EXPRESSION OF ENDOPLASMIC RETICULUM OXIDOREDUCTASES (EROS) AND THEIR ROLE IN THE GI TRACT

WATSON, GRAEME

CHAPTER 1

INTRODUCTION

1.1 Overview

The endoplasmic reticulum (ER) is a network of membranous sacs and tubular branches extending from a perinuclear location into the cytoplasm of eukaryotic cells. It provides a crèche with a regulated redox-potential for disulphide bond catalysis in nascent polypeptides destined for secretion or residency within the secretory pathway. The full extent of mechanisms that maintain the redox differential between the ER and cytoplasm remain to be fully elucidated. Within this protein nursery, a range of chaperones and catalytic enzymes facilitate oxidative protein folding. It has previously been shown that some ER redox enzymes are differentially expressed in stomach and oesophagus tissue (Dias-Gunasekara *et al.*, 2005). The gastrointestinal (GI) system therefore represents a novel domain in which the roles and responses of human ER oxidoreductases can be studied in GI tissue and models of acid reflux respectively. Changes in the redox environment and pH of the stomach and oesophagus may influence gene expression, particularly during acid reflux, which is a major risk factor in developing Barrett's oesophagus (Fass *et al.*, 2001, Atherfold and Jankowski, 2006). Therefore the work in this thesis was to understand more about the expression of ER oxidoreductases in the GI tract, and to develop tools to explore the potential of ER oxidoreductases as biomarkers for the early diagnosis of Barrett's oesophagus and its progression through dysplasia and metaplasia to oesophageal adenocarcinoma.

1.2 Protein synthesis and pre-ER events

Proteins are a linear polymer of amino acids, which fold into a three dimensional form. These amino acids are bound together by peptide bonds formed between carboxyl and amino groups of adjacent amino acid residues.

Protein synthesis is a multi-stage process whereby the genetic code of DNA is used by the cellular machinery to build proteins. In eukaryotes, genes are first transcribed by RNA polymerase into pre-messenger RNA within the nucleus to form a primary transcript. In eukaryotes this requires further processing, such as RNA splicing, 5' end capping and the covalent linking of a polyadenyl moiety, typically to the 3' end, before it can become so-called mature mRNA. This is not the case for bacteria and archea. The mature mRNA can then be transported from the nucleus across the nuclear membrane to the cytoplasm, the site of protein translation. In eukaryotes, mature mRNA may be translated by cytoplasmic 70S ribosomes, or be directed to the ER by the signal recognition particle. The 70S ribosome comprises a large 50S subunit, and a smaller 30S subunit, themselves made of a number of protein and RNA components (Cabrita *et al.*, 2010). In contrast, prokaryotic translation happens co-transcriptionally, as there is no processing or transport of the original mRNA transcript.

An appreciation of the thermodynamic principles of protein folding reveals the importance of conserved folding pathways, and chaperone stabilisation during the folding process.

It was first suggested that a protein's native state is of lower entropy than the unfolded or misfolded states it could become, which would be of higher entropy (see Mirsky and Pauling, 1936, reviewed in Wolynes, 2005). Subsequently, energy landscapes have been developed to visualise the journey of a polypeptide to its nascent folded state (See Figure 1.1).

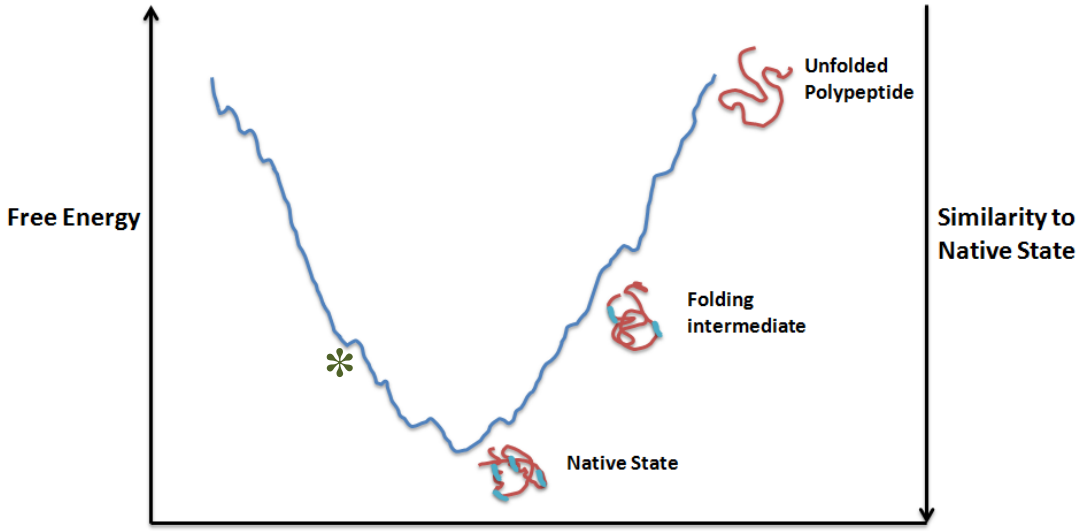

Figure 1.1: Energy Landscapes in Protein Folding

An energy landscape can highlight intermediate steps that may be prone to misfolding or aggregation, which occur at points on the landscape with little free energy, so-called "kinetic traps" (*) (Leopold *et al.*, 1992, further reviews by Jahn and Radford, 2005 and Wolynes, 2005). The resultant funnel-shape is unique for any given polypeptide, based on its thermodynamic and kinetic properties, in any unique environment. The open, upper-end of the funnel represents the high entropy of denatured states, which decreases down to the tip of the funnel, and represents the nascent, correctly folded protein. The correctly folded protein also has a smaller amount of stabilisation energy, relative to unfolded or misfolded states (Bryngelson and Wolynes, 1987).

The energy landscape infers that the folding of a polypeptide occurs down a diagrammatic funnel of free energy, with the native structure representing the lowest energy state. Provided the energy landscape favours native-like structures, the company of denatured polypeptides follows a downward course down the funnel.

Smaller proteins (less than ~100 amino acids) fold very quickly (<1s) with few intermediate steps, giving rise to a so-called "smooth" energy landscape funnel. In comparison, folding of larger proteins (~300 amino acids) exposes non-folded aggregation-sensitive sections for a number seconds, whilst translation continues. It is the role of molecular chaperones to bind to the polypeptide chain to prevent premature folding or misfolding, in "kinetic-traps" on the downward journey (Cabrita *et al.*, 2010). The energy landscape funnel in the folding of larger proteins is described as "rough", owing to the presence of a number of intermediate folding steps, and hence kinetic traps.

Cellular chaperones ensure an efficient protein folding pathway to achieve folding to the nascent state, although they can also assist in re-folding of misfolded intermediates, intracellular transport or proteolytic destruction. Some proteins, such as actins and tubulins cannot fold without chaperone assistance (see Hartl and Hayer-Hartl, 2009), and chaperones can assist in folding proteins even in the presence of potentially deleterious mutations.

1.3 Protein entry into the Endoplasmic Reticulum

The ER is the principle site for folding, assembly and quality control of membrane and secretory proteins, as well as ER resident proteins in the secretory pathway (Anelli and Sitia, 2008). The majority of these proteins are glycoproteins (Apweiler *et al.*, 1999). The ER is comprised of two morphological subtypes; the rough and smooth ER (RER and SER respectively). The RER is continuous with the nuclear membrane and so named because of the presence of attached ribosomes whereas the SER is less extensive and located at the periphery of the RER. It is important also to distinguish the functional difference between the two types of ER. The RER is the site of synthesis for secretory or membranous proteins, and the smooth ER, the site of lipid production and precursors of carbohydrate metabolism (Lodish *et al.*, 2007). Here, the term ER will be used to refer to specifically the RER for simplicity, unless otherwise stated.

1.3.1 Protein entry into the ER; the SRP and the translocon

The secretory protein polypeptide is directed to the ER along with the ribosome by way of the signal recognition particle (SRP), a process known as co-translational translocation (Walter *et al.*, 1981). In post-translational translocation, the protein passes into the ER after complete ribosomal synthesis. Translation itself ceases until the SRP/polypeptide/ribosome complex is transported to the ER. The SRP cycle is reviewed extensively (Nagai *et al.*, 2003, Wild *et al.*, 2004), but will be summarised here. First, the emergent translating polypeptide is met by the SRP, which binds to the ribosome and the non-polar signal sequence (signal peptide), which is at the N-terminus of the early transcript. Once the SRP has bound the translation complex, further translation and elongation of the polypeptide is prevented; further folding is prevented by binding of cytosolic chaperones. One such group of chaperones are the Heat Shock

Proteins (HSPs), which although structurally unrelated, are named in accordance with their molecular weight, such as Hsp40, Hsp60, Hsp70, Hsp90, Hsp100 and "small" Hsp proteins (Chang *et al.*, 2007, Tang *et al.*, 2007). Chaperones exert an effect both in stabilising and initiating folding of nascent polypeptides (Kramer *et al.*, 2009), and downstream components that complete the folding process (Albanèse *et al.*, 2006). The Hsp70 chaperone proteins are constitutively expressed, as well as being stress inducible (Chang *et al.*, 2007), and typically collaborate with Hsp40/DnaJ chaperones and nucleotide-exchange factors. In *de novo* protein folding, Hsp70 binds to nascent peptide chains, in an ATP-regulated binding and release mechanism; an N-terminal ATPase domain and a C-terminal peptide binding domain (PBD) couple together. The β-sandwich domain of the PBD recognises ~7 residue hydrophobic amino acid segments, which occur every 50-100 residues (Rüdiger *et al.*, 1997). By shielding these hydrophobic segments, the number of aggregate prone species decreases.

Hsp40 mediates delivery of substrate to ATP-bound HSP70. ATP-ADP hydrolysis and Hsp40 close the α-helical lid of the HSP70 PBD, and hence ensures tight binding of substrate to HSP70. Following this, Hsp40 dissociates and this step is catalysed by NEF (nucleotide exchange factor). Substrate release is induced by ATP-binding, and the α-helical lid opens. When Hsp70 releases its substrate, the protein has the chance to continue folding and bury the hydrophobic segments within the core; if not, further Hsp70 binding ensures that folding proceeds minimising risk of aggregation (Hartl and Hayer-Hartl, 2009).

The SRP-translation complex is transported to the membrane of the ER (or plasma membrane in prokaryotes), which is recognised by the SRP receptor (SR). The binding of the SRP to its receptor initiates the transfer of the polypeptide across the ER membrane, by way of a translocation channel, the translocon, a process which is

dependent on the hydrolysis of GTP to GDP. The signal sequence is cleaved by a signal peptidase, leaving the polypeptide free within the lumen of the ER.

SRP has been characterised in pro- and eukaryotes. The mammalian SRP comprises six proteins (SRP9, SRP14, SRP19, SRP54, SRP68 and SRP72) and a ~300 nucleotide RNA molecule, termed the 7SL RNA (Lutcke, 1995). The SRP contains two domains; the *Alu* and S domain, as shown following nuclease treatment (Gundelfinger *et al.*, 1983). The *Alu* domain has been shown to be responsible for the elongation arrest of the nascent polypeptide, and is formed from a SRP9/14 heterodimer, bound to the 5' and 3'end of the 7SL RNA. The S domain has been shown to be responsible for signal sequence binding, and comprises the central region of the 7SL RNA and SRP19, SRP54, SRP68 andSRP72 (Krieg *et al.*, 1986, Kurzchalia *et al.*, 1986). The S domain is also responsible for the GTP-dependent interaction with the SRP receptor (Connolly *et al.*, 1991).

After binding to the SRP receptor, only found on the membrane of the RER, the polypeptide is targeted to the translocon; in mammals, this is formed from a heterotrimer of Sec61α, Sec61β, Sec61γ (prokaryotes: SecY, SecG, SecE), the translocating nascent chain–associated membrane protein, TRAM (Gorlich *et al.*, 1992). Additonal accessory components are signal peptidase and oligosaccharyltransferase, OST, which glycosylates the polypeptide (see Johnson and van Waes, 1999, Johnson, 2009 and Rapoport, 2007). The eukaryotic translocon is alternatively named the Sec61 complex, while in bacteria and archaea, it is called the SecY complex (Brundage *et al.*, 1990). Central to the translocon is an aqueous, ion-conducting pore; in bacteria, this has been estimated to have a diameter of around ~15 Å (Breyton *et al.*, 2002, Wirth *et al.*, 2003). Previously, using fluorescent-labelled nascent chains, and hydrophilic quenching agents of known molecular weight, pore diameter had been estimated to be ~40-60 Å (Hamman *et al.*, 1997). This discrepancy

was resolved, as the size of the pore in the ribosomal-free electron microscopy experiment is smaller than in the ribosome-bound state (Hamman *et al.*, 1998).

The aqueous nature of the pore was inferred from the fluorescence lifetime of water-sensitive probes incorporated into nascent chains of translocon intermediates (Crowley *et al.*, 1993, Crowley *et al.*, 1994). The same quenching experiments (Crowley *et al.*, 1994) also showed that the ribosome was responsible for maintaining an ion-tight permeability barrier, whilst in the non-functioning, ribosome-free state, this barrier is maintained by BiP, (HSP70 family; a.k.a. Grp78), an ER-resident chaperone with a KDEL-retention sequence (Munro and Pelham, 1987, Haas and Wabl, 1983, Hamman *et al.*, 1998).

Previously, the ribosomal and translocon elements of this co-translational machinery were thought of as separate entities, although it has been shown recently that they are indeed linked (Pool, 2009). When the ribosome-nascent chain complex binds to the translocon, the exit tunnel of the ribosome aligns with the pore of the translocon; the signal sequence engages a binding site on Sec61α (Crowley *et al.*, 1993, Crowley *et al.*, 1994). As the nascent polypeptide moves through the large ribosomal subunit, the nascent chain transmembrane segment (TMS) determines the destination of the polypeptide, be it the cytosolic or lumenal side of the ER, or beyond (Liao *et al.*, 1997).

