterisk. The positive controls CHOP and Herp are upregulated 20.6 and 26.5-fold respectively. b) The 5 genes significantly upregulated by 12 h tunicamycin treatment. Error bars are mean +/- SEM.
To determine if any of the proteins of unknown function interacted with tyrosinase, a co-immunoprecipitation approach was used. Tyrosinase and several unknown proteins were expressed in BHK cells for 48 hours followed by immunoprecipitation with anti-FLAG and tyrosinase (His) Western blotting. Here, tyrosinase interacted strongly with nicalin and TMX2s, weakly with stexin and erlin-1, and not at all with sterlin, JAGN1, or sterubin (Figure 2.6). Importantly, only the immature ER form of tyrosinase was observed to interact with the ER-localized unknown proteins shown. No tyrosinase running at the molecular weight of the mature (post-ER) form was observed to interact with any unknown proteins (Figure 2.6). Therefore, stexin, nicalin, erlin-1, and TMX2s interact with tyrosinase but do not form a part of its final mature complex.

Discussion

In Chapter 1 it was found that known proteins functioning in cargo processing exhibit distinct localizations in the proteomics resource. These localizations overlap with other general functions in the ER such as detoxification and lipid metabolism. Therefore, to determine the function of unknown proteins, it is necessary to screen for proteins specifically involved in cargo processing. This chapter describes a strategy to identify poorly characterized ER proteins that function in cargo processing.

First, a bioinformatics approach was taken, which gives domain and feature information about unknown proteins on account of amino acid sequence similarity to known proteins. Of the proteins interrogated, two were selected: Nudilin has a nudix domain and sequence similarity to an RNA binding protein. TMX2 has sequence similarity to the protein disulfide isomerase family of protein folding enzymes.
Fig. 2.5: Tyrosinase is a model cargo protein trafficking through the ER. BHK cells transfected with myc/his tagged tyrosinase were pulsed for 15 min. with 35S methionine, followed by chase from 0 to 4 hrs. as indicated. Newly synthesized glycosylated (CHO+) tyrosinase shifts from an ER form to a mature glycosylated form (post-Golgi) on SDS-PAGE during chase. An unglycosylated form (CHO-) is also visible using this protocol and is rapidly degraded. Addition of the proline analogue azetidine 2-carboxylic acid (Azc) to the pulse medium (right panel) causes tyrosinase misfolding and the shift to the mature glycosylated form does not occur.
Fig. 2.6: Interaction between four resident ER proteins and the cargo protein tyrosinase

a) BHK cells co-expressing tyrosinase and the unknown proteins indicated were immunoprecipitated using anti-FLAG (unknown protein) and Western blotted using anti-His (tyrosinase). The far right lane is a Western blot of tyrosinase using 10% of the material used for the immunoprecipitations shown in the other lanes. b) Expression of FLAG tagged proteins of unknown function. The seven proteins indicated were tagged at the C-terminus with 3xFLAG and expressed in BHK cells, followed by anti-FLAG Western blotting.
Second, predicted transmembrane proteins were selected for a screen looking for genes upregulated by ER misfolding stress, since UPR upregulated genes are enriched for cargo processing functions. 33 genes were tested, of which 7 were found to have significant upregulation using a combination of Q-PCR and endpoint RT-PCR.

Finally, candidates were tested for interaction with the model cargo substrate tyrosinase. Tyrosinase has many features useful as a cargo substrate. It exhibits several determinants of folding since it is heavily disulfide bonded, heavily glycosylated, and moderately sized for potential hydrophobic domain exposure. In heterologous cells it is synthesized at the ER and transported through the Golgi, exhibiting a molecular weight shift on SDS-PAGE due to glycosylation maturation, before being transported to a lysosomal/endosomal compartment. Thus, tyrosinase maturation can be followed by SDS-PAGE. Co-expression of several unknown proteins with tyrosinase followed by a co-immunoprecipitation protocol found that nicalin, stexin, erlin-1, and TMX2s all interact with the immature but not mature form of tyrosinase. This suggests that these proteins function in cargo processing at the protein level, in the ER.

In summary, nudilin, nicalin, stexin, erlin-1, TMX2, and nudilin represent candidate proteins for functions in cargo processing in the ER. Subsequent chapters will explore their sublocalization in the proteomics resource, which predicts the specific step in cargo processing at which each protein is found to function.

**Materials and methods**

**Bioinformatics**
Domain annotations were obtained from NCBI entrez, either from the Gene and UniGene databases, or the automated domain finder associated with NCBI Blast. Signal sequences were determined using SignalP. Transmembrane domains were
determined by TMHMM. See Gilchrist et al. 2006 for details. Sequence alignments were generated using Clustal W.

**Generation of cDNA sets**
The human liver hepatoma cell line HepG2 was treated with control or tunicamycin (1 μg/ml), dithiothreitol (2.5 mM), or thapsigargin (1 μM) for 12 hours. RNA was extracted using the RNeasy tissue kit from Qiagen under RNAse free conditions according to manufacturer’s specifications. cDNA was generated using Superscript III reverse-transcriptase from Invitrogen according to manufacturer’s specifications. The poly-T primer was used here to only amplify mRNA.

**Primer design**
Human sequences were chosen based upon the highest BLAST score with the identified rat sequence from proteomics. Primer design utilized the software Primer 3 and utilized human sequences. High-scoring primer sets were then aligned with the rat and mouse orthologues of the unknown protein using Clustal-W. Primer sequences with significant deviation between humans and rodents were discarded to ensure functionality across species. Primers were synthesized by AlphaDNA, Inc.. Primer sequences used are shown in Table 2.1.

**Endpoint RT-PCR**
The cDNA sets were first amplified with primer sequences against the controls GAPDH and beta-actin to determine the relative amounts needed to normalize the control signal. Next, all other primers were amplified at various numbers of rounds of PCR ranging from 18 - 30 rounds in order to find the linear range of amplification for each primer set. The primers of interest were then amplified into the linear range and compared between cDNA sets (i.e. stress vs. control) for relative signal. Signal was assessed after separation on a 1% agarose gel containing ethidium bromide, using photography upon UV visualizaton.

**Q-PCR**
The PCR reaction was set up using 384 well plates. Each sample was split into three technical replicates and an n = 3 of cDNA sets from control vs. tunicamycin-treated cells was used. Only the primers that functioned appropriately on endpoint PCR (i.e. a single visible band below 30 rounds of amplification) were used for Q-PCR analysis. Q-PCR used the ABI 7900HT system with Rox and SYBR green kindly provided by the McGill Microarray Platform lab. SDS2.1(ABI) software was used for the quantitative analysis. Statistical assessment utilized Student’s T-test from Microsoft Excel.

**Antibodies**

Monoclonal mouse M2 anti-FLAG was purchased from Sigma, either as antibody alone or conjugated to agarose beads. Rabbit anti-His was from Bethyl. Secondary antibodies were goat anti-rabbit, anti-mouse, or anti-rat conjugated to HRP. Dilutions were optimized for minimum amounts necessary and ranged from 1:2500 to 1:10 000 for Western blotting.

**Generation of FLAG-tagged proteins of unknown function**

Proteins of unknown function were cloned into the Invitrogen Gateway system entry vector by Dr. H. Nagaya. For subcloning into the 3xFLAG vector, the LR recombination reaction (Invitrogen) was used, resulting in gene insertion with a C-terminal 3xFLAG tag. Subclones were then restriction mapped and sequenced for verification. Vectors were maintained in transformed XL-1 blue E. coli and DNA for transfections were generated using Qiagen miniprep and midiprep kits.

**Transfections and FLAG immunoprecipitation**

The tyrosinase construct was tagged at the C-terminus with myc/his and was a gift from Dr. D. Hebert (Univ. Massachusetts – Amherst). The described FLAG-tagged proteins were co-expressed with tyrosinase in BHK cells using Fugene 6 reagent. 48 hours after transfection cells were washed with PBS and lysed in lysis buffer (1% Triton, 50 mM Tris pH 7.5, 150 mM NaCl, 1 mM EDTA, Roche Complete protease inhibitors). The insoluble portion was removed by a 14 000 g spin at 4OC for 10 min.
Immunoprecipitation was carried out using Sigma EZ-view agarose M2 anti-FLAG beads overnight at 4OC. Beads were then washed 3 x 5 minutes with TBS before being heated in Laemmli buffer containing DTT for 5 min. at 95OC. Lysates were then processed for SDS-PAGE.

**Western blotting**
Following transfer, the nitrocellulose membrane was stained with Ponceau red and monitored for efficient transfer. The membrane was blocked with 5% milk in TWA (Tris, NaCl and Tween-20) for a minimum of 30 min. at room temperature, followed by overnight incubation at 4OC using primary antibody in 1% milk/TWA. Next, the membrane was washed 5 x 5 min. in 1% milk/TWA and incubated with secondary antibody for 1 hr. at RT. A final 5 x 5 min. wash in 1% milk/TWA was used, followed by processing for development using chemiluminescence.

**Pulse-chase**
48 hours after transfection, cells were prepared for radiolabelling. One hr. prior to labelling cells were washed once with methionine negative [met (-)] medium then replaced with met(-) medium containing the normal amount of dialyzed FBS. For cells treated with azetidine-2-carboxylase (Azc), 5 mM Azc was added to the medium here. Following 1 hr. methionine starvation, 250 μCi 35S-met was added for 15 min. (pulse). For the chase, cells were washed once with met(+) medium and then replaced with the same medium before being returned to the incubator for 1 – 8 hours. Following chase, cells were washed once with PBS and harvested in lysis buffer before processing for IP.
Table 2.1: RT and Q-PCR primer sequences. Indicated are the protein name, forward (F1) and reverse (R1) primer sequences, melting temperature of the indicated primers (Tm), and the expected PCR product length for each pair.

