Laura Silvennoinen

ERp57—CHARACTERIZATION OF ITS DOMAINS AND DETERMINATION OF SOLUTION STRUCTURES OF THE CATALYTIC DOMAINS
LAURA SILVENNOINEN

ERp57—CHARACTERIZATION OF ITS DOMAINS AND DETERMINATION OF SOLUTION STRUCTURES OF THE CATALYTIC DOMAINS

Academic Dissertation to be presented with the assent of the Faculty of Medicine, University of Oulu, for public discussion in Auditorium of the Medipolis (Kiviharjuntie 11), on April 21st, 2006, at 12 noon

OULUN YLIOPISTO, OULU 2006
Abstract

The correct three dimensional structures of proteins are essential for their ability to function properly. Proteins start to fold as soon as they are synthesized in the ribosomes from activated amino acids. Many secreted, cell-surface, secretory pathway and endoplasmic reticulum (ER) luminal proteins have in their amino acid sequence cysteine residues which form intra- and intermolecular disulfide bridges that stabilize the overall fold of the proteins and protein complexes. The formation of correct disulfide bonds is a complex process which takes place within the ER.

Protein disulfide isomerase (PDI) is the key enzyme in the formation and rearrangement of correct disulfide bonds in the ER. It is an archetypal and the best studied member of the PDI family, i.e. a group of ER proteins that resemble thioredoxin (TRX), a protein reductase, in their structure. PDI has a four domain \( a-b-b'-a' \) structure the \( a \) and \( a' \) domains having the catalytic activity and amino acid sequence similarity to TRX. In addition to its function as a thiol-disulfide oxidoreductase, PDI acts as the \( \beta \) subunit in two protein complexes: collagen prolyl 4-hydroxylase (C-P4H) and microsomal triglyceride transfer protein (MTP).

The closest homologue of PDI is the multifunctional enzyme and chaperone ERp57 that functions in concert with two lectins, calnexin (CNX) and calreticulin (CRT) specifically in the folding of proteins that have sugar moieties linked to them. ERp57 is 56% similar to PDI in its amino acid sequence and has also the four-domain architecture. Despite the high similarity in their structures ERp57 cannot substitute for PDI as the \( \beta \) subunit of C-P4H. The minimum requirement for the C-P4H tetramer assembly is fulfilled by domains \( b' \) and \( a' \) of PDI, while domains \( a \) and \( b \) enhance this function and can be substituted in part by those of ERp57.

Until very recently the structural information of any of the PDI family members, which contains the TRX active site was limited to solution structures of human PDI domains \( a \) and \( b \). In this research the domain boundaries of the full length ERp57 were defined and the individual domains characterized. Furthermore the solution structures of the catalytically active domains \( a \) and \( a' \) of ERp57 were studied by nuclear magnetic resonance (NMR).

Keywords: biomolecular NMR, disulfide bond, PDI family, protein folding, protein structure
Acknowledgements

This work was carried out at the Collagen Research Unit of the Department of Medical Biochemistry and Molecular Biology, University of Oulu, and Biocenter Oulu, and in the National Biological NMR Laboratory of the Institute of Biotechnology, University of Helsinki.

I wish to express my deepest gratitude to my supervisor, Docent Johanna Myllyharju for her continuous encouragement and never-ending optimism throughout these years. It has been a privilege, under your guidance, to have the opportunity to work in such a wonderful research group and at the same time to have the possibility to carry out the research in collaboration with other experts whenever needed. I am most grateful to my other supervisor, Academy Professor Emeritus Kari Kivirikko for his enthusiasm towards and extensive knowledge of science and absolutely astonishing ability to see the importance of each little finding every time I was unable to see any light at the end of the tunnel. I would like to express my gratitude to Professor Taina Pihlajaniemi for her admirable dedication to making high science and providing such excellent facilities for the research groups at our department and in Biocenter Oulu to work at. Professor Peppi Karpinnen deserves my appreciation for her positive attitude and help as workmate, co-author and head of our department. I also wish to express my respect to Professor Emeritus Ilmo Hassinen, Professor Leena Ala-Kokko and Professor Seppo Vainio for their distinguished scientific work.

I would like to thank Professor Johannes Buchner and Professor Eric Quemeneur for their critical review of the thesis and valuable comments regarding it. Peter Cura is acknowledged for careful revision of the English language of the thesis.

I thank Professor Ilkka Kilpeläinen for introducing me to the fascinating, although many times so frustrating world of bioNMR and setting up wonderful facilities at the Viikki campus in Helsinki. If there was, and there definitely is, one person without whom I wouldn’t be here today, he is without a doubt Docent Perttu Permi. I’m indebted to you for your extraordinary ability to be patient with me and for all the help with this work. I would also like to thank Docent Perttu Permi along with Professor Arto Annila for the ever so inspiring discussions at the coffee room in Viikki and during conference trips concerning the Finnish science policy and the current state of NMR research. I will always remember them fondly with a smile. ☺ I would also like to thank the director of
the Institute of Biotechnology, Professor Mart Saarma for allowing me to use the Institute’s excellent facilities.

I am very happy I have had co-authors Päivi Riihimaa and Helena Tossavainen working in these projects with me, Päivi at the very beginning and Helena taking part in the last few but long meters. Thank you both for patiently helping me.

All the members of the NMR team in Helsinki deserve my thanks for helping me, putting up with my sudden appearances in the Viikki campus and making me feel at home during all those short and long visits. Thank you Atro, Hannu, Kai, Kimmo, Maarit, Maija-Liisa and Tiia.

I would like to acknowledge all the past and present workmates form the Department of Medical Biochemistry and Molecular Biology. Especially I would like to thank Reija Hieta, Joni Mäki, Ritva Nissi, Outi Pakkanen, Kati Rautavuoma, Annamari Ruddock and Kati Takaluoma for making every single work day amusing. It truly has been great fun! Both at work and especially outside the lab. The computer support team, in particular Risto Helminen, and the people who have helped with the practical matters, i.e. Marja-Leena Karjalainen, Auli Kinnunen, Marja-Leena Kivelä, Seppo Lähdesmäki and Perti Vuokila are also acknowledged. I am very grateful for the expert technical assistance from Outi Mänty, Merja Nissilä, Minna Siurua and Jaana Träskelin.

This time in Oulu has been the best in my life. So far. The biggest thanks go to my friends. I wish to thank Ritva and Vesa for always being there for me. Through laughter and tears. True friends accept each other as they are, nothing added or nothing taken away. Friends like you are rare. Rare friends are also Kati and Marko. Thank you for your unconditional friendship. I wish we never lose contact no matter where life takes us. Special thanks go also to Reija and Onni for all the fun moments and discussions relating to motorcycles, or wining and dining, or young R.P., or… I just wish I knew how to whistle. 😊 A very special ‘tänks’ to Outi. The list of hilarious times and also serious girly discussions at any given hour is endless. Now my friend, we’re off to never-never land! I also wish to thank a friend and former member of our research group, Leanne Laajoki for all the help with NMR matters and also for the warm hospitality she and her family showed me before and after the 42195 m through Sydney in 2003.

My mother, Aino and my late father, Jouko deserve my deepest thanks for their love and support, and my two “little” brothers, Olli and Matti, for making this life such an adventure. Keep on rocking! I would also like to thank all relatives and friends who have kept me busy with everything else but science during all these years.

This work was supported by the Finnish Centre of Excellence Programme 2000-2005 of the Academy of Finland, the Finnish Cultural Foundation and Biocenter Oulu.

Hailuoto, April 2006

Laura Silvennoinen
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>aa</td>
<td>amino acid</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>α1,2-M I</td>
<td>α1,2-mannosidase I</td>
</tr>
<tr>
<td>β2m</td>
<td>β2-microglobulin</td>
</tr>
<tr>
<td>CD</td>
<td>circular dichroism</td>
</tr>
<tr>
<td>CNX</td>
<td>calnexin</td>
</tr>
<tr>
<td>C-P4H</td>
<td>collagen prolyl 4-hydroxylase</td>
</tr>
<tr>
<td>CRT</td>
<td>calreticulin</td>
</tr>
<tr>
<td>CSI</td>
<td>chemical shift index</td>
</tr>
<tr>
<td>-CXXC-</td>
<td>cysteine-any amino acid-any amino acid-cysteine</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>EDEM</td>
<td>ER degradation enhancing α-mannosidase-like protein</td>
</tr>
<tr>
<td>ER</td>
<td>endoplasmic reticulum</td>
</tr>
<tr>
<td>ERAD</td>
<td>ER associated degradation</td>
</tr>
<tr>
<td>Ero1</td>
<td>ER oxidoreductin 1</td>
</tr>
<tr>
<td>FAD</td>
<td>flavin adenine dinucleotide</td>
</tr>
<tr>
<td>G I/II</td>
<td>glucosidase I/II</td>
</tr>
<tr>
<td>GSH</td>
<td>reduced glutathione</td>
</tr>
<tr>
<td>GSSG</td>
<td>oxidized glutathione</td>
</tr>
<tr>
<td>HC</td>
<td>heavy chain</td>
</tr>
<tr>
<td>IPR1</td>
<td>inositol 1,4,5-trisphosphate receptor type 1</td>
</tr>
<tr>
<td>kDa</td>
<td>kilo dalton</td>
</tr>
<tr>
<td>LDL</td>
<td>low-density lipoprotein</td>
</tr>
<tr>
<td>MHC</td>
<td>major histocompatibility complex</td>
</tr>
<tr>
<td>MTP</td>
<td>microsomal triglyceride transfer protein</td>
</tr>
<tr>
<td>N-linked</td>
<td>asparagine linked</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>PC-TRP</td>
<td>plasma cell thioredoxin-related protein</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>PDI</td>
<td>protein disulfide isomerase</td>
</tr>
<tr>
<td>QC</td>
<td>quality control</td>
</tr>
<tr>
<td>RNase</td>
<td>ribonuclease</td>
</tr>
<tr>
<td>SAXS</td>
<td>small angle X-ray scattering</td>
</tr>
<tr>
<td>SERCA</td>
<td>sarco endoplasmic reticulum calcium ATPase</td>
</tr>
<tr>
<td>TRX</td>
<td>thioredoxin</td>
</tr>
<tr>
<td>UDP</td>
<td>uracil diphosphate</td>
</tr>
<tr>
<td>UGGT</td>
<td>UDP-glucose:glycoprotein glucosyltransferase</td>
</tr>
<tr>
<td>UPR</td>
<td>unfolded protein response</td>
</tr>
</tbody>
</table>
List of original articles

This thesis is based on the following articles, which are referred to in the text by their Roman numerals:


In addition, some data which are not included in the original articles are presented.
Contents

Abstract
Acknowledgements
Abbreviations
List of original articles
Contents
1 Introduction ................................................................................................................... 13
2 Review of the literature ................................................................................................. 14
  2.1 Protein folding in the ER ........................................................................................14
    2.1.1 Disulfide bond formation in eukaryotes .........................................................15
    2.1.2 Glycoprotein folding ......................................................................................17
  2.2 PDI........................................................................................................................ ..17
    2.2.1 Structure ..........................................................................................................18
    2.2.1.1 The thioredoxin fold .................................................................................22
    2.2.2 Function...........................................................................................................23
    2.2.2.1 The β subunit of collagen prolyl 4-hydroxylases and microsomal triglyceride transfer protein .........................................................................................25
  2.3 ERp57.....................................................................................................................26
    2.3.1 Structure ..........................................................................................................26
    2.3.2 Function...........................................................................................................27
  2.4 Other PDI family members.....................................................................................29
3 Outlines of the present study ......................................................................................... 33
4 Materials and methods................................................................................................... 34
  4.1 Materials.................................................................................................................34
  4.2 Methods ..................................................................................................................35
5 Results ........................................................................................................................... 38
  5.1 Characterization of the domains of human ERp57 (I) ............................................38
  5.2 1H, 13C and 15N resonance assignments and secondary structures of the a and a’ domains of human ERp57 (II, IV) .................................................................40
  5.3 Solution structure determination of the N-terminal domain a of human ERp57 (III) ...................................................................................................................41
6 Discussion .................................................................................................................... 45
References
Original articles
Introduction

In order to become biologically functional a protein needs to be correctly folded. If a polypeptide is misfolded it becomes nonfunctional and is degraded or sometimes accumulates as large insoluble aggregates, which can result in diseases like Alzheimer’s and cystic fibrosis. The folding and assembly of a newly translated polypeptide in a cell is controlled and assisted by numerous catalysts that, for example, accelerate its folding and inhibit it from aggregating. These catalysts are collectively known as molecular chaperones.