In summary, the nature of the nascent polypeptide determines its final destination in elegant fashion; secretory proteins are fully extended, and translocate through the pore (Johnson, 2004, Woolhead *et al.*, 2004, Sadlish *et al.*, 2005). The transmembrane segment of a polypeptide destined to become a membrane protein folds into an α-helix by a weakly non-polar constriction in the tunnel of the large ribosomal tunnel, which is formed by subunits Rpl4 and Rpl17. When the transmembrane segment reaches this constriction, the membrane protein RAMP4 is recruited to the complex, and cross-links with the transmembrane segment and Rpl17. This is coincidental to BiP-mediated

closure of the aqueous pore of the translocon, and opening of the ribosome translocon junction (Pool, 2009), though it has been suggested in a crystal structure of an archaebacterial Sec61 homolog that the displacement of a helical plug is responsible for the opening of the translocon pore (van den Berg *et al.*, 2004). In simple eukaryotes (and by inference, mammals), the polypeptide moves through the channel, where BiP, released from ADP, binds to the incoming polypeptide chain, which continues repeatedly until complete polypeptide translocation is achieved. The translocon complex does not disassemble when the ribosome leaves, though the pore diameter contracts from ~40-60 Å to ~15 Å (Hamman *et al.*, 1998). This model however, remains a topic of debate for explaining membrane integration of polytopic proteins. Simple single-spanning membrane proteins with transmembrane segments are laterally dislocated from the ER translocon into the lipid phase of the membrane. This is thermodynamically favoured due to the side chains present on the transmembrane segment (White and von Heijne, 2008). The presence of hydrophobic residues promotes efficient egress from the ER translocon and incorporation of the transmembrane segment into the cell membrane (Hessa *et al.*, 2005). The Sec61 translocon has been shown to simultaneously accommodate (at least) two membrane-spanning peptides and multiple transmembrane segments (Kida *et al.*, 2007). The final C-terminal transmembrane segment of opsin has been shown to remain directly adjacent to the ER translocon for an extended time, possibly to facilitate assembly of two folding domains (Ismail *et al.*, 2006, Ismail *et al.*, 2008). A similar delay has also been seen in the eighth transmembrane segment of the cystic fibrosis transmembrane conductance regulator (CFTR), which is believed to be regulated by the Sec61 complex and requires the presence of a charged residue within this eighth TM domain (Pitonzo *et al.*, 2009).

1.3.1.1 Signal peptide cleavage

As the nascent chain grows in length, it is subject to enzymatic action; signal peptidase cleaves the signal sequence when the chain reaches an approximate length of ~150 amino acids (Mothes *et al.*, 1994, Nicchitta *et al.*, 1995), while OST glycosylates any N-glycan sequence extant from the lumenal side of the translocon between 12-14 residues (Nilsson and von Heijne, 1993, Popov *et al.*, 1997).

1.3.1.2 ER protein retention and retrieval

ER resident lumenal proteins, such as BiP, maintain residency as the result of a C-terminal KDEL (Lysine-Aspartate-Glutamate-Leucine) tetrapeptide (Munro and Pelham, 1987) . In yeast, this function is performed by the HDEL tetrapeptide. ER localisation of membrane proteins is maintained via a KKXX (di-lysine) motif (Nilsson *et al.*, 1989, Townsley and Pelham, 1994). The localisation of soluble ER resident proteins requires KDEL receptor-mediated protein retrieval from the intermediate compartment or Golgi apparatus (Pelham and Munro, 1993). The concept was proved by adding the KDEL motif to lysozyme protein, which remained unsecreted and resident in the ER (Pelham, 1988).

The KDEL receptor is a seven transmembrane domain protein, and to date, three classes have been identified; KDELR1, KDELR2, KDELR3a/b (Capitani and Sallese, 2009). The KDEL receptors demonstrate a significant degree of conservation, sequence identities range from ~65.0% to ~83.5%, and their homologies range from ~80% to ~94% (Capitani and Sallese, 2009). The Ruddock lab showed each of the KDELR isoforms bound specific KDEL-like motifs preferentially, using bimolecular-fluorescence-complementation screening (Raykhel *et al.*, 2007). They showed that KDELR1 and KDELR3 are less able to bind/retrieve KDEL-motif-like sequences than KDELR2. KDELR1 preferentially binds chaperones bearing the KDEL-motif rather

than the HDEL-motif; the converse of KDELR3. KDELR2 preferentially recognises HDEL-motif variants over any other. The functional significance of these different binding properties remains undefined (Capitani and Sallese, 2009).

The yeast HDEL receptor was identified by studying ER-retention defective (erd) yeast mutants, which did not retain tagged invertase protein or endogenous GRP78; genetic complementation studies identified the ERD1 and ERD2 genes (Hardwick *et al.*, 1990, Semenza *et al.*, 1990). Deletion of ERD1, an integral membrane protein, leads to secretion of HDEL tagged proteins and defects in their N-glycans (Hardwick *et al.*, 1990). The ERD2 gene also encodes an integral membrane protein, localised to the ER and Golgi complex, and similarly deletion of this gene leads to secretion of HDEL tagged proteins (Semenza *et al.*, 1990). ERD2-deleted yeast also shows an accumulation of intracellular membrane, and defective secretory protein transport throughout the Golgi complex (Semenza *et al.*, 1990).

1.3.1.3 Glycoprotein folding, calnexin and calreticulin

During the protein folding process, some amino acid residues can undergo post-translational modification, which can alter the properties of folding, stability, activity and function. The majority of proteins that fold within the ER are glycoproteins, 90% of which carry N-linked glycans.

Glycans are large hydrophilic sugar polymers, which increase solubility and confer some protease resistance to a protein. They are covalently attached co-translationally in the translocon complex (Nilsson and von Heijne, 1993, Helenius and Aebi, 2004) and during post-translational modification by OST. Glycans are usually added to asparagine residues in the Asn-Xaa-Ser/Thr sequence (Apweiler *et al.*, 1999, Helenius and Aebi, 2004).

Some modifications, such as phosphorylation and methylation, confer negative charge or mask charge respectively, while glycans have been shown to enhance protein stability, solubility, folding kinetics and secretion (Olden *et al.*, 1982, Sola *et al.*, 2007, Shental-Bechor and Levy, 2008). In addition, glycans have been shown to act as recognition markers, modulate immune response and protein turnover (reviewed in Helenius and Aebi, 2004).

The importance of glycans can be shown experimentally with tunicamycin (from *Streptomyces lysosuperficus*), a mix of homologous nucleoside antibiotics, which indirectly inhibits glycoprotein synthesis, causes cell cycle arrest in G1 phase and induces the unfolded protein response (Olden *et al.*, 1978). Glycoprotein synthesis is arrested at the first step whereby tunicamycin inhibits the GlcNAc phosphotransferase enzyme (GPT), which catalyses transfer of N-linked acetylglucosamine-1-phosphate from UDP-N-acetylglucosamine to dolichol phosphate. Furthermore, some glycoproteins have been shown to partially misfold, secrete and remain functional in the absence of glycans. This was exemplified by non-glycosylated MHC class I molecules, which continued to be recognised by cytotoxic T cells, albeit with reduced cell surface expression (Miyazaki *et al.*, 1986), resulting from impaired intracellular transport.

The eukaryotic calnexin/calreticulin cycle is an important lectin-based chaperone system that assists in the folding of glycoproteins and also acts as a quality control mechanism (Hammond *et al*., 1994, Hebert *et al.*, 2005).

Calnexin is a 90 kDa type 1 ER membrane protein, with the majority of the structure located within the ER lumen (Wada *et al.*, 1991). Calreticulin is the soluble paralog of calnexin, and is a 60 kDa ER lumenal protein, retained by a KDEL retention sequence (Fliegel *et al.*, 1989). Retention and retrieval mechanisms work in concert to keep the lectin calreticulin localised within the ER (Sonnichsen *et al.*, 1994). Both calreticulin

and calnexin have been shown to bind ATP, Ca^2+, Zn^2+ and ERp57, a thiol-oxidoreductase of the human PDI family (Baksh *et al.*, 1995, Oliver *et al.*, 1997, Williams, 2006).

Calnexin and calreticulin both recognise the initial glycan transferred to asparagine residues on nascent glycoproteins, Glc1Man9GlcNAc2, formed via glucosidase I and II action on Glc3Man9GlcNAc2 (Hammond *et al*., 1994, Spiro *et al.*, 1996, Ware *et al.*, 1995). Glucosidase II removes the innermost glucose residues, *l* and *n* (see Figure 1.2). The ER soluble glycosyltransferase re-glycosylates the glycan in a cycle that ceases when the glycoprotein reaches its native state and proceeds through the secretory pathway (see Caramelo and Parodi, 2009 for review). The presence of these lectin like chaperones promotes folding efficiency and reduction, isomerisation of disulphide bonds through association with ERp57. Glycosyltransferase acts as a protein-conformation sensor, and will not re-glycosylate glycans in the cycle unless the protein displays non-native 3D structures (Sousa *et al.*, 1992, Ritter *et al.*, 2005).

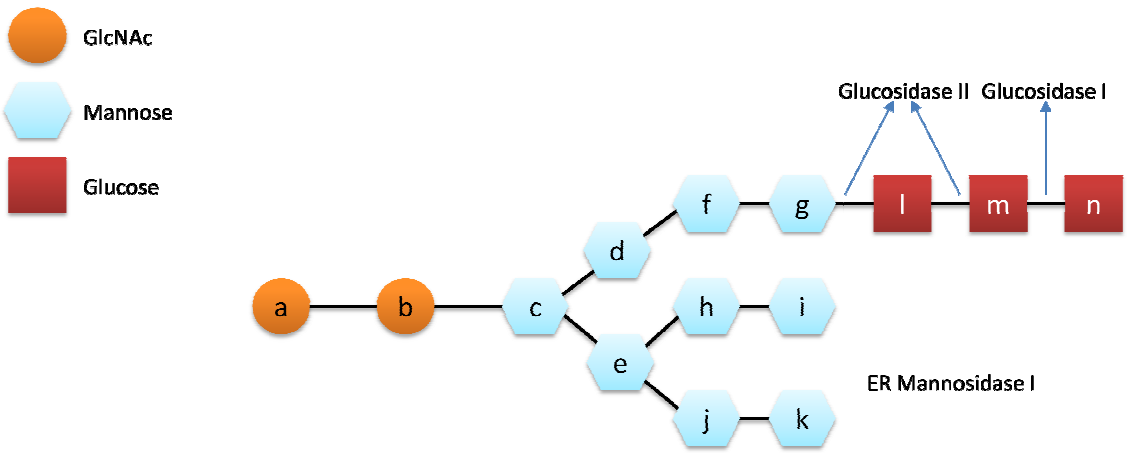

Figure 1.2: Structure of an N- glycan

The letters (a–n) show the order in which monosaccharides are added to the glycan chain. Glucosidase I removes residue n, and Glucosidase II removes residues l and m. (Figure based on a similar figure by Caramelo and Parodi, 2009).

Structurally, the lumenal (major) portion of calnexin has a globular β-sandwich domain and a proline-arm domain, comprising a hairpin of two β-strands (Schrag *et al.*, 2001). Only the proline-arm domain of (rat) calreticulin has been solved to date by NMR studies (Ellgaard *et al.*, 2001), and calreticulin contains three proline-rich tandem repeats, compared to the four of calnexin. The globular domain of calnexin, and calreticulin, has been shown to contain the oligosaccharide binding site, recognising $Glc_1Man_9GlcNAc_2$ and three additional mannose residues (Spiro *et al.*, 1996, Vassilakos *et al.*, 1998). Binding to ERp57 is via the tip of the proline-arm domains, shown by mutational studies in the distal part of the domain (Leach *et al.*, 2002). Similarly, the crystal structure of the non-catalytic **b** and **b'** domains of ERp57 have shown that the calnexin interaction is sited at the basic C-terminus of the **b'** domain (Kozlov *et al.*, 2006). The polypeptide binding region is also predicted to reside in the globular domain (Leach *et al.*, 2002).

Studies of calnexin/calreticulin function can be performed experimentally with glucosidase inhibitors, which inhibit glucosidase I and II, and calnexin/calreticulin deficient cell lines. Using glucosidase inhibitors, it has been shown the majority of glycoproteins require the action of calnexin/calreticulin (Danilczyk and Williams, 2001). Eukaryotic cell lines remain viable following genetic deletion of either calnexin or calreticulin, but show defects in phagocytosis (Muller-Taubenberger *et al.*, 2001), though deletion of the calnexin homolog gene in *Schizosaccharomyces pombe* is fatal (Jannatipour and Rokeach, 1995). Calreticulin negative mice expire from cardiac defects at embryonic day 18 (Mesaeli *et al.*, 1999), and calnexin negative mice show severe neurological problems (Denzel *et al.*, 2002). As cell lines remain viable and mice experience life-incompatible phenotypes, this suggests a role for these lectins in development, beyond that of folding proteins, including calcium homeostasis and

signalling (Michalak *et al.*, 2002), modulation of gene expression (Michalak *et al.*, 1999) and cell adhesion (Johnson *et al.*, 2001).

Studies in calnexin/calreticulin deficient cell lines focussing on viral glycoproteins such as Semliki Forest viral proteins E1 and p62, influenza haemagglutinin (HA) and also MHC class I have shown the differential dependence of these glycoproteins on calnexin and calreticulin for folding rate and efficiency (reviewed in Williams, 2006, Caramelo and Parodi, 2009). In infected host cells, the Semliki Forest viral glycoproteins E1 and p62 are cleaved from a precursor by signal peptidase, before forming heterodimers (Andersson and Garoff, 2003). In calreticulin negative cells, Semliki Forest viral protein folding is accelerated, though is less efficient (Molinari *et al.*, 2004). In calnexin negative fibroblasts however, maturation is unaffected, though E1 and p62 associate more with calreticulin.