<table>
<thead>
<tr>
<th>Protein Name</th>
<th>Primer F1</th>
<th>Tm F1</th>
<th>Primer R1</th>
<th>Tm R1</th>
<th>PCR product length</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAPDH</td>
<td>ACCACATTGGAGAAGGCTGG</td>
<td>64.5</td>
<td>CTCAGTGTAGCCCAAGATGC</td>
<td>61</td>
<td>528</td>
</tr>
<tr>
<td>beta-actin</td>
<td>TTCTCAATGAGCTCGTGTG</td>
<td>60</td>
<td>ACATCTGCTGGAAGTTGGAC</td>
<td>60</td>
<td>799</td>
</tr>
<tr>
<td>herp</td>
<td>TTTTTGCTCTGCTTGCCA</td>
<td>59</td>
<td>TTTCAACAGAAATGGAATTT</td>
<td>58.5</td>
<td>438</td>
</tr>
<tr>
<td>ChOP</td>
<td>GGAACCTAGAGAGAGAGTGTTTC</td>
<td>59</td>
<td>CACCTCTGTTCTCCCTTCCT</td>
<td>59</td>
<td>423</td>
</tr>
<tr>
<td>Sterlin</td>
<td>AATTTTCCACTCCGGCCTT</td>
<td>61</td>
<td>TAAATTGCTGCGGCACTGCTT</td>
<td>60</td>
<td>501</td>
</tr>
<tr>
<td>A2B6</td>
<td>TTGTGTACTGCGGAGAGAC</td>
<td>60</td>
<td>CAGGAGCATAGGAGGAGAG</td>
<td>60</td>
<td>551</td>
</tr>
<tr>
<td>C14orf1</td>
<td>AAGGAGCTAGACAGAGAGAAGA</td>
<td>57.5</td>
<td>AGAGGAGCAGACTCTTTCCAA</td>
<td>58</td>
<td>624</td>
</tr>
<tr>
<td>CDC56</td>
<td>GCCCTGGTGTTAGTTTGA</td>
<td>60</td>
<td>GCTCAAAACACTGCAAGACG</td>
<td>60</td>
<td>496</td>
</tr>
<tr>
<td>MSO32</td>
<td>GCTGTGCTTTGCTTCAAGA</td>
<td>60</td>
<td>GTCGCCGTTTTCTGGAGAT</td>
<td>60</td>
<td>400</td>
</tr>
<tr>
<td>ZCD1</td>
<td>CAGCTGACTGAGTGTGAGGA</td>
<td>60</td>
<td>CAGCAAGAGTTCTGAGCACA</td>
<td>60</td>
<td>400</td>
</tr>
<tr>
<td>TMEM14C</td>
<td>GACCACCTGCACTTCTCTCCT</td>
<td>60</td>
<td>CAAACTGGCACTCAGGGGAA</td>
<td>60</td>
<td>400</td>
</tr>
<tr>
<td>TMEM168</td>
<td>CAGGACATGCCCTTTTCTTA</td>
<td>60</td>
<td>GCCCTGCAACAGATAGTTT</td>
<td>60</td>
<td>400</td>
</tr>
<tr>
<td>TMEM111</td>
<td>CACCAACACAGAGGTGATTG</td>
<td>60</td>
<td>TCCAAAGCTCCACTCTGCT</td>
<td>60</td>
<td>433</td>
</tr>
<tr>
<td>SCPPDH</td>
<td>CTAGAAGCTGGAGAACGCAAC</td>
<td>60</td>
<td>ATGAAATGCTCAACCCCTCA</td>
<td>60</td>
<td>400</td>
</tr>
<tr>
<td>JAGN1</td>
<td>TCGTCTCGTTGTCGTAAGTA</td>
<td>60</td>
<td>TCTCTCTCGTTGCTGCTCTG</td>
<td>60</td>
<td>400</td>
</tr>
<tr>
<td>DC2</td>
<td>CCAACCTGAGCTCAGAAGA</td>
<td>60</td>
<td>GATCGGTCAGGATTATGAA</td>
<td>59</td>
<td>278</td>
</tr>
<tr>
<td>Stevin</td>
<td>AGCTTCGACTATGCGATGAA</td>
<td>59.5</td>
<td>GATGAGGTGAGTCAAGCTT</td>
<td>55.5</td>
<td>440</td>
</tr>
<tr>
<td>MDSC2</td>
<td>TCTGATGCTTTGGAGGAGGATAC</td>
<td>60.5</td>
<td>TCATCACCCATCCCGATCAACAG</td>
<td>60</td>
<td>278</td>
</tr>
<tr>
<td>TMEM97</td>
<td>CACCTCCCTCTCTGAGTTTCT</td>
<td>60</td>
<td>TGGCTCTGAGCCACCTTCTCCT</td>
<td>60</td>
<td>503</td>
</tr>
<tr>
<td>Nicalin</td>
<td>ACCCTGACCCACACATCCTGCTCCT</td>
<td>60</td>
<td>TCTCTGCTGATCTGCACTCT</td>
<td>60</td>
<td>476</td>
</tr>
<tr>
<td>Erbin-1</td>
<td>TACGGCTCCTCCACACAAAG</td>
<td>60</td>
<td>TATGCTTCTGGAGATTTGTTGG</td>
<td>60</td>
<td>480</td>
</tr>
<tr>
<td>C17orf52</td>
<td>AGCTCTGAGGCGGAGGAAAAA</td>
<td>60</td>
<td>AGGCTGGCTTCCGCCTTACCC</td>
<td>60</td>
<td>503</td>
</tr>
<tr>
<td>C18orf19</td>
<td>GTTCAGTCTGAGCAGCAGATCA</td>
<td>60</td>
<td>GCAATGGGCTAAATGCGCTT</td>
<td>60</td>
<td>550</td>
</tr>
<tr>
<td>Lyrin</td>
<td>AGCTGAAAGCTGCTGCAAGA</td>
<td>60</td>
<td>ATCGAGAAACCTGCTGAGCA</td>
<td>59</td>
<td>513</td>
</tr>
<tr>
<td>IER3orf1</td>
<td>GGAAGCGATTCCCTCACAAGA</td>
<td>60</td>
<td>TATCAATGCGACTCTCATACG</td>
<td>61</td>
<td>133</td>
</tr>
<tr>
<td>C10orf35</td>
<td>AGGCTCCTCATATTTCCTGTG</td>
<td>60</td>
<td>ACCAGATCTCCCTCAATA</td>
<td>60</td>
<td>506</td>
</tr>
<tr>
<td>ERGIC1</td>
<td>TCTACAGGAAGGTGCAACAG</td>
<td>60</td>
<td>GGGTTGGAGGTGAGTCTGTC</td>
<td>60</td>
<td>514</td>
</tr>
<tr>
<td>TMEM4</td>
<td>GACCCAAAGAAGAACATCTCCT</td>
<td>60</td>
<td>TCTACACGCAAACTGGAGG</td>
<td>60</td>
<td>280</td>
</tr>
<tr>
<td>C10orf58</td>
<td>ATAGAAGCTGAAAACACTGGAGAA</td>
<td>57</td>
<td>TCTTTTCTCTGCTGCTCAAGA</td>
<td>60</td>
<td>452</td>
</tr>
<tr>
<td>C15orf24</td>
<td>GTCGGTATACGGAGAAGAGG</td>
<td>60</td>
<td>GCTCTGACCTACCTCCTTG</td>
<td>55</td>
<td>535</td>
</tr>
<tr>
<td>TMEM85</td>
<td>CCTCTACACGCTTGCTTGATG</td>
<td>60</td>
<td>CATGGAGACTGCACTTGAA</td>
<td>59.5</td>
<td>340</td>
</tr>
<tr>
<td>ATP13A1</td>
<td>GACCAAGGTGCTGCAAGCATTG</td>
<td>60</td>
<td>GGCCTACAGTGCTGAGAGA</td>
<td>80</td>
<td>400</td>
</tr>
<tr>
<td>TMEM87B</td>
<td>CATCAGATTGGCCCTATGTTT</td>
<td>60</td>
<td>CTCCTCATCAGCCCTTCAACA</td>
<td>60</td>
<td>133</td>
</tr>
<tr>
<td>TMCO1</td>
<td>GAGTCAAGCTGGTGTCGAACA</td>
<td>60</td>
<td>ACTCCCAAGAGAGTTGGTG</td>
<td>60</td>
<td>400</td>
</tr>
<tr>
<td>METTL7A</td>
<td>GCAAGCTGACCAAGAGAGAG</td>
<td>60</td>
<td>GGAAGTCAGAGGTGGGAAGAAT</td>
<td>60</td>
<td>399</td>
</tr>
<tr>
<td>TMX2</td>
<td>CCTACATCTGCTGCTTTCA</td>
<td>60</td>
<td>AGCTTTGGAAGAAACCATT</td>
<td>60</td>
<td>502</td>
</tr>
<tr>
<td>SCYL2</td>
<td>CTGGCTAGTCTGCCAAGAAT</td>
<td>60</td>
<td>GCAGGAGGAAACACCTCAAAGC</td>
<td>60</td>
<td>502</td>
</tr>
<tr>
<td>SterolB</td>
<td>TTTTCTCCTCTGCTCCCTCCT</td>
<td>60</td>
<td>GTGGGCGATTCAAGGAAACT</td>
<td>60</td>
<td>499</td>
</tr>
<tr>
<td>Tyrosinase</td>
<td>ATGGATTGTCGCGCTGCTGCT</td>
<td>60</td>
<td>ATTTGCTAGCTGCTGCTG</td>
<td>60</td>
<td>501</td>
</tr>
</tbody>
</table>
CHAPTER 3: Nudilin is a P-body constituent that regulates secretory traffic

Abstract

The spatial organization of the ER predicts that the function of unknown proteins can be predicted by their co-localization with known proteins in the proteomics resource. The following chapters describe the use of the proteomics resource to predict functional features of poorly characterized proteins. In this chapter, nudilin (nudt16) is predicted to have a function in cargo processing on account of its nudix domain and amino acid sequence similarity to RNA binding proteins. In the proteomics resource, nudilin is localized to the cytosolic face of the rough ER, similar to the localization of known transcription and translational proteins. Localization of nudilin by immunofluorescence shows localization to the P-body, a site for mRNA metabolism in the cytoplasm. The structure of nudilin shows an RNA-binding groove, in agreement with the function of nudilin on mRNA metabolism. To determine if nudilin has a function related to the secretory pathway, the effect of nudilin overexpression and siRNA knockdown on tyrosinase levels was assessed. Overexpression of nudilin increased tyrosinase levels while siRNA knockdown had the opposite effect. Hence, nudilin is a cytoplasmic protein that affects cargo in the secretory pathway as predicted by its localization in the proteomics resource.

Introduction

In Chapter 1 the ER is shown to be organized into spatial subdomains with specific functions in cargo processing. In Chapter 2, candidate proteins of unknown function were identified to be involved in cargo processing. Based on these observations, the specific function of unknown proteins should be predicted by their sublocalization in the proteomics resource. That is, where unknown proteins co-localize with known proteins, a common function is shared.
This chapter describes the protein we have named nudilin. Nudilin is previously known as nudt16 and contains a conserved nudix domain. The nudix superfamily is found from bacteria to humans and encodes pyrophosphate hydrolases of varying specificity (McLennan, 2006). In humans, several of the nudix domain proteins act as RNA binding and decapping proteins (McLennan, 2006). The core catalytic residues of the nudix domain are conserved in nudilin (Figure 2.1a).

Nudilin is homologous to the Xenopus laevis protein X29 and both enzymes been shown to bind snoRNA and act as RNA decapping enzymes in vitro (Ghosh et al., 2004). In the course of this work, further studies into the specificity of X29 and nudilin were published, showing that nudilin has a wider substrate RNA specificity than X29; in vitro, nudilin is capable of hydrolyzing snoRNA, mRNA, and rRNA using various metal ion concentrations, while X29 has more specificity for snoRNA (Peculis et al., 2007).

In vivo, RNA metabolism occurs partly in specialized cytoplasmic organelles known as P-bodies. The constituents of P-bodies include mRNA decapping proteins (a process that leads to RNA decay), proteins involved in nonsense-mediated decay, constituents of the endogenous RNA interference mechanism, other RNA-binding proteins, and proteins involved in translation (Eulalio et al., 2007a). Correspondingly, targeting of mRNAs to P-bodies results in transcriptional and translational repression.

A link between P-bodies and the ER has been previously uncovered, involving two proteins from the Scd6/Lsm family of RNA binding proteins. Trailer hitch is a Drosophila protein that localizes to P-bodies and apparently co-localizes with the ER while Car1 is a C. elegans RNP granule protein (functionally related to P-bodies) that localizes to the ER at various phases of cytokinesis (Decker and Parker, 2006). Both affect ER morphology. Deletion of trailer hitch causes disorganization of ER exit sites, while deletion of Car1 results in a dispersed ER
during formation of the mitotic spindle (Squirrell et al., 2006; Wilhelm et al., 2005). Whether these effects are on transcription or translation of ER factors or if a more direct interaction is involved is at present unclear.

This chapter shows that nudilin is localized to the cytoplasm of the rough ER in the proteomics resource, and to P-bodies in the cytoplasm by immunofluorescence. The structure of nudilin shows structural homology to the Xenopus RNA decapping protein X29. The ER function of nudilin was assessed by overexpression and siRNA knockdown effects on the cargo protein tyrosinase. Overexpression increased tyrosinase levels, while siRNA decreased tyrosinase. These effects occurred proximal to protein synthesis. Thus, nudilin functions to affect cargo upstream of protein synthesis, as predicted by its sublocalization in the proteomics resource.

**Results**

The amino acid sequence of nudilin shows homology to syndesmos, a plasma membrane protein, and X29, an RNA binding protein (Figure 2.1a). The identification of nudilin in the proteomics resource suggests an ER function. Specifically, nudilin is enriched in the salt wash fraction of the rough ER in the proteomics resource (Figure 3.1a and b). The removal of nudilin by a high salt wash suggests a cytoplasmic protein associating with the rough ER membrane. Known proteins in the proteomics resource with this distribution include RNA metabolism and translation-associated proteins (Chapter 1). The co-distribution of nudilin predicts a similar function.

An antibody was raised against a recombinant nudilin protein derived from mouse sequences. In mouse NIH3T3 cells, immunofluorescence using anti-nudilin shows diffuse staining in the cytoplasm and nucleus, as well as in punctate bodies distributed in the cytoplasm (Figure 3.1c). This localization is distinct to that of calnexin, a transmembrane chaperone facing the luminal ER (Figure 3.1c). Some
of the punctate bodies in the cytoplasm are P-bodies, since they co-localize with punctate structures marked by the P-body marker Dcp1 (Figure 3.1c).

The structure of nudilin has been solved by the Cygler laboratory using X-ray crystallography of the recombinant nudilin protein. Some features of this work are shown here. The structure of nudilin shows an RNA binding groove (Figure 3.2a). Notably, the structure of nudilin superimposes on X29 and the RNA decapping enzyme Dep2p (Figure 3.2b). Hence, nudilin has both amino acid sequence and structural similarity at catalytic residues for RNA binding proteins.

The proteomics resource predicts that nudilin localizes to the ER to affect cargo processing. The fluorescence localization of nudilin does not support a primary localization to this compartment. To test the integrity of the proteomics prediction, the effect of nudilin overexpression and knockdown on the secretory cargo protein tyrosinase was assessed in human 293T cells. Overexpression of nudilin caused an increase in total levels of tyrosinase, but not that of calnexin (Figure 3.3a). Pulse chase studies show that the effect of nudilin on tyrosinase results in increased tyrosinase synthesis and does not affect the rate of tyrosinase degradation (Figure 3.3b). This is consistent with the proteomics prediction, sequence analysis, and structural evidence on nudilin function, which support an effect on cargo proximal to ER entry.

Knockdown of nudilin by siRNA had the opposite effect, and led to a decrease in the amount of tyrosinase synthesized (Figure 3.4). The use of scrambled siRNA oligos (control) was compared to oligos against nudilin and the chaperone calnexin. Calnexin knockdown is predicted to result in enhanced degradation of tyrosinase, based upon studies of calnexin knockdown or inhibition on certain cargo, alongside the known interaction between calnexin and tyrosinase (Pieren et al., 2005; Toyofuku et al., 1999). Similar to calnexin, knockdown of nudilin resulted in diminished tyrosinase levels (Figure 3.4a). Pulse chase analysis following nudilin knockdown showed that the amount of tyrosinase synthesized
was diminished by nudilin knockdown, while the degradation rate of tyrosinase was once again unaffected (Figure 3.4b).

Discussion

Nudilin is a cytoplasmic protein localizing to the cytoplasmic face of the rough ER by proteomics and to the P-body and cytoplasm by immunofluorescence. How can these differing localizations be resolved? For other P-body proteins that affect ER function, a co-localization between P-bodies and ER has been observed. Therefore, these organelles are in close proximity at various times in the cell cycle. Moreover, it is a feature of fluorescence localization that highly bright objects (in this case the localization of nudilin to the P-body and cytoplasm) can obscure less bright objects (in this case the localization of nudilin to the ER), similar to the difficulty of looking at objects near bright stars in the sky (Bell et al., 2007b). It is unlikely that the proteomics localization of nudilin represents P-body contamination of ER fractions since of the 29 confirmed P-body proteins only 1 (staufen) was found in our proteomics resource (Eulalio et al., 2007a; Gilchrist et al., 2006). Hence, it is unlikely that P-bodies converge with the ER under the steady state conditions of rat liver used for proteomics and nudilin is specifically enriched in the ER compared to other P-body constituents.

The amino acid sequence of nudilin suggests a decapping function on account of its similarity and the conservation of catalytic residues with Xenopus X29. The structure of nudilin also agrees with this function as it has structural homology with decapping proteins as well as due to the presence of a distinct RNA binding groove.

Overexpression of nudilin increased tyrosinase levels while siRNA knockdown had the opposite effect. This effect was visible immediately after a 15 min. pulse, suggesting that these effects occurred upstream of synthesis. If nudilin were directly decapping and contributing to the degradation of tyrosinase mRNA, the opposite effect on nudilin levels would be expected where nudilin overexpression
should enhance nudilin transcript degradation. Like Car1 and trailer hitch, the mechanism of how nudilin is affecting cargo is therefore not determined. Nudilin may have an effect on transcript levels of other ER proteins important to cargo processing or a more direct effect may be implicated.

Without the localization of nudilin in the proteomics resource, the effect of nudilin on cargo would not have been predicted. Moreover, the step at which nudilin affects cargo (not folding or degradation but upstream of these processes) shows an agreement between the proteomics prediction and that suggested by sequence, structure, overexpression and knockdown analysis. Hence, the model of a spatially organized ER gives insight into the function of the previously uncharacterized protein nudilin.