Protein disulfide isomerase (PDI) is an essential folding catalyst and chaperone of the eukaryotic cell. It introduces disulfide bonds into proteins and catalyzes rearrangement of incorrect disulfide bonds. PDI is archetypal of the PDI family of proteins and is the best studied member of this group. The proteins of the PDI family are located within the endoplasmic reticulum (ER), and have one or more structural units resembling the fold of thioredoxin, a protein reductase. PDI also acts as the $\beta$ subunit in two enzymes, namely collagen prolyl 4-hydroxylase (C-P4H), an $\alpha_2\beta_2$ tetramer, and microsomal triglyceride transfer protein (MTP), an $\alpha\beta$ dimer.

The closest homologue of PDI is the multifunctional enzyme and chaperone ERp57 that functions in concert with two lectins, calnexin (CNX) and calreticulin (CRT), specifically in the folding of glycoproteins. Both PDI and ERp57 have a four domain $a$-$b$-$b'$-$a'$ structure, the $a$ and $a'$ domains being the catalytically active domains. The $b'$ domains of both enzymes contain the primary binding site reserved for substrate binding in PDI and lectin binding in ERp57. ERp57, although very similar in its amino acid sequence and domain architecture to PDI, cannot substitute for PDI as the $\beta$ subunit of C-P4H. The minimum requirement for the C-P4H tetramer assembly is fulfilled by domains $b$ and $a'$ of PDI, while domains $a$ and $b$ enhance this function and can be substituted in part by those of ERp57.

Until very recently, when the crystal structure of a yeast PDI was published, the only high resolution structures of PDI or ERp57 were the NMR solution structures of domains $a$ and $b$ of PDI. To better understand the roles of the various domains in the C-P4H tetramer assembly and other functions of PDI and ERp57, the biochemical and structural studies described in this thesis were started.
2 Review of the literature

2.1 Protein folding in the ER

After proteins have been synthesized from activated amino acids in the ribosomes they must fold and assemble into unique three-dimensional structures in order to become functionally active. The maturation of nascent proteins into their fully functional three-dimensional conformation in vivo requires many steps in which numerous proteins assist in their folding and inhibit them from aggregating and folding incorrectly (Anken et al. 2005). In the case of extracellular proteins, i.e. proteins that are secreted out of the cell and cell-surface proteins, folding into native and functional three-dimensional structures takes place inside the endoplasmic reticulum (ER). In the ER molecular chaperones prevent the newly synthesized proteins from aggregating and misfolding and foldases accelerate their folding (Gething & Sambrook 1992). Proteins that have not reached their correct tertiary or quaternary structure are subjected to ER quality control (QC) (Ellgaard & Helenius 2003) which is also mediated by the ER chaperones and foldases. Ultimately the fate of a protein is either to pass the ER QC and exit the ER in COPII coated vesicles (Stephens et al. 2000) or if maintained misfolded to be retro-translocated to the cytosol for degradation by the proteasome (Tsai et al. 2002).

Extracellular proteins are largely N-glycosylated and form intra- and intermolecular disulfide bridges, which stabilize their native structures. These two modifications happen exclusively in the ER. Oligosaccharyltransferase is the enzyme that mediates the N-glycosylation while the key enzyme in the formation of correct disulfide bonds in proteins is the oxidoreductase protein disulfide isomerase (PDI). As both modifications occur as soon as the nascent polypeptide chain enters the lumen of the ER a closeness of cysteines and glycosylation sites can lead to competition between the disulfide bond formation and glycosylation (Allen et al. 1995).
2.1.1 Disulfide bond formation in eukaryotes

Disulfide bonds stabilize the three-dimensional structure of the newly synthesized, correctly folded proteins. The formation of native disulfide bonds is a redox reaction that takes place in the ER of eukaryotic cells and in the periplasm of prokaryotic cells and requires appropriate conditions for oxidation, reduction and isomerization or rearrangement of the bonds. It was previously thought that the correct conditions in the ER of eukaryotes are maintained primarily by the major redox buffer glutathione, a glutamate-cysteine-glycine tripeptide (Hwang et al. 1992). It was proposed that the redox potential of the cell was altered by changes in the total concentration of glutathione and the ratio of the oxidized (GSSG) and reduced (GSH) form of the tripeptide (Jones 2002). In cytosol the GSH:GSSG ratio is ca. 100:1 and in the ER ca. 3:1 (Hwang et al. 1992), the ER environment being more favorable for disulfide bond formation.

The formation of disulfide bonds is a two-electron reaction that requires oxidizing and reducing equivalents. During bond formation, when PDI acts as the oxidizing agent, disulfide bonded cysteines in its active site motif -CXXC- accept two electrons from the substrate polypeptide and PDI is reduced to its dithiol form. The general mechanism of the reaction is shown in Fig. 1A. In order for PDI to act as an oxidase again, it must be re-oxidized. For a long time it was believed that glutathione acted as the counterpart in re-oxidizing PDI, but studies in the yeast Saccharomyces cerevisiae identified a new member of the oxidative pathway, namely ER oxidoreductin 1 protein (Ero1p) as the electron acceptor coupled with PDI (Pollard et al. 1998, Frand & Kaiser 1999, Frand et al. 2000, Tu & Weissman, 2004). Today it is clear that in yeast the enzyme Ero1p is first oxidized by molecular oxygen after which it acts as a specific oxidant for PDI (Tu et al. 2000).

Ero1p is a flavin adenine dinucleotide (FAD) -dependent, glycosylated luminal ER protein, which is tightly associated with the ER membrane (Pollard et al. 1998, Frand & Kaiser 1999). In the Ero1p-mediated reaction pathway as Ero1p is oxidized, FAD is reduced to FADH₂. Cellular levels of FAD affect the Ero1p-PDI pathway so that high levels of FAD accelerate the formation of disulfide bonds in the final substrate and low levels slow it down (Tu & Weissman 2002). Mutational analyses of Ero1p have shown that four cysteines, namely C100, C105, C352 and C355, are essential for the oxidative activity of the protein, as mutation of any of the four cysteines disrupted the thiol-disulfide exchange between Ero1p and PDI (Frand & Kaiser 2000).

A second oxidative pathway in yeast involves a small ER oxidase known as Erv2p (Sevier at al. 2001). Like Ero1p, Erv2p uses molecular oxygen as an electron acceptor and oxidizes PDI in vivo (Sevier at al. 2001). It has also been shown that recombinant Erv2p is able to oxidize proteins directly in vitro (Gerber et al. 2001). Crystal structures of Ero1p (Gross et al. 2004) and Erv2p (Gross et al. 2002) show that these proteins are also structurally, not only functionally, related, despite the lack of similarity in their primary structure. Fig. 1B shows a simplified mechanism of the formation, breakage and rearrangement of disulfide bonds, i.e. oxidative protein folding in yeast with Ero1p and Erv2p as the oxidizing agents.

Two homologues of the yeast Ero1p have been identified in mammalian cells, namely Ero1α (Cabibbo et al. 2000) which is up-regulated by hypoxia (Gess et al. 2003) and
Ero1β which is up-regulated by the unfolded protein response (UPR) (Pagani et al. 2000). Both human Eros have been found to selectively oxidize human PDI (Benham et al. 2000, Mezghrani et al. 2001). Ero1α also forms complexes with some other members of the PDI protein family (see section 2.2) while Ero1β forms homodimers and mixed heterodimers with Ero1α (Dias-Gunasekara et al. 2005).

![Diagram of disulfide bond exchange and oxidative protein folding in the yeast ER.]

The formation and rearrangement of disulfide bonds require different redox conditions. If eukaryotic Eros primarily re-oxidize PDI from its reduced form, how is the reduced form of PDI, which is needed for reduction and isomerization obtained and maintained? To understand this reductive pathway that exists in the ER parallel to the oxidative pathway, the role of glutathione has been studied in both yeast and mammalian cells (Cuozzo & Kaiser 1999, Jessop & Bulleid 2004, Molteni et al. 2004). Elimination of glutathione from yeast resulted in hyperoxidation of substrate molecules suggesting that glutathione provides the required reducing equivalents for essential reduction (Cuozzo & Kaiser 1999). In semipermeabilized mammalian cells that overexpressed human Ero1α, an increased level of reduced glutathione (GSH) was needed to balance the oxidative folding (Molteni et al. 2004). A model that used PDI homologue ERp57 as the thiol-oxidoreductase showed that the component responsible for the reduction of the oxidoreductase is reduced glutathione (GSH) and that the reductive pathway is cytosol-dependent as reduction of ERp57 in oxidized microsomes only occurs in the presence of cytosol, or more specifically GSH (Jessop & Bulleid 2004). In the reductive pathway
reduced glutathione GSH is transported from the cytosol into the ER, reducing ERp57 and being oxidized itself to GSSG. How the trafficking of the GSH and GSSG molecules happens in eukaryotic cells, is still largely unknown.

2.1.2 Glycoprotein folding

The majority of the vertebrate extracellular proteins have glycans attached to an asparagine residue that occurs in an -NXS/T- consensus sequence, where X is any amino acid except proline (Gavel & von Heijne 1990). N-linked glycans are added to polypeptide chains as they enter the ER lumen. The addition of the branched, triglucosylated core oligosaccharide Glc$_3$Man$_9$GlcNAc$_2$ is catalyzed by the oligosaccharyl transferase complex. As soon as the N-linked glycan is added to the folding polypeptide chain, glucosidases I and II trim the oligosaccharide moiety of the glycoprotein by sequentially removing the two end glucose leaving one glucose moiety attached to the nascent polypeptide chain. This monoglycosylated glycoprotein is a substrate for the lectins calnexin (CNX) and calreticulin (CRT) (Helenius et al. 1997, Trombetta & Helenius 1998). The disulfide bonds in glycoproteins are formed by the PDI family member ERp57 (see section 2.3) (Oliver et al. 1997, Zapun et al. 1998), which is found noncovalently bound to both CNX and CRT (Molinari & Helenius 1999). Glucosidase II cleaves the last glucose and terminates the interaction of the glycoprotein with CNX or CRT. If the glycoprotein is correctly folded into its native conformation it will then be transported to the Golgi complex and to the secretory pathway. If no native functional form has been reached the enzyme UDP-glucose:glycoprotein glucosyltransferase (UGGT), which recognizes unfolded proteins, adds a glucose residue to the N-linked glycan (Helenius 1994, Sousa & Parodi 1995). During the folding event glycoproteins go through cycles of glucose trimming that mediate their interaction with CNX and CRT. This cycle is called the calnexin/calreticulin cycle. If the protein fails to fold correctly, α$_1$,$\alpha_2$-mannosidase I removes the mannose moiety from the middle branch of the core oligosaccharide and targets the unfolded protein to ER associated degradation (ERAD) via EDEM, an ER degradation enhancing α-mannosidase-like protein. A schematic presentation of glycoprotein folding is shown in Fig. 2.

2.2 PDI

PDI (EC 5.3.4.1) was the earliest identified member of the protein family now called the PDI family. It was initially extracted and purified from rat liver microsomes and shown to catalyze the reactivation of reduced RNase (Goldberger et al. 1964). Further studies of the same group suggested that the enzyme catalyzed a disulfide interchange in substrates rather than a net oxidation (Givol et al. 1964). This is the best characterized member of the PDI family and it is found widely in many species like plants (Shimoni et al. 1995, Li & Larkins 1996), fungi (Kajino et al. 1994, Ngiam et al. 1997), yeast (LaMantia et al. 1991) and animals (Edman et al. 1985, Gong et al. 1988). Due to its high expression
and unusual stability it is one of the most abundant proteins in the ER (Lyles & Gilbert 1991) having a half-life of approximately seven days in rat liver (Ohba et al. 1981).

![Glycoprotein folding pathway](image)

**Fig. 2.** Glycoprotein folding pathway. The N-linked oligosaccharide is shown in more detail: the dark grey polygons indicate N-acetylglucosamine, light grey circles mannose and light grey squares glucose residues. The glucose (G) and mannose (M) residues, which are cleaved off by glucosidases I and II (GI and GII) and α1,2-mannosidase I (α1,2-M I), respectively, are indicated. For clarity CNX is omitted and only folding pathway proceeding through ERp57-CRT complex is shown. The figure is compiled from the articles referred to in the text.