Influenza virus haemagglutinin (HA) is another model glycoprotein assembled in the ER with a solved crystal structure (Wilson *et al.*, 1981) and well-characterised folding pathway (Braakman *et al.*, 1991). Mature HA is a homotrimer, comprising monomers of HA1 and HA2 chains linked by a disulphide bond, with N-glycans being recognised by calnexin, calreticulin and ERp57 upon emergence from the translocon (Daniels *et al.*, 2003). In calreticulin-negative cells, folding efficiency of HA is compromised by formation of disulphide-linked complexes, though folding rate, trimerisation and ER export is increased. In calnexin-negative cells, folding of HA to the fully oxidised state is 70% less than in wild-type control cells, and decreased export is apparent (Molinari *et al.*, 2004).

MHC class I assembly, cell-surface expression and antigen-presenting function is impaired in calreticulin negative cells and cannot be rescued with a soluble form of calnexin (Gao *et al.*, 2002). MHC class I processing in calnexin deficient cells is

unaffected (Scott and Dawson, 1995) . Calreticulin can also bind directly to non-glycosylated peptide sequences in cells and *in vitro* (Saito *et al.*, 1999).

Quantifying the biophysical effects of glycans on glycoproteins is an emerging field; only around 3.5% of proteins in the Protein Database (PDB) have covalent N-glycans, with few having solved glycan structures (Shental-Bechor and Levy, 2008, Shental-Bechor and Levy, 2009). This is in part due to the heterogeneity of the structures formed from the 9 monosaccharides that make up an N-glycan, and that these glycans themselves hinder protein crystal determination.

A number of authors have published papers which show N-glycans to enhance protein thermodynamic stability (e.g. CD2: Wyss *et al.*, 1995, serine protease inhibitors: DeKoster and Robertson, 1997, trypsin and chymotrypsin: Pham *et al.*, 2008), while further roles of glycans were modelled by computational analysis (Shental-Bechor and Levy, 2008). Shental-Bechor and Levy designed an *in silico* model, using the 64 amino acid residue src homology 3 (SH3) domain, which exposes 35 residues to the solvent for addition of [Man3GlcNAc2] or [Man9-GlcNAc2] glycans. Although not normally glycosylated the SH3 domain is well characterised in computational folding models. The position of single glycans was shown to affect protein folding stability between 25-52% of the norm, at a temperature where the control construct would be 50% folded. In terms of the energy landscape, Shental-Bechor and Levy showed that each glycan contributed to a 4% increase in folding, relative to control, at a temperature where it would be 50% folded; therefore the folding rate accelerates, and the unfolding rate decelerates.

The glycosylation of a 14-monosaccharide variant ($Glc_3Man_9GlcNAc_2$) of the immune receptor CD2s Ig adhesion domain was shown to confer enhanced stability over that seen in stable CD2 homolog mutants (Hanson *et al.*, 2009).

In light of these studies, it has been proposed that N-glycans confer a form of rigid stability, which is responsible for a decrease in energy (enthalpy) and dynamics (entropy) that cancel each other out, leaving a negligible change in free energy in the folded state (Shental-Bechor and Levy, 2009).

Small angle x-ray scattering measurements have shown interesting comparisons between unfolded and folded *Pheniophora lycii* phytase, that has either been glycosylated or not (Hoiberg-Nielsen *et al.*, 2009). Unfolded, glycosylated phytase is 32% larger than its unglycosylated counterpart, whilst folded, glycosylated phytase is 11% larger than it's non glycosylated form. The presence of such glycans prevents local interactions between residues, and the glycans are believed to be responsible for this expansion, with the first sugars of any given glycan affecting stability. The length of the glycan has little additional effect, with no significant long-range communication between glycans (Shental-Bechor and Levy, 2009). These studies show that N-glycans confer specific stabilising properties to glycoproteins, as well as enhancing solubility, folding kinetics and secretion.

1.4 Disulphide bond formation

Disuphide bridges are common structural elements of secreted proteins. They are covalent bonds between thiol groups of cysteine residues, and their formation by thiol-disulphide exchange is particularly susceptible to perturbations of the redox potential within the ER lumen (for recent reviews, see Mamathambika and Bardwell, 2008, Riemer *et al.*, 2009). They are formed in the oxidising environment of the rough ER as opposed to the reducing environment of the cellular cytosol, which contains reduced glutathione in relatively high abundance (10-14 mM; Ostergaard *et al.*, 2004). Disulphide bond containing proteins are typically found in secreted proteins, lysosomal proteins and the extracellular domains of membrane proteins. The bonds themselves

may be reversible, which is important in the function of redox-regulated proteins, including the cytosolic chaperone HSP33, the OxyR and YAP1 transcription factors, and the disulfide oxidoreductase Ero1p (Linke and Jakob, 2003, Sevier *et al.*, 2007). Disulphide bonds formed via thiol-disulphide exchanges confer high thermodynamic stability and have a high equilibrium constant (K_{eq}) of 10^5, with the folded protein stabilising the disulphide bond as much the protein itself is stabilised as a result of that bond (Creighton, 1990, Darby and Creighton, 1993).

Studies *in vitro* of disulphide bonded proteins such as RNAse A and bovine pancreatic trypsin inhibitor have revealed mechanisms of disulphide bonding (see Mamathambika and Bardwell, 2008). The thiol-disulfide exchange reaction between a protein and a redox reagent is actually a two-stage process, which results in the net transfer of a disulphide bond from a redox reagent to the protein. In the reduced protein, a thiolate anion (S^-) displaces a sulphur atom in the redox reagent, forming a so-called mixed disulfide. Next, the remaining thiol anion of the redox reagent displaces the mixed disulfide via intramolecular attack. This leads to the formation of the oxidized protein (Figure 1.3).

Isomerisation or disulphide shuffling can also occur when a cysteine thiolate attacks a disulphide within the protein; in the eukaryotic ER, this is catalysed by PDI. In these *in vitro* studies, the pH of the refolding buffer has a dramatic effect on oxidative folding, which further modulates the relativistic concentrations of reactive thiolate to thiol (see Mamathambika and Bardwell, 2008). The pK_a of unfolded cysteine residues is 8.7; this is the pH wherein 50% of the cysteine residues will be ionised. As a result refolding is rapid above pH 9, and decreases accordingly with a lowering of pH, as less cysteine residues will be ionised. In cells at fixed physiological pH, the oxidoreductases which transfer disulphide bonds have a lower pK_a for the unreactive thiols (Arolas *et al.*, 2006 Mamathambika and Bardwell, 2008).

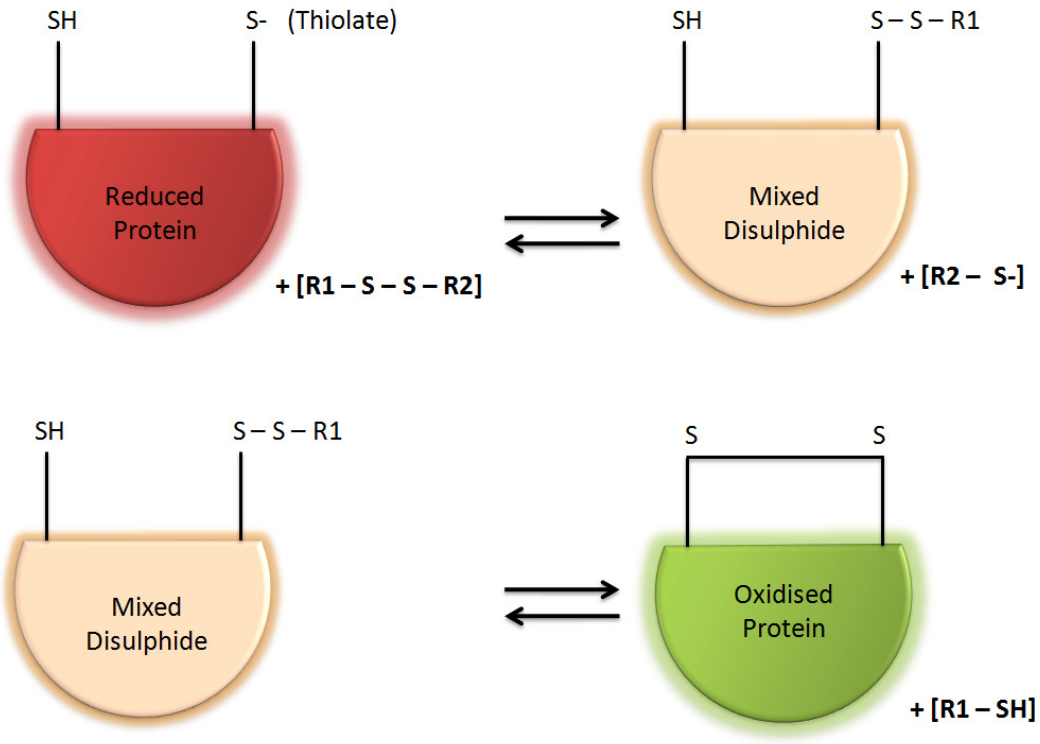

Figure 1.3: Disulphide Bond Formation

A thiolate anion (S^-) displaces a sulphur atom in the reduced protein, forming a mixed disulfide. Next, the remaining thiol anion of the redox reagent displaces the mixed disulfide via intramolecular attack. This leads to the formation of the oxidised protein

1.4.1 Disulphide bond formation in Prokaryotes

In gram-negative prokaryotes such as *Escherichia coli*, there are two well-characterised pathways in the establishment of disulphide bonds in proteins; the DsbA/DsbB oxidative pathway (Bardwell *et al.*, 1991, Akiyama *et al.*, 1992), and the DsbC/DsbD isomerisation pathway (McCarthy *et al.*, 2000). DsbA is the primary donor of disulphides via thiol-disulphide exchange to secreted proteins within the periplasm of *Escherichia coli*, although the introduction of these disulphides is non-specific. DsbB is a membrane protein which reoxidises DsbA. Following introduction of disulphides, they are re-shuffled (isomerised) by DsbC, which is maintained in a reduced state by DsbG, a homodimeric inner membrane protein. DsbG shares a ~24% sequence identity with DsbC (Heras *et al.*, 2004).

DsbA is a 21kDa, soluble, highly oxidising protein and features a redox-active CXXC motif that allows conversion from dithiol to disulphide, and a *cis*-proline motif that interacts with substrate (Kadokura *et al.*, 2004). When DsbA is disulphide bonded (oxidised), it has a high redox potential of – 120 mV and Cys30 ($C_{27}X_{28}X_{29}C_{30}$) has a pKa of only 3, which ensures catalytic activity at physiological pH. At such pH, this cysteine will be for the most part in the thiolate anion state, and mutations at nearby His32 reduce the oxidising power of Cys30, which is thought to be stabilising (Zapun *et al.*, 1993, Grauschopf *et al.*, 1995, Guddat *et al.*, 1997, Guddat *et al.*, 1998). The reduced form of the *Vibrio cholerae* DsbA active site (NMR; Horne *et al.*, 2007), and a Cys33Ala mutant (Crystal Structure; Ondo-Mbele *et al.*, 2005) show that the reduced (inactive) form of DsbA is more flexible than the oxidised (active) form; Cys33Ala mutation changes the structure of the CXXC active site by disrupting helix α1. The peptide groove binding region and the active site tetrapeptide shifts dramatically in this mutant.

DsbB is a functional analogue of the eukaryotic oxidoreductases Erv2p and Ero1p, and (based on sequence analysis) features four anti-parallel α-helices, a CXXC motif, cofactor-binding residues and another cysteine pair acting as an electron "shuttle" in dithiol exchange between DsbA and DsbB active sites (Sevier *et al.*, 2005). Analysis of a crystal structure of DsbA-DsbB shows that DsbA is reactivated by thiol-disulphide exchange with the DsbB active site and confirms the presence of the anti-parallel α-helices (Inaba *et al.*, 2006). In addition, the redox-active Cys104 of DsbB resides on a mobile periplasmic loop, which decouples from its own redox partner Cys130 to form a mixed disulphide with Cys30 of the DsbA active site.

In dithiol exchange with DsbA, the terminal electron acceptor employed by DsbB in aerobic conditions is ubiquinone and molecular oxygen via the electron transfer system of the respiratory chain (Bader *et al.*, 1999). In anaerobic conditions, electrons are transferred initially by menaquinone, and transferred to nitrate, nitrite, and fumarate (Bader *et al.*, 1999, Bader *et al.*, 2000, Kobayashi and Ito, 1999, Xie *et al.*, 2002). DsbA is such a strong oxidase that disulphide bonds are introduced non-specifically into nascent proteins, such that they require re-shuffling or isomerisation so that the protein may reach its native conformational state (Rietsch *et al.*, 1996). The DsbC/DsbD isomerisation pathway in prokaryotes fulfils this need.

DsbC is a ~23 kDa protein which is essential for folding of the periplasmic phytase proteins, AppA and MepA, as well as RNase I (Hiniker and Bardwell, 2004 Berkmen *et al.*, 2005). DsbC homodimerises to form a V-shaped structure (see crystal structure: McCarthy *et al.*, 2000); each arm of this 'V' contains a C-terminal thioredoxin-like domain linked via an α-helix to an N-terminus dimerisation domain, which forms the dimeric apical joint of the 'V'. The cleft of the 'V' is uncharged, and is believed to be the interface whereby hydrophobic patches on non-native protein substrate conformations are detected for disulphide isomerisation. Crystal structures of reduced

(*E. coli*: Banaszak *et al.*, 2004) and oxidised (*H. influenzae*: Zhang *et al.*, 2004) DsbC suggest the cleft can flex to accommodate differently-sized substrates. DsbC functions elegantly to isomerise non-native disulphide conformations by one of two alternate mechanisms (Nakamoto and Bardwell, 2004). In the first, the nucleophilic Cys98 of the reduced DsbC CXXC active site attacks an incorrect disulphide in a substrate protein, and forms a mixed disulphide with it; this is in turn is attacked by a reduced cysteine thiolate in the substrate. DsbC is returned to its reduced state, and the non-native disulphide is thus reshuffled. In an alternate mechanism, DsbC can exit the mixed-disulphide with the substrate by employing the remaining cysteine thiolate of its active site in a thiol-disulphide intramolecular attack upon it. DsbC exits in an oxidised form, while the reduced protein can now be subject to re-oxidation via DsbA.