**Materials and Methods**

**YFP visualization and immunofluorescence**

Mouse NIH3T3 cells were cultured on coverslips and transfected with YFP-tagged Dcp1p. Mouse cells were used because the nudilin antibody was raised against mouse recombinant protein. Thirty-six hours after transfection, cells were fixed with methanol for 5 min at 20°C. The fixed and permeabilized cells were blocked in 1% bovine serum albumin (BSA) in PIPES based buffer for 30 min and then incubated with anti-calnexin or anti-nudilin antibody in the blocking buffer for 1 hour at room temperature. Cells were washed 3 x 5 min. with blocking buffer and then incubated with Alexa 594-conjugated anti-Rabbit IgG antibody (Molecular Probes, Eugene, OR) for 30 min. at room temperature. The cells were again washed 3 x 5 min with blocking buffer, and the coverslips were mounted onto ProLong Gold (Invitrogen). Images were taken with an inverted confocal laser scanning microscopy (Zeiss LSM 510 META) with a 63× oil objective lens and processed by LSM Image Browser (Zeiss).
Fig. 3.1: Localization of nudilin

a) Nudilin is predominantly localized in the rough microsomes salt wash fraction, indicating a cytosolic protein associated with the rough ER membrane. b) Nudilin predominantly localizes to the rough ER in the total fractions. c) Localization of nudilin to the punctate and diffuse structures in the nucleus and cytosol. Top, nudilin immunofluorescence in mouse NIH3T3 fibroblasts using a polyclonal antibody raised against mouse sequences. Co-localization of the punctate structures of nudilin in the cytosol with the P-body marker Dcp1b is shown. Bottom, the ER marker calnexin does not co-localize with the Dcp1b positive P-body structures. Scale bar: 10 microns.
Fig. 3.2: Nudilin is similar to the RNA decapping protein X29

a) The structure of nudilin (provided by the Cygler laboratory) shows a cleft suitable for RNA binding. b) Superposition of nudilin (yellow) with its closest structural homologues X29 (blue, PDB code 1U20), nudix hydrolase DR1025 from *D. radiodurans* (red, PDB code 1SJY) and *S. pombe* mRNA decapping enzyme Dcp2P (green, PDB code 2A6T), shows a high structural similarity between these proteins.
Fig. 3.3: Overexpression of nudilin increases the level of tyrosinase

Overexpression of nudilin increases tyrosinase levels. 293T cells were transfected for 48h with tyrosinase and the amount of nudilin shown, followed by Western blotting as indicated. 1 μg tyrosinase causes an increase in the amount of tyrosinase observed. The nudilin antibody shown here was raised against a mouse recombinant protein and is ineffective at recognizing endogenous nudilin in human 293T cells, although it does recognize the overexpressed human nudilin used here. b) Consistent with its localization, nudilin increases the level of tyrosinase at the 0 h time point but does not affect the rate of degradation thereafter. 293T cells were pulsed for 15 min. with radiolabel followed by chase as indicated.
Fig. 3.4: siRNA knockdown of nudilin decreases tyrosinase levels

a) siRNA of nudilin and calnexin in 293T cells. siRNA of both calnexin and nudilin diminish tyrosinase levels. Although the knockdown of calnexin is incomplete after two days (due to the long half life of calnexin), a decrease in tyrosinase expression is seen, consistent with studies on other secretory cargo (see text). Neither calnexin nor nudilin knockdown diminishes the expression of the expressed cytosolic YFP protein.

b) siRNA of nudilin does not affect the degradation kinetics of tyrosinase. Consistent with the localization of nudilin to a translation-associated function, knockdown of nudilin affects the amount of tyrosinase seen at 0 h chase, but does not affect the rate of degradation thereafter. As previously, cells were transfected with siRNA for 48 h, followed by 15 min. radiolabelling and chase as indicated.
Pulse-chase and Western blotting

Pulse-chase and Western blotting experiments using tyrosinase were performed as described in Chapter 2. Briefly, human 293T cells were co-transfected using Fugene 6 (Roche) with His-tagged tyrosinase and FLAG-tagged nudilin for 48 h prior to processing for pulse-chase or Western blotting. For siRNA experiments, 293T cells were co-transfected with siRNA constructs (Dharmacon) and tyrosinase using Lipofectamine 2000 (Invitrogen). Cells were then incubated with the siRNA constructs for 48 h prior to processing for pulse-chase or Western blotting. The sequences used are as follows. Control: CGUCAACAAAGGUAAGAGUUU. Calnexin: GACGAUACCGAUGAUGAAUU. Nudilin: siGenome smart pool FLJ31265.
CHAPTER 4: Nicalin is localized in proximity to the translocon

Abstract

Earlier in this thesis, nicalin was found to be upregulated by protein misfolding stress and to interact with the immature form of the secretory cargo protein tyrosinase, describing a role in cargo processing. In the proteomics resource, nicalin is found preferentially in the rough ER in the detergent phase of Triton X-114. This predicts that nicalin is a transmembrane protein and distributes similar to translocon-associated proteins. In this chapter, nicalin is found to localize and function in proximity to the translocon. First, antibodies raised against nicalin show increased abundance in rough ER fractions compared to smooth ER. Second, nicalin interacts with signal peptide peptidase, an aspartic protease known to function in proximity to the ER. Third, overexpression of nicalin decreases the abundance of the cargo protein tyrosinase in transfected cells. Nicalin does not affect tyrosinase mRNA levels, nor does it affect tyrosinase degradation. Instead, nicalin overexpression decreases the amount of tyrosinase that enters into the ER and becomes glycosylated. Finally, deletion studies of nicalin show the surprising result that the N-terminal signal sequence of nicalin is sufficient to diminish tyrosinase levels. Pulse-chase analysis shows that the nicalin signal sequence is degraded slowly in the ER. Together these results suggest a model whereby nicalin inhibits tyrosinase translocation into the ER, with evidence suggesting a long-lived nicalin signal sequence as a mechanism.

Introduction

In the course of this work, nicalin was named on the basis of its amino sequence similarity to nicastrin, a protein located at the plasma membrane which works in concert with presenilin in regulated intramembrane proteolysis (Haffner et al., 2004; Shah et al., 2005). Nicalin is found to interact with another unknown
protein named nomo, and when both proteins were deleted in a zebrafish model, nodal signalling and mesoendodermal patterning were affected (Haffner et al., 2004). The cellular process underlying this phenomenon was not determined. Although nicalin and nicastrin have amino acid sequence similarity, nicalin localizes to the ER and has no effect on regulated intramembrane proteolysis by nicastrin (Haffner et al., 2004). For gamma secretase, the regulated intramembrane proteolysis process involves the interaction between nicastrin and the aspartic protease presenilin, found at the plasma membrane. There is also an aspartic protease found in the ER membrane, known as signal peptide peptidase (SPP), predicting that nicalin may function in concert with signal peptide peptidase.

Signal sequences are found at the N-terminus of cargo proteins in order to target these proteins to the ER and to assure their correct orientation in the membrane (Martoglio and Dobberstein, 1998). Once cargo proteins are fully translocated into the ER, the signal sequence is cleaved off the nascent protein by a protease known as signal peptidase. This liberates the free N-terminal end of the cargo protein to release into the lumen of the ER, leaving behind a transmembrane “stub” in the ER membrane. These stubs are then rapidly removed by intramembrane proteolysis using signal peptide peptidase (Martoglio, 2003). The resulting degradation products of the signal sequence stub emanating from the ER membrane could also be biologically active, as is known for the degradation products of intramembrane proteolysis found at the plasma membrane. The best described use of this system is in immunity. In this case, the immune system uses SPP to generate HLA-E epitopes from MHC I for presentation to natural killer cells (Martoglio, 2003). This reports on the proper MHC I synthesis inside of cells, thereby preventing elimination by natural killer cells.

Signal peptide peptidase has also been recently shown to have a role in protein degradation (Lilley and Ploegh, 2004; Loureiro et al., 2006). In the context of this thesis, it is hard to reconcile the rough ER role of signal peptide peptidase with
the smooth ER localization of the ERAD machinery and proteasome. However, Ploegh has recently proposed a degradation pathway that does not involve a translocation channel, but instead a vesicular system (Ploegh, 2007). This idea is currently only conceptual, but it points to the possibility of segregation between a vesicular degradation system and a retrotranslocon channel based system. The vesicular system cannot be localized in the proteomics resource because the constituents of this system are not known, but may well emanate from a location separate to the typical pathway which we have localized to smooth ER. Nevertheless, signal peptide peptidase has a well-defined function in proximity to the translocon to degrade signal sequences. Its role in retrotranslocation and degradation is less well understood.

In this thesis, nicalin has been shown to function in cargo processing. It interacts with the immature ER form of tyrosinase and is upregulated by ER stress. In Chapter 1, known proteins functioning in cargo processing are shown to have a segregated spatial distribution in the proteomics resource. Localization of nicalin in the proteomics resource is predicted to determine the specific step at which nicalin functions.

Here, nicalin is shown to localize to the detergent phase of the rough ER membrane, suggesting a translocon-associated function. Consistent with this localization, nicalin interacts with signal peptide peptidase. Overexpression of nicalin decreases tyrosinase expression; however, neither mRNA levels, nor the degradation rate of tyrosinase are affected. Instead, nicalin diminishes the amount of tyrosinase able to be translocated into the ER, as monitored by the N-glycosylation state of tyrosinase. Glycosylated tyrosinase (N-glycosylation occurs in the ER) is decreased by nicalin, while unglycosylated tyrosinase is unaffected or increased. This suggests that nicalin participates in a pre-emptive quality control system that prevents entry of cargo into the ER during times of stress. Surprisingly, the signal sequence of nicalin is alone sufficient to reduce tyrosinase
levels and is cleared slowly. Together, these results support the proteomics prediction that nicalin functions at the level of the translocon.

Results

In the total fractions, nicalin was found to be localized predominantly in rough ER (Figure 4.1a). In the biochemical fractions, nicalin was independently found in the rough ER, with its greatest abundance in the detergent phase (Figure 4.1b). In addition, expression of nicalin tagged at its C-terminus to YFP displayed an ER localization (Figure 4.1c). To determine the orientation of nicalin in the membrane, the use of consensus N-glycosylation sites found in the nicalin amino acid sequence was assessed. Nicatin is predicted to have an N-terminal signal sequence and a C-terminal transmembrane domain. If signal sequence cleavage results in the N-terminus being exposed to the ER lumen, with the C-terminal transmembrane domain anchoring it to the membrane, then two consensus sites for N-glycosylation (NXS/T) should be used. A polyclonal antibody was raised against nicalin peptide sequences; this antibody recognizes a single band at the predicted molecular weight for nicalin in purified fractions of rat liver ER (Figure 4.1d). When rat liver ER fractions were digested with a deglycosylating enzyme (PNGase F), followed by SDS-PAGE and Western blotting, nicalin showed a small molecular weight shift (Figure 4.1d). Thus, nicalin is a glycoprotein with the bulk of its sequence oriented towards the lumen of the rough ER.

In Chapter 2, nicalin was observed to interact with the immature ER form of tyrosinase. To determine the time course of nicalin interaction with tyrosinase, a pulse-chase sequential immunoprecipitation protocol was used (Figure 4.2). This protocol isolates the pool of tyrosinase interacting with nicalin and has been previously used to capture cargo interactions with the chaperone calnexin (Ou et al., 1993). FLAG-tagged nicalin and His-tagged tyrosinase were co-expressed in BHK cells for 48 h, given a 15 min. pulse with 35-S methionine and chased from 0 to 8 h. Nicatin associated transiently with the immature form of tyrosinase. To determine if the conformational state of tyrosinase affected the nicalin interaction,
azetidine 2-carboxylic acid (Azc) was added for 60 min. prior to pulse (75 min. total treatment), then washed out in the chase. This compound is a bulky proline analogue that is incorporated into newly synthesized proteins and through steric hindrance inhibits formation of stable conformational folding. When Azc was used, the interaction between tyrosinase and nicalin was extended in time. This is characteristic of molecular chaperones, which make transient interactions with immature cargo proteins in a manner dependent upon conformational states. Hence, nicalin displays chaperone-like properties, consistent with being an ER secretory cargo processing protein.

In 293T cells, overexpression of nicalin but not stexin or TMX2 led to a decrease in the amount of expressed tyrosinase visualized after Western blotting (Figure 4.3a). This was selective for cargo, as nicalin overexpression did not affect the endogenous cytoplasmic protein p97, the endogenous ER membrane chaperone calnexin, nor cytoplasmic expressed YFP at the protein level (Figure 4.3a and b). Nor did nicalin affect the degradation of tyrosinase. Here, the rate of the total pool of tyrosinase was assessed by pulse-chase analysis and single immunoprecipitation of the His-tagged tyrosinase (Figure 4.3c). This showed that while the initial (t = 0) amount of tyrosinase was diminished after nicalin expression, the rate of degradation during the chase was identical when comparing nicalin vs. vector expression (Figure 4.3c). Nicalin also had no effect on the mRNA expression of tyrosinase, as assessed by endpoint RT-PCR (Figure 4.3d). Together these results show that the effect of nicalin is distal to mRNA metabolism, but proximal to protein degradation; nicalin thus functions either in protein synthesis or translocation into the ER.

The proteomics localization predicts that nicalin functions in translocation. In Figure 4.3c it was noticed that immediately post-pulse (t = 0), nicalin overexpression decreased the higher molecular weight forms of tyrosinase corresponding to the use of 7, 6, and 5 N-glycosylation sites (and thus showing 3 bands on the gel; Figure 2.5) (Ujvari et al., 2001). However, a lower molecular
weight form of tyrosinase appeared not to be affected by nicalin overexpression. Recently, a phenomenon known as pre-emptive quality control (pQC) has been described (Kang et al., 2006). In pQC, newly synthesized proteins in the secretory pathway are prevented from entering into the ER, in order to prevent further stress on an already overloaded ER. The secretory proteins subjected to pQC are distinguished from proteins entering the ER by their lack of glycosylation, since the addition of the N-glycan tree occurs in the lumen of the ER (Kang et al., 2006). To determine if tyrosinase is subjected to pQC, cells expressing tyrosinase were treated for 30 min. with dithiothreitol (DTT; to induce ER stress) and pulsed for 10 min. (conditions shown to optimize pQC visualization), before processing for tyrosinase immunoprecipitation and SDS-PAGE. Figure 4.4a shows that tyrosinase is indeed subject to pQC. Addition of DTT diminished greatly the visualization of the N-glycosylated forms of tyrosinase, with a much less effect on the lower molecular weight unglycosylated form of tyrosinase. That this lower molecular weight form is unglycosylated was confirmed by the use of tunicamycin, which inhibits glycosylation and under these conditions caused a partial shift from the higher molecular weight glycoforms of tyrosinase to the unglycosylated form of tyrosinase. Thus, tyrosinase is subject to pQC.