### 2.2.1 Structure

PDI is a 55 kDa polypeptide that was first sequenced from rat liver (Edman et al. 1985). The mature human protein after cleavage of the 17-amino-acid signal peptide has 491 residues with an ER retention signal -KDEL. Sedimentation equilibrium and velocity centrifugation experiments suggest that the full length PDI is a monomer and has an elongated shape with an axial ratio of 5.7 (Solovyov & Gilbert 2004). A very recent small
angle X-ray scattering study in solution suggests that the full length PDI is a short and roughly elliptical cylinder with dimensions of 105 Å × 65 Å × 40 Å (Li et al. 2006). When PDI was sequenced it was already proposed, based on homology searches, that it had a four-domain architecture (Fig. 3). Its N-terminal and C-terminal parts, which are designated as domains a and a’, have a thioredoxin-like active site -CGHC- sequence and show a high sequence similarity to the oxidoreductase thioredoxin. The region between these two domains shows no homology to thioredoxin nor to the N- and C-terminal domains, but according to internal sequence homology is composed of two domains named b and b’ (Edman et al. 1985). Domains a and a’ are 47% identical in their amino acid sequences while b and b’ share a 28% identity.

![Diagram of domain organization of human PDI](image)

Fig. 3. Domain organization of human PDI. The length of the processed polypeptide is shown on the right. The amino acids comprising the structural domains a, b, b’ , and a’ and the catalytic motifs are indicated. The linker region x between the b’ and a’ domains is shown in black.

The domain boundaries were first defined by sequence alignments (Freedman et al. 1994) and later refined experimentally by a combination of protein engineering and limited proteolysis studies (Darby et al. 1996). It has also been suggested that the protein is comprised of six individual domains, namely a, e, b, b’, a’ and c (Freedman et al. 1994). The e domain between domains a and b was introduced due to the homology of PDI to the estrogen-binding domain of human estrogen receptor (Tsibris et al. 1989). The a’ domain was thought to end at residue 435 leaving 56 residues to form a highly acidic putative Ca2+ binding region, domain e (Freedman et al. 1994). Further studies excluded domains e and c from the domain architecture and confirmed that PDI consisted of four structural domains, namely a, b, b’ and a’, and an acidic C-terminal extension (Freedman et al. 1998, Darby et al. 1999). Recently, by using sequence alignments, a large number of domain constructs of six redox active PDI family members, namely PDI, PDIp, PDIR, P5, ERp57 and ERp72 (see section 2.4) were generated and the domain boundaries were refined (Alanen et al. 2003a). The structural domains of human PDI were shown to be slightly longer than originally thought (Alanen et al. 2003a). A linker region between domains b’ and a’, region x, was recently identified to consist of 19 amino acids and to comprise residues 333-351 (Pirneskoski et al. 2004). It has been suggested that this unstructured area allows more flexibility between domains b’ and a’ than between the other domains.

Although attempts to determine the three dimensional structure of a full length PDI have failed until very recently, attempts to define the structures of the domains individually have been successful. So far solution structures of domains a and b of human PDI have been solved by NMR (Fig. 4A and B) (Kemmink et al. 1995, 1996, 1999) and although they share no sequence homology they both were found to have a typical
thioredoxin fold with four $\beta$-strands surrounded by $\alpha$-helices. According to a preliminary NMR study domain $a'$ of human PDI also has a thioredoxin fold (Dijkstra et al. 1999) and based on amino acid sequence homology domain $b'$ is likely to have this same fold.

The very recent crystal structure of PDI from the yeast $S.\ cer\textit{er}v\textit{i}ss\textit{iae}$ confirmed that the protein contains four structural thioredoxin-like domains and showed them to be arranged in the shape of a twisted “U” (Fig. 4C) (Tian et al. 2006). The interior of the U shape is enriched with hydrophobic residues and the catalytically active domains $a$ and $a'$ are facing each other, although there is a large cleft separating them. The dimensions of the U shaped molecule are ca. 80 Å in width and 60 Å in height, which differs somewhat from those obtained with SAXS (Li et al. 2006). All individual domains fold into a thioredoxin fold (see section 2.2.1.1) having some minor variations in their secondary structure elements (Tian et al. 2006). The interface between domains $b$ and $b'$, which has a buried surface area of ca. 700 Å$^2$, is much larger than the interfaces between the $a$ and $b$ domains and the $b'$ and $a'$ domains. These each have a buried surface area of ca. 200 Å$^2$ suggesting that the $b$ and $b'$ domains form a rigid base allowing more flexibility in the $a$ and $a'$ domains (Tian et al. 2006). This flexibility suggests that the molecule is able to adopt the elongated conformation described by Solovyov and Gilbert in 2004. In the yeast PDI the residues belonging to the C-terminal extension form an $\alpha$-helix, which is thought to stabilize the $a'$ domain and to play an important role in the activity of the enzyme, as deletion of this extension decreases the refolding of both scrambled RNase and reduced RNase (Tian et al. 2006). In contrast, the C-terminal extension of mammalian PDI most likely does not have an important role in PDI functions (Darby et al. 1998, Koivunen et al. 1999).
Fig. 4. Ribbon models of the structures of the human PDI domains a (A, PDB code 1MEK, Kemmink et al. 1996) and b (B, PDB code 1BJX, Kemmink et al. 1999) determined by NMR, and of the full length S. cerevisiae PDI (C, PDB code 2B5E, Tian et al. 2006) determined by X-ray crystallography. The active site cysteines are marked in blue in domain a of human PDI (C36 and C39) and in domains a (C61 and C64) and a’ (C406 and C409) of yeast PDI.
2.2.1.1 The thioredoxin fold

The thioredoxin fold is a distinct structural motif that was named after the protein in which it was first observed, *E. coli* thioredoxin. It is a one-domain, 10 kDa-protein with a -CGPC- active site motif. The X-ray structure of oxidized *E. coli* thioredoxin was first determined at 2.8 Å resolution in 1975 (Holmgren *et al.* 1975) and later refined at 1.68 Å (Katti *et al.* 1990). The structure was found to consist of a central core of five β-strands surrounded by four α-helices.

In addition to its presence in thioredoxin, the thioredoxin fold has also been found in at least seven different protein families with a variety of different functions: protein disulfide isomerases that also belong to the thioredoxin-superfamily (Kemmink *et al.* 1997), DsbAs and glutaredoxins that catalyze thiol-disulfide exchange reactions (Martin 1995), glutathione S-transferases that catalyze the formation of glutathione conjugates from electrophilic substrates (Martin 1995), calsequestrins that bind Ca²⁺ ions (Wang *et al.* 1998) and glutathione peroxidases and peroxiredoxins that catalyze reduction of peroxides (Martin 1995, Wood *et al.* 2003).

The prototypical thioredoxin fold consists of four β-strands and three α-helices. Moreover, the thioredoxin fold can be divided into an N-terminal βαβ motif and a C-terminal ββα motif that are connected by a loop that incorporates the third helix (Fig. 5). The β-strands in the N-terminal motif run parallel while the β-strands in the C-terminal motif run antiparallel to each other. The active site motif is found at the N terminal end of the first α-helix. Another structural feature common to thioredoxin fold containing proteins is a cis-proline buried behind the active site in the three-dimensional structure. In many thioredoxin fold containing structures there is also an additional β-strand and an additional α-helix in the complete globular fold.

![Fig. 5. Schematic architecture of the thioredoxin fold. α-helices are marked with grey squares and β-strands with black arrows.](image-url)
2.2.2 Function


The two conserved -$\text{CGHC}$- active sites, one in the N-terminal domain $a$ and the other in the C-terminal domain $a'$ make it possible for PDI to function in the formation, breakage and rearrangement of native disulfide bonds. The isolated domains $a$ and $a'$ themselves are capable of catalyzing simple disulfide exchange reactions, but for an efficient catalysis of disulfide bond isomerization where unfavorable conformational changes in the substrate molecule occur, all domains are required (Darby \textit{et al.} 1998). Like other -$\text{CXXC}$- containing proteins PDI can form an intramolecular disulfide bond or exist in a reduced dithiol form. The redox state of the active sites \textit{in vitro} is maintained by a redox buffer, usually glutathione (GSH:GSSG) or the addition of a small molecule reducing agent dithiothreitol (DTT) in the buffer.

Residues that influence the function of the enzyme, i.e. whether it acts as a reductase or as an oxidase, are the active site cysteines, the residues between them and residues that are close to the active site in the tertiary structure of the enzyme. The N-terminal cysteine of each -$\text{CXXC}$- motif is exposed on the surface of the molecule and reacts with substrates forming a mixed disulfide (see Fig. 1A), while the C-terminal cysteine is buried behind it and reacts only with the N-terminal nucleophilic cysteine (Walker \textit{et al.} 1996). The two residues between the active site cysteines have a major role in determining the redox potential of the enzyme, i.e. the equilibrium between the dithiol and disulfide states. For example, mutation of the -$\text{CGPC}$- sequence of thioredoxin to a -$\text{CGHC}$- corresponding to the active sites of PDI made thioredoxin a 10-fold better oxidant than its natural form (Lundström \textit{et al.} 1992). The redox potential of PDI varies from -110 mV to -180 mV (Hawkins \textit{et al.} 1991, Lundström \& Holmgren 1993) making it a mild oxidant while the redox potential of reduced glutathione is -240 mV (Rost \& Rapoport 1964). An arginine residue that is conserved between the PDI family members was recently shown to be involved in the catalytic cycle of PDI, specifically in the reoxidation (Lappi \textit{et al.} 2004). This involvement was shown to be due to movement of the side-chain of the residue into and out of the active site region, thus affecting the pK$_a$ values of the cysteine residues (Lappi \textit{et al.} 2004). In addition, a charged glutamic acid-lysine pair buried in the molecule behind the active site has been shown to be important for the catalytic activity of thioredoxin and for the oxidative activity of PDI and ERp57 (Dyson \textit{et al.} 1997, Ellgaard and Ruddock 2005). This pair is widely found in the PDI family (Ellgaard \& Ruddock 2005).
The individual catalytic domains of PDI have very low isomerase activities (Darby et al. 1998, Xiao et al. 2001) while the oxidase activities are close to those of the full length PDIs that have one of the active sites mutated (Darby et al. 1996). The minimum requirement for catalytic activity is a 21 kDa fragment consisting of approximately half of domain $b'$, domain $a'$ and the C-terminal extension (Puig et al. 1997). Isomerase activity of a construct containing the C-terminal part of the enzyme starting from domain $b'$ was similar to that of a full length PDI where the active site cysteines were mutated to serine (Darby et al. 1996, 1998, Xiao et al. 2001).

Although studied extensively, the specific role of PDI in the formation and, conversely, in the breakage and rearrangement of disulfide bonds is still unclear. A recent study suggests distinct roles for the two catalytic PDI domains in Ero1p-mediated oxidative folding, one domain acting as a disulfide isomerase and the other as a disulfide oxidase (Kulp et al. 2006). The individual catalytic domains $a$ and $a'$ were oxidized at the same rate by Ero1p, but in the case of the full length PDI asymmetry in the rate of oxidation was observed: the $a'$ domain was oxidized at an enhanced rate while a substrate-mediated inhibition of oxidation was observed for the $a$ domain (Kulp et al. 2006).

PDI has also been shown to have a role as a chaperone inhibiting the aggregation of misfolded proteins, some of which do not even contain disulfide bonds (Cai et al. 1994, Puig & Gilbert 1994, Song & Wang 1995, Primm et al. 1996). The chaperone activity was shown to be independent of its function as a disulfide oxidoreductase, as alkylating the active sites did not affect its ability to inhibit aggregation (Quan et al. 1995). The deletion of 51 C-terminal residues eliminated the chaperone activity of PDI but did not affect the isomerase activity (Dai & Wang 1997).

In addition to its function as a chaperone, PDI has been shown to possess anti-chaperone activity, i.e. to facilitate misfolding and aggregation under certain conditions (Puig & Gilbert 1994). The influence of PDI on the folding process depends on its concentration and whether the substrate is soluble or aggregation-prone. At high concentrations PDI inhibits aggregation, but at concentrations of 0.1-0.2 compared to that of the substrate, PDI precipitates many normally soluble proteins (Sideraki & Gilbert 2000).

The peptide binding site of PDI was initially misidentified to be located at the very end of the C-terminal region of the enzyme (Noiva et al. 1993). However, the primary site for the binding of peptides and misfolded proteins to PDI was later shown to be located in the $b'$ domain (Klappa et al. 1998a). The $b'$ domain in itself is sufficient for binding of small peptides but other domains, especially domains $a$ and $a'$, are also required for efficient folding of longer substrates (Klappa et al. 1998a, 2000). A molecular model of domain $b'$ revealed a small hydrophobic pocket comprising mainly residues L242, L244, F258 and I272, a region corresponding to the region of the active site of domain $a$ (Pirneskoski et al. 2004). Mutation of any of the four residues affected the peptide binding ability of PDI, the greatest impact being from the mutations of residue I272 (Pirneskoski et al. 2004). The binding specificity of PDI is still largely unknown, while one member of the PDI family, the pancreas specific homologue PDIp (see section 2.3), has been reported to specifically bind tyrosine and tryptophan residues (Ruddock et al. 2000, Klappa et al. 2001) and another member of the family, ERp57, is
known to interact specifically in concert with the lectins calnexin and calreticulin with N-glycosylated proteins (Elliot et al. 1997, Oliver et al. 1997).