DsbD maintains DsbC and DsbG in a catalytically-active reduced form. DsbD has three domains of cysteine pairs, the catalytic cysteine has a higher than typical pK_a (Mavridou *et al.*, 2007). The central transmembrane domain is linked to an Ig-like N-terminal domain and a thioredoxin-like C-terminal domain. Electron transfer occurs from cytoplasmic thioredoxin, to the transmembrane domain, the C- and N-terminals, before being used to reduce DsbC.

DsbG shares 26% sequence identity with DsbC. Like DsbC, it is also a V-shaped dimer with a thioredoxin fold and a CXXC motif. It is maintained in a reduced state by DsbD. Depuydt and colleagues studied potential substrates of DsbG, and showed an interaction between DsbG and YbiS, an L,D-transpeptidase which preferentially depends on DsbG rather than DsbC for reduction of it's sole cysteine (Depuydt *et al.*, 2009). The wider implications of this is that the negatively charged surface of DsbG interacts with folded proteins, and the authors propose that this acts to prevent sulphenic acid formation by rescue of sulphenated orphan species in folding proteins. The authors show that DsbC may also assist in this task, with its hydrophobic inner

surface interacting with unfolded proteins and correct non-native disulphide conformations.

1.4.2 Disulphide bond formation in Eukaryotes

The ER is the major site of disulphide bond formation in eukaryotes, but before considering the ER, it is worth noting that the intermembrane space of mitochondria contains proteins with structural disulphides (Mesecke *et al.*, 2005). Briefly, the mitochondrial disulphide relay system utilises two principle proteins; Mia40 and Erv1 (see Riemer *et al.*, 2009 for review). Mia40 is a conserved protein which contains a CXC redox-active motif, which oxidises cysteine residues of polypeptides. Once stably folded, these polypeptides are unable to cross the outer mitochondrial membrane (Lu *et al.*, 2004). Erv1, an FAD-containing sulphydryl oxidase (Mesecke *et al.*, 2005), reoxidises Mia40, and is in turn oxidised by cytochrome c, and electrons are transferred to molecular oxygen (Bihlmaier *et al.*, 2007, Dabir *et al.*, 2007). This link with the mitochondrial respiratory chain produces water, and prevents formation of hydrogen peroxide.

In eukaryotes, the Ero1-protein disulphide isomerise (PDI) relay is responsible for the formation and isomerisation of disulphide bonds in nascent proteins within the ER (Sevier and Kaiser, 2002). Disulphides are provided to nascent proteins during the translocation process. The principle proteins involved in eukaryotic disulphide bond formation are members of the protein disulphide isomerase family (PDIs; see section 1.4.2.2) and ER oxidoreductase (Ero1; see section 1.2.4.1), which facilitate the correct folding and oxidation of disulphide bonded proteins under oxidising conditions within the ER lumen. These conditions must be tightly regulated as too oxidising an environment leads to misfolding and misoxidisation of nascent polypeptides (Marquadt *et al.*, 1993), while a too reducing environment is incompatible with the formation of

disulphide bonds (Braakman *et al.*, 1992), although a specific pathway for disulphide bond formation exists based on the presence of reduced glutathione (Chakravarthi *et al*., 2006).

1.4.2.1 Endoplasmic Reticulum Oxidoreductase

The yeast and mammalian Eros will be introduced, and then discussed in turn. The conserved yeast gene ERO1, located on chromosome XIII, and its protein product Ero1p, was first described in 1998 (Frand and Kaiser, 1998, Pollard *et al*., 1998),
In humans, there are two Eros; Ero1α and Ero1β, located at chromosomes 14q22.2 and 1q42.2-q43 respectively (Cabibbo *et al*., 2000, Pagani *et al*., 2000,).
Yeast Ero1p is retained in the ER by 127 residues in the C-terminal tail (Pagani *et al.*, 2001), although the mammalian Eros lack both the tail and a retention motif, such as KDEL. It is currently believed that mammalian Eros associate with the inner ER membrane and the PDI homologue ERp44, and may be retained indirectly by the ERp44 KDEL (RDEL) retention signal (Pagani *et al*., 2000, Anelli *et al.*, 2002, Anelli *et al.*, 2003). The disulphide production and donation pathway is well-conserved between eukaryotic organisms. The flavoenzyme Ero1 maintains the oxidised state of PDI, which supplies disulphides for S-S bond formation in newly synthesised proteins (Frand and Kaiser, 1998, Frand and Kaiser, 1999, Pollard et al., 1998, Cabibbo *et al*., 2000, Pagani *et al*., 2000). PDI in turn supplies these disulphides for S-S bond formation in nascent proteins, in a manner analogous to the DsbA-DsbB pathway described above (Tu and Weissman, 2002, Gross *et al*., 2004, Gross *et al*., 2006).
The redox potential of the ER can be modulated by altering the levels of Ero1 or glutathione (Cuozzo and Kaiser, 1999), and reduced glutathione is the principle reductant which aids PDI-mediated isomerisation. Glutathione is a linear peptidic redox reagent existing as either reduced glutathione (GSH) or oxidised glutathione (GSSG).

1.4.2.2 The yeast Ero, Ero1p

Ero1p was identified following a screen of temperature-sensitive *S. cerevisiae* mutants (Hartwell *et al.*, 1973) for defective export of ER secretory proteins. The CKY599 *ero1-1* strain failed to grow at temperatures above 36°C, and maturation of the model protein carboxypeptidase Y (CPY) was halted. The ERO1 gene was discovered with the aid of a high-copy genomic plasmid library. From 32,000 transformants, only one plasmid conferred resistance to a plate containing 8mM DTT, compared to a normal limit of 4mM DTT in untransformed yeast. This plasmid contained an insert from yeast chromosome 13, and open-reading frame analysis confirmed that the ERO1 gene was responsible for DTT tolerance.

An *S. cerevisiae* transformant from a high copy plasmid library was selected from a high DTT treatment series and was found to overexpress express ERO1; this expression could be enhanced by deleting the open reading frame of the plasmid insert. Mutation in ERO1 resulted in extreme sensitivity to DTT, and an enhanced UPR (Pollard *et al*., 1998). A high copy plasmid expressing Ero1p with a myc-tag motif on the c-terminus (96kDa) was also able to rescue ERO1-Δ spores. EndoH treatment, which removes N-linked glycans, established that Ero1p-myc was a glycosylated ER resident protein with 5 (predicted) N-linked oligosaccharides. A location in the Golgi was ruled out as anti-alpha1, 6-mannose could not immunopreciptate Ero1p-myc, suggesting that there was no Golgi modification of the protein. The role of glutathione as an essential provider of oxidising equivalents was also questioned, as CPY processing proceeded in a gsh1-Δ mutant, where the key glutathione enzyme gamma-glutamylcysteine synthetase, was deleted (Cuozzo and Kaiser, 1999).

The CPY model protein was also used to show the interaction of Ero1p with yeast PDI, Pdi1p (Frand and Kaiser, 1999). Trichloroacetic acid (TCA) was used to trap mixed disulphides, by lowering intracellular pH and preventing thiol-disulphide exchange

reactions. The C-terminal cysteine of both Pdi1p active sites was mutated, and swapped for serine. This prevents resolution of any mixed disulphides formed as the active site cysteine is unavailable for intramolecular attack. Cells transfected to overproduce Ero1p and wildtype Pdi1p or the serine mutant were labelled with ^{35}S cys/met, and a combination of TCA treatment and immunoprecipitation experiments showed that Ero1p and Pdi1p formed mixed disulphides. Similar data were obtained for the yeast Pdi1p homologue, Mpd2p, but mixed disulphides between CPY and Ero1p could not be detected. The *in vivo* oxidation state of Pdi1p and Mpd2p was examined following TCA treatment to prevent further thiol-disulphide exchange, and alkylation with the thiol-conjugating reagent AMS. Data showed that Pdi1p was mostly oxidised, which suggested function as an oxidase. The smaller reduced portion was believed to reflect isomerase activity, which is dependent on dithiols (Laboissière *et al.*, 1995). Oxidase function was supported as CPY was fully reduced when synthesised in the absence of PDI, and the capture of mixed disulphides between Pdi1p and the CPY precursor, p1. In addition, Ero1p was also mostly oxidised *in vivo*, even in a Pdi1p-depleted yeast strain, which first suggested a flow of oxidising equivalents between Ero1p to PDI (Frand and Kaiser, 2000). Yeast Ero1p has 14 cysteines, 7 of which are absolutely conserved in other species homologues. These are at positions 90, 100, 105, 208, 349, 352, and 355 in yeast Ero1p. The three C-terminal cysteines are in a highly conserved region of the Ero1p protein, which is 65% similar to the human homologue. Cysteine-alanine Ero1p mutants, analysed on the ire1Δ yeast background (which prevents UPR induction of Ero1p) showed that disulphide bonds between Cys100-Cys105 and Cys352–Cys355, (which became known as shuttle and active site cysteines respectively), were crucial for disulphide bond formation and cell viability. When the Cys100-Cys105 bond is removed with a Cys100Ala substitution, the previously observed capture of Ero1p-Pdi1p or Ero1p-Mpd2p mixed disulphides was impeded,

and Pdi1p oxidation was blocked *in vivo* (Frand and Kaiser, 1999, Frand and Kaiser, 2000). Cys352Ala or Cys355Ala substitution was also shown to prevent Ero1p reoxidation *in vivo*.

1.4.2.3 The Mammalian Eros; Ero1α and Ero1β

The identification of Ero1p led to the discovery of the human genes ERO1-Lα, first published as ERO1-L, for *S. cerevisiae* Ero1-like; (Cabibbo *et al*., 2000) and ERO1-Lβ (Pagani *et al*., 2000), Figure 1.4. A structural comparison of Ero1α with Ero1p is shown in Figure 1.5. The terms Ero1α and Ero1β will be used here to describe their protein product. Ero1α was discovered through searching genomic databases using the yeast ERO1 gene as a starting point, and contained the highly conserved redox active CXXC motif. Similar highly conserved sequences were also discovered for *Mus musculus*, *Arabidopsis thaliana*, *Drosophila melanogaster*, and *Schizosaccharomyces pombe*. ER residency of transfected Ero1α-myc was shown by co-staining with fluorescent ConA, a glycoprotein binding lectin. A gel mobility shift following EndoH treatment and *in vitro* translation studies showed that Ero1α was an N-linked glycoprotein (Pollard *et al*., 1998). When an *ero1-1* yeast mutant was transfected with human Ero1α, the yeast was able to grow at 37 °C and confered a level of resistance to DTT, but less than that of an Ero1p transfectant. The (then) presumed redox-active motif $C_{391}XXC_{394}XXC_{397}$ was also shown to be essential for Ero1α activity.

The sequence of ERO1-Lα was used to screen sequence databases for similar sequences. The ERO1-Lβ gene was discovered, which has a ~75% similarity and ~65% identity to ERO1-Lα (~50%/~75% for yeast ERO1). One of the main conserved features was a redox-active CXXCXXC motif, with one of the notable key differences to Ero1α being the presence of 4 N-linked glycosylation sites. ER residency of Ero1β-myc was shown in transfected cells, with Ero1β co-localising with PDI and calnexin.

(Cabibbo *et al*., 2000, Pagani *et al*., 2000). As shown previously with Ero1α, Ero1β was also able to partially complement the ero1-1 mutant, and allow growth at 37 °C (Pagani *et al*., 2000). In yeast, tunicamycin and DTT can induce the UPR, and concomitant expression of Ero1p; such stresses when applied to mammalian cells over 6 hours failed to show upregulation of Ero1α, although the converse was true for Ero1β.

The CXXCXXC C-terminal redox-active motif was shown to have a role in the structure, folding and stability of human Ero1α in a series of mutation and pulse chase experiments (Benham *et al*., 2000). Transfected Ero1α exists in three states; a reduced state, R and two oxidised states OX1 and OX2. Cysteine mutants of the CXXCXXC motif showed distinct folding patterns, and the Cys391Ala mutant loses its abilty to interact with PDI. Non-reducing gel analysis showed that Ero1α was tranlsated as a reduced glycosylated form, R, before assuming one of two oxidised states, OX1 or OX2. Ero1α formed complexes with PDI, since Ero1α could be immunopreciptated with a PDI anti-serum.

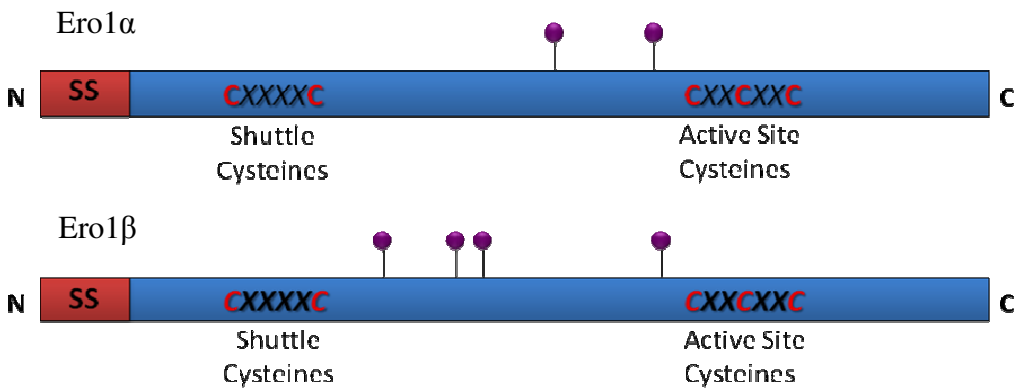

Figure 1.4: Stick diagrams representing Ero1α and Ero1β

Ero1α and Ero1β share structural similarities such as the shuttle and active site cysteines. Note that Ero1α has only two glycosylation sites, whereas Ero1β has four. 'SS' denotes signal sequence. Each stick represents an N-glycan. Adapted from Dias-Gunasekara and Benham, 2005.