To determine if nicalin overexpression was having a pQC effect on tyrosinase, the forms of tyrosinase were monitored using pQC conditions. After 30 min. methionine starvation and a 10 min. pulse, cells overexpressing nicalin were found to have decreased amounts of the glycosylated forms of tyrosinase, while the unglycosylated form was unaffected or increased (Figure 4.4b). Treatment with DTT not only affects pQC, but also decreases protein translation in the acute phase. To rule out that the effect of nicalin resembled that of protein synthesis inhibition, cycloheximide was used as a pharmacological inhibitor of protein translation. Unlike the effect of nicalin, cycloheximide caused a decrease in both the glycosylated and unglycosylated forms of tyrosinase (Figure 4.4b). The quantitation shown in Figure 4.4c demonstrates the varying effects of DTT (pQC + translation inhibition), nicalin (pQC), and cycloheximide (translation inhibition)
on tyrosinase expression. These results are as predicted if nicalin functions in pre-emptive quality control.

Nicalin has sequence similarity to nicastrin, a protein that combines with presenilin in the gamma-secretase complex. To determine if nicalin interacts with the aspartic protease of the ER, signal peptide peptidase (SPP), nicalin was expressed in BHK cells and co-immunoprecipitated with a commercial SPP antibody. As seen in Figure 4.5a, the SDS-resistant dimer (∼95 kDa) form of SPP interacted preferentially with nicalin compared to stexin, TMX2, and sterubin. In support of this interaction, SPP is found preferentially in the membrane of rough ER in the proteomics resource, similar to nicalin (Figure 4.5b). Western blotting of total rat ER fractions for nicalin, SPP and the cargo protein transferrin showed that nicalin and SPP are both enriched in rough ER, while transferrin distributes equally across rough and smooth ER (Figure 4.5c). Hence, nicalin and signal peptide peptidase are found in the same spatial location of the ER and are observed to interact by co-immunoprecipitation. To rule out that the interaction between SPP and nicalin is an enzyme-substrate interaction, a pharmacological inhibitor of SPP, ZLL-2-ketone (Weihofen et al., 2003) was used. The N-terminus corresponding to the signal sequence of nicalin (residues 1 – 51 tagged to YFP) was expressed in BHK cells and after a 15 min. pulse, the signal sequence cleaved and uncleaved forms of nicalin are visible. Treatment of cells with the SPP inhibitor did not prevent cleavage (Figure 4.5d). Therefore, SPP does not cleave nicalin and the interaction between SPP and nicalin is not an enzyme-substrate interaction.

Next, an attempt was made to determine the domain of nicalin responsible for the pre-emptive control function. A series of C-terminal nicalin deletion mutants was constructed and tested for the ability to decrease tyrosinase levels, as assessed by co-expression in 293T cells and Western blotting. All of the mutants tested were capable of decreasing tyrosinase levels, including the 1 – 51 a.a. construct tagged to YFP (Figure 4.6a). When the expression of the various constructs was assessed
by Western blot, the 1-51 a.a. construct migrated on SDS-PAGE at a nearly identical rate to YFP alone (Figure 4.6 a, arrow). In conjunction with Figure 4.5d, this suggested that this construct was expressed having already had its signal sequence cleaved.

An amino acid sequence analysis of the N-terminus of nicalin showed two possible start methionines, a potential transmembrane anchor domain, and a predicted cleavage site after residue 41 (Figure 4.6b). A further set of deletion mutants was constructed. At low expression, residues 1-20 did not affect tyrosinase levels, but at a higher expression even this construct was able to diminish tyrosinase (Figure 4.6c). All other N-terminal constructs diminished tyrosinase levels, however, detection of their expression by YFP Western blotting was variable. This suggests that even unstable forms of the nicalin signal sequence, here degraded, are capable of preventing tyrosinase entry into the ER. To determine the stability of the nicalin signal sequence, a pulse chase protocol was used. Normally, signal sequences are cleaved within minutes of entry of the nascent protein into the ER (Lyko et al., 1995), however the signal sequence of nicalin persisted for greater than 30 min. and was still visible at 90 min. after synthesis (Figure 4.6d).

Discussion

Interrogation of the organellar proteomics resource of Gilchrist et al. 2006 shows that the ER is spatially segregated into functional groups of proteins. The function of unknown proteins can be predicted by their co-distribution with known proteins in the resource. Nicalin is found to be enriched in the detergent phase of the rough ER, similar to translocon-associated proteins. Nicalin is predicted to be a translocon-associated protein. Mapping of its N-glycans shows that it is oriented towards the lumen of the ER. Moreover, Western blotting of nicalin in rat liver fractions using a polyclonal antibody confirms its localization.
That nicalin is a translocon-associated protein is consistent with the following evidence. First, nicalin interacts with signal peptide peptidase, an aspartic protease that is known to cleave signal peptides by intramembrane proteolysis. This is consistent with the localization of nicalin and SPP to the same space, as shown both by proteomics and Western blot. Also, SPP does not cleave the nicalin signal sequence in place of signal peptidase; pharmacological inhibition of SPP does not affect nicalin signal sequence cleavage. Second, overexpression of nicalin decreases the amount of tyrosinase observed. This effect is distal to mRNA and proximal to ERAD, since the transcript levels and degradation rate of tyrosinase was unaffected by nicalin. Third, overexpression of nicalin preferentially diminishes N-glycosylated forms of tyrosinase, but does not affect or increases the unglycosylated form. Since glycosylation occurs in the lumen of the ER, this suggests that nicalin is preventing entry of tyrosinase into the ER lumen. It is unlikely that glycosylation itself is affected; in pQC experiments conducted by Hegde et al., the signal sequence of the cargo protein entering the ER determined the amount of pQC under similar stress conditions (Kang et al., 2006). In an ER stress situation, upregulation of nicalin is here suggested to be protective by enacting pQC and preventing additional entry into an already overloaded ER.

When deletion studies were carried out to determine the domain of nicalin necessary for the effect of tyrosinase, the N-terminal signal sequence was found to be sufficient for this action. Even forms of the nicalin signal sequence with expression undetectable by Western blot were able to diminish tyrosinase levels. Further, the longer detectable forms are slowly cleaved such that very little sequence is left tagged to YFP. Thus, the model for this function is that translocation of newly synthesized nicalin itself is enacting the pre-emptive quality control. This could occur in several ways: translocation of nicalin non-competitively competes with other substrates for translocon access as it is itself translocated, or secondly that the long residence time of the nicalin signal sequence prevents further access of other cargo to enter the ER. It is interesting
that the signal sequence of the cargo entering the ER is also a determinant of pQC – the signal sequence of nicalin may activate pQC \textit{in trans} to regulate other cargo entering the ER. It cannot be determined if the YFP constructs that are not expressed are being degraded while the signal sequence is maintained, since it is not possible to track the signal sequence itself in these assays, only the cleavage event. That the signal sequence cleavage event is occurring slowly shows that the signal sequence is maintained in the membrane for at least this amount of time. Recently, a pharmacological inhibitor of translocation was described (Garrison et al., 2005). This inhibitor has a physical resemblance to a signal sequence folded on itself (R. Hegde, personal communication). Therefore it is plausible that a signal peptide from a specific protein could act as a translocation inhibitor. The chaperone characteristics of nicalin may be separate to this function or may be related to the pQC process, as Kang et al. found that chaperones at the luminal face of the translocon relate to pQC in an undefined manner.

This chapter describes nicalin to be localized to the translocon by proteomics. Studies into the function of nicalin are in agreement with this prediction. The finding that the signal sequence of nicalin has a specific function is a novel concept and from preliminary studies appears to be enacted by an unexpected process that will require further work to untangle.
Fig. 4.1: Localization of nicalin. *a*) Localization of nicalin to the rough ER in the proteomics resource. The distribution of nicalin in rough ER, smooth ER, and Golgi fractions is plotted as a proportion of the total nicalin abundance observed. RM, rough microsomes; SM, smooth microsomes; Golgi, Golgi fractions. *b*) Localization of nicalin to the detergent fraction of the rough ER in the proteomics resource. As described in Gilchrist et al. 2006 and the main text, ER fractions were extended by biochemical subfractionation. The distribution of nicalin in these fractions is plotted as a proportion of the total nicalin observed. SLT, salt wash fraction; INS, Triton X-114 insoluble fraction; DET, Triton X-114 soluble fraction; AQU, aqueous phase of the Triton X-114 soluble fraction. *c*) Localization of nicalin tagged to YFP. Left, HeLa cells transfected with nicalin-YFP. Right, merged image showing the colocalization (yellow) of nicalin-YFP (green) with anti-calnexin immunofluorescence (red). Scale: 10 microns. *d*) Nicalin is N-glycosylated, a process requiring entry into the ER. Purified ER fractions from rat liver were treated with PNGase F (+) or mock (-) and Western blotted (WB) as indicated. Both transferrin and a polyclonal antibody raised against a nicalin-specific peptide show a shift after PNGase F treatment, while calnexin does not.
Fig. 4.2: Time course of nicalin interaction with tyrosinase
BHK cells co-expressing tyrosinase and nicalin-FLAG were pulse labelled for 15 min. with 35S methionine and chased up to 8 hrs. Lysates were subjected to a sequential immunoprecipitation protocol to isolate the pool of tyrosinase interacting with nicalin, rather than the total tyrosinase pool. Kinetics are shown under normal and Azc misfolding conditions. The asterisk (*) indicates nicalin-FLAG that tightly co-associates. Quantitation by densitometry is shown as a % of the maximal interaction observed.

Sequential IP protocol
1st IP: FLAG (nicalin)
Wash beads to remove all protein not bound to FLAG-nicalin
Boil beads in 1% SDS
Quench SDS with 10 volumes cold 1% Triton X-100
Clear with protein A beads to remove FLAG antibody
2nd IP: His to isolate tyrosinase previously bound to FLAG-nicalin
Fig. 4.3: Overexpression of nicalin diminishes tyrosinase levels

a) Overexpression of nicalin decreases tyrosinase expression. FLAG-tagged vector (V), stexin, nicalin, or TMX2 were co-expressed with tyrosinase in 293T cells and Western blotted for tyrosinase using anti-His. Only nicalin diminishes tyrosinase content, with no effect on p97 content (lower panel).

b) No effect of nicalin overexpression on endogenous (calnexin) and cytosolic (YFP) proteins. Nicalin was co-expressed with YFP for 48 h and Western blotted as indicated.

c) Effect of nicalin overexpression on initial expression of tyrosinase, but not on tyrosinase degradation kinetics. Left, 293T cells expressing nicalin-FLAG or FLAG vector and His-tyrosinase were pulsed for 15 min. followed by chase as indicated. The asterisk indicates FLAG-tagged nicalin that co-associates with tyrosinase Right, Quantitation of tyrosinase levels. d) No effect of nicalin overexpression on tyrosinase RNA levels. Tyrosinase was co-expressed with YFP or nicalin-YFP for 48 h in 293T cells, followed by RNA isolation and endpoint RT-PCR as indicated.
Fig. 4.4: Nicalin functions in proximity to translocon in pre-emptive quality control

**a)** Tyrosinase is a target of the pre-emptive quality control system in stressed cells. 293T cells expressing tyrosinase were incubated with indicated drugs for 30 min., prior to 10 min. pulse labeling and IP for tyrosinase (anti-His). DTT: dithiothreitol; Tm: tunicamycin. The arrow indicates the mobility of the unglycosylated form (CHO - ) of tyrosinase.

**b)** Effects of nicalin overexpression and cycloheximide treatment (CHX) on glycosylated (CHO +) and unglycosylated forms (CHO - ) of tyrosinase. 293T cells co-expressing tyrosinase and the constructs indicated were treated for 30 min. +/- CHX (0.1 μg/ml), pulsed for 10 min. and analyzed by anti-His (tyrosinase) IP. Asterisks indicate the tagged forms of nicalin co-associated with tyrosinase.

**c)** Quantitation of the glycosylated (+ CHO) and unglycosylated (-CHO) forms of tyrosinase after experiments as indicated. Results are expressed as a ratio between treated (DTT, CHX) or nicalin overexpression (nicalin-YFP) and untreated (no DTT, no CHX) or YFP overexpression (YFP). Results represent at least 3 independent experiments. P-value is as determined by T-test. DTT: dithiothreitol. CHX: cycloheximide.
Fig. 4.5: Nicalin interacts with signal peptide peptidase, found in proximity to the translocon

a) Nicalin interacts with signal peptide peptidase (SPP) as deduced by co-IP and western blot with anti-SPP antibody. Lanes 1-5: FLAG-tagged vector (V), Stexin, nicalin, and sterubin were immunoprecipitated with FLAG and western blotted for SPP. Right lane (lane 6): SPP western blot of 1/20 of the protein used for the IPs. b) Localization of signal peptide peptidase in the proteomics resource. c) Western blotting shows that nicalin and signal peptide peptidase show a similar enrichment in rough ER. Rat liver ER fractions were Western blotted against nicalin, SPP, and the cargo protein transferrin as indicated. d) The interaction between nicalin and SPP is not due to a specific cleavage event. 293T cells expressing YFP or the first 51 amino acids of nicalin tagged to YFP were treated with DMSO vehicle or an SPP inhibitor (ZLL ketone) for 3 hrs. prior to 15 min. pulse with 35S methionine and processing for YFP immunoprecipitation. While the level of both the uncleaved and cleaved forms of the nicalin signal sequence are increased by ZLL ketone, their relative ratio is unchanged.
Fig. 4.6: The signal sequence of nicalin is sufficient to diminish tyrosinase levels

a) The N terminal residues of nicalin are sufficient to diminish tyrosinase levels. 293T cells were transfected with deletion mutants of nicalin tagged to YFP, with the remaining N-terminal residues indicated. Top, Western blotting for tyrosinase shows that all C-terminal truncations of nicalin efficiently reduce tyrosinase expression. Bottom, expression of the nicalin deletion mutants tagged to YFP. The arrow shows that the migration of residues 1-51 of nicalin tagged to YFP migrates at the same molecular weight as the YFP vector alone.

b) The amino acid sequence of nicalin residues 1-51. Red indicates the two N-terminal methionines. Blue indicates the predicted transmembrane domain if the signal sequence is a signal anchor. Green indicates the amino acids left in the 1-51 construct if the signal sequence is cleaved.

c) Further N-terminal truncations of nicalin are sufficient to reduce tyrosinase levels. Top, Western blotting of tyrosinase after expression of yet shorter N-terminal nicalin truncations. Bottom, Expression of the N-terminal truncations. Western blotting for YFP reveals that
the construct with residues 1-20 is partially cleaved while the longer constructs run at slightly lower molecular weights, indicative of cleavage. Many show faint expression, possibly due to degradation. d) Cleavage of nicalin 1 - 41 occurs slowly. 293T cells were pulse labelled for 15 min. with 35-S methionine and chased as indicated.