2.2.2.1 The β subunit of collagen prolyl 4-hydroxylases and microsomal triglyceride transfer protein

The enzymes collagen prolyl 4-hydroxylases (C-P4Hs, EC 1.14.11.2) act on Y position proline residues in the Gly-X-Y repeats in collagens and more than 20 other proteins that have collagen-like domains. This hydroxylation is essential for the stability of the collagen triple helix (for general references, see Kivirikko & Myllyharju 1998, Kivirikko & Pihlajaniemi 1998, Myllyharju 2003, Myllyharju & Kivirikko 2004).

The vertebrate C-P4Hs are α2β2 tetramers where the β subunit is identical to PDI. The α subunit has a molecular mass of ca. 61 kDa while the molecular mass of the tetramer is ca. 232 kDa. Three isoforms of the vertebrate C-P4H α subunit are known today (Helaakoski et al. 1989, 1995, Annunen et al. 1997, Kukkola et al. 2003, Van Den Diepstraten et al. 2003). Each of these α subunits forms an α2β2 tetramer where PDI acts as the β subunit. The function of the β/PDI subunits is to keep the highly insoluble α subunits in a non-aggregated catalytically active conformation (Vuori et al. 1992a, John et al. 1993, Veijola et al. 1994, Helaakoski et al. 1995). In addition, PDI helps to keep the tetramer in the ER as the α subunit has no ER retrieval signal (Vuori et al. 1992b). Site-directed mutagenesis studies showed that the catalytic activity of PDI is not required for the tetramer assembly nor for the tetramer to be a fully active C-P4H (Vuori et al. 1992b). PDI also serves as the β subunit of C-P4Hs in D. melanogaster and C. elegans and its function in these enzymes is identical to that in the vertebrate C-P4Hs (Myllyharju 2003, Myllyharju & Kivirikko 2004).

Although the catalytic activity of PDI is not required in the C-P4H assembly, PDI is an essential part of the tetramer. It has been shown that co-expression of recombinant human C-P4H α subunit with another ER chaperone BiP leads to formation of a soluble tetramer, which has no P4H activity (Veijola et al. 1996). The closest homologue of PDI, namely ERp57, is 29% identical and 56% similar to PDI in its primary structure (see section 2.3), but can not replace PDI as the β subunit in the C-P4H assembly (Koivunen et al. 1996). The ERp57 domains a and b can in part substitute for the corresponding domains of PDI while domains b’ and a’ of PDI are crucial for the formation of an active C-P4H tetramer (Pirneskoski et al. 2001). Furthermore, the acidic C-terminal extension of PDI has been shown to play no role in the assembly (Koivunen et al. 1999). Recently, essential residues for the assembly of an active C-P4H tetramer were identified in the thioredoxin-like domains a and a’ and in the peptide-binding domain b’ of PDI (Koivunen et al. 2005). Mutations in the peptide-binding domain b’ alone did not reduce the C-P4H tetramer assembly or activity. When additional mutations in residues, which probably contribute to a putative peptide binding groove in domains a and a’, were introduced, a decrease in the assembly and activity of the tetramer was seen, and a clear cooperative effect between the binding sites in the three PDI domains was observed (Koivunen et al. 2005).

The PDI activity of the β subunit in the C-P4H tetramer has been shown to be approximately half that of the free human PDI (Koivu et al. 1987). When mutations of
the active site cysteines were introduced into the PDI subunits of the $\alpha_2\beta_2$ tetramer, clear differences were observed in the PDI activity levels of the mutant $\beta$ subunits (Vuori et al. 1992b). Inactivation of the active site of the $\alpha$ domain had no effect on the activity while inactivation of the active site of the $\alpha'$ domain decreased the PDI activity level significantly (Vuori et al. 1992b). This suggests that the catalytic site of the PDI $\alpha$ domain is not functional in the C-P4H tetramer.

The microsomal triglyceride transfer protein (MTP) is an enzyme known to catalyze the transfer of triglycerides, cholesteryl esters and phospholipids between synthetic membranes or plasma lipoproteins. It is an $\alpha\beta$ dimer, the $\alpha$ subunit having a molecular mass of 97 kDa, and the $\beta$ subunit being identical to PDI (Wetterau et al. 1990). As with C-P4H, the role of the PDI/$\beta$ subunit is to keep the $\alpha$ subunit of MTP in a soluble catalytically active form and, because the $\alpha$ subunit lacks the ER retention signal, to retain it within the ER (Wetterau et al. 1991, Lamberg et al. 1996). The $\beta$ subunit of MTP differs from that of the C-P4H tetramer in that it has no PDI activity in the $\alpha\beta$ dimer (Wetterau et al. 1991). However, PDI activity is recovered after the dimer dissociates.

### 2.3 ERp57

ERp57 (EC 5.3.4.1), also known as ERp60, ERp61, ER58, ER-60, GRP58, P58, HIP-70 and Q2, is the closest homologue of PDI with an overall 29% identity and 56% similarity at the amino acid level (Koivunen et al. 1996). It is a 54 kDa protein that was at first misidentified as a phospholipase C-$\alpha$ (Bennet et al. 1988) and later as a cysteine protease (Urade et al. 1992). Immunohistochemical studies have shown it to be widely distributed in tissues that actively synthesize and secrete proteins containing cysteine residues (Kozaki et al. 1994, Iida et al. 1996, Marcus et al. 1996).

#### 2.3.1 Structure

Like PDI, the monomeric ERp57 is thought to have an elongated shape with an axial ratio of 4.9 (Frickel et al. 2004). The mature protein contains 481 amino acids after the cleavage of a 24-residue signal sequence (Bourdi et al. 1995, Hirano et al. 1995, Koivunen et al. 1996). The high sequence similarity to PDI suggests that it likewise consists of four domains in the same order as PDI, namely $a-b-b'-a'$, but in contrast to PDI the C-terminal extension of ERp57 is rich in basic lysine residues (Ferrari & Söling 1999). ERp57 has a -QDEL ER retention signal variant (Bennet et al. 1988, Mazzarella et al. 1994).

As in PDI, the $a$ and $a'$ domains of ERp57 have the thioredoxin-like -CGHC- active site sequences, the overall amino acid sequence identities being 50% and 46% to the corresponding domains of PDI. In addition to the active site cysteines, the ERp57 $a$ domain has two additional cysteines, namely C61 and C68, which form an intradomain disulfide bond (Frickel et al. 2004), the non-catalytic $b'$ domain having one additional cysteine, C220. The $b$ and $b'$ domains are only 17-20% identical to those of PDI. The
high sequence similarity suggests that at least domains $a$ and $a'$ have the common thioredoxin fold.

### 2.3.2 Function

ERp57 has been shown to possess disulfide oxidoreductase and isomerase activity \textit{in vitro} in numerous studies (Srivastava \textit{et al.} 1993, Bourdi \textit{et al.} 1995, Hirano \textit{et al.} 1995, Antoniou \textit{et al.} 2002, Frickel \textit{et al.} 2004, Mayer \textit{et al.} 2004). The reductase and isomerase activities of ERp57 are notably lower than those of PDI (Bourdi \textit{et al.} 1995, Frickel \textit{et al.} 2004, Mayer \textit{et al.} 2004). The presence of either of the lectin chaperones calreticulin (CRT) or calnexin (CNX) increases the activity level of ERp57 while the activity of PDI is decreased in the presence of these lectins (Zapun \textit{et al.} 1998). For the individual catalytic $a$ and $a'$ domains of ERp57, the redox potentials have been determined to be -0.167 V and -0.156 V, respectively (Frickel \textit{et al.} 2004), comparing well to that of the full length PDI (–0.175 V) (Lundström & Holmgren 1993).

ERp57 functions as a specific glycoprotein oxidoreductase (Elliot \textit{et al.} 1997, Oliver \textit{et al.} 1997, Van der Wal \textit{et al.} 1998) although it also interacts with glycoproteins with no cysteine residues (Elliot \textit{et al.} 1997) suggesting that its role is broader than that of just an oxidoreductase. A very recent study showed that ERp57 binds PDI and CRT in a competitive manner (Kimura \textit{et al.} 2005). Low concentrations of ERp57 enhanced the chaperone activity of PDI while high concentrations interfered with it. ERp57 had no influence on the isomerase activity of PDI (Kimura \textit{et al.} 2005). The exact pathway of reactions where ERp57 is involved is still partly unknown as, for example, ERp57 does not form complexes with either of the mammalian Eros that are the oxidizing agents of PDI (Mezghrani \textit{et al.} 2001). No other cellular oxidants for ERp57 are known either.

ERp57 associates with the membrane bound lectin CNX and its soluble form CRT (Oliver \textit{et al.} 1999, Freedman \textit{et al.} 2002, Ellgaard & Frickel 2003, Ellgaard & Helenius 2003, Ellgaard & Ruddock 2005) in the calnexin/calreticulin cycle (see section 2.1.2) and catalyzes disulfide bond formation in glycoproteins (Molinari & Helenius 1999). The glycoproteins do not interact directly with ERp57 but are first bound to the lectins, which in turn are bound to ERp57. NMR studies have shown that the P-domain of CRT has a unique hairpin or arm-like structure (Ellgaard \textit{et al.} 2001), and the crystal structure of CNX revealed it also contained a P-domain in addition to a globular lectin domain (Schrag \textit{et al.} 2001). The ERp57 binding site for both lectins has been shown to reside in the P-domain (Frickel \textit{et al.} 2002, Leach \textit{et al.} 2002).

The best studied glycoprotein folding event that ERp57 is involved in is the folding and assembly of major histocompatibility complex (MHC) class I molecules (Hughes & Cresswell 1998, Morrice & Powis 1998). The function of MHC class I molecules is to bind short antigenic peptides derived from cytosol and present them on the cell surface to cytotoxic T lymphocytes (CD8$^+$), permitting the detection and elimination of pathogen-infected cells. Mature MHC class I molecules consist of two subunits: a transmembrane heavy chain (HC) and a small soluble protein, $\beta_2$-microglobulin ($\beta_2$m). The early stages of folding of the MHC class I heavy chain involves glycosylation and binding to CNX (Rajagopalan & Brenner 1994, Vassilakos \textit{et al.} 1996). Before the glycosylated heavy
chain dissociates from CNX and associates with β₂m, the heavy chain forms disulfide-bonded intermediates with ERp57 and becomes fully oxidized (Lindquist et al. 1998, 2001, Farmery et al. 2000). Once oxidized, properly folded and associated with the β₂m, the HC/β₂m heterodimer interacts with CRT, ERp57 and the transmembrane glycoprotein tapasin. This complex is associated with TAP, i.e. the transporter associated with antigen processing, through the C-terminal end of the tapasin molecule. The MHC class I HC, β₂m, CRT, ERp57, tapasin and TAP form the MHC class I peptide loading complex (Cresswell 2000, Wright et al. 2004). It was recently shown that of the proteins participating in the peptide loading complex tapasin is the preferred substrate of ERp57 within the cell, and that the cysteines C33 of ERp57 and C95 of tapasin form a stable disulfide bond (Peaper et al. 2005). This interaction of ERp57 and tapasin is quite unique as interactions of ERp57 with for example CNX and CRT are rather weak and require special chemical cross-linkers to be detected (Elliot et al. 1997, Oliver et al. 1997, van der Wal et al. 1998, Daniels et al. 2003). While in the stable heterodimer form, noncovalent interactions between ERp57 and tapasin were shown to affect the reductase activity of the α domain of ERp57, abolishing it completely (Peaper et al. 2005). The activity of ERp57 was regained as the interactions were disrupted by denaturation (Peaper et al. 2005). The role of ERp57 in the peptide loading complex seems to be structural rather than functional as in ERp57 deficient mice the interaction of MHC class I molecules with the peptide-loading complex was affected and found to be short-lived while the redox state of the MHC class I molecules were not influenced (Garbi et al. 2006).