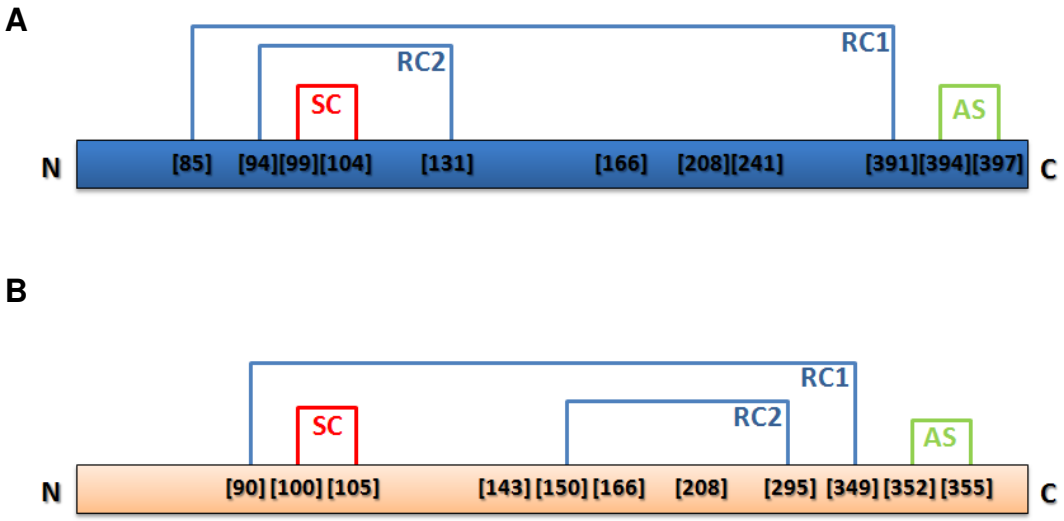

Figure 1.5: Stick diagrams comparing Ero1α and Ero1p structural disulphide components

Structural disulphides of Ero1α (A) and Ero1p (B). RC1 Regulatory cysteines 1; RC2 Regulatory cysteines 2; SC Shuttle cysteines; AS Active site cysteines.

When Ero1α was translated in the absence of an ER membrane, the same three oxidation states, R, OX1 and OX2 were seen on gel, and could be resolved to a single band following reduction by DTT. This shows that these states are dependent on the presence of specific disulphide bonds. Cys/Ala mutants of the Ero1α $C_{391}XXC_{394}XXC_{397}$ motif showed a varying propensity for forming the three oxidation states. Cys391Ala showed the R form, and no distinct OX1 or OX2, but a smear of intermediates. Cys394Ala showed R, OX1 and OX2. Cys397Ala showed mostly R and OX1, with little progression to OX2, the native form of Ero1α. Cys391Ala and Cys394Ala were unstable, whilst most of the C397Ala reached the OX2 state and formed a stable complex with PDI.

Flavin adenine dinucleotide (FAD) is an electron-acceptor in the Ero1p-Pdi1p pathway (Tu *et al.*, 2000, Tu and Weissman, 2002, Papp *et al.*, 2005). Using yeast *ero1-1* and *fad-1* strains and crosses, the temperature sensitivity of the *ero1-1* strain was suppressed in response to increased expression of FAD1 (Tu *et al.*, 2000), as opposed to total FAD, which was protein bound (Tu and Weissman, 2002). The FAD-Ero1p interaction was confirmed *in vitro* using a labelled FAD analogue, azido-FAD. Using scrambled RNAse A as a model client, physiological levels of FAD could augment folding, judged by reconstitution of RNAse A activity. Taken together, this showed that Ero1p is a flavoenzyme. Under aerobic conditions, a time-dependent consumption of molecular oxygen by Ero1p was seen, and two disulphides were formed per oxygen molecule. This was dependent on Ero1p, PDI and FAD. Under anaerobic conditions, an Ero1p PDI reaction mixture supplemented with 5μM FAD did not result in efficient oxidative processing of RNAse A after 20 minutes, nor was this affected by increasing FAD concentration. This shows that FAD alone, without the presence of oxygen, is unable to accept electrons. Bacitracin can block oxidative folding by inhibiting the oxidation of PDI and microsomal proteins, but not Ero1 (Papp *et al.*, 2005).

Two crystal structures of a functional Ero1p fragment, termed Ero1p-c have been solved (Gross *et al*., 2004). The structures showed that the CXXCXXC active site is in contact with the isoalloxazine ring of a non-covalently bound FAD molecule, which is located between the α2 and α3 helices in a bent conformation. The Ero1p-c crystal revealed that the Cys100=Cys105 bond forms a flexible loop which lacks a secondary structure and which can adopt one of two conformations. These are in fact the "shuttle" cysteines (CXXXC motif) which are responsible for the direct oxidation of PDI. These then transfer electrons to the CXXCXXC active site. It has been proposed that the polypeptide loop is in fact regulated by the differential oxidation or reduction of Cys90-Cys349 and Cys150-Cys295 (Sevier *et al.*, 2007, Sevier and Kaiser, 2008). This is based on experiments with Ero1p-c (truncated form) and reduced thioredoxin (TRX1; from *Escherichia coli*). Oxygen consumption was increased three-fold in a Cys150Ala-Cys295Ala (regulatory cysteines) mutant. The observation that the first reduced form of Ero1 co-migrates on SDS-PAGE with the Cys150Ala-Cys295Ala mutant suggests that the Cys150-Cys295 is reduced first (step 1), which is then followed by the concomitant reduction of Cys90-Cys349 (step 2). Step 1 has been shown to occur independently of glutathione, unlike step 2 (Sevier *et al.*, 2007). However, the reductant needed for completion of step 1 has not been identified *in vivo* (Sevier *et al.*, 2007). Therefore, following the two-stage reduction of the regulatory cysteines Cys150-Cys295 and Cys90-Cys349, the polypeptide loop is retracted, unshielding the shuttle cysteines, Cys100-Cys105. In this way, "activated" Ero1 directly oxidises PDI, taking electrons away from PDI to the CXXXXC motif of the shuttle cysteines, thus reducing them. The shuttle cysteines are then re-oxidised by an internal dithiol-disulphide exchange to the CXXCXXC motif of the active site cysteines. These in turn are then re-oxidised when the electrons are then transferred to FAD and molecular oxygen, in a

process which produces hydrogen peroxide (Sevier *et al.*, 2007, Sevier and Kaiser, 2008).

Human Ero1α is regulated by Cys85-391 and Cys94-131, but the cysteines involved are different from those regulating yeast Ero1p (Cys90-349 and Cys150-295). Mutation analysis of Ero1α showed the importance of these regulatory disulphides for function and that breaking them causes conformational changes (Baker *et al.*, 2008, Appenzeller-Herzog *et al.*, 2008). Recombinant Ero1α-HIS (HIS tag) was shown to be reactive with thioredoxin, resulting in complete oxidation within 10 minutes, and near-complete oxidation of mammalian PDI. The interaction between Ero1α-HIS and PDI was sustained only in the presence of glutathione. Using thioredoxin assays the reduction potential of the Ero1α regulatory disulphides was calculated to be -275mV, and therefore are highly stable upon reaction with PDI (-180mV), which could not fully oxidise Ero1α-HIS. The rate of reformation of these regulatory disulphides was slower than the rate of substrate oxidation, which suggested that Ero1α could perform a number of oxidation reactions with PDI before becoming inactive again, and that the Ero1α disulphide pair C94-C99 was required to directly engage with PDI (Baker *et al.*, 2008)

Two crystal structures at 0.9Å of "hyperactive" Ero1α OX1 (C104A, C131A) and "inactive" Ero1α OX2 (C99A, C104A) have been published (Inaba *et al.*, 2010). Crystals of full-length Ero1α could not be solved without selectively deleting these functional bonds, due to the number of oxidative states that Ero1α can adopt. These structures showed that the FAD co-factor comes within 3.30Å of the active site of Ero1α. The authors inferred from this that the mechanism of *de novo* disulphide bond formation in Ero1α is a result of the thiolate anion of the active site, C397, forming a charge-transfer complex and a covalent bond with FAD (Figure 1.6).

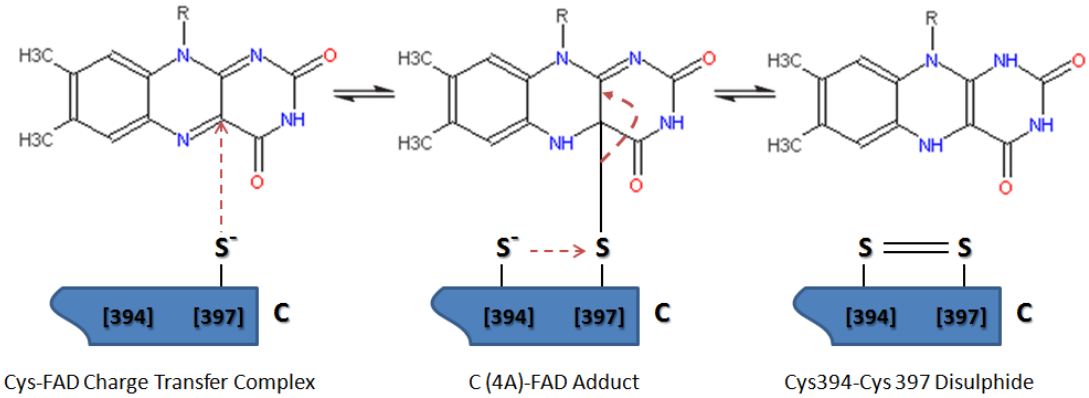

Figure 1.6: Diagram showing proposed mechanism of disulphide formation in Ero1α

The proposed mechanism of disulphide bond formation in Ero1α is a result of the thiolate anion of the active site, C397, forming a charge-transfer complex and a covalent bond with FAD. A carbon-sulphur adduct then undergoes nucleophilic attack from C394, yielding a C394-C397 disulphide bond, and reduced FAD (adapted from Inaba *et al.*, 2010).

A carbon-sulphur adduct then undergoes nucleophilic attack from C394, yielding a C394-C397 disulphide bond, and reduced FAD.

The Ero1α OX2 state is the predominant form found in living cells (Benham *et al*., 2000). The Ero1α OX1/OX2 ratio was diminished in PDI knockouts, suggesting that the Ero1α regulatory cysteines are modulated by PDI, which causes an increase in the amount of expressed Ero1α OX1 (Appenzeller-Herzog *et al.*, 2008, Inaba *et al.*, 2010).

The structure, activity and regulation of Ero1β has been less well studied than Ero1α. Ero1β-PDI complexes, at least in transfected HeLa cells, can depend on both covalent and non-covalent interactions (Dias-Gunasekara et al., 2005). Cysteine to alanine substitutions in the Ero1β active CXXCXXC site showed that Cys393Ala and Cys396Ala could weakly dimerise with PDI, but not Cys390Ala. The Cys390Ala mutant was still able to complex with non-PDI proteins, based on the continuing presence of alternate high molecular weight bands following removal of PDI by immunprecipitation. SDS-PAGE analysis confirmed that Ero1β was capable of forming homodimeric complexes, dependent on the CXXCXXC motif. Interestingly, Ero1β Cys390Ala and Cys393Ala mutants were still able to dimerise with wild type Ero1β, whereas Cys396Ala homodimers could not be shown. Modelling of Ero1β on the Ero1p-c crystal structure (Gross et al., 2004) showed that the two FAD moieties of each counterpart of the Ero1β-Ero1β dimer might come into close contact, and that mutation of Cys396 could disrupt FAD binding and homodimerisation. Ero1α and Ero1β can also form heterodimers in transfected cells, that are partly dependent on the CXXCXXC motifs of both Ero1α/Ero1β (Dias-Gunasekara *et al.*, 2005).

In addition to the Ero1-PDI disulphide relay, the flavin-linked quiescin-sulphydryl oxidase (QSOX) enzyme also serves to introduce disulphides into reduced proteins. *In vitro* studies have shown that QSOX can install disulphide bonds at a rate of hundreds or thousands per minute. Unfolded, reduced protein substrates supply their thiolate

anions to a PDI-like CGHC motif to the N-terminal thioredoxin domain of QSOX. Electrons are then transferred to the Erv/ALR-like domain, and on to the reduction of molecular oxygen producing hydrogen peroxide (Rancy and Thorpe, 2008).

The peroxide produced during disulphide bond formation is a source of reactive oxygen species, and hence oxidative cell damage. There exists a number of mechanisms to break down peroxide, including direct reaction with glutathione, catalase, and glutathione peroxidises, which can degrade free peroxide and lipid hyperoxides (Valko *et al.*, 2007). The six isoforms of mammalian peroxiredoxins (Prx) (Hofmann *et al.*, 2002, Wood *et al.*, 2003) also serve to degrade peroxide and each have specific locations throughout the cell, including the cytosol, nucleus, mitochondria and peroxisomes. Prx IV has also been shown to reside in the ER; despite having no known ER retention signal (Tavender *et al.*, 2008). This may mean it is retained in a manner similar to that of Ero1, via interations with PDI and Erp44 (Otsu *et al.*, 2006). Prx I, II, III, and IV are termed "two-cysteine" Prx, referring to the peroxidatic and so-called "resolving" cysteines. Their mode of action is to eliminate peroxide, and in doing so, the peroxidatic cysteine becomes a sulphenic acid species, and becomes part of a stable bond upon reacting with the resolving cysteine partner (Ellis and Poole, 1997, Hirotsu *et al.*, 1999). This bond may be re-oxidised by a putative as yet unknown cell-specific disulfide reductase (Tavender *et al.*, 2008)

1.4.2.4 Protein Disulphide Isomerase, PDI

PDI is a 57 kDa soluble thiol-disulphide isomerase, which is glycosylated in yeast, but not in higher order eukaryotes (Figure 1.7). It is retained in the ER by the C-terminal retention sequence, KDEL and is normally found in a stable, oxidised state. The function of PDI was initially believed to be solely dependent on glutathione, although the publication and characterisation of the essential oxidoreductase, Ero1p showed this

not to be the whole story (Frand and Kaiser, 1998). The relationship between the four thioredoxin-like domains of PDI, **a, a', b, b'**, was first characterised by limited proteolysis experiments (Freedman *et al.*, 1998, Darby *et al.*, 1996), and has since been shown in the crystal structure of yeast PDI, Pdi1p (Tian *et al.*, 2006). This landmark paper showed that Pdi1p assumes a U-shaped formation, with the **a** and **a'** domains facing each other at the tip of the arms, similar to the V-shaped arrangement seen in prokaryotic analogue, DsbC (Tian *et al.*, 2006, Heras *et al.*, 2007). In 2008, it was shown that PDI can exist in an alternate "open boat" conformation, and that this conformational flexibility is crucial for catalytic activity (Tian *et al.*, 2008).