Materials and Methods
Reagents
Tunicamycin, dithiothreitol, cycloheximide, azetidine-2-carboxylic acid, and PNGase F were purchased from Sigma. ZLL2 ketone was from Calbiochem. Anti-signal peptide peptidase was purchased from Abcam, while nicalin antibody was produced in-house by rabbit immunization with purified peptides, followed by affinity purification of the crude serum.

Deletion Mutants
For nicalin deletion mutants, the original full length nicalin-YFP was used as a template for PCR amplification of sequences corresponding to the sequence of interest (i.e. amino acids 1 – 51). The primers used were engineered to include HindIII and BamHI sites at each end, which after PCR were digested and subcloned into a monomeric YFP vector to generate the tagged deletion mutants. All constructs were validated by sequencing (McGill Core Sequencing Centre).

PNGase F digestion assay
Purified rat liver microsomes were treated with PNGase F (Sigma) or vehicle overnight at 37 OC, followed by processing for SDS-PAGE and Western blotting as described in Chapter 2, using the antibodies indicated.

Sequential immunoprecipitation
Pulse-chase experiments were performed in BHK cells and are as described in Chapter 2. For sequential immunoprecipitations, radiolabelled lysates were precleared with protein A sepharose beads for 30 min. at 4 OC, followed by incubation with primary antibody (anti-FLAG) for 30 min., then addition of protein A sepharose beads for 2 h while rotating at 4 OC. Lysates and beads were then washed 3 x 5 min. in TBS. Subsequently, SDS was added to 1% final
concentration and heated for 10 min. at 95°C to separate bound FLAG protein, antibody, and beads. Samples were cooled on ice and quenched with a 1% final volume of Triton X-100. The antibody/beads were removed by centrifugation at 16,000g for 10 min. The lysates were pre-cleared again at room temperature with protein A sepharose, followed by an overnight incubation with anti-His (tyrosinase) antibody. The following day, lysates were washed 3 x 5 min. in room temperature TBS and processed for SDS-PAGE as described, followed by visualization using autoradiography.

**Pre-emptive quality control**

For experiments optimized to observe pre-emptive quality control, the conditions for pulse-chase were modified. Transfected 293T cells were incubated with drug or vehicle as indicated for 30 min. in methionine-free medium prior to a 10 min. pulse, and were harvested immediately after the pulse. Samples were processed for SDS-PAGE and autoradiography as described in Chapter 2. Films were quantitated using the BioRad image quantitation system and analyzed statistically using Microsoft Excel.

**Other**

YFP expression and immunofluorescence are as described in Chapter 2. FLAG expression and co-immunoprecipitation with Western blotting are also described in Chapter 2. Endpoint RT-PCR is as described in Chapter 2, using nicalin-specific primers. Pulse-chase immunoprecipitation and visualization are as described in Chapter 2.
CHAPTER 5: Stexin, erlin-1 and TMX2 interact with cargo in a temporal order predicted by the proteomics resource

Abstract

In Chapter 2, stexin, erlin-1, and TMX2 were found to interact with the immature form of the cargo protein tyrosinase. In the proteomics resource, stexin is found in the membrane of the rough ER, similar to translocation-associated proteins. Erlin-1 is found in the membrane of smooth ER, similar to degradation-associated proteins. TMX2 is of low abundance and cannot be assessed for localization, however, the presence of an atypical thioredoxin domain required for tyrosinase interaction suggests a protein folding function. In Chapter 1, quality control proteins found in proximity to the rough ER interact with cargo earlier than proteins found in the smooth ER. Here, the N-glycosylation system in the ER is also found to show such a distribution, with earlier modifications occurring in proximity to rough ER, with later modifications more enriched in smooth ER. Based on these findings, proteomics predicts that cargo will interact first with stexin, followed by TMX2 and then erlin-1. The trimming of the N-glycans on tyrosinase is used to follow the temporal order of stexin, TMX2, and erlin-1 in their interaction with cargo. Here, stexin is shown to interact with an earlier N-glycosylated form of tyrosinase compared to erlin-1 and TMX2. Hence, the spatial and temporal ordering of cargo processing in the proteomics resource is found to predict the specific step at which unknown proteins act in the secretory pathway.

Introduction

Newly synthesized glycoprotein cargo are synthesized on ribosomes at the rough ER membrane and imported into the lumen of the ER. As the nascent chain of the newly synthesized protein exits the Sec61 import channel, N-glycosylation at consensus NXS/T sites is conferred by the oligosaccharyltransferase complex (Ruddock and Molinari, 2006). As seen in Chapter 1 and in agreement with
previous studies (Chavan and Lennarz, 2006), the oligosaccharyltransferase complex is tightly associated with the Sec61 channel and is in the membrane of the rough ER. From there, the terminal glucose residues are rapidly cleaved by glucosidase I to yield a monoglycosylated N-glycan structure. The lectin chaperone calnexin binds to this species of glycan, until this final glucose is cleaved by glucosidase II. The protein folding sensor UGGT can add the glucose back to the N-glycosylated cargo protein, allowing it to be bound again by calnexin. Cargo leaves the calnexin cycle if it is productively folded, or if it is targeted for degradation. Thus, a temporal sequence of N-linked glycan processing occurs in the ER and it is intimately linked to cargo processing.

Newly synthesized disulfide-bonded cargo are folded by thioredoxin family members in the ER (Ellgaard and Ruddock, 2005). The thioredoxin family members oxidize disulfide bonds in misfolded cargo, offering the opportunity to refold. Redox balance is maintained in the ER by Ero1, which oxidizes the thioredoxin family members so that they may be revitalized for further disulfide bond oxidation of cargo (Tu and Weissman, 2004). Disulfide bond reformation does not appear to follow any temporal sequence and occurs diffusely throughout the rough and smooth ER.

Here, the specific functions of stexin, TMX2, and erlin-1 in cargo processing are predicted by their localization in the proteomics resource. The oligosaccharyltransferase complex, glucosidase I, and stexin are found in the rough ER. UGGT and thioredoxin family members are found in both the rough and smooth ER. Erlin-1 and glucosidase II are found predominantly in the smooth ER. The order of glycoprotein processing is therefore captured by the proteomics resource and predicts the order of cargo interaction by stexin, TMX2 and erlin-1. In agreement with this prediction, stexin is found to interact with an early N-glycosylated form of tyrosinase that precedes glucosidase II cleavage, while TMX2 and erlin-1 interact with later forms of tyrosinase.
Results

In chapter 2, stexin, erlin-1, and TMX2 were found to be candidates for functioning in cargo processing because they interact with the immature form of the cargo protein tyrosinase. To predict their specific functioning in cargo processing, these proteins were localized in the proteomics resource. Stexin is found predominantly in rough ER and associated with the detergent phase, indicating a membrane protein (Figure 5.1a and b). TMX2 could not be localized by proteomics due to its low abundance, but thioredoxin family members in the ER are found in the aqueous phase of the rough and smooth ER as they are mostly luminal proteins (Figure 5.1a and b). Erlin-1 is found predominantly in the detergent phase of the smooth ER (Figure 5.1a and b). In addition, YFP expression of stexin, erlin-1 and two forms of TMX2 showed ER localizations (Figure 5.1c).

In Chapter 2, TMX2 was found to have amino acid sequence similarity to thioredoxin family members (Figure 2.1). The active site of thioredoxin domains contains a canonical CXXC motif, with the first C involved in creating intermolecular disulfide bonds between protein disulfide isomerase (PDI) and cargo (mixed disulfides), and the second C causing a break in the mixed disulfide bond to allow cargo release (Ellgaard and Ruddock, 2005). TMX2 contains an atypical SXXC motif at the conserved site, which is biochemical unlikely to have function as a PDI protein. To determine if TMX2 functions in protein folding, the contribution of the thioredoxin domain to cargo interaction was assessed.

The form of TMX2 found to interact with tyrosinase utilized a 258 amino acid sequence cloned from human liver cDNA and that exists abundantly in EST databases. However, it was noted that TMX2 is also observed in NCBI to exist as a 296 amino acid form. This form was cloned from HepG2 cell mRNAs and is designated here as TMX2L (long), with the 258 amino acid form as TMX2s (short). These two mutants differ by amino acid sequences containing a predicted
transmembrane domain, corresponding to alternate splicing of a single exon (Figure 5.2a). This transmembrane domain greatly changes the prediction for the topology of the protein. TMX2s is predicted to be entirely luminal and its thioredoxin domain is therefore in the lumen of the ER. TMX2L on the other hand, is predicted to span the membrane, placing the thioredoxin domain and the KKXX retention motif in the cytosol (Figure 5.2b). Despite these differences, YFP chimeras of both forms are found to localize to the ER (Figure 5.1c) and both forms interact with the immature form of tyrosinase (Figure 5.2c).

To determine if the thioredoxin domain is responsible for TMX2 interaction with cargo, FLAG-tagged deletion mutants of TMX2L were assessed for interaction with tyrosinase by co-immunoprecipitation. TMX2L constructs deleted for the thioredoxin domain did not interact with tyrosinase (Figure 5.2c).

Thioredoxin family members form mixed disulfides with cargo through their CXXC motifs. To determine if the atypical SXXC thioredoxin domain of TMX2 forms mixed disulfides with cargo, the interaction between TMX2s and tyrosinase was monitored under non-reducing conditions in order to trap mixed disulfides. Tyrosinase is a highly disulfide-bonded substrate, with 8 possible pairs. Cells co-expressing tyrosinase and TMX2 were lysed in the presence of N-ethylmaleimide (NEM) to freeze disulfide bonds. If TMX2 forms a mixed disulfide with tyrosinase, the following qualities are expected. First, Western blot of tyrosinase or TMX2 under non-reducing conditions should show higher molecular weight bands compared to cells expressing either alone. These bands would represent any mixed disulfides between TMX2 and tyrosinase. This did not occur. Tyrosinase formed intermolecular disulfide bonds with itself, as shown by the higher molecular weight forms of tyrosinase seen under non-reducing versus reducing conditions (Figure 5.3a lane 3 vs. 7); however, no additional bands were seen when TMX2s was co-expressed (Figure 5.3a lane 7 vs. 8). Similarly, Western blotting against TMX2 did not trap higher molecular weight bands when tyrosinase was co-expressed (Figure 5.3b lane 4 vs. 8). Second,
immunoprecipitation of TMX2 followed by Western blotting for tyrosinase to detect interactions should show bands at a higher molecular weight, again corresponding to the mixed disulfide. Based on the SDS-PAGE migration of tyrosinase and TMX2s, a 1:1 mixed disulfide should run between 100 and 150 kDa. No such bands were observed on co-immunoprecipitation (Figure 5.3a and b lanes 6; the band seen in 5.3b lane 6 is also observed in lane 8 and corresponds to TMX2s). Together these results suggest that, although TMX2 interacts strongly with tyrosinase, it does not do so through the formation of mixed disulfide bonding.

An attempt was made to determine if the atypical SNDC motif found in the thioredoxin domain of TMX2 forms mixed disulfides weakly, as might be suggested by the substitution of the classical cysteine found at the first position with an atypical serine. If TMX2 forms mixed disulfides with substrates, the mutation of SNDC to SNDA should prevent the mixed disulfide from disassociation by prohibiting the release of cargo. Figure 5.3c shows that cells expressing wild type and mutant TMX2 display identical behaviour after Western blot under non-reducing conditions; that is, no mixed disulfides are observed. Furthermore, this mutant does not have a dominant negative effect that blocks the maturation or degradation of tyrosinase, indicating that tyrosinase is not being trapped on the SNDA mutant form of TMX2.

Next, the time-course of the interaction between tyrosinase and TMX2, stexin, or erlin-1 was assessed. FLAG-tagged constructs of TMX2, stexin, and erlin-1 were co-expressed with tyrosinase in BHK cells and subjected to a pulse-chase and co-immunoprecipitation protocol to isolate the pool of tyrosinase interacting with the FLAG-tagged protein. All 3 proteins interacted with tyrosinase transiently (Figure 5.4). When the proline analogue azetidine-2-carboxylic acid (Azc) was added for 60 min. prior to 15 min. pulse and washed out in the chase (total 75 min. of Azc), the time course of interaction was extended, suggesting chaperone-like characteristics for these proteins. Stexin was noted to interact with two forms of
tyrosinase. First, a higher molecular weight (slow mobility) form of tyrosinase was observed to interact with stexin immediately post-pulse, followed by a lower molecular weight (faster mobility) form during the chase (Figure 5.4). The higher molecular weight form was not observed in association with erlin-1, nor with TMX2.

The higher molecular weight form of tyrosinase interacting with stexin may represent a specific glycoform of tyrosinase. Since stexin was found to be associated with the rough ER in the proteomics resource, it suggests that this form of tyrosinase is an early glycoform relative to those interacting with TMX2 and erlin-1. N-glycosylation is processed in a temporal sequence of steps within the ER. To determine if there is a spatial relationship between N-glycosylation enzymes, their localization in the proteomics resource was assessed (Figure 5.5a). Like synthesis, folding, and degradation, N-glycosylation is spatially organized in the proteomics resource and is directed from rough to smooth ER. Figure 5.5a shows that the sequence of steps in N-glycosylation, from transfer of the oligosaccharide chain to glucose trimming, are spatially organized and proceed from the rough towards the smooth ER. Notably, there is a division whereby the oligosaccharyltransferase complex and glucosidase I are mainly associated with rough ER, while glucosidase II and UGGT have localization in both rough and smooth ER. The proteomics localization of stexin is similar to that of the oligosaccharyltransferase complex and glucosidase I, and is closer to the translocon than glucosidase II (Figure 5.1a versus 5.5a). Based upon this pattern, it is predicted that stexin interacts with cargo prior to glucosidase II trimming of N-linked glycans, while TMX2 and erlin-1 interact after this trimming event has occurred.