In addition to participating in the formation of the disulfides in the heavy chain of the MHC class I complex, ERp57 has been reported to affect the folding of numerous other glycoproteins like the human CD1d heavy chain (Kang & Cresswell 2002), influenza hemagglutinin (Daniels et al. 2003), asialoglycoprotein receptor chains H2a and H2b (Frenkel et al. 2004), sarco endoplasmic reticulum calcium ATPase (SERCA) (Li & Camacho 2004) and low-density lipoprotein (LDL) receptor-related protein (McCormick et al. 2005). The ER provides the best environment for ERp57 to perform its disulfide bond forming function, but this oxidoreductase has also been found in non-ER locations like the cytosol and nucleus (Turano et al. 2002). The way in which some of the ER located PDI family members escape the ER is not known and their presence in locations other than the ER has often been questioned. However, immunohistochemical studies have located ERp57 in the nuclei of rat spermatids and spermatozoa (Ohtani et al. 1993), chicken embryo fibroblasts (Alti et al. 1993) and HeLa cells (Coppari et al. 2002). ERp57 has also been found in nuclear STAT3-DNA complexes (Eufemi et al. 2004) but the role of ERp57 in this complex is not yet understood.
2.4 Other PDI family members

Proteins that contain one or more thioredoxin-like domains in their structures and are located within the ER are generally characterized as members of the PDI family (Fig. 6). This domain possesses a characteristic -CXXC- active site, which allows the protein to cycle between its thiol and disulfide forms. To date altogether 15 PDI family members have been reported in higher eukaryotes: PDI, PDIp, PDIR, PDILT, P5, ERp19, ERp28, ERp44, ERp46, ERp57, ERp72, ERdj5, TMX, TMX2 and TMX3 (Fig. 6) (Ellgaard & Ruddock 2005). Two additional ones, namely ERp27 and TMX4, are found in public databases although they have not been characterized in detail so far (Ellgaard & Ruddock 2005).

Fig. 6. Domain architecture of the PDI family members found in the ER. All proteins are human except P5 which is from rat. Numbers after the names refer to the numbers of amino acid residues of the processed full length proteins except in ERdj5/JPDI and TMX2 where the amino acids of the signal sequence are also included as the position of the signal peptide cleavage site is not known. The catalytic domains are indicated in grey and their catalytic site sequences are also indicated within the grey area. The b-like domains are indicated in white, the b'-like domains by thick diagonal lines and the liker region x between domains b' and a' with black square. Non-homologous regions are shown by thin diagonal lines and the transmembrane regions of TMX, TMX2 and TMX3 with a black ellipse. The ER retention sequences are included at the C-termini. (Modified from Ellgaard & Ruddock 2005)
The pancreas acinar cell-specific isoform of PDI called PDIp has a 46% identity and 66% similarity to PDI at the amino acid level (Desilva et al. 1996, 1997). This 56 kDa-protein has two domains with the active site motif -CXXC- and two catalytically inactive domains (Desilva et al. 1996). In the N-terminal domain the active site sequence is identical to that of PDI, but in the C-terminal domain the glycine residue following the N-terminal cysteine is replaced by a hydrophilic threonine (Desilva et al. 1996). PDIp has been shown to catalyze reductive cleavage of insulin and renaturation of reduced RNaseA (Desilva et al. 1996) and is thought to be involved in the late stages of the maturation of newly translocated proteins (Volkmer et al. 1997). It interacts specifically with tyrosine and tryptophan residues in substrate peptides (Ruddock et al. 2000) and hydroxyaryl groups in non-peptide ligands (Klappa et al. 2001). Certain hydroxyaryl group-containing xenoestrogens and phytoestrogens efficiently inhibit the interaction of peptides and misfolded proteins with PDIp (Klappa et al. 1998b, 2001).

The only member of the PDI family that has, in addition to the typical a-b-b'-a domain organization, an additional catalytic domain, a0, preceding the four domains, is the 71 kDa-protein ERp72, which was originally misidentified as a deoxycytidine kinase (Huang et al. 1989). ERp72 is also known as calcium binding protein 2, CaBP2, and it has been shown to catalyze in vitro the reduction of insulin (Van et al. 1993) and oxidation and rearrangement of disulfide bonds of denatured reduced RNase AIII (Rupp et al. 1994, Kramer et al. 2001). Recent functional characterization of the three thioredoxin-like domains suggests that the first two contribute to the catalysis whereas the third has a role in the binding of the substrates (Satoh et al. 2005).

The 64-kDa PDI-like protein of the testis, PDILT, was characterized as the first eukaryotic ER protein with an -SXXC- motif corresponding to the -CXXC- motif of PDI (van Lith et al. 2005). It has an a-b-b'-a' domain architecture, and shares 27% identity with human PDI at the amino acid level. In particular the a' domain of PDILT shows a high identity, 48%, to that of PDI (van Lith et al. 2005). The amino acid sequences, which align with the -CXXC- catalytic sites of PDI are -SKQS- in the a domain and -SKKC- in the a' domain of PDILT. As the a domain lacks cysteine residues in the site corresponding to the active sites of the catalytic domains of PDI it is redox inactive. In spite of the lack of an N-terminal surface-exposed cysteine in the -SXXC- motif, PDILT has been shown to interact with Ero1α in a redox-dependent way suggesting that the N-terminal cysteine of the active site is not required for binding to these ER oxidants (van Lith et al. 2005).

The PDI-related protein, PDIR, is a 57-kDa protein that has three thioredoxin-like -CXXC- motif-containing domains (Hayano & Kikuchi 1995) with a b-a-b-a'-a architecture (Ferrari & Söling 1999). Each of the thioredoxin-like domains has a diverse variant of the active site sequence: -CSMC-, -CGHC- and -CPHC- (Hayano & Kikuchi 1995). These three motifs were found to contribute to the isomerase activity of the protein, with the order of their contribution importance being CGHC > CPHC > CSMC, while the rate of oxidative refolding was affected in the opposite order (Horibe et al. 2004a). The chaperone activity of PDIR was shown to increase when the N-terminal b domain was replaced with the b' domain of human PDI (Horibe et al. 2004b).

P5 is a 46-kDa member of the PDI family that is also known as the calcium binding protein 1, CaBP1 (Chaudhuri et al. 1992, Fullekrug et al. 1994). It has two thioredoxin-like domains both having the -CGHC- site. In addition it has a C-terminal domain which
is rich in acidic residues and has been reported to resemble domain b (Ferrari & Söling 1999), a statement that has since been challenged (Alanen et al. 2003a). Like ERp72/CaBP2, P5/CaBP1 has been reported to have reductase and isomerase activities in vitro (Rupp et al. 1994, Kramer et al. 2001).

Two independent groups recently reported identification of a J domain-containing PDI-like protein (Cunnea et al. 2003, Hosoda et al. 2003). The J-domain is a ca. 70-amino acid domain through which J-proteins interact with their Hsp70 chaperone partners to stimulate ATP hydrolysis (Laufen et al. 1999). Both the human protein, ERdj5 (Cunnea et al. 2003), and the mouse protein, JPDI (Hosoda et al. 2003), contain an N-terminal J-like domain, also called DnaJ, in addition to four thioredoxin-like domains with -CSHC-, -CPPC-, -CHPC- and -CGPC- catalytic sequences. This 90 kDa protein was shown to interact through its J-domain with the chaperone BiP, a member of the heat shock protein family, in an ATP-dependent manner (Cunnea et al. 2003, Hosoda et al. 2003) suggesting that it is involved in assisting protein folding and quality control in the ER.

Mouse ERp46 was identified together with ERp19 as a new ER lumenal protein enriched in purified liver vesicles (Knoblach et al. 2003) and endothelial cells (Sullivan et al. 2003), and as a plasma cell thioredoxin-related protein (PC-TRP) (Wrammert et al. 2004). The mature protein has a molecular mass of 44 kDa and contains three thioredoxin-like domains that all share the catalytically active -CHGC- amino acid sequence. It has been shown to possess in vitro thioredoxin activity and to reduce insulin disulfide bonds (Wrammert et al. 2004). It also complements PDI deficiency in yeast (Knoblach et al. 2003).

ERp44 is a 44-kDa ER resident protein that has an N-terminal thioredoxin-like domain with a -CRFS- motif corresponding to the catalytic site of PDI (Anelli et al. 2002). It has been shown to form mixed disulfides with both human Eros, Ero1α and Ero1β (Anelli et al. 2002, 2003). Specifically, the cysteine residue in the -CRFS- motif interacts with Ero1α mediating its intracellular retention, as the human Eros lack any ER retention signal (Anelli et al. 2003). Recently ERp44 was shown to also bind the third luminal loop of inositol 1,4,5-trisphosphate receptor type 1 (IP₃R1) (Higo et al. 2005). By doing so, ERp44 inhibits the calcium channel activity of IP₃R1 in a pH, calcium and redox-dependent manner.

The 146-amino-acid ER resident protein ERp19 (also known as ERp18) is composed of a single thioredoxin-like domain and a C-terminal -EDEL ER retention signal variant (Alanen et al. 2003b, Knoblach et al. 2003). The catalytic domain contains an unusual active site motif -CGAC-. The protein has 15% of the oxidase activity that the a domain of PDI has and it is thus thought to participate in the formation rather than the reduction of the disulfide bonds in substrate molecules (Alanen et al. 2003b).

ERp28 is a soluble ER lumenal protein that lacks a thioredoxin active site (Ferrari et al. 1998). It is a close homologue of the protein ERp29 that is widely expressed in rat tissues (Demmer et al. 1997, Mkrchian et al. 1998, Hubbard et al. 2000). The solution structure of ERp29, determined by NMR, showed that it contains two domains, the N-terminal domain resembling that of the PDI b domain and the C-terminal domain having a novel five-helix fold (Liepinsh et al. 2001). ERp29 does not seem to possess the classical PDI activities, nor calcium-binding activities (Hubbard et al. 2004), but rather it acts as a foldase in the processing of secretory proteins within the ER (Hermann et al. 2005, Baryshev et al. 2006).
Three members of the PDI family, TMX, TMX2 and TMX3, have a transmembrane domain (Matsuo et al. 2001, Meng et al. 2003, Haugstetter et al. 2005). TMX has a unique -CPAC- active site, but no ER retention signal, although it has been suggested that it is localized in the ER membrane as it colocalizes with the ER lectin CNX (Matsuo et al. 2004). It may participate in various redox reactions through reversible oxidation of its active centre dithiol to a disulfide and catalyze dithiol-disulfide exchange reactions. TMX is highly expressed in the kidney, liver, placenta and lung (Matsuo et al. 2004). The second thioredoxin-related transmembrane protein, TMX2, was identified from a human fetal cDNA library (Meng et al. 2003). The widely expressed unprocessed 372-residue protein contains an N-terminal signal sequence, a potential transmembrane domain and a -KKDK- ER retention signal at its C-terminal end (Meng et al. 2003). TMX2 has been so far quite poorly characterized. The mature TMX3 is a 49-kDa ER protein with a -CGHC- active site motif, a possible transmembrane region and a -KKKD ER retention signal (Haugstetter et al. 2005). The protein may act as a dithiol oxidase and it has a redox potential similar to those of PDI and ERp57 (Haugstetter et al. 2005).
3 Outlines of the present study

ERp57 is a multifunctional enzyme that functions as a thiol-disulfide oxidoreductase and a chaperone in glycoprotein folding with the lectins calnexin and calreticulin. At the amino acid sequence level it is the closest homologue of the enzyme and chaperone PDI. PDI, in addition to its primary function as an oxidoreductase, acts as the \( \beta \) subunit in collagen prolyl 4-hydroxylase (C-P4H), an \( \alpha_2\beta_2 \) tetramer where its role is to keep the highly insoluble \( \alpha \) subunits in a catalytically active nonaggregated conformation and to retain the tetramer within the ER.

Both enzymes, PDI and ERp57, have the domain architecture \( a-b-b'-a' \), where the \( a \) and \( a' \) domains have the catalytic site -CGHC- motif, and the noncatalytic \( b \) and \( b' \) domains have roles in substrate binding and contributing to the overall tertiary structure of the full length proteins. ERp57 cannot replace PDI in the C-P4H assembly although PDI domains \( a \) and \( b \) can in part be substituted by those of ERp57. When this research was started structural information of any of the PDI family members was limited to the solution structures of the PDI domains \( a \) and \( b \). To better understand the roles of the various domains in the C-P4H tetramer assembly and other functions of PDI and ERp57, the biochemical and structural studies presented in this thesis were started.

The aims of this research were:

1. to purify recombinant human ERp57, to determine its domain boundaries and to identify and characterize the individual domains
2. to produce \( ^{15}\text{N} \) and \( ^{15}\text{N}\text{/^{13}}\text{C} \) -labeled thioredoxin-like domains \( a \) and \( a' \) of human ERp57 in \( E.coli \) and
3. to study the structures of the two ERp57 domains by NMR spectroscopy
4 Materials and methods

4.1 Materials

The materials used are shown summarized below in the tables 1-2. They have been described in detail in the original publications (I-IV).