The **a** and **a'** domains house redox active CXXC motifs, and have a measured redox potential of -188 mV and -152 mV respectively, in yeast Pdi1p (Tian *et al.*, 2006). An earlier experiment using bovine liver-derived PDI measured the redox potential of the active domains at around-175 mV. The residues that separate the two cysteines of the **a** and **a'** CXXC motifs modulate redox potential, and define its role as either a thioldisulphide reductase, oxidase or isomerase. (Grauschopf *et al.*, 1995). The **a** domain is believed to subserve the isomerase function by forming a dithiol, whilst the **a'** domain subserves the disulphide oxidase function, by forming a disulphide (Kulp *et al.*, 2006). This was shown using a reconstituted Ero1p folding pathway, and mutational analysis of either the **a** or **a'** redox active site cysteines switched to alanines. This activity of PDI is partly dependent on the asymmetrical arrangement of the **a** and **a'** domains on either arms of the U-shaped 3D structure, and the presence of the two remaining redox-inactive **b** and **b'** domains.

The **b** domain is believed to have a structural role, despite containing a thioredoxin fold (Kemmink *et al.*, 1997). Truncation studies of the **b** domain in yeast Pdi1p show little change in the folding kinetics of reduced RNase, but this did decrease the rate of scrambled RNase refolding (Tian *et al.*, 2006). Using *in vitro* enzymology and

truncation assays, the **b'** domain has been shown to be the principle peptide binding site, and is essential for simple isomerisation reactions, though all four thioredoxin domains are required for complex isomerisations (Darby *et al.*, 1996, Klappa *et al.*, 1998). The crystal structure confirmed previous biochemical data regarding the nature of a hydrophobic substrate binding pocket in the **b'** domain (Pirneskoski *et al.*, 2004). Mutations designed to affect the structure of this pocket in human PDI (Pirneskoski *et al.*, 2004) showed a marked reduction in known substrate binding (based on Pdi1p in Ruddock *et al.*, 2000).

In order for PDI to catalyse disulphide-isomerisation reactions with a substrate, a degree of mobility is inferred from PDIs dual catalytic and chaperone roles. To date, no full-length structure of PDI has been published, suggesting a degree of heterogeneity in structural conformation (Darby *et al.*, 1998, Freedman *et al.*, 2002). Proteolysis of native PDI with proteinase K, chymotrypsin or trypsin, consistently targets the x-linker and **a'** domains in the C-terminus region, leaving the N-terminus mostly intact (Wang *et al.*, 2010). The x-linker region can adopt alternate confirmations. These are the so-called "capped" or "uncapped" conformations, hence shielding or exposing the **b'** domain ligand binding site. This was shown using tryptophan residue W347 in the x-linker region as a reporter of intrinsic fluorescence emission spectra. The mutants I272A and L343A of a **b'**x construct force the x-linker region to adopt the capped or uncapped conformation respectively (Nguyen *et al.*, 2008), and when reconstructed in full-length PDI, the **a'** region was shown to be responsible for moderating this ability of the x-linker (Wang *et al.*, 2010). Denatured, reduced rhodanese can be used to measure the chaperone properties of a given protein, with aggregation analysed by light scattering. The capped conformation of the x-linker showed a lower rhodenase chaperone activity (i.e. suppression of aggregation is less). In RNAse reactivation experiments, a measure of the isomerase activity of PDI, the rate is similar between

wild type PDI and the capped/uncapped mutants, I272A and L343A. In the reduction of 130 μM insulin by 0.1 mM DTT, the uncapped conformation (I272A) had an enhanced reductase activity, relative to regular PDI, and the capped conformation showed decreased activity.

PDI is known to have at least 19 homologues in humans, often referred to as the "PDI family" (Table 1.1; for a review see Ellgaard and Ruddock, 2005). Not all members of the family are capable of efficiently catalysing disulphide bond formation, and vary in their similarities and differences. The main four main determinants of PDI activity are the active site sequence (CXXC, CXHC), the presence/absence of an arginine residue shown to modulate the pK_a of the active-site cysteines (Lappi *et al.*, 2004), the presence or absence of a glutamic acid-lysine charged pair subserving proton transfer, and the presence of a high-affinity binding site in order to catalyse reactions. Other PDI family members are less well studied but show variation in the **a** and **b** domain arrangements and have divergent ER retention sequences (summarised in Ellgaard and Ruddock, 2005).

The now classical yeast studies of Ero1 and PDI show that Ero1 directly oxidises PDI, which in turn directly oxidises substrate. Ero1 is crucial for the oxidation and folding of the model protein RNAse A via PDI, without the presence of GSH or GSSG, or even in an excess GSH environment (Tu *et al.*, 2000). A link between Ero1 and glutathione is believed to exist, and that this serves in the reduction of non-native disulphides prior to isomerisation. GSSG regeneration is slower in an *ero1-1* mutant, and enhanced in an Ero1 transfectant (Cuozzo and Kaiser, 1999). Despite this, *in vitro*, glutathione is a poor substrate for Ero1 (Tu and Weissman, 2002). Two studies both show that in a low-glutathione system, non-native disulphide bond formation increases in model proteins such as tissue-type plasminogen activator (tPA), ultimately leading to an increase isomerisation time. By inference, this suggests a role for a glutathione system as part of

the normal isomerisation process (Chakravarthi and Bulleid, 2004, Molteni *et al.*, 2004). This may be achieved through other members of the PDI family, such as ERp57, which has been shown to form mixed disulphides with biotinylated glutathione, and can be reduced by GSH, in microsomes derived from the fibrosarcoma cell-line, HT1080 (Jessop and Bulleid, 2004).

The relationship of Ero1α with other PDI family members is not clear. This was explored by comparing the rates of oxidation of PDI, ERp57, and **b'**-domain chimeras, by Ero1α. The reaction of Ero1α with wildtype ERp57 is relatively slow compared to that of wildtype PDI. The reaction rate of ERp57 could be increased by swapping its **b'**-domain with that of PDI, and conversely, the reaction rate with PDI could be slowed by swapping its **b'**-domain with that of ERp57. The authors concluded that interaction with Ero1α was via the **b'**-domain of PDI, and that this activity is modulated by disulphide bond rearrangements amongst the regulatory cysteines.

Recent data, using reductive challenge experiments with DTT, shows that reoxidation of PDI and reduced glutathione occurs very quickly, and that this is dependent on the presence of Ero1 and PDI (Appenzeller-Herzog *et al.*, 2010). Despite such a challenge, accumulation of oxidised glutathione in the ER is relatively short lived, and the ratio of oxidised/reduced glutathione returns to a baseline level after a few minutes. The authors note that this cannot be explained by export of oxidised glutathione, shown to be slow by the Banhegyi group (Banhegyi *et al.*, 1999), or by the import of reduced glutathione. Appenzeller-Herzog *et al*. suggest that reduced glutathione can compete with PDI for reactive substrates, resulting in the production of oxidised glutathione, and that low availability of reduced PDI will result in large-scale inactivation of Ero1α and shutdown of its oxidation pathway (Appenzeller-Herzog *et al.*, 2008, Appenzeller-Herzog *et al.*, 2010).

Using purified recombinant Ero1α, the activity of Ero1α on PDI required glutathione, as seen by measuring oxygen consumption (Baker *et al.*, 2008). Mechanisms of glutathione homeostasis and transport of glutathione across the ER membrane are unclear, although the principles of potential mechanisms have been reviewed by Chakravarthi *et al*., 2006.

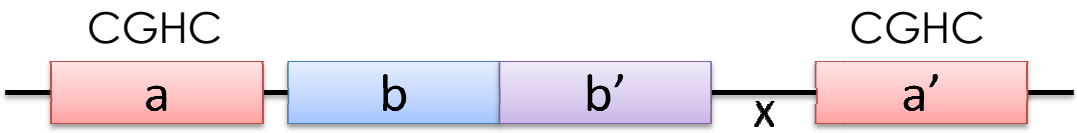

Figure 1.7: Schematic representation of human PDI

The catalytic thioredoxin-like domains are **a** and **a'**. The **b** and **b'** domains are non-catalytic thioredoxin-like domains. 'X' denotes the x-linker region.

(Adapted from: Wang *et al.*, 2010)

Table 1.1 A summary of the 19 PDI family members (known to date)

Member Name	Amino Acid Length	Number of a domains	Domain composition
Hag 3	165	1	a
ERp18	172	1	a
Hag 2	175	1	a
ERp28/29	261	0	b-D*
ERp27	273	0	b-b'
ERp44	406	1	a-b-b'
ERp46	432	3	a°-a-a'
P5	440	2	a°-a-b
ERp57	505	2	a-b-b'-a'
PDI	508	2	a-b-b'-a'
PDIr	519	3	b-a°-a-a'
PDIp	525	2	a-b-b'-a'
PDILT	584	2	a-b-b'-a'
ERp72	645	3	a°-a-b-b'-a'
ERdj5	793	4	J^{+}-a''-b-a°-a-a'
TMX	280	1	a
TMX2	296	1	a
TMX4	349	1	a
TMX3	454	1	a-b-b'

The table shows the amino acid length of the family member, as well as domain-specific information (adapted from Ellgaard and Ruddock, 2005 and Appenzeller-Herzog and Ellgaard, 2008).

***** PDI-D members (i.e ERp28/ERp29) have one or two thioredoxin-like domains followed by the novel α-helical C terminal D-domain (Ferrari and Soling, 1999, Ma *et al.*, 2003).

$^{+}$ The PDI-J protein, ERdj5, contains a DnaJ domain as well as PDI and thioredoxin domains. ERdj5 interacts with BiP in an ATP-dependent manner, via the DnaJ domain (Cunnea *et al.*, 2002).

1.4.3 The Unfolded Protein Response and ERAD

The process of protein folding within the ER is susceptible to simple perturbations of the cellular environment, including glucose deprivation and hypoxia (Lee, 2001, Gess *et al.*, 2003, Rutkowski and Kaufman, 2004). The homeostatic response to the presence of misfolded proteins, either as a result of enhanced protein load, or environmental perturbation is the unfolded protein response (UPR). Misfolded proteins either accumulate in the ER lumen, or can be sent for degradation via the ER-associated degradation pathway (ERAD) (Rutkowski and Kaufman, 2004, Schröder and Kaufman, 2005).

The three main ER sensors of stress are IRE1, PERK and ATF6. IRE1 (α and β) is both a kinase and an endoribonuclease, PERK is a kinase and ATF6 (α and β) is a basic leucine zipper transcription factor (Haze *et al.*, 1999, Shen *et al.*, 2002, Liu *et al.*, 2003). The ER chaperone, BiP is bound to all three of these sensors to their proximal, lumenal facing regions, and in this way keeps them in a state of deactivation (Schröder and Kaufman, 2005).

BiP is an abundant lumenal ER chaperone, which contains an N-terminal ATPase, and a C-terminal substrate binding domain (Gething, 1999). When exposed, hydrophobic regions on proteins undergoing the folding process bind to the C-terminus of BiP, the ATPase is activated, and ATP is hydrolysed to produce an ADP-bound version of BiP, which acquires a high affinity for its substrate (Gething, 1999). When ADP is exchanged with ATP, BiP releases its substrate to undergo further folding.

If a large number of unfolded proteins accumulate in the ER lumen, then BiP is titrated away from the lumenal facing regions of the ER stress sensors IRE1, PERK and ATF6, and preferentially binds to the unfolded proteins (Schröder and Kaufman, 2005).

When relinquished from BiP, PERK phosphorylates the α-subunit of eIF2α, which is a translation initiation factor involved in 80S ribosome assembly, and as a result translation is repressed.

When freed from BiP, ATF6 travels to the Golgi apparatus, and is cleaved by the Site-1 serine protease and the Site-2 metalloprotease, releasing the transcription factor moiety which binds to CRE and ERSE-1/2 promoter regions of UPR responsive genes in the nucleus (Wang *et al.*, 2000, Kokame *et al.*, 2001).

IRE1 is the slowest acting component of the triumvirate of ER stress sensors, given the lag time needed for ATF6 activation and XBP1 expression. (Yoshida, 2003). XBP1 mRNA is the substrate of the IRE1 nuclease domain, which removes an intron from XBP1 mRNA, causing a frameshift which yields an active transcription factor (Shen *et al.*, 2001, Yoshida *et al.*, 2001, Calfon, 2002, Lee *et al.*, 2002). XBP-1 is a bZIP transcription factor of the ATF/CREB family and controls X-box genes.

The end result of the activation of UPR sensors is a downregulation of genes controlling protein translation and, in parallel, the upregulation of protein clearance mechanisms, to ensure a decrease in folding demand (Harding *et al.*, 1999). The three arms also activate the transcription factors ATF6, XBP-1, ATF4, and Nrf2. In addition, ER luminal surface area increases, and genes encoding ER resident chaperones and foldases are upregulated (Schröder and Kaufman, 2005). Genes induced by the UPR are extensive (reviewed by van Laar *et al.*, 2001). These include Herp/Mif1via ATF4/ATF6 (Ma and Hendershot, 2004), which recruits the 26S proteasome to degrade misfolded proteins, and EDEM (via XBP1 and ATF6, Fewell *et al.*, 2001) which is a lectin-like ER resident protein with a role in recognising proteins for degradation.