The form of tyrosinase interacting with stexin (Figure 5.4) has the three bands corresponding to the 7, 6, and 5 sites of occupied glycosylation, therefore it is interacting after the addition of N-glycans by the oligosaccharyltransferase complex. As shown, glucosidases I and II trim the number of glucose residues on
each N-glycan tree. The hypothesis is that the slightly higher molecular weight form of tyrosinase interacting with stexin has a larger number of glucoses on the 5-7 N-glycan trees of tyrosinase compared to the forms interacting with erlin-1 or TMX2. Strictly speaking, the prediction based on proteomics is that the post-pulse stexin-interacting form contains 2 terminal glucoses on each N-glycan (prior to glucosidase II cleavage), while the TMX2 and erlin-1 interacting forms contain 1 or 0 terminal glucoses (after glucosidase II cleavage).

To test this hypothesis several enzymatic modifiers of the N-glycan tree were used to digest the post-pulse interacting pools of tyrosinase. These are diagrammed in Figure 5.5b. PNGase F is used in vitro to entirely cleave N-glycans from the polypeptide backbone (Hirsch et al., 2003). Castanospermine is an in vivo inhibitor of glucosidase I, with the result that all N-glycans are kept in the triglucosylated state (Helenius and Aebi, 2001). Jack bean α-mannosidase (JBαm) differentiates between N-glycan trees with 1-3 terminal glucoses from those that have 0 terminal glucoses (Hammond et al., 1994). With JBαm, N-glycans with terminal glucoses are relatively protected from cleavage, while N-glycan chains that do not have terminal glucoses are fully cleaved.

The composition of the N-glycan chain of the low mobility form of tyrosinase interacting with stexin was assessed using these reagents (Figure 5.5c). BHK cells co-transfected with tyrosinase and FLAG-tagged stexin, erlin-1 or TMX2 were pulsed for 10 min. then processed for co-immunoprecipitation. The higher molecular weight form of tyrosinase in association with stexin was abolished by castanospermine treatment and PNGase F but not by jack bean mannosidase (Fig. 5.5c). Since PNGase F removes all N-glycan trees from proteins, the abolishment of the molecular weight difference in tyrosinase means that this difference is due to a variation in the N-glycan tree. Since castanospermine forces N-glycan trees to have three terminal glucoses, the abolishment of the molecular weight difference by castanospermine means that the modification causing the molecular weight difference is downstream of cleavage of the 3rd terminal glucose. That is,
restoring the three terminal glucoses with castanospermine abolishes the difference and the higher molecular weight form of tyrosinase interacting with stexin is downstream of glucosidase I. The failure of JB mannosidase to affect the molecular weight difference means that the N-glycan tree on the interacting tyrosinase contains at least 1 terminal glucose at the post-pulse time point. That is, the difference is upstream of the action of glucosidase II to completely remove terminal glucoses.

The effects of jack bean mannosidase are further shown in Figure 5.5d. Here, JB-mannosidase cleavage causes the stexin, erlin-1, and TMX2 interacting forms of tyrosinase to migrate at a lower molecular weight on SDS-PAGE (Figure 5.5d; compare lanes 1 – 4 and 9 to lanes 5 – 8). However, as seen in Figure 5.5c, JB-mannosidase does not eliminate the molecular weight differences between tyrosinase interacting with stexin compared to that interacting with erlin-1 and TMX2 (Figure 5.5d; compare lane 6 to lanes 7 and 8). Together these results show that the post-pulse stexin-interacting form of tyrosinase contains more terminal glucoses than the post-pulse TMX2- and erlin-1-interacting forms. Since the removal of the terminal glucoses is a proximal step mediated by glucosidases, it is concluded that stexin interacts with an earlier intermediate of tyrosinase compared to erlin-1 and TMX2, one that appears to occur between glucosidase I and glucosidase II action.

Discussion

The spatial organization of the ER suggests a directionality of cargo processing from synthesis to degradation, proceeding from the rough ER towards to the smooth ER (Chapter 1). N-glycans serve as signals for protein folding and degradation and proceed in a series of sequential steps in the ER. Here it is shown that the order of N-glycan processing is also spatially segregated and proceeds from the rough ER towards the smooth ER. This unifies the spatial and temporal aspects of cargo processing; the ordered sequence of N-glycan processing is segregated in space. Since N-glycans represent signals for quality control
decision-making, it is satisfying that the enzymes which add the N-glycan tree and do early trimming are restricted to the rough ER, while the enzymes signalling folding and degradation are found also in the smooth ER.

That glycan processing converges with cargo processing is supported by the proteomics localization of 3 unknown proteins which interact with various glycoforms of tyrosinase. Stexin is highly enriched in the rough ER and has a membrane distribution suggesting a tight association with the translocon. N-glycosylation enzymes with this distribution include the oligosaccharyltransferase complex and glucosidase I. TMX2 could not be localized, however it was found to have chaperone-like characteristics, although it does not appear to act as a classical protein disulfide isomerase enzyme. Erlin-1 is localized largely in the membrane of the smooth ER. Later N-glycosylation enzymes include glucosidase II and UGGT; these proteins are found with a relatively higher distribution in smooth ER. The form of tyrosinase interacting with stexin contained a higher number of terminal glucoses (i.e. an earlier form in glucosidase trimming) compared to TMX2 and erlin-1. Thus, the stexin interacts first with tyrosinase (in both space and time) while TMX2 and erlin-1 interact later. This organization is as predicted by localization in the proteomics resource.

Stexin is a previously uncharacterized predicted transmembrane protein. It localizes to the detergent phase of the rough ER in the proteomics resource, similar to the oligosaccharyltransferase complex and translocon-associated proteins. By YFP expression, it localizes to the ER. In Chapter 2 it was found to be upregulated at the mRNA level by protein misfolding stress and to interact with the immature (ER) form of the secretory cargo protein tyrosinase. Stexin was found to interact with a glycoform of tyrosinase found early in cargo maturation. Hence, the spatial and temporal action of stexin is in proximity to the translocon shortly after cargo entry into the ER. Although stexin interacts with an early glycoform of tyrosinase it does not appear to itself be a glycan-modifying enzyme: at later chase points stexin was found in association with the glucose
trimmed form of tyrosinase (Figure 5.4) and overexpression of stexin did not generate the slow mobility form of tyrosinase in the total tyrosinase pool (data not shown). A clue to its function is the chaperone-like characteristics displayed by stexin, where stexin interacts with immature tyrosinase for extended time when Azc was added to induce protein misfolding. The determinants for stexin recognition of cargo and how stexin is maintained in proximity to the translocon (possibly through protein interactions) merit further investigation. Confirmation of the interaction between stexin and N-glycans of the type shown here has been shown in the course of this work; by NMR, malectin (the *Xenopus* stexin orthologue) is found to interact specifically with Glc2 N-glycosylated proteins (Schallus et al., 2008).

TMX2 is an atypical member of the protein disulfide isomerase family (Ellgaard and Ruddock, 2005) that contains a single thioredoxin domain with a non-classical SXXC motif. Although TMX2 associated transiently and strongly with only the immature ER form of newly synthesized tyrosinase, no evidence for mixed disulfide interactions with tyrosinase could be found. Proteomics was unable to distinguish whether the long or short form of TMX2 was characterized in our isolated ER fractions. The EST database suggests that both forms are expressed in human and mouse. Hence, both forms were tested and both were found to be ER-associated (YFP colocalization with calnexin) and both were found to associate with the immature ER form of tyrosinase. Mutational analysis indicated that the predicted cytosolic exposed C-terminal region of TMX2 (containing the non-classical SXXC thioredoxin motif) is needed for its cargo binding. TMX2 interacted with a later glycoform of tyrosinase compared to stexin, indicating a more distal function in protein folding. The significance of the splice variants encoding a luminal and a cytosolic protein, its importance given its low abundance in rat liver, and its ability to interact with tyrosinase in spite of its inactive thioredoxin domain are all open questions.
Erlin-1 is shown here to interact with chaperone-like characteristics to a later glycoform of tyrosinase compared to stexin. It is upregulated at the mRNA level by protein misfolding stress, and it resides in the smooth ER alongside of ERAD constituents. Erlin-1 has 89% amino acid identity to another recently identified protein, named SPFH or erlin-2. Erlin-1 and 2 were named on the basis of a putative location in lipid rafts of the ER (Browman et al., 2006). Recently, SPFH was found to be a constituent of ERAD (Pearce et al., 2007). This is consistent with its proteomics localization, which places both erlin-1 and erlin-2 in the membrane of the smooth ER, along with aspects of the ERAD machinery. Based upon this amino acid sequence similarity as well as the confluence of the proteomics data with the work of Pearce et al., erlin-1 is also predicted to function in ERAD of unstable cargo. How erlins interact with cargo and specifically function in ERAD are further mechanistic questions.

In summary, this chapter shows that the spatial organization of the proteomics resource has a correlate in the temporal ordering of cargo maturation. This is supported by the finding that stexin, which interacts with tyrosinase glycoforms indicative of early maturation, is a rough ER protein. TMX2 and erlin-1 interact with later glycoforms of tyrosinase and are found also in the smooth ER. The function of unknown proteins such as stexin, TMX2 and erlin-1 can thereby be predicted from the spatiotemporal localization designated by the proteomics resource. All 3 proteins have chaperone-like qualities: stexin interacts with early cargo intermediates near the translocon, while erlin-1 is likely involved in ERAD in proximity to the smooth ER. TMX2 has chaperone-like characteristics in spite of its non-classical thioredoxin domain that does not form mixed disulfides with cargo. Thus, the spatial segregation of the ER explains how N-glycan processing is ordered in time and can be used to gain insight into the function of individual ER resident proteins.
Fig. 5.1: Localization of stexin, erlin-1, and TMX2

a) Localization of stexin, erlin-1, and thioredoxin family members in the total fractions of the proteomics resource. Since TMX2 was of too low abundance to be localized on its own, an average of the localizations of known thioredoxin family members identified is shown. RM, rough microsomes; SM, smooth microsomes; G, Golgi fractions.

b) Localization of stexin, erlin-1, and thioredoxin family members in the biochemical fractions of the proteomics resource. SLT, salt wash; INS, insoluble in Triton X-114; DET, soluble in Triton X-114; AQU, aqueous phase of Triton X-114.

c) Localization of YFP chimeras of stexin, erlin-1, and two forms of TMX2 by expression of YFP chimeras in HeLa cells. Left panels: Localization of the indicated YFP chimeras. Right panels: Co-localization of the indicated YFP chimeras (green) with immunofluorescence against the ER marker calnexin (red). Scale bar: 10 microns.
Fig. 5.2: The thioredoxin domain of TMX2 is required for cargo binding

a) Amino acid sequence comparison of the two forms of TMX2 described in this thesis. Underlined is the predicted transmembrane domain present in TMX2L (long) but lacking in TMX2s (short).

b) Schematic representation of the two forms. Green, predicted signal sequence. Blue, predicted transmembrane domain. Yellow, thioredoxin domain. The length in amino acids is indicated by the numbering underneath.

c) Interaction between deletion mutants of TMX2 and tyrosinase. Left, TMX2 mutants comprising the amino acids indicated were expressed with tyrosinase in BHK cells and tested for co-immunoprecipitation. Right, expression of the mutant constructs indicated. FL, full length. The decreased interaction between TMX2L 1 - 259 aa compared to the full length (compare lanes 4 and 5 on the left) is likely due to the decreased expression of this construct (lanes 4 vs. 5 on the right).
Fig. 5.3: TMX2 does not form mixed disulfides with substrate

a), b) Association of TMX2 with tyrosinase under reducing and non-reducing conditions. BHK cells co-expressing tyrosinase and TMX2 or vector were processed under conditions to preserve disulfide bonds (see materials and methods). Immunoprecipitating and Western blotting antibodies are indicated. In particular, panel (a) lane 6 shows that TMX2 interacts with disulfide bonded forms of tyrosinase, but no band is present indicative of a mixed disulfide interaction (i.e. seen at the appropriate MW in lane 6 but not lane 7). Asterisks indicate intermolecular tyrosinase disulfide bonding. c) Mutation of the TMX2 SNDC motif to SNDA does not trap substrate proteins. The indicated TMX2 forms were expressed in 293T cells and were processed under non-reducing conditions prior to anti-FLAG Western blotting.
Fig. 5.4: Stexin, erlin-1 and TMX2 show characteristics of chaperones
BHK cells were co-transfected with tyrosinase and the FLAG-tagged protein indicated (stexin, erlin-1 or TMX2s) for 48 h, pulsed for 15 min. with radiolabel and chased for the number of hours indicated. A sequential IP protocol (as diagrammed) was used to isolate the pool of tyrosinase interacting with stexin, erlin-1 or TMX2s. Azc indicates the inclusion of the misfolding agent azetidine-2-carboxylic acid for 75 min. (1 h prior to pulse + 15 min. pulse) followed by washout for the chase. Quantitation of the signal seen on the gels is seen in the line graphs on the right, expressed as a % of maximal signal.