Table 1. The recombinant baculovirus and E.coli constructs used.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Description</th>
<th>Used in</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baculovirus constructs, used in Sf9 and HighFive insect cells</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIS-ERp57 in pVL1392 (Invitrogen)</td>
<td>aas 1-481 of mature hERp57 with 6×His in the N-terminus</td>
<td>I</td>
</tr>
<tr>
<td>ERp111-115 in pQE-30 (Qiagen) and pET21(+) (Novagen)</td>
<td>aas 1-115 of mature hERp57</td>
<td>I, II, III</td>
</tr>
<tr>
<td>ERp111-123 in pET32a (Novagen)</td>
<td>aas 1-123 of mature hERp57</td>
<td>I</td>
</tr>
<tr>
<td>CRT p-domain in pET32 (Novagen)</td>
<td>aas 198-308 of mature rat calreticulin</td>
<td>I</td>
</tr>
<tr>
<td>E.coli strain SG13009[pREP4] (Qiagen)</td>
<td>aas 349-468 of mature hERp57</td>
<td>I, IV</td>
</tr>
<tr>
<td>E.coli constructs, used in strains BL21(DE3) or Origami(DE3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERp1-111 in pQE-30 (Qiagen) and pET21(+) (Novagen)</td>
<td>aas 1-111 of mature hERp57</td>
<td>I</td>
</tr>
<tr>
<td>ERp111-123 in pET32a (Novagen)</td>
<td>aas 1-123 of mature hERp57</td>
<td>I</td>
</tr>
<tr>
<td>CRT p-domain in pET32 (Novagen)</td>
<td>aas 198-308 of mature rat calreticulin</td>
<td>I</td>
</tr>
</tbody>
</table>

Abbreviations: amino acids (aas), human ERp57 (hERp57), human PDI (hPDI)

* Reference: Pirneskoski et al. 2001
Table 2. NMR samples and their conditions used in the structure determination.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Used in</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERp1-113, 15N and 13N/13C labeled in NaH2PO4 pH 6.3 with 10% D2O</td>
<td>II, III</td>
</tr>
<tr>
<td>ERp349-468, 15N and 13N/13C labeled in NaH2PO4 pH 6.16 with 10 mM DTT and 9% D2O</td>
<td>IV</td>
</tr>
</tbody>
</table>

4.2 Methods

The methods used are summarized in tables 3-5. They have been described in detail in the original publications (I-IV).

Table 3. Biochemical methods used.

<table>
<thead>
<tr>
<th>Method</th>
<th>Used in</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generation of recombinant expression vectors</td>
<td>I, IV</td>
</tr>
<tr>
<td>Recombinant protein expression in insect cells</td>
<td>I</td>
</tr>
<tr>
<td>Recombinant protein expression in E.coli</td>
<td>I, II, III, IV</td>
</tr>
<tr>
<td>Protein purification</td>
<td>I, II, III, IV</td>
</tr>
<tr>
<td>Protein analysis</td>
<td></td>
</tr>
<tr>
<td>Limited proteolysis</td>
<td>I</td>
</tr>
<tr>
<td>Transverse urea gradient gel electrophoresis</td>
<td>I</td>
</tr>
<tr>
<td>Circular dichroism spectrum analysis</td>
<td>I</td>
</tr>
<tr>
<td>RNase folding analysis</td>
<td>I</td>
</tr>
<tr>
<td>Chemical cross-linking</td>
<td>I</td>
</tr>
</tbody>
</table>
Table 4. The NMR spectra used in the structure determinations and relaxation delay values used to determine $^{15}$N longitudinal ($T_1$) and transversal ($T_2$) relaxation times. The spectra were acquired with Varian Unity Inova 600 and 800 MHz spectrometers equipped with either a $^1$H, $^{13}$C, $^{15}$N triple-resonance probehead or a $^1$H, $^{13}$C, $^{15}$N triple-resonance coldprobe.

<table>
<thead>
<tr>
<th>Spectrum</th>
<th>Reference</th>
<th>Used in</th>
</tr>
</thead>
<tbody>
<tr>
<td>Backbone assignment</td>
<td>Sattler et al. 1999, Permi &amp; Annila 2002</td>
<td></td>
</tr>
<tr>
<td>2D $^1$H-$^15$N-HSQC</td>
<td>II, IV</td>
<td></td>
</tr>
<tr>
<td>3D HNCA</td>
<td>II</td>
<td></td>
</tr>
<tr>
<td>3D HNCA CB</td>
<td>II, IV</td>
<td></td>
</tr>
<tr>
<td>3D CBCA(CO)NH</td>
<td>II, IV</td>
<td></td>
</tr>
<tr>
<td>3D HN(CO)CA</td>
<td>II</td>
<td></td>
</tr>
<tr>
<td>3D HNCO</td>
<td>II, IV</td>
<td></td>
</tr>
<tr>
<td>Sidechain assignment</td>
<td>Sattler et al. 1999</td>
<td></td>
</tr>
<tr>
<td>3D H(CCCO)NH</td>
<td>II</td>
<td></td>
</tr>
<tr>
<td>3D (H)CC(CO)NH</td>
<td>II, IV</td>
<td></td>
</tr>
<tr>
<td>$^{15}$N-TOCSY-HSQC</td>
<td>II, IV</td>
<td></td>
</tr>
<tr>
<td>3D H(C)CH-COSY</td>
<td>II, IV</td>
<td></td>
</tr>
<tr>
<td>3D H(C)CH-TOSY</td>
<td>II, IV</td>
<td></td>
</tr>
<tr>
<td>2D CT-$^{13}$C-HSQC</td>
<td>II</td>
<td></td>
</tr>
<tr>
<td>3D NOESY-$^{15}$N-HSQC</td>
<td>IV</td>
<td></td>
</tr>
<tr>
<td>3D NOESY-$^{13}$C-HSQC</td>
<td>IV</td>
<td></td>
</tr>
<tr>
<td>Assignment of aromatic protons</td>
<td>Sattler et al. 1999</td>
<td></td>
</tr>
<tr>
<td>2D CT-$^{13}$C-TOSY</td>
<td>II</td>
<td></td>
</tr>
<tr>
<td>2D (HB)CB(CGCD)HD</td>
<td>II</td>
<td></td>
</tr>
<tr>
<td>2D (HB)CB(CGCDCE)HE</td>
<td>II</td>
<td></td>
</tr>
<tr>
<td>3D $^{13}$C-NOESY-HSQC</td>
<td>II</td>
<td></td>
</tr>
<tr>
<td>Distance restraints</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3D $^{15}$N-NOESY-HSQC</td>
<td>III</td>
<td></td>
</tr>
<tr>
<td>3D $^{13}$C-NOESY-HSQC</td>
<td>III</td>
<td></td>
</tr>
<tr>
<td>$^{15}$N-relaxation delay values (ms)</td>
<td>Farrow et al. 1997</td>
<td></td>
</tr>
<tr>
<td>$T_1$: 10, 50, 110, 210, 410, 750, 1100, 1500, 2100</td>
<td>III</td>
<td></td>
</tr>
<tr>
<td>$T_2$: 10, 30, 50, 70, 90, 110, 150, 190, 250</td>
<td>III</td>
<td></td>
</tr>
</tbody>
</table>
Table 5. Programs used in the structure determination and evaluation. Detailed description of the parameters used in the structure calculations are found in the original publication III.

<table>
<thead>
<tr>
<th>Program</th>
<th>Reference</th>
<th>Used in</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vnmr6.1B-C</td>
<td>Varian Inc., Palo Alto, CA, USA</td>
<td>II, III, IV</td>
</tr>
<tr>
<td>FELIX97</td>
<td>Accelrys Software Inc.</td>
<td>IV</td>
</tr>
<tr>
<td>XEASY</td>
<td>Bartels et al. 1995</td>
<td>IV</td>
</tr>
<tr>
<td>Sparky 3.110</td>
<td>Goddard &amp; Kneller 2002</td>
<td>II, III</td>
</tr>
<tr>
<td>CSI</td>
<td>Wishart &amp; Sykes 1994</td>
<td>II, III, IV</td>
</tr>
<tr>
<td>TALOS</td>
<td>Cornilescu et al. 1999</td>
<td>III</td>
</tr>
<tr>
<td>CYANA2.0</td>
<td>Güntert et al. 1997, Herrmann et al. 2002</td>
<td>III</td>
</tr>
<tr>
<td>AMBER 8</td>
<td>Pearlman et al. 1995</td>
<td>III</td>
</tr>
<tr>
<td>PROCHECK-NMR</td>
<td>Laskowski et al. 1996</td>
<td>III</td>
</tr>
<tr>
<td>WHATIF</td>
<td><a href="http://swift.cmbi.kun.nl/WIWWW">http://swift.cmbi.kun.nl/WIWWW</a></td>
<td>III</td>
</tr>
</tbody>
</table>
5 Results

5.1 Characterization of the domains of human ERp57 (I)

Early gel-filtration experiments with PDI suggested it was a homodimer while analytical ultracentrifugation studies implied it was a monomer (Lambert & Freedman 1983). Recently PDI was shown to form stable dimers in the presence of physiological levels of thiophilic divalent cations, such as Zn\(^{2+}\) (Solovyov & Gilbert 2004). The oligomeric status of ERp57 was studied here by expressing a recombinant human ERp57 in S9 insect cells and applying the soluble fraction of the cell lysate to a calibrated gel filtration column. SDS-PAGE and non-denaturing PAGE analyses of the elution fractions indicated that the ERp57 polypeptide eluted from the column as a monomer with a molecular mass of about 60 kDa.

To study the domain organization of human ERp57 by limited proteolysis a recombinant ERp57 with a hexa-histidine tag in its N-terminus was generated and expressed in HighFive insect cells. The His-ERp57 polypeptide was purified with Ni-affinity chromatographic methods and subjected to limited proteolysis performed with trypsin, thermolysin, proteinase K, chymotrypsin, and V8 proteases. The protease-resistant fragments, called domains from hereon, were analysed by SDS-PAGE under reducing conditions and their N-termini identified by sequencing (see Fig. 2 in I). These analyses showed that the His-ERp57 polypeptide has three major protease-sensitive regions indicating that it consists of four structural domains, termed \(a\), \(b\), \(b'\) and \(a'\) (see Fig 3 in I). The boundaries of these domains mainly correspond to those of PDI with minor differences. The first protease-sensitive region, between domains \(a\) and \(b\), was identified by the molecular masses of the protease-resistant fragments that started from the N-terminal His-tag and by the N-termini of two fragments which started with F121 and F147. This region thus locates between residues E120 and F150. Similarly, a protease-sensitive region between domains \(b\) and \(b'\) was identified between residues Q201 and N215, and the third region between domains \(b'\) and \(a'\) was identified between E326 and L341. Interestingly, all other domains except domain \(b\) were found in the form of a single protease-resistant fragment. The \(b\) domain was only identified together with
either the third or the third and fourth protease-resistant domains, i.e. as a double domain 
bb' or a triple domain bb'a'.

Several recombinant polypeptides with an N-terminal hexa-histidine tag were 
generated, expressed in *E.coli* (see Table 1) and characterized in order to confirm that the 
protease-resistant fragments obtained from the full length ERp57 represented individual 
protein domains: five constructs coding for domain *a*, one for domain *b*, three for domain  
*b'*, one for domain *a'* and one of each coding for *bb'* and *b'a'. A construct coding for the 
full length ERp57 with a hexa-histidine tag in its C-terminus was also generated and 
expressed in *E.coli* as a control. Polypeptides coding for domains *a* (S1-V113), *b* (Q107-
I216), *a'* (S349-E468), the double domain *b'a'* (F211-E468) and the full length ERp57 with the C-terminal histidine tag were found in the soluble cell fraction and were purified 
with Ni-affinity chromatography until essentially homogenous for further analyses. The 
polypeptide coding for domain *b'* (F211-E348) was only partially soluble, but could be 
purified for further analysis.

Transverse urea gradient gel electrophoresis and circular dichroism (CD) spectroscopy 
analysis in the far UV region were performed on the purified ERp57 domains, and in 
addition, CD spectroscopy analysis was performed on the full length His-ERp57 and the 
double domain *b'a'* (see Fig.5 and 6 in I). The analyses suggested that the individual 
domains and the double domain were in a folded conformation. CD analysis suggested 
that the noncatalytic domains *b* and *b'* have minor differences in their secondary structure 
composition when compared to each other and to the catalytic domains.