Two ER-stress induced routes exist that result in apoptosis; the intrinsic (responding intracellular insults, receptor-independent) and extrinsic (responding to extracellular stimuli, receptor dependent) pathway. The intrinsic pathway is triggered by the

insertion of proapoptotic proteins Bak and Bax into the ER membrane, and results in calcium ion efflux from the ER (Zong *et al.*, 2003, Scorrano *et al.*, 2003). This in turn activates calpain, which cleaves and hence activates caspase-12 (Nakagawa and Yuan, 2000). Active caspase-12 cleaves and activates caspase-9, which in turn activates the "executioner" caspase-3, resulting in apoptosis (Rao *et al.*, 2002). The ER calcium efflux described previously can also lead to a collapse of the mitochondrial membrane potential, the formation and activation of the apoptosome, and the operation of the permeability transition pore that releases cytochrome c into the cytoplasm (Crompton, 1999).

The long term consequences of ER stress can extend beyond the ER to the organism itself, with the aggregation of degradation resistant proteins the basis of many diseases, including Alzheimer's disease, Huntington's disease and ankylosing spondylitis (Kaufman, 2002, Ramos and Lopez de Castro, 2002).

1.5 Barrett's Oesophagus and Oesophageal Adenocarcinoma

The human oesophagus is a muscular conduit from the pharynx to the stomach, lined with stratified squamous epithelium and is bounded by the upper and lower oesophageal sphincters. The lower oesophageal sphincter is located at the transition from low intra-thoracic pressure to high intra-abdominal pressure, and normally prevents acidic stomach refluxate from entering the oesophagus, although a number of neurohumoral and exogenous factors can either increase of decrease lower oesophageal sphincter pressure (McPhee *et al.*, 1995).

Barrett's oesophagus is a pre-malignant metaplastic disease in which the normal stratified squamous epithelium of the oesophageal tract changes to become like stomach or intestinal epithelium, or a mixture of both (Barrett, 1950, Allison and

Johnstone, 1953). This in turn is associated with an annual conversion of between 1-5% into oesophageal cancer (Schnell *et al.*, 2001, Jankowski *et al.*, 2002).

The aetiopathogenesis of Barrett's adenocarcinoma and gastric cancer are not well defined, although patients with gastro-oesophageal reflux disease (GORD) are known to be at risk. A causal agent of adenocarcinoma in the GI tract is believed to be gastric refluxate, which in 48% of 8-week old male Sprague-Dawley rats over 16 weeks, was shown to induce oesophageal cancer in the absence of any other agents (Fein *et al.*, 1998). In addition, cell culture experiments varying pH and exposure to bile acid preparations have been used to search for putative biochemical markers and alterations in gene transcription (Morgan *et al.*, 2004, Hu *et al.*, 2007). Ultimately, biomarkers are needed in order to diagnose these conditions earlier, and to improve survival (Fitzgerald, 2006).

The precise mechanism of how exposure to reflux agents leads through a progression of metaplasia to neoplasia in the upper GI tract is uncertain, not least because of the heterogeneous nature of disease manifestations between patients, and their refluxate constituents (Atherfold and Jankowski, 2006). It was previously believed that acid alone was the principle causative agent, via a mechanism of caustic erosion, and patients with GORD are commonly prescribed with acid-suppressing drugs. Whilst this improves symptoms of GORD, there is evidence to suggest that this does not affect the incidence trends of Barrett's oesophagus (Prach *et al.*, 1997, Pohl and Welch, 2005), and may actually contribute to the increase in incidence of adenocarcinoma (Stamp, 2002). This has considerable implications for current prescription practices.

Although the precise composition of reflux constituents varies between patients (and hence publications), "bile" has been shown to present in the oesophagus and stomach of around ~86% (of 37) patients with GORD, at a concentration of around 200 μM (Gillen *et al.*, 1988, Gotley *et al.*, 1988, Kauer *et al.*, 1997). In normal volunteers, 58%

of this group also had bile acid in gastric aspirate. Secondary bile acids are associated with extensive mucosal damage, and the presence of these along with acid (so called mixed-reflux) is more damaging than acid alone (Nehra *et al.*, 1999).

1.5.1 Bile acids

Bile salts are strong ionic amphipathic detergents, and have the potential to disrupt cell membranes, and even enter cells and cellular organelles, because of these properties (Batzri *et al.*, 1991).

Bile is released from the bile duct following the entry of digesting food into the duodenum, and serves to emulsify digested fats by forming mixed micelles, and keep fat-soluble compounds such as vitamins or cholesterol in solution for absorption or excretion respectively. It is important to differentiate between bile salts, bile acids and bile itself, which is a complex mix of bile salts, cholesterol, phospholipids and bilirubin. Bile acids are cholesterol-derived insoluble polar lipids in a protonated form, irrespective of conjugation with glycine or taurine; these may be natural compounds, or structurally similar artificial compounds. Bile salts are bile acids compounded with a cation, such as Na^+, K^+ or Ca^+ and are highly soluble in water (Fini *et al.*, 1985).

In the physiological environment of gastric reflux, it is believed that the bile component could maintain an ability to enter cells in a fluctuating pH environment. Cell entry is achievable when bile is either uncharged (above pKa value), protonated (at normal stomach pH) or if there is a low critical micelle concentration. The other putative suspects in gastric refluxate are hydrochloric acid and pepsin (active at <pH2), which would be neutralised by the combined influences of oesophageal bicarbonate, pancreatic and duodenal juice (Stamp, 2002).

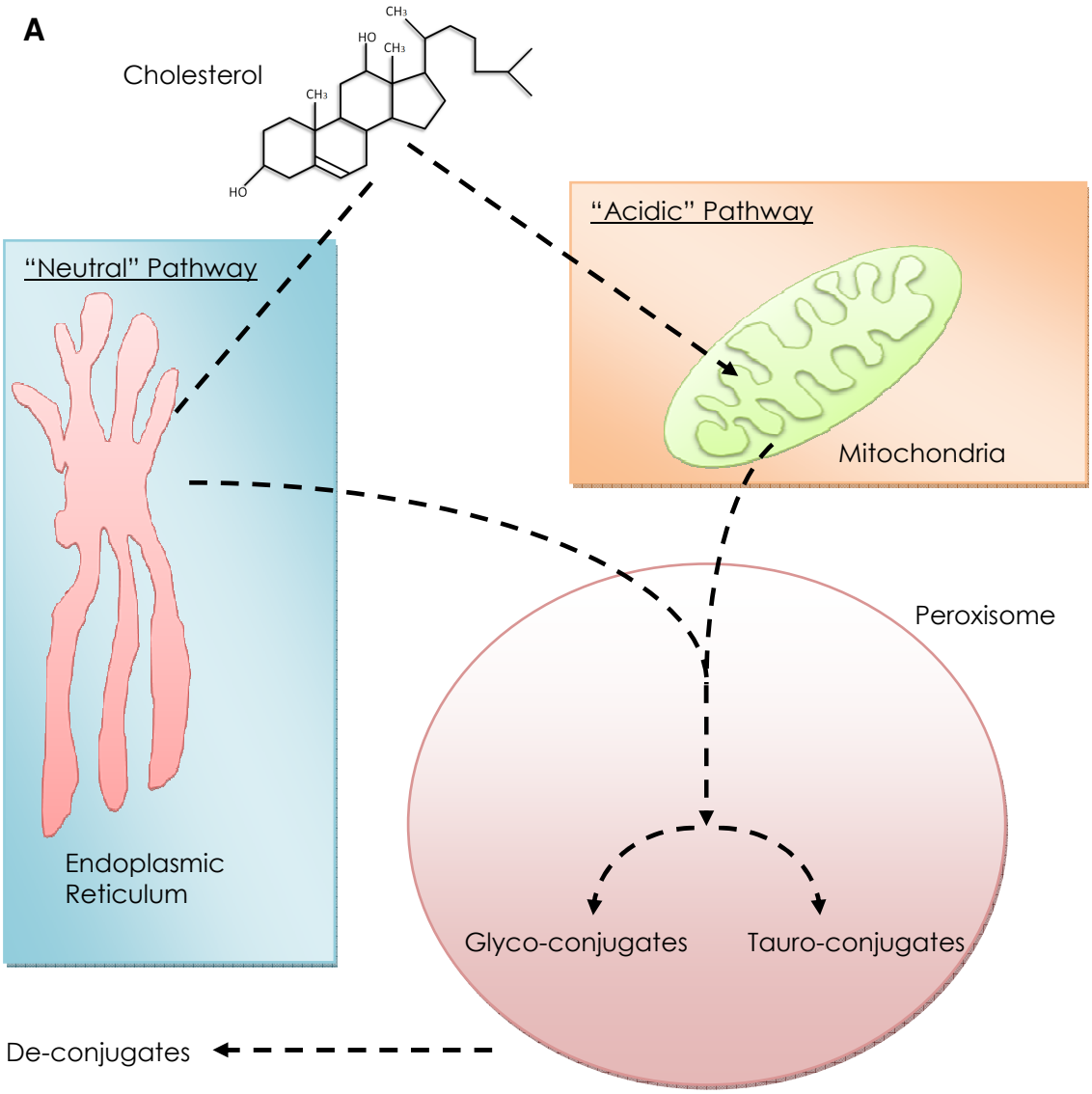

Figure 1.8 Overview of bile acid synthesis

Bile salts are synthesized in hepatocytes. The main entrant into the either the neutral or acidic pathway is cholesterol. The rate-limiting enzyme of the ER "neutral" pathway is CYP7A1. The rate-limiting enzyme of the mitochondrial "acidic" pathway is CYP27A1. Subsequent conversions are carried out by enzymes residing in the ER, cytoplasm and peroxisome. Glyco- or Tauro- conjugation is restricted to the peroxisome. De-conjugation of bile-acids occurs in the intestine, although bile acids can be re-conjugated by re-gaining entry into the peroxisome (simplified from Pellicoro and Faber, 2007).

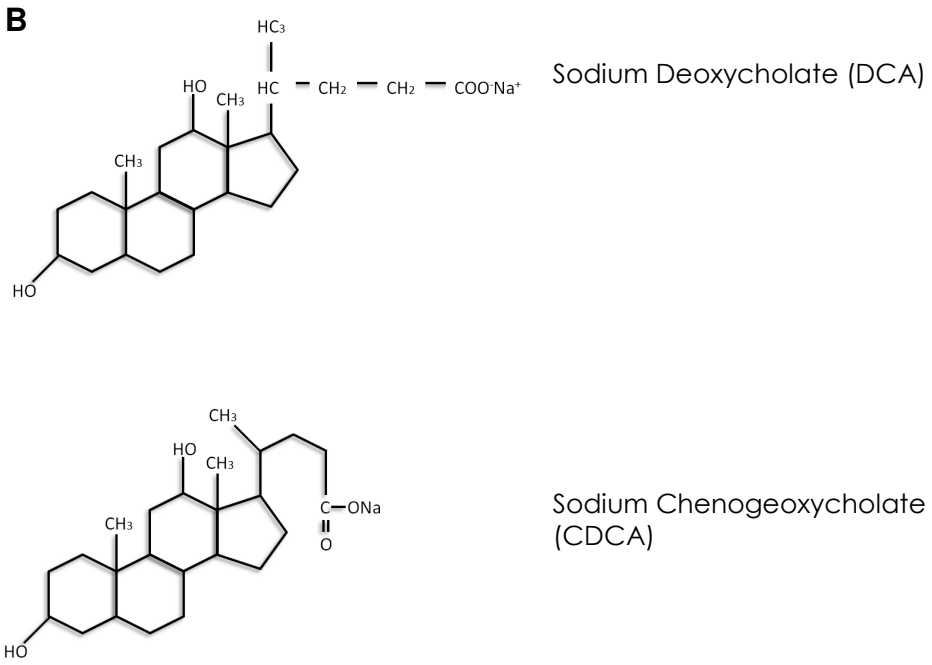

Figure 1.8 (Cont) Overview of bile acid synthesis

B: Structures of the sodium salts of deoxycholic and chenodeoxycholic acid.

1.5.2 Synthesis, structure and properties of bile acids

The synthesis of bile salts involves the ER, peroxisome and mitochondria (see Figure 1.8a). Most bile salts are made through a cholyl CoA intermediate, following the conversion of cholesterol into trihydroxycoprostanoate. Glycocholate is the major bile salt, created when the activated carboxyl carbon of cholyl CoA reacts with the amino group of glycine. Similarly cholyl CoA may react with the amino group of taurine, which is derived from cysteine, to form taurocholate. This process is known as conjugation (for review see Pellicoro and Faber, 2007). Three major modifications are required to convert cholesterol into cholic acid or chenodeoxycholic acid: hydroxylation, side-chain shortening, and glycine/taurine conjugation. These reactions occur in different cellular compartments, in either the neutral (classical) or acidic pathway, differentiated by the order of transformative reactions and cellular location (Pellicoro and Faber, 2007).

The neutral pathway commences in the ER of hepatocytes, with the 7α-hydroxylation of the steroid nucleus (C7) by cholesterol 7α-hydroxylase, CYP7A1 (Myant and Mitropoulos, 1977). In contrast, the acidic pathway commences in the mitochondria, with the hydroxylation of the cholesterol side chain at C27 by sterol 27-hydroxylase, CYP27A1 (Cali and Russell, 1991). The CYP27A1 product, 5-cholesten-3b-27-diol, cannot be hydroxylated by CYP7A1, though this is achieved via the alternate enzyme, CYP7B1, within the ER (Toll *et al.*, 1994). At this point, the pathways effectively merge. The ring modification proceeds with catalysis by 3b-hydroxy C27-steroid dehydrogenase/isomerase (3bHSD/HSD3B7) and sterol 12-a-hydroxylase (CYP8B1) in the ER (Gafvels *et al.*, 1999, Schwarz *et al.*, 2000). CYP8B1 hydroxylates carbon 12, and as such, determines the ratio between double-hydroxylated chenodeoxycholic acid (CDCA) and triple-hydroxylated cholic acid (CA) produced.