Note the increased duration of binding when Azc is added. “Slow” indicates a slow mobility form of tyrosinase noticed in the stexin interacting pool at 0 min. chase. “Fast” indicates the tyrosinase mobility normally observed. Where the FLAG-tagged protein was seen to tightly associate with tyrosinase, an asterisk (*) is placed.
Fig. 5.5: Stexin interacts with tyrosinase early in its folding

a) Localization of the enzymes involved in N-linked glycosylation and trimming in the ER. Their localization is consistent with their known timing of N-glycan processing. The effect of each enzyme on the N-glycan tree is shown in a schematic below. The oligosaccharyltransferase (OST) complex adds the entire N-glycan tree to the NXST consensus site of glycoproteins. The tree figures have been modified from Ruddock and Molinari J. Cell Sci. 2006.

b) Schematic representation of the in vitro effects of Jack Bean alpha-mannosidase (JBαM) and peptide N-glycanase F (PNGaseF) on glycan trees containing (orange triangles) or lacking terminal glucosylation.

c) The Glc1 inhibitor castanospermine as well as PNGaseF abolish the molecular weight difference between tyrosinase bound to stexin and tyrosinase bound to TMX2 or erlin-1. Treatment with JBαM cleaves tyrosinase, but the molecular weight difference is preserved.
Materials and Methods

N-glycan experiments
Castanospermine (Sigma) was added to the medium of 293T cells 1 h. prior to 10 min. pulse labelling, prior to sequential immunoprecipitation and SDS-PAGE, as described in Chapters 2 and 4. For jack bean α-mannosidase and PNGaseF, 293T cells were pulse labelled without drug for 10 min. followed by pulse-labelling and sequential immunoprecipitation as described in Chapter 4. Prior to SDS-PAGE, cellular lysates were treated overnight at 37°C with vehicle, JB mannosidase, or PNGaseF as indicated.

Deletion and point mutant constructs
For TMX2L deletion mutants, the sequences desired (i.e. amino acids 1 – 109) were amplified from the full length TMX2 target vector using primers with HindIII and BamHI sites engineered into each end. The fragments were restriction digested and subsequently ligated into p3xFLAG-CMV14 (Sigma) to generate the FLAG-tagged deletion mutants. For the SNDC to SNDA point mutation construct, the mutation was introduced using the QuickChange protocol (Stratagene) using the sets of complimentary oligonucleotides (5'-TTGCCAATTTGCTATTGAAGCCCGAATCATTTG-3' and 5'-GTCATCTGGTTAAGCCCAAGATACTCCAC -3'). All chimeras and mutants were verified by sequencing (McGill Core Sequencing Centre).

Experiments under non-reducing conditions
For experiments examining TMX2 under reducing and non-reducing conditions, N-ethylmaleimide (NEM; Sigma) was added to the wash and lysis buffers at a concentration of 2.5 mg/ml. For SDS-PAGE samples were boiled in Laemmli buffer with β- mercaptoethanol not included. Instead, samples were boiled with (reducing) or without (non-reducing) 5mM dithiothreitol (Sigma). Otherwise co-immunoprecipitation, Western blotting, and pulse-chase were performed as described in Chapters 2 and 4.
Other
Transfection, YFP visualization, immunofluorescence, co-immunoprecipitation, pulse-chase, sequential immunoprecipitation, and localization by proteomics were as described in Chapters 1 – 4.
DISCUSSION

Organellar proteomics to recapitulate the spatiotemporal organization of the secretory pathway

The proteomics resource used in this thesis (Gilchrist et al., 2006) describes the mapping of subcellular fractions enriched for the endoplasmic reticulum (ER) and Golgi apparatus. The ER was separated into fractions representing rough and smooth ER. In independent experiments the rough and smooth ER fractions were further separated into biochemical fractions. These biochemical fractions separated a salt wash fraction enriched for cytosolic-associated proteins of the ER, a Triton X-114 detergent fraction enriched for integral membrane proteins, and an aqueous fraction enriched for ER luminal proteins with some cytosolic proteins tightly associated with the ER membrane (i.e. resistant to salt wash).

The functional difference between the rough ER and the smooth ER is demonstrated by cell types that have disproportionate amounts of each. Electron micrographs of dedicated secretory cells such as plasma cells or pancreatic acinar cells show highly developed rough ER membranes (Federovitch et al., 2005). Smooth ER is found most abundantly in hepatocytes, which contain abundant levels of detoxification and modification enzymes. Upregulation of the detoxification systems by compounds such as barbiturates yield gross proliferation of smooth ER membranes (Federovitch et al., 2005).

This thesis starts with identifying in the proteomics resource known resident proteins of the ER that function to process cargo. This includes mRNA metabolism, protein synthesis, import into the ER (translocation), folding, and proteasomal degradation (ERAD). Remarkably, these functional categories showed distinct localizations in the proteomics resource. However, instead of relying upon a marker or enzymatic assay to define these profiles, close to all of the proteins known to function in a category could be interrogated. In total, the segregation of the cargo processing pathway was based on the individual
localization profiles of over 100 known proteins. Moreover, the process used to
define the distinct functional categories is robust; the profiles were found after
cell fractionation, proteomics mapping, and unbiased selection of proteins known
to function in cargo processing. This suggests a dedicated subdomain-function
relationship to the ER whereby spatial localization defines function.

In the proteomics resource, RNA metabolism and protein translation were
localized to the cytoplasmic face of the rough ER, as expected. Translocation and
the associated oligosaccharyltransferase complex localized to the membrane of
rough ER as expected. Protein folding enzymes resident to the ER were localized
to the rough and smooth ER equally, a supportable finding given the free
diffusion in the ER of some marker chaperones (Nikonov et al., 2002; Snapp et
al., 2006). Unexpectedly, cytoplasmic folding enzymes associated with the
smooth ER and constituents of ER-associated degradation (ERAD) and the
proteasome were also found restricted to the smooth ER. Previous localizations of
the proteasome by Western blotting of rat liver fractions also show enrichment in
the smooth ER versus rough ER (Palmer et al., 1996). This suggests a model for
the ER whereby newly synthesized cargo proceed from the rough ER to the
smooth ER through subdomains of spatially restricted ER resident proteins. These
subdomains have specific functions and this organizes the cargo processing
pathway. The organization is in space as well as time, since the enzymes that
process glycoprotein cargo in a determined temporal order (from the
oligosaccharyltransferase to glucosidase trimming) were found to also be spatially
restricted, proceeding from rough to smooth ER.

The spatial organization of the ER provides a foundation for the sublocalization-
function relationship of the ER. It appears that the ER is similar to the Golgi
apparatus, which restricts the resident enzymes of cargo processing in a
spatiotemporal manner proceeding from early processing at the cis Golgi to late
processing at the trans Golgi. This view of the ER is consistent with the
aforementioned diffusion studies of ER resident proteins which find translocons
to be immobile and chaperones to diffuse freely. It is also consistent with previous reports localizing proteasomes specifically to sublocations of the ER [reviewed in (Wojcik and DeMartino, 2003)]. Combined with measures of protein abundance, the spatial organization of the ER is also consistent with explanations of ER function based on protein energetics (Wiseman et al., 2007a; Wiseman et al., 2007b). Such studies suggest a competition between secretory exit, ERAF and ERAD, which we find here to be spatially segregated, with ERAD at the lowest abundance. That is, the conceptual divisions imposed by protein energetics studies have a physical correlate uncovered by the work of this thesis. Altogether, the total conceptual/functional organization of the ER preceding this work has been found by this work to be mirrored in physical space.

**Strategies to characterize unknown resident ER proteins**

The view of the ER as being organized into spatial subdomains corresponding to function is productive. The function of uncharacterized proteins can be predicted on the basis of their subdomain localization. In this thesis, examples from 5 poorly characterized proteins are used to demonstrate the agreement between the proteomics prediction of function and that validated by biochemical, genetic and structural investigation. A current goal of cell biology is to annotate the function of uncharacterized proteins and this work suggests that organellar proteomics is a productive method in this regard. As a comparator, a brief review of other approaches to the annotation of uncharacterized proteins follows.

**Functional genomics**

A functional genomics screen by Malhotra and colleagues (Bard et al., 2006) utilized RNAi in Drosophila S2 cells in order to identify cellular constituents that affect the secretory pathway. After the systematic knockdown of genes using greater than 22 000 individual dsRNAs, the authors identified 1133 different targets that impaired secretion of the model product horseradish peroxidase (HRP). A second screen validated 130 of these targets. Of these targets important
to protein secretion, 100 proteins of unknown function were identified and named TANGO1-100; YFP localization of a subset of these proteins identified 14 TANGO proteins that localized to the cytosol, ER and/or Golgi (TANGO1-14).

Of the 20 proteins localized by Bard et al. (including TANGO1-14), the organellar proteomics resource of Gilchrist et al. 2006 identified only one; this protein (Use1p) has a fairly well-defined function in the literature as an ER SNARE (Dilcher et al., 2003). Of the 14 TANGO proteins localized by YFP, 5 represent Drosophila-specific genes without reasonable mammalian homologues. In the initial set of 1133 genes identified in the first siRNA screen, 1003 (88%) were removed as genes known to function indirectly in the inhibition of HRP secretion (i.e. cell survival, cytoskeletal genes, etc.), rather than representing true secretory pathway constituents. Of their list of unknown proteins, one would expect that 90% of these would also mediate indirect effects on secretion, and may explain their low detection by proteomics of rat liver. From the portion of their work functioning in the annotation of unknown proteins, Bard et al. therefore find a total of 7 Drosophila homologues of proteins of unknown function that localize to the ER and/or Golgi and affect secretory transport and/or Golgi organization. None of these proteins were independently localized to be in the secretory pathway by our proteomics effort on rat liver.

A separate siRNA screening project in human cells is in progress (Simpson et al., 2007). In this approach, a GFP-tagged secretory cargo protein (VSV-G) is used to monitor trafficking following siRNA knockdown of an assortment of genes. While the full project is yet unpublished, it has been reported that 90% of the targets found to affect VSV-G trafficking were found to be in ER or Golgi by the proteomics resource of Gilchrist et al. 2006 (Donaldson and McPherson, in press). This illustrates the complementary value of such projects.

**Protein-protein interactions**
It is difficult to directly assess the usefulness of large scale protein-protein interaction studies on ER protein function. This is because assessments of the predictive power of these datasets specifically for ER proteins are lacking. There is reason to believe that such datasets are less effective for determining the function of ER proteins relative to proteins localized elsewhere in the cell. For instance, one method of determining protein interactions relies on yeast-2-hybrid screens, which is particularly poor for membrane and ER luminal proteins on account of the classical requirement for protein complex translocation to the nucleus to activate a transcription factor (Wang et al., 2004). Another method using TAP-tagging that has been recently published shows that ER proteins are more difficult to purify (60% vs. 90% of cytosolic proteins) for use in protein interaction screens (Krogan et al., 2006). That said, various works have used data from protein interaction screens to make inferences on ER proteins (Behnia et al., 2007; Scott et al., 2007). However, it is difficult to assess from the published literature the proportion of the time the protein interaction predictions are complete and accurate for ER proteins, versus those inaccurate or missed interactions that go unpublished.

**Epistatic miniarray profiling**

Another strategy, known as epistatic miniarray profiling (E-map), has recently been used to study the yeast early secretory pathway (Schuldiner et al., 2005). Here, the authors first defined a list of 424 essential and non-essential genes, including 160 of unknown function, that have been found to localize to the early secretory pathway in previous genome-wide GFP tagging projects in yeast (Huh et al., 2003). For the non-essential genes, deletion mutants were obtained, while for the essential genes, a strategy to make hypomorphic alleles through disruption of 3’UTRs was validated. This allowed for epistatic interactions among the entire set to be profiled. In this case, predicted growth rates for double mutants (based on the growth rate of each alone) were compared to the actual growth rate, and compared for epistatic interactions via hierarchical clustering. Next, this
information was used to predict physical complexes and pathways. An observation was made that two genes that function in related, but dissimilar processes often have negative epistatic interactions when combined. That is, the double mutants have more severe growth impairment than predicted from adding the impairments of the two individual mutations. However, two genes that function in the same or highly similar processes often have positive epistatic interactions, consistent with the concept that knockout of two members of a coherent protein machine should often be no worse than the effect of knocking out each individually (i.e. because they function so strongly together). With this in mind, several known complexes were able to be recapitulated through analysis of their epistatic interactions. With respect to unknown proteins, the function of 3 unknown proteins, identified as part of a Golgi-ER transport (GET) complex, was able to be predicted. Accordingly, deletion analysis suggested that the 3 members in the GET complex function to retrieve ER luminal proteins from the Golgi and return them to the ER. Follow up studies showed these proteins to indeed function as a complex, although to not to have a specific role in ER-Golgi transport. Rather, the GET proteins function to insert a specific type of secretory cargo, tail-anchored proteins, into the ER membrane (Schuldiner et al., 2008; Stefanovic and Hegde, 2007). Subsequent to this step, these proteins are trafficked to their final destination in the cell (Bulbarelli et al., 2002). Therefore, as a consequence of the E-map analysis of known and unknown proteins, epistatic information was able to be extracted, with predictive power towards the function of three proteins in a complex.

Characterization of proteins of unknown function in the endoplasmic reticulum

The characterization of unknown proteins by proteomics has several purposes in this thesis. First, as described above, it represents a test of the organellar proteomics strategy for proteome annotation. Based on the review conducted above this strategy appears to compare favourably to other high-throughput approaches for identifying and analyzing unknown proteins. Second, the
validation of the unknown proteins represents a test of the integrity of the resource. That is, interrogation of the proteomics resource suggests that the ER is composed of spatially segregated subdomains that link structure to specified functions. This model is productive in that it explains aspects of known ER function as well as makes predictions that unknown proteins should function as specified by their spatial localization. Here, a cytoplasmic ER-associated protein (nudilin) was found to function upstream of protein synthesis in cargo processing, as predicted by its proteomics localization. A rough ER membrane protein (nicalin) was found to be in a complex with signal peptide peptidase and to function in proximity to protein translocation, also as predicted. The temporal ordering of cargo interactions for 3 proteins with chaperone-like characteristics was determined, with the rough ER protein (stexin) interacting prior to the rough and smooth ER proteins (TMX2, erlin-1), as predicted.

**Nudilin as a P-body constituent that affects secretory cargo**

Nudilin was selected and named on the basis of its conserved nudix domain which suggested a role in RNA metabolism. Nudilin was localized in the proteomics resource to the cytoplasmic face of the rough ER, suggesting a function in RNA metabolism or translation. The structure of nudilin supports the function in RNA metabolism by featuring a prominent RNA binding groove and 3D similarity to RNA decapping proteins.

Previously, the *Xenopus* orthologue of nudilin, X29, was found to be a nucleolar protein that utilizes its nudix domain to decap nuclear RNAs such as U8 and U3 ribosomal RNAs as well as mRNA (Ghosh et al., 2004). Nudilin (called H29K) was also tested for *in vitro* decapping activity of U8 snoRNA and found to have positive activity, albeit with much lower efficiency than that of X29. Subsequently, the structure of X29 was solved (Scarsdale et al., 2006), and is found to be similar to that of nudilin shown here. Most recently, these investigators have found that the human nudilin has a much broader *in vitro*
specificity for RNAs (with increased specificity for mRNA) compared to X29, especially in the presence of certain metals (Peculis et al., 2007).