The catalytic activity of the full length recombinant human ERp57 and the catalytic 
domains *a* and *a'* were studied by measuring the relative amounts of folding 
intermediates formed from denatured RNase A as a function of time (see Fig.7 in I). The 
spontaneous disappearance of the fully reduced form (8 thiols) of RNase A occurred in 20 
minutes without a catalyst (Ruoppolo et al. 1997). With His-ERp57 as the catalyst the 
fully reduced species disappeared within the first 4 minutes and after 10 minutes the fully 
oxidized species (4 disulfide bonds) predominated, but reached an abundance of only 
90% after 24 hours. In contrast, in the folding of denatured RNase A catalyzed by PDI, 
the fully oxidized species became the most prominent in 20 minutes and reached an 
abundance of 100% in 40 minutes (Ruoppolo et al. 2003). Thus the full length ERp57 
catalyzed the refolding of reduced denatured RNase A but not as effectively as PDI. The 
isolated domains *a* and *a'* of ERp57 catalyzed the disappearance of the fully reduced 
RNase A down to a relative amount of 5-10% in 4 minutes, but this species did not fully 
disappear even after 420 minutes. The fully oxidized species became the predominant one 
in 150 minutes with domain *a* and in 80 minutes with domain *a'*. The overall folding 
pathway observed in the presence of the single domains *a* and *a'* of ERp57 resembled 
that of the uncatalyzed reaction. The corresponding domains of PDI catalyze the 
disappearance of the fully reduced species and formation of disulfide bonded 
intermediates, but a point at which the fully oxidized species predominates is not reached 
with the PDI domains (Ruoppolo et al. 2003). Thus the individual catalytic domains of 
ERp57 are more efficient catalysts in the refolding of denatured reduced RNase A than 
the corresponding domains of PDI.

As ERp57 functions in the folding of glycoproteins with the lectins CRT and CNX 
(Ellgaard & Frickel 2003), and moreover, as the tip of the arm-like P-domain of these 
lectins is known to interact with ERp57 (Ellgaard *et al.* 2001), the association of
recombinant rat CRT P-domain with the individual ERp57 domains, the double domain b’a’ and ERp57-PDI hybrid polypeptides (see Table 1) was investigated by chemical cross-linking. None of the individual domains nor the b’a’ double domain associated with the P-domain. The hybrid polypeptide containing domains a and b of ERp57 and b’ and a’ of PDI with the C-terminal extension c of PDI did not associate with the P-domain either, while the hybrid polypeptide containing domains a, b and b’ of ERp57 and a’ and the C-terminal extension c of PDI was shown to interact with the P-domain (see Fig. 8 in I). These data hence demonstrate that several domains of ERp57 are required in the interaction with the P-domain of CRT.

5.2 1H, 13C and 15N resonance assignments and secondary structures of the a and a’ domains of human ERp57 (II, IV)

Constructs encoding the human ERp57 residues S1-V113 (domain a) and S348-E469 (domain a’) were expressed in E. coli and 15N and 13C/15N labelled samples were prepared and purified for nuclear magnetic resonance (NMR) studies (see Table 2). The a domain samples were analyzed at concentrations of 0.8-2.5 mM in sodium phosphate buffer, pH 6.3, with 10% D2O added, while the a’ domain samples at concentrations of 1.4-2.6 mM were in sodium phosphate buffer, pH 6.16, with 9% D2O and 10 mM dithiothreitol (DTT). To obtain the resonance assignments of 1H, 13C and 15N nuclei, several multidimensional heteronuclear NMR spectra were recorded for the a domain at 298K and for the a’ domain at 303K with Varian Unity Inova 600 and 800 MHz spectrometers equipped with either a 1H, 13C, 15N triple-resonance probehead or a 1H, 13C, 15N triple-resonance coldprobe (see Table 4). The spectra were processed with either Vnmr6.1B-C (both domains) or Felix97 (domain a’) and analyzed with Sparky (domain a) and XEASY (domain a’).

For the a domain, backbone amide correlations of 103 out of the 107 non-proline amino acids could be assigned, the missing correlations in the [1H,15N]-HSQC spectrum (see Fig.1 in II) being those of the three catalytic site amino acids W32, C33, H35, and residue G93. The side chain Hδ2 and Hε2 resonances of all four asparagines and the single glutamine, respectively, were assigned, as well as the side chain H proton of all five arginines. Almost 98% of the backbone and side chain nuclei were assigned, the missing resonances being some of the aromatic protons of F28, F29, W32, F82 and H103 and protons from the residues that could not be detected in the well-resolved [1H,15N]-HSQC spectrum. Secondary structure elements were defined by the chemical shift indices (CSI) of Hα, Cα, Cβ and C using the program TALOS and confirmed with HN-HN and HN-Hα NOE connectivities. The data suggest that the a domain of ERp57 forms a typical thioredoxin fold with a βαβαβαβαβα architecture, as was to be expected in view of its sequence similarity to the corresponding domain of PDI and thioredoxin. The chemical shifts of all assigned 1H, 13C and 15N nuclei have been deposited in BioMagResBank (http://www.bmrb.wisc.edu) with accession code BMRB 6308.

Of the 109 non-proline backbone amide correlations from domain a’ of ERp57, 105 could be detected in a well-resolved 800 MHz [1H,15N]-HSQC spectrum (see Fig.1 in IV), the missing 15N-1H correlations being those from E369, N370, K435 and N439. A
nearly complete backbone assignment was obtained by triple resonance methods, and the chemical shifts of all assigned ¹H, ¹³C and ¹⁵N nuclei were deposited in BioMagResBank with entry number BMRB 4908. Hα, Cα and C' chemical shift indices indicated that the secondary structure of the a’ domain of ERp57 consists of five β-strands and three α-helices most probably forming a thioredoxin fold (see Fig.1 in IV). The CSI of the a’ domain of PDI suggests it contains secondary structure elements in a similar order (Dijkstra et al. 1999). A minor difference was found between the domains according to their CSIs, as in the a’ domain of PDI a fourth α-helix between the third and fourth β-strands was predicted to be present.

5.3 Solution structure determination of the N-terminal domain a of human ERp57 (III)

For the tertiary structure determination of the a domain of human ERp57, in addition to the spectral analysis used for resonance assignment of the ¹H, ¹³C and ¹⁵N nuclei, ¹⁵N and ¹³C-edited NOESY spectra were analyzed with SPARKY. For structure determination the iterative automated NOE assignment approach CANDID in the software package CYANA was used. From the two NOESY spectra a total of 2293 NOE-derived distance restraints were generated by CANDID/CYANA. In addition, 136 approximated dihedral angle restraints (68 φ and 68 ψ) generated by TALOS and marked good were added to the calculations (see Table 1 in III). As the Cβ chemical shifts and side chain NOE data of C61 and C68 suggested that they were in an oxidized state, six disulfide bridge restraints, composed of lower and upper distance restraints between Cβ and Sγ atoms, were included in the calculations. The two NOESY peak lists were used along with the TALOS angle predictions for the first set of calculations. In the final set of calculations for 300 conformers, the disulfide restraints were also included and the peptide bond between Y76 and P77, which is conserved and in the cis conformation in the thioredoxin-like folds (Eklund et al. 1984), was changed to a cis conformation. The presence of a cis-bond was apparent as no NOE between the Hα of Y76 and Hβ of P77, which would be characteristic of a trans peptide bond, was found. Ultimately, the 30 best conformers with the lowest target function values were selected from the final round of calculations for structural refinement with AMBER 8. The generalized Born solvation model was used for refinement with the final set of restraints obtained from CYANA. The quality of the structures was evaluated with PROCHECK-NMR and WHATIF and the atomic coordinates of the 20 best structures were deposited in the Protein Data Bank with entry code 2ALB.

The average RMSD for the mean structure of all 113 residues was calculated to be 1.8 ± 0.42 Å for the backbone atoms and 2.22 ± 0.38 Å for the heavy atoms (see Table 1 in III). The well-ordered regions consisting of residues 23-30, 39-87 and 97-107 form a compact structure with an average backbone RMSD of 0.71 Å. The solution structure of the a domain consists of four helical regions composed of residues F11-D17 (α1), C36-R49 (α2), T65-Y71 (α3) and A97-A108 (α4) and five β-strands composed of L4-E5 (β1), L23-F29 (β2), L56-D60 (β3), T78-R83 (β4) and E86-A90 (β5). The structure shows a typical thioredoxin fold with a core made up of four parallel β strands and one
antiparallel surrounded by the helices (see Fig. 1B in III). The first helix is composed of an α-helical part (F11-R14) and a 3₁₀-helical part (I15-D17), while the others are clearly right-handed α-helices. Between β-strands 4 and 5, two residues, namely D84 and G85, form a very tightly packed 3₁₀ helical region. The active site cysteines C33 and C36 are located in a loop area preceding helix α2 and as the first residue of α2, respectively. The two other cysteines, C61 and C68, which are in their oxidized state and form a disulfide bridge as evidenced by their Cβ chemical shifts, are positioned right after the β3-strand in the loop area and in the α3-helix, respectively. These two cysteine pairs are 11.4-13.4 Å apart in the mean structure and are thus very unlikely to form mixed disulfides. In addition, the relaxation times indicate that the disulfide bridge between C61 and C68 is stable, as the times in this region do not markedly differ from the mean value (see Fig.1C in III).

The backbone mobility of the a domain was studied by measuring the 15N relaxation times for all non-proline residues, excluding the ones that were not observable in the 15N-HSQC, namely W32, C33, H35 and G93 (see Fig.1C in III). In addition, A57 was excluded from the relaxation analysis as no reliable fit for the 15N relaxation time was obtained. The average rotational correlation time of the domain is 6.1 ns, indicating that the polypeptide tumbles as a monomer under the conditions studied. Closer inspection of the relaxation data shows that the core of the domain is motionally restricted. In addition to the flexible C-terminal region, flexibility of the backbone is apparent in the regions composed of residues D17-G22, P31-K37, S74-Y76 and Y91-R95, showing decreased $T_1/T_2$ ratios. We regard this as an indication of increased internal mobility in the loop regions. The loops are mainly characterized by only a small number of short and medium-range NOEs, yielding poorly defined structures in these regions, including the active site.

Two of the flexible regions participate in the redox activity of the domain, namely the region composed of residues P31-K37 which contains the active site motif -CGHC-, and the region composed of residues Y91-R95, where R95 participates in the redox activity, probably with assistance from the preceding loop residues (Lappi et al. 2004). The S74-Y76 flexible region is followed by a cis-proline that is conserved between thioredoxin-like domains and is essential for their function and stability (Charbonnier et al. 1999). The only flexible region that is not close to the active site in the three-dimensional structure is the D17-G22 loop between α1 and β2.

The 16 residues that are fully conserved between the ERp57 and PDI α domains and thioredoxin (see Fig.2A in III) cluster mainly into one area, around the catalytic site, the only exceptions being F82 and G85, which are found in a hydrophobic patch on the opposite side of the molecule (Fig. 7). Four residues that are identical between these three polypeptides are buried, namely V26, F28, A30 and V59. The ERp57 α domain contains eleven aromatic residues, ten of which are located on the surface, the only buried one being F28. Most of the aromatic residues are found on the surface face opposite that containing the β1 strand.
Fig. 7. Distribution of conserved residues on the surface of the ERp57 a domain. The positively charged K37 is shown in blue, the negatively charged D60 in red, active site cysteines C33 and C36 in yellow and other 12 conserved residues in orange. In the upper left orientation the β1 strand is facing towards the viewer. The other views correspond to successive 90 degree rotations to the right around the y-axis.

As could be expected from the amino acid sequence alignment of the ERp57 and PDI a domains, their Cα traces are very similar, with the residues participating in catalysis in identical positions (see Fig.1D in III). There are some distinct differences in structure, however. The loop region between α1 and β2 in the ERp57 a domain is composed of five residues, T18-G22, while the corresponding loop region in the PDI a domain consists of only two. This particular loop region in ERp57 was found to be flexible according to 15N relaxation measurements (see above). Another difference exists in the loop regions between the α2 and β3 of the ERp57 and PDI a domains, the loop in the PDI domain being longer. In the ERp57 a domain the β4 strand is followed by a tight turn that forms a short two-residue 3_10 helix, which is followed by the β5 strand. In contrast, the β4 and β5 strands of the PDI a domain are separated by a loop of eight residues.