Further modifications of the carbon ring are performed by D4-3-ketosteroid-5-b-reductase (AKR1D1; Zhu *et al.*, 1996, Crystal structure; Di Costanzo *et al.*, 2008) and 3a-hydroxysteroid dehydrogenase (AKR1C4; Usui *et al.*, 1994) within the cytoplasm. The final production of conjugated primary bile salts occurs in hepatocytic peroxisomes, but the cytosol-peroxisome transport mechanism is yet to be determined. The sterol intermediate is activated following coenzyme A conjugation via bile acid coenzyme A ligase (Falany *et al.*, 2002). The bile acid precursors are transformed from 25R to 25S isomers by 2-methylacyl-coenzyme A racemase (AMACR; Cuebas *et al.*, 2002), before becoming a 24,25-*trans*-unsaturated derivative via acyl CoA oxidase (ACOX2; Baumgart *et al.*, 1996). A C24-oxo product is catalysed via a hydration/oxidation reaction at the D24 bond by the bifunctional enzyme, D-3-hydroxyacyl- CoA dehydratase/D-3-hydroxyacyl-CoA dehydrogenase (Kurosawa *et al.*, 1997). Thiolase 2 cleaves the side chain at C24-C25, leading to the production of a C24-CoA bile acid intermediate and propionyl-CoA (Seedorf and Assmann, 1991) Finally, either the amino acid glycine or taurine is added to C24 by peroxisomal bile acid coenzyme A:amino acid N-acyltransferase (BAAT; Falany *et al.*, 1994). Conjugation of bile acids increases their amphipathicity and solubility, which makes them impermeable to cell membranes. Following production, bile salt is actively transported against a steep concentration gradient from hepatocytes into bile canaliculi, before being stored in the gallbladder. One of the main bile salt export mechanisms in hepatocytes is the Bile Salt Export Pump (BSEP/ABCB11; Gerloff *et al.*, 1998) which belongs to the ATP-binding cassette transporter superfamily (ABC).

Around 97% of the bile released during digestion is reabsorbed by the epithelial cells of the ileum, in the small intestine. In a rabbit model, it was discovered that these ilial cells were unable to accumulate bile acids (Schweitzer *et al.*, 1986). The mechanics behind this remain to be fully elucidated, though it should be noted that cancer of the

small intestine is relatively rare within the GI tract, and evidently the small intestine possesses an unknown mechanism that prevents bile-acid mediated toxicity. The remaining bile acids undergo dehydroxylation and deconjugation by the colonic flora, leading to the formation of the secondary bile acids, deoxycholic acid (DCA) from cholic acid, and lithocholic acid (LCA) from chenodeoxcholic acid. There is also evidence to suggest these flora can produce free bile acids and cholesterol compounds with structural similarities to polycyclic aromatic hydrocarbons, which are carcinogenic (Stamp, 2002).

Bile acids (or salts) are capable of forming micelles, which is essential to their function. Micelles are spherical structures, with the hydrophobic portion of the molecules facing inward, the hydrophilic portion facing outward. Micelles are formed when the concentration of bile salt is equal to its critical micelle concentration, CMC, which is dependent on the balance of the hydrophobic and hydrophilic parts of the anion, in particular, a back-to-back interaction between hydrophobic moieties (Small *et al.*, 1969).

The hydroxyl groups are a key element in driving self-aggregation in water. When a 3, 7, or 12 hydroxyl is epimerized from α to β orientation, the critical micelle concentration increases two to three fold as a result of the diminished hydrophobic area of the β face (Roda *et al.*, 1983).

The pKa properties of bile acids normally act to protect sensitive mucosal areas. When charged ≤pKa, they cannot cross the mucosa, although when uncharged, above their pKa values, they are free to enter epithelial cells. The second way bile acids can enter mucosal cells is when they are in a protonated state, which occurs at physiological pH values. Bile acids at a concentration equal to or below their critical micelle concentrations enter the mucosa by coating membrane-spanning regions of proteins (see Stamp, 2002). At a concentration above their CMC, bile salts solubilise cellular

membranes and membrane proteins, forming mixed micelles of detergent and phospholipid together.

Recently, bile acids have been shown to be signalling molecules and the discovery of a bile activated G-protein coupled receptor, TGR5, has led to research to characterise such pathways with respect to gastric cancers (Maruyama *et al.*, 2002, Kawamata *et al.*, 2003, Yasuda *et al.*, 2007). TGR5 is known to be responsive to lithocholic acid, deoxycholic acid, chenodeoxycholic acid, ursodeoxycholic acid and cholic acid. TGR5 activation results in the downstream launch of MAPK pathways and cAMP synthesis (Houten *et al.*, 2006). Activation of MAPK pathways are responsible for invasive and metastatic properties of gastric cancers; these can be inhibited experimentally in human AGS gastric cancer cells with Phenethyl isothiocyanate (Yang *et al.*, 2010). TGR5 modulation of calcium influx/efflux from the ER has also been seen in hepatocytes, though how this may affect other tissues remains unclear (Bouscarel *et al.*, 1999, Nguyen and Bouscarel, 2008).

1.5.3 Studying gastric reflux using model systems

The study of gastric refluxate as a causative agent frequently uses surgical animal models (i.e. gastroduodenoesophageal anastomosis, GDEA), and more recently, cell culture techniques. The use of these techniques is subject to a number of caveats, not least that the squamous epithelium of the rat oesophagus is keratinising, which may confer limited protection to caustic stimuli, and library cells are grown in an monoculture environment without a basal cell layer, although the use of fibroblasts as a feeder-layer of cell signalling is beginning to address this issue.

The GDEA rat is a surgical model that results in the reflux of gastroduodenal juices into the oesophagus (Goldstein *et al.*, 1997). This GDEA procedure eradicates the

normal preventative function of the pyloric and cardia valves of the stomach, and allows gastric reflux into the oesophagus or the duodenum.

In gastrectomy (stomach removal) experiments performed on GDEA rats, there was three-times the amount of adenocarcinoma than those rats with intact stomachs. By inference, it was concluded that hydrochloric acid/stomach pepsin may have a protective role, although the mainstay of clinical practice (to date) is to prescribe acid-suppressant drugs (Ireland *et al.*, 1996). Clinical trials focus on the therapeutic benefit (i.e. reduction of symptoms), rather than the efficacy in Barrett's/adenocarcinoma reduction (Lagergren *et al.*, 1999). The biochemical significance of increasing stomach acid pH with acid-suppressants is as follows: increasing pH from ~pH4 to ~pH5 removes the charge from glycine-conjugated bile acids (which represent ~60% of total bile), allowing them to cross the epithelial barrier in the manner described above (see Stamp, 2002). This also may explain why the 58% of "normal" individuals have bile acids present in gastric juices, and yet do not have Barrett's or adenocarcinoma. Development of these conditions may also be in conjuction with altered pH.

Using radioactive bile acids and thin-layer chromatography, rabbit oesophageal epithelial cells have been shown to accumulate taurocholic acid and chenodeoxycholic acid up to seven times that of the original treatment medium. After three hours, ~50% of the bile salts remained undegraded within the cells (Schweitzer *et al.*, 1986).

It should also be noted that bile itself may also be a causative or associated agent in the aetiopathogenesis of stomach or colon cancers. In cultured stomach cells, at pH5, glycocholic acid is protonated, can travel across the gastric mucosa to the pH neutral epithelial cells, remain trapped, and accumulate up to eight times the concentration of the treatment medium (Schweitzer *et al.*, 1986, Batzri *et al.*, 1991). Stomach cancers are also often associated with *H. pylori* infection (*cag A* strain present in ~30–50% in people of the developed world), and this may itself contribute to de-protonation of bile

acids, and their enhanced toxicity. *H. pylori* persist in the stomach because they convert stomach urea into ammonia, which neutralises stomach acid. This in combination with acid suppression treatment can lead to a pH of ~8.5, and by inference further enhance bile acid toxicity (Scott *et al.*, 1998, Christensen, 1999).

Frequently, many cell culture experiments have used the secondary bile acid, deoxycholic acid (DCA) to test hypotheses relating to bile as a causative agent of gastrointestinal cancers, which has been shown to be present in patient refluxate (Gotley *et al.*, 1988, Kauer *et al.*, 1997).

Cell culture studies have also been able to demonstrate concordance with histological studies of Barrett's tissue. Crucially, such studies are an accepted model in current Barrett's research. A notable example is the demonstration in cell culture that the caudal-related homeobox gene, CDX2, can be upregulated following bile acid treatment (Hu *et al.*, 2007). CDX2 has a role in the early differentiation and maintenance of intestinal epithelium, is upregulated in Barrett's tissue showing intestinal-type metaplasia (Eda *et al.*, 2003, Phillips *et al.*, 2003, Groisman *et al.*, 2004, Lord *et al.*, 2005), and upregulated following bile acid treatment in rat keratinocytes (Kazumori *et al.*, 2006). Hu *et al*. showed a dose and time dependent relationship between deoxycholic and chenodeoxycholic acid treatment (100–1,000 μM range, 1/24 hours) and CDX2 expression. It should be noted that these cell culture treatment models are very simple representations of a complex reflux environment, involving multiple bile acids, variable pH, and exposure to other agents including stomach acid and foodstuffs.

1.6 Hypoxia and ROS in gastric cancers

In the complex environment of upper GI tumours, perturbations of the redox environment in which EROs operate may be responsible for upregulation of Ero1 expression. Changes in Ero1 oxidation state remain to be detected.

Using hypoxic animal models, cell culture, and mRNA expression analysis, Ero1α was shown to be upregulated as a result of hypoxia in all tissues. A rat cell line deficient in hypoxia inducible factor, HIF, was unable to induce Ero1α expression following hypoxia. Ero1β was not highly expressed following 4.5 hours of hypoxic conditions, but could be induced by the UPR, as expected (Gess *et al*., 2003, Pagani *et al*., 2000).

Using a rat hypoxia cDNA library, Ero1 was shown to be upregulated in global cerebral ischaemia, particularly the hippocampus and cortical areas, localised to the dendritic microdomain of neuronal cells. Initially, Ero1 was erroneously identified as Ero1-L/Giig11. Ero1 upregulation was determined by Northern blot, *in situ* hybridisation and cell culture assays (Chen *et al.*, 2003).

Reactive oxygen species are implicated in the activation of NF-κB (Jenkins *et al.*, 2004, Jenkins *et al.*, 2008b). These studies are of particular note for showing NF-κB activation by DCA at neutral pH, and the induction of IL-8. NF-κB activation is known to be associated with a variety of cancer types. This work showed that exposure of OE33 cells (a Barrett's adenocarcinoma cell line) to 100-300 μM DCA activated NF-κB and induced expression of downstream target genes, including IκB and IL-8 (Jenkins *et al.*, 2004). Notably, this was only at neutral pH, which is interesting given that bile acids are believed to have enhanced toxicity around this pH range, and the authors postulated that NF-κB activation could be driven by the bile acid in the acid-suppressed patient. In later work, NF-κB activation occurred following 100 μM DCA treatment, but not less. The mechanism is believed to involve ROS, as NF-κB

activation is abrogated with the addition of antioxidants (epigallocatechin gallate, resveratrol and vitamin C). Similarly, doses at this level or higher were capable of causing DNA damage as detected by a micronucleus assay (Jenkins *et al.*, 2008a).

In comparing basal superoxide levels of two non-malignant (CP-A, CP-C) and malignant cell lines (BE-3 and SK-4) both were shown to be similar (Chen *et al.*, 2009a). However, each of these cell lines experienced a ten-fold increase in the production of ROS following CDC treatment. Approximately 90% cell death was observed following CDC treatment, but only 20% cell death was observed in the malignant cell lines, up to a 200μM treatment. In CP-A cells, CDC treatment was associated with a 38% and 80% drop in cellular respiration at 3 hours and 8 hours respectively. In the malignant cells, cellular respiration dropped by approximately 10%. Expression of mitochondrial superoxide dismutase 2 (SOD2) and thioredoxin-2 (Trx-2) increased, and cellular glutathione levels were increased by approximately 400% (Chen *et al.*, 2009b). The authors believed the glutathione increase to be a component in the enhanced stability of the malignant cells to CDC treatment. They were able to selectively target malignant cells with honokiol (an anti-tumorigenic drug, derived from Magnolia grandifloris), which disrupted the cyclosporine-A sensitive mitochondrial permeability transition pore, via a putative interaction with cyclophilin D.

The glutathione S-transferase (GST) and glutathione peroxidise (GPX) families were compared in normal and aberrant oesophageal tissue samples obtained from the same patients (Peng *et al.*, 2009). GSTs normal function is to conjugate glutathione with electrophilic compounds, including ROS. GPX catalyses the glutathione-mediated reduction of hydrogen peroxide. The genetic analysis showed that a number of members of each family were inactivated by epigenetic hypermethylation, and as such may be responsible for the decreased ability to defend against defensible oxidative

stressors. How this relates to the observations of Chen *et al*. (2009) is unclear, given in their model a ~400% increase in glutathione was observed in malignant or non-malignant cells following CDCA treatment. The work of Peng *et al*., (2009) could suggest that glutathione functions independently of the protective GST and GPX systems.

1.7 Aims of this thesis

The main objectives of this project were:

- ★ To study the expression of Ero1α, Ero1β and other ER chaperone proteins in cell lines derived from two distinct human oesophageal cancers (a squamous carcinoma and an adenocarcinoma).
- ★ To study any potential changes in Ero1 expression as a result of exposure to different pH media and various concentrations of bile salts, the postulated causative agents of Barrett's oesophagus.
- ★ To produce recombinant Ero1β protein to assess functional properties *in vitro.*
- ★ To generate monoclonal antibodies specific to Ero1β, generate possible redox state specific antibodies for Ero1α.