In this work, nudilin was localized by immunofluorescence to the P-body, as well as cytoplasmic and nuclear structures. In addition to the broader substrate specificity describing the human protein, it appears that nudilin also exhibits a broader localization within the cell compared to X29. Nudilin was also found at reliable abundance in the ER by proteomics. This is consistent with the suggested ability of proteomics to resolve localizations of some proteins better than that possible by fluorescence microscopy, since brighter structures can mask signals from adjacent less bright membranes (Bell et al., 2007a). This is exemplified by nudilin, which is identified in the ER by proteomics but not by fluorescence, and it is concluded that nudilin is partially an ER protein with a primary localization to the P-body.

Consistent with this proteomics prediction, manipulation of nudilin levels affected secretory cargo: overexpression of nudilin caused an increase in tyrosinase levels while siRNA knockdown caused a decrease in tyrosinase levels. That these effects occurred proximal to protein synthesis is consistent with the localization of nudilin to the cytosolic face of the rough ER by proteomics.

**Nicalin as a pre-emptive quality control constituent**

Nicalin was upregulated in response to thapsigargin and DTT but not tunicamycin. This difference may reflect its pathway of activation since certain protein misfolding sensors (i.e. Perk) are less sensitive to tunicamycin than the other misfolding agents (DuRose et al., 2006). Nicalin also interacts with the immature ER form of the cargo protein tyrosinase. These aspects suggested a cargo processing function for nicalin. A previous study found an effect of a nicalin orthologue on nodal signalling and axial patterning in zebrafish (Haffner et al., 2004). A fundamental function of nicalin in the ER is the likely basis of this particular developmental finding.
Nicalin is localized to the membrane of the rough ER in the proteomics resource. Cargo processing-related proteins that have this localization profile includes members of the ER import (translocation) machinery. That nicalin functions in proximity to the translocon is suggested by two main lines of evidence.

First, nicalin interacts with signal peptide peptidase (SPP). This was predicted based on the amino acid sequence similarity of nicalin with nicastrin, a member of the gamma-secretase complex. The association of nicalin with SPP is noteworthy, since alone of all the proteins we studied, this association was only found for nicalin. The function of SPP is to take the remnant transmembrane “stubs” left over after signal sequence cleavage and to cause their intramembrane proteolytic degradation (Weihofen et al., 2002). As SPP functions in proximity to the translocon, the association of nicalin with SPP predicts a similar localization.

Second, overexpression of nicalin diminished tyrosinase levels at a step proximal to cargo entry into the ER. Nicalin did not increase tyrosinase ERAD since the rate of degradation was unchanged. Nicalin also had no effect on the mRNA level of tyrosinase. Thus, the effect of nicalin was distal to RNA metabolism and proximal to protein degradation. Specifically, nicalin overexpression affected N-glycosylated tyrosinase, with the smaller amounts of non-ER translocated unglycosylated tyrosinase remaining largely unaffected. These are the properties expected for a constituent of the ER pre-emptive quality control system and a co-translocational inhibition phenotype (Kang et al., 2006; Oyadomari et al., 2006).

In pre-emptive quality control, newly synthesized cargo proteins are prevented from entering the ER lumen in order to prevent overloading of a stressed ER. Since N-glycosylation occurs within the lumen of the ER, unglycosylated tyrosinase represents the tyrosinase that is synthesized but does not enter the ER.

Surprisingly, the signal sequence of nicalin alone was able reduce tyrosinase levels. Since the signal sequence of nicalin is cleaved off of the rest of the protein
at a relatively slow rate, this posits a model whereby nicalin remains in the translocon and thereby blocks additional proteins from entering into the ER. This is a novel function for a signal peptide, but is compatible with the structure of a pharmacological inhibitor of translocation (Garrison et al., 2005), which apparently resembles a circularized signal peptide (R. Hegde, personal communication). The ER stress upregulation of nicalin could therefore tune the amount of protein allowed to enter into an already overloaded ER by increasing the amount of nicalin signal peptide blocking the channel.

Currently, the relationship between this signal peptide function and the interaction between SPP and nicalin is unclear. Moreover, additional work is needed to solidify the evidence linking the nicalin signal sequence to pre-emptive quality control. Nonetheless, all of the assayed functions of nicalin are consistent with a role in cargo processing at the level of the translocon; this is as predicted by its localization in the proteomics resource.

**Stexin interacts with cargo early after synthesis**

Like nicalin, stexin was upregulated in response to thapsigargin and DTT, but not tunicamycin. Also, as for nicalin, stexin transiently associated with the secretory cargo tyrosinase and only with the immature ER form. Also noted was a consistent interaction between stexin and a slightly higher molecular weight version of tyrosinase visible immediately after pulse radiolabeling. Since this tyrosinase version shifted downward upon subsequent chase, this appears to correspond to a early glycoform of tyrosinase. In the proteomics resource, stexin localized predominantly to the membrane of the rough ER, similar to the translocon and the associated oligosaccharyltransferase (OST) complex. When the enzymes of N-linked glycosylation were localized in the proteomics resource, OST and glucosidase I were found predominantly in rough ER, while glucosidase II and UGGT were localized further along in the rough and smooth ER. This predicts that the early glycoform of tyrosinase interacting with stexin should be proximal to glucosidase II trimming.
Analysis of the glycoform of tyrosinase preferentially interacting with stexin showed that it is indeed a glycoform occurring prior to glucosidase II trimming. This analysis compared the higher molecular weight form interacting with stexin to the slightly lower molecular weight form interacting with erlin-1 or TMX2. That the molecular weight difference is attributable to variation in glycosylation was shown when PNGase F (a deglycosylating enzyme) abolished the molecular weight difference. That the effect did not occur prior to glucosidase I treatment was shown when treatment with the glucosidase I inhibitor castanospermine increased the molecular weight of that interacting with stexin as well as that interacting with TMX2 or erlin-1 and abolished the molecular weight difference. Treatment with jack bean α-mannosidase did not change the molecular weight difference dramatically, meaning that the form of tyrosinase interacting with all three proteins had at least one terminal glucose. Putting these observations together, the form of tyrosinase interacting with stexin is an early glycoform downstream of the OST complex but upstream of glucosidase II trimming, likely proximal to glucosidase I activity. This is in agreement with the proteomics localization of the N-glycan processing enzymes of the ER, as well as that determined for stexin, TMX2 and erlin-1. Stexin therefore represents a transmembrane domain-containing resident ER constituent that interacts with secretory proteins early into their synthesis. Recently, the *Xenopus* orthologue of stexin is described as malectin and is shown to interact with diglucosylated N-glycans by NMR (Schallus et al., 2008), agreeing with the data shown here.

**Erlin-1**

Erlin-1 is an abundant transmembrane ER protein with a prohibitin-like domain (PHB). Erlin-1 was upregulated in response to tunicamycin and associated transiently (but weakly) with the immature form of tyrosinase. Erlin-1 has been described to localize to a lipid raft-like microdomain of the ER in a cholesterol-dependent manner (Browman et al., 2006). Erlin-1 was found in this work to interact with a later glycoform of tyrosinase than stexin, in keeping with its
localization in the proteomics resource to predominantly smooth ER. Erlin-1 has 83% amino acid sequence identity to erlin-2, which also localizes to predominantly smooth ER in the proteomics resource. Independently, erlin-2 has been shown to function in ER-associated degradation of activated IP3 receptors in the ER (Pearce et al., 2007). The ERAD function of erlin-2 is as predicted by the proteomics resource. The function of erlin-1 is predicted to be similar to erlin-2 based upon amino acid sequence similarity and proteomics localization. PHB domain proteins in mitochondria form rings closely apposed to the membrane (Tatsuta et al., 2005), which may be part of the mechanism of action of erlins in ERAD.

**TMX2**

This atypical member of the protein disulfide isomerase family (Ellgaard and Ruddock, 2005), contains a single thioredoxin domain with a non-classical SXXC motif. Although TMX2 associated transiently and strongly with only the immature ER form of newly synthesized tyrosinase, there was no evidence for mixed disulfide bonding with tyrosinase. Mutational analysis indicated that the predicted cytosolically exposed C-terminal region of TMX2 (containing the non-classical SXXC thioredoxin motif) is needed for its cargo binding activity. Further work will be required to understand the mechanism by which TMX2 interacts with cargo, given its requirement for a thioredoxin domain that does not form disulfide bonds.

Due to the low abundance of TMX2 in the isolated ER fractions, proteomics was unable to distinguish whether the long or short form of TMX2 was characterized. The EST database suggests that both forms are expressed in human and mouse. Hence, both forms were tested and both were found to be ER-associated (YFP colocalization with calnexin) and both were found to associate with the immature ER form of tyrosinase. TMX2 also could not be ascribed a localization profile by proteomics. However, based upon the interaction with a later glycoform of
tyrosinase compared to stexin and the typical localization of thioredoxin family members to both rough and smooth ER, TMX2 is predicted to have a protein folding function in the ER.

**New targets of the mammalian unfolded protein response**

Four other proteomics-uncovered proteins were found to increase transcription after tunicamycin treatment: sterlin, lyric, cvak104, and sterubin. Sterlin did not associate with tyrosinase and had no effect when overexpressed (not shown). The previously identified protein lyric has no known function (Britt et al., 2004; Sutherland et al., 2004). Cvak104 has been suggested to be a constituent of the clathrin-coated vesicle machinery (Conner and Schmid, 2005; Duwel and Ungewickell, 2006). We were unable to express the protein as a YFP construct and the protein was found in the biochemical fractions but not the total fractions. The finding of this protein in the ER may represent contamination or a link between the CCV machinery and ER functions. The upregulation of this protein by ER stress predicts the latter, but work has not been done to substantiate this idea. Sterubin, with a UBX domain, was a smooth ER protein also upregulated by stress but did not associate with tyrosinase. Consistent with the proteomics prediction, this protein is now described as UBXD8 with a function in ERAD (Mueller et al., 2008). Therefore, a total of 7 proteins showed increased transcription by stress, with 4 of these stress-responders shown to have functions in ER quality control in this thesis (nicalin, stexin, erlin-1) or recent literature (sterubin).

**A model of the secretory function of the ER**

In this thesis, a spatial model of the ER is proposed. It is proposed that newly synthesized cargo traverse the ER through subdomains restricted by the divisions of rough and smooth ER as well as topology across the membrane. Cargo start synthesis at the cytoplasmic face of the rough ER. As they are being translated they are imported across the membrane of the rough ER. After translation and
translocation, these proteins diffuse into the rough and smooth ER (either lumen or membrane) and are met by protein folding enzymes that are also unrestricted. If the proteins are stably folded enough to be secreted, they leave the ER by ER exit sites. If they are too unstable they traffic to the smooth ER-located ERAD machinery, transferring the misfolded protein to smooth ER-associated proteasomes.

It is proposed that this process can similarly be understood in terms of the N-glycosylation machinery. N-glycans control the recognition machinery for protein degradation. Upon entry into the ER, at the rough ER, newly synthesized secretory cargo are glycosylated with a Glc3-Man9-(GlcNac)2 glycan using the oligosaccharyltransferase complex. As a secretory cargo protein begins to diffuse between rough and smooth ER it is cleaved by glucosidase I (very early) followed by glucosidase II (at about steady state in the rough and smooth ER). The secretory cargo N-glycans then become Glc1 or 2-Man9-(GlcNac)2 (after glucosidase I) and Man9-(GlcNac)2 (after glucosidase II). If the protein is not folded correctly it is recognized by UGGT and a glucose is added so that the protein becomes Glc1-Man9-(GlcNac)2. Calnexin binds the Glc1 form, attempting to refold the protein. However, if the protein remains improperly folded it cannot leave the ER. The longer it spends in the lumen of the ER the more likely it is to be recognized for degradation. These degradation proteins include EDEM which creates a GlcNMan7-(GlcNac)2. This is recognized by the lectin Yos9, which delivers the protein to the smooth ER ERAD machinery. Here it is proposed that an ordered sequence of modifications to the N-glycan chain culminates in degradation if the secretory cargo protein does not leave the ER first. The model described in this thesis shows that the early enzymes of N-glycan modification are in the rough ER and the sequentially later enzymes are in the smooth ER. Therefore, the sequential action of these enzymes is organized in the space of the ER and from these sites dictate the degradation decision.
It is also proposed that this model can be understood in terms of the competing secretory protein energetics of ER-associated folding (ERAF), ERAD and secretory exit. ERAF occurs in rough and smooth ER where chaperones freely diffuse. ERAD is at much lower abundance and occurs in the smooth ER. The high local concentrations of ERAD constituents at retrotranslocation sites in smooth ER ensure immediate and complete degradation of unstable cargo. Otherwise, cargo that escape ERAF by achieving good protein energetics interact with ER exit sites localized to different parts of smooth ER and are transported to the Golgi for secretion. The machinery involved in synthesis, ERAF and ERAD cluster into subdomains of the ER such that the ER is composed of a series of functional units segregated in space.

Together this thesis describes the interrogation of a proteomics resource to make a global model of the secretory function of the ER. This model predicts the function of individual proteins. The predictions of the model were then tested by work on the function of 5 individual proteins. Since the predictions are supported through further work, the integrity of the model itself is supported. This thesis highlights the utility of organellar proteomics projects to function at two levels of inquiry. First, the acquisition of information at the large scale using observations on hundreds of proteins simultaneously gives an idea about how a biological system is functioning. Second, the quality of the information is such that it is informative at the reductionist level of inquiry into individual proteins involved in the process. In sum, this thesis argues that organellar proteomics is useful both in top-down and bottom-up approaches in cell biology. To study the organization of cells is to find out about that which defines living things. Through new tools and research, cell biologists further the understanding of the very essence of life.


Sarnow, P. (1989). Translation of glucose-regulated protein 78/immunoglobulin heavy-chain binding protein mRNA is increased in poliovirus-infected cells at a time when cap-
dependent translation of cellular mRNAs is inhibited. Proceedings of the National Academy of Sciences of the United States of America 86, 5795-5799.