Although the Cα traces of the ERp57 and PDI a domains are highly homologous (see Fig.1D in III), distinct differences were found between their surface potential (see Fig.2B and 2C in III). Negatively charged residues form clusters on two surface faces of the a domain of ERp57, one contains the N-terminal β1 strand, and the other is the face adjacent to this when the domain is rotated 90° to the right around the vertical axis.
Positively charged residues are scattered around the entire surface of the domain and do not form clusters. A definite uncharged hydrophobic region is present on the surface opposite to that containing the \( \beta_1 \) strand. In contrast charged residues form distinct clusters on all the surface faces of the PDI \( \alpha \) domain. In both domains the faces containing the \( \beta_1 \) strands are mainly negatively charged, with some individual positive side chains pointing away from the surface. When the molecules are turned 90 degrees to the right around their vertical axis, a major difference is observed. The surface of the ERp57 \( \alpha \) domain is negatively charged on this face, while the surface of the PDI \( \alpha \) domain has a positively charged cluster and an uncharged area. Another difference in the distribution of electropotential between these two domains is found on the face opposite that containing the \( \beta_1 \) strand. This surface of the ERp57 \( \alpha \) domain is mainly uncharged, while the corresponding face of the PDI domain has definite positively and negatively charged clusters. The PDI residues forming the positively charged cluster are R89, R115, K97, K113 and K114. Although the ERp57 residues R83, K105 and K106 correspond to the PDI residues R89, K113 and K114, the ERp57 residues are not clustered.
Discussion

ERp57 is the closest homologue of PDI with a 29% identity and 56% similarity in their amino acid sequence (Bourdi et al. 1995, Hirano et al. 1995, Koivunen et al. 1996), which suggests that like PDI, ERp57 may possess a four-domain \(a-b-b'-a'\) architecture. The limited proteolysis studies performed in this study provide biochemical evidence for this suggestion. The domain organization has recently been confirmed, and the domain boundaries have also been defined by three other groups, by limited proteolysis studies (Frickel et al. 2004, Urade et al. 2004) and by sequence homology studies (Alanen et al. 2003a). Although there are some differences in the results obtained from the limited proteolysis studies performed by us and by Frickel and co-authors and Urade and co-authors, the \(E.\ coli\) domain constructs generated and conclusions about the domain boundaries are compatible.

Although our results confirm the four-domain structure of ERp57, some minor differences were found between ERp57 and PDI in the locations of the protease-sensitive regions and the presumed lengths of some of the domains (see Fig.3 in I). The catalytic domains \(a\) and \(a'\) of ERp57 correspond very well to those of PDI, while domains \(b\) and \(b'\) seem to have a greater variation in their lengths. \(E.\ coli\) constructs encoding polypeptides consisting of ERp57 amino acids S1-V113 (\(a\) domain), Q107-I216 (\(b\) domain), F211-E348 (\(b'\) domain) and S349-E468 (\(a'\) domain) were expressed in sufficient amounts as soluble, folded polypeptides to enable further analysis. The NMR studies performed in this study defined the domain boundaries further and showed that the catalytic domains \(a\) and \(a'\) consist of residues L4-A108 and V354-R458, respectively.

Secondary structure predictions using two programs suggest that the first \(\alpha\)-helix of domain \(b\) of human ERp57 begins around E118, which would indicate a ten residue linker region between domains \(a\) and \(b\). The first protease-sensitive region separating domains \(a\) and \(b\) extends from residue E120 until the region containing residues F140-A150. In PDI the last secondary structure element of domain \(a\) ends at T116 (Kemmink et al. 1996) and the first of domain \(b\) begins at residue A120 (Kemmink et al. 1999), thus leaving only a three-residue \(-\text{GPA}-\) linker between the domains. This \(-\text{GPA}-\) tripeptide is conserved between the human ERp57 and PDI. Thus, according to the present data, the exact length of the linker region between domains \(a\) and \(b\) in human ERp57 and the N-
terminus of the $b$ domain remains to be determined. It is however unlikely that there could be an unstructured linker containing ten or more residues.

In the limited proteolysis studies, domain $b$ was obtained only together with domain $b'$ or the rest of the polypeptide, but a protease-sensitive region from Q201 to N215 was identified, suggesting that the C terminus of domain $b$ is located around residue E200. However, the secondary structure predictions suggested that the last $\alpha$-helix of domain $b$ extends to residues E214 and N215, contradicting the results obtained by the proteolysis studies. The C terminus of domain $b$ of PDI is around residue N215 which aligns with residue N215 of ERp57. Nevertheless, the limited proteolysis studies performed on human PDI revealed a protease-sensitive region after N215 (Freedman et al. 1998) thus differing from the proteolysis studies on ERp57. It is thus likely that domain $b$ of human ERp57 may be slightly shorter than the corresponding PDI domain, due to shortening of its C-terminal end.

A folded polypeptide corresponding to domain $b'$ of ERp57 was found to be expressed with the construct coding for the ERp57 residues 211-348. The proteolysis studies and secondary structure predictions suggest that the N terminus is located around residue I219. The C terminus of domain $b'$ is somewhat unclear as a protease-sensitive region was found to be located between residues 313 and 341, in particular in the interval 326-341. Secondary structure predictions suggest that the last $\alpha$-helix of domain $b'$ may end at F333 or D334.

A 19-residue linker region, $x$, has been identified between domains $b'$ and $a'$ of human PDI (Pirneskoski et al. 2004). The recent crystal structure of yeast PDI identified this same $x$ region as comprising 17 residues and being mainly in an extended conformation (Tian et al. 2006). As the first secondary structure element of domain $a'$ of ERp57 was found to begin at V354, it is thus very likely that region $x$ is present also in human ERp57. NMR data indicated that the last $\alpha$-helix of domain $a'$ ends at R458, human ERp57 thus having 23 additional residues forming the C terminal region of this domain. This stretch is distinctly shorter than the 37-residue extension $c$ forming the C terminal region of human PDI. The C-terminal tail of ERp57 contains a stretch of basic lysines, while the C-terminal extension of PDI is composed of mainly acidic residues. Surprisingly, the acidic C-terminal extension of yeast PDI forms an $\alpha$-helix and stabilizes the structure of the $a'$ domain. The presence of a corresponding $\alpha$-helix in human PDI is unclear as the amino acid identity between these two species in this region is low. The secondary structure prediction of ERp57 however, suggests the presence of a C-terminal $\alpha$-helix (see Fig3. in I).

According to the data presented here, domains $a$ and $a'$ of ERp57 have the thioredoxin fold, like domains $a$ and $b$ of PDI (Kemmink et al. 1996, 1999) and all four domains of yeast PDI (Tian et al. 2006). The data for domain $a'$ suggests that it differs slightly from the $a$ domain as the third $\alpha$-helix seems to be missing, and, in addition to the conserved cis-proline peptide bond, another cis-proline appears to be present in the structure (data not shown). Although the CD spectra of domains $b$ and $b'$ of human ERp57 suggested only minor differences in their secondary structure composition when compared to domains $a$ and $a'$, it is nevertheless very likely that there are specific differences in the structures of the corresponding domains between ERp57 and PDI, as ERp57 domains $b'$ and $a'$ cannot substitute for the corresponding PDI domains in the assembly of the collagen prolyl 4-hydroxylase $\alpha_2\beta_2$ tetramer, and as ERp57 domains $a$ and $b$ can only in...
part substitute for the corresponding PDI domains in this assembly (Pirneskoski et al. 2001).

ERp57 is characterized by its specificity for glycoproteins. Its substrate selection is facilitated by the binding of nascent glycoproteins to calnexin or calreticulin, which then associate with ERp57 through their arm-like P-domain. NMR spectroscopy analysis and a membrane yeast two-hybrid system showed that the P-domain of CNX interacts with the basic C-terminal end of ERp57 (Pollock et al. 2004). In this study no association of the P-domain of rat CRT with any single human ERp57 domain was found, nor with the double domain b’a’, indicating that this association requires several domains. Cross-linking studies with the hybrids ERpabPDI'a'c and ERpabPDIb’a’c showed that the b’ domain is an absolute requirement for complex formation and cannot be replaced by the corresponding domain of PDI. Similar data has also been obtained by other groups, showing that domains b and b’ are the minimum requirement for the ERp57-lectin complex formation, the C-terminal extension enhancing the interaction to the level of intact ERp57 (Russell et al. 2004, Urade et al. 2004).

The solution structure of the human ERp57 a domain revealed that it has a compact thioredoxin fold structure, which is highly similar to that of the PDI a domain (Kemmnink et al. 1996) and of other members of the thioredoxin superfamily, the active site being located in identical positions at the N-terminal end of the α2 helix. Other features in common with the structures of the ERp57 and PDI a domains and thioredoxin are the cis-bond preceding a conserved proline at the N-terminal end of the β4 strand close to the active site (P77 in ERp57), an acidic residue buried in a pocket behind the cis-proline (E27 in ERp57), and a conserved lysine close to the acidic residue (K58 in ERp57). These glutamic acid and lysine residues form a charge pair that is involved in proton transfer and is located in a binding pocket in the active site (Ellgaard & Ruddock 2005). In addition, the α2 helix is distorted by a conserved proline (P41 in ERp57) in its middle in all three structures. The structure of the yeast PDI a domain suggests that the N-terminal region of the domain folds back over the active site and stabilizes the conformation with hydrogen bonds. As in the case of the human PDI a domain, this stabilization does not occur in the a domain of ERp57 either, but instead a distinct binding pocket can be observed on the surface. In contrast to thioredoxin, the PDI a and a’ domains and the ERp57 a’ domain, the ERp57 a domain has two additional cysteines, which were found to form an intradomain disulfide bridge. This is in accordance with previous assays showing that only one free cysteine is detected in an oxidized full-length recombinant ERp57 under denaturing conditions, and none are found in its a domain, indicating that these cysteines form a disulfide bridge (Frickel et al. 2004). The a domain of yeast PDI also has two additional cysteines in corresponding positions likewise forming a disulfide bond (Tian et al. 2006), which destabilizes the oxidized state of the active-site cysteines (Wilkinson et al. 2005).

Like PDI, ERp57 has been shown to catalyze reduction, isomerization and oxidation in vitro in this study and others (Srivastava et al. 1993, Bourdi et al. 1995, Hirano et al. 1995, Frickel et al. 2004, Mayer et al. 2004). The ERp57 a domain has a redox potential of -0.167 V, which is comparable to that of full-length PDI, and intermediate between those of bacterial thioredoxin, which is a reductase, and DsbA, an oxidase (Frickel et al. 2004). The nature of the residues between the active site cysteines is a critical determinant in the redox properties of these enzymes, e.g. a change from the proline at
the active site of thioredoxin to the histidine present at those of the PDI and ERp57 a domains accounts for half of the difference (Lundström & Holmgren 1993, Huber-Wunderlich & Glockshuber 1998). In addition, other residues that come close to the active site in the tertiary structure influence the pKₐ values of the active site cysteines. The PDI and ERp57 a domain residues R103 and R95, respectively, which are located in the loop between the β₄ strand and the α₃ helix, have been shown to be critical for the catalysis of oxidation (Lappi et al. 2004). It has been suggested that this residue may function by moving into and out of the active site and may thus modulate the pKₐ values of its cysteines (Lappi et al. 2004). This residue is conserved in most PDI family members (Ellgaard & Ruddock 2005), but is replaced by a glutamate in thioredoxin. The R95 was shown here to be located in a flexible loop in the ERp57 a domain, thus supporting the suggested mode of action. The in vitro redox and isomerase activities of ERp57 are in most cases lower than those of PDI, as shown in this study and by other groups (Srivastava et al. 1993, Bourdi et al. 1995, Hirano et al. 1995, Frickel et al. 2004, Mayer et al. 2004). It has been proposed that this is due to the direct efficient interaction of PDI with its substrates through the principal binding site in its b’ domain, while the corresponding site in ERp57 is adapted for calnexin/calreticulin association. However, there are also differences in catalytic behaviour between the single a domains in ERp57 and PDI, so that some of the intrinsic properties of the domains must be responsible for these differences despite their highly similar structures of the domains.

Site-directed mutagenesis studies have shown that a region of the ERp57 b’ domain that is equivalent to the primary binding site in the PDI b’ domain is required for calreticulin binding, although the actual interaction site encompasses a larger region, and the a domain appears to be entirely dispensable as far as this binding is concerned (Russell et al. 2004). Despite this, it is likely that the ERp57 a domain contains a site that contributes to protein-protein interactions. Binding sites in three PDI domains, a, b’, and a’, are required for efficient C-P4H tetramer assembly (Koivunen et al. 2005), and the fact that the PDI a domain in this assembly can in part be replaced functionally by the ERp57 a domain indicates that a corresponding binding site exists in the ERp57 a domain.

Hydrophobic regions in proteins provide potential sites for noncovalent interactions. A large hydrophobic patch was found to be present here in the ERp57 a domain on the surface opposite to that containing the N-terminal β1 strand. Residues corresponding to those found in the hydrophobic binding groove leading to the active sites of DsbA and thioredoxin (Qin et al. 1996, Guddat et al. 1997), and those in the PDI a domain that have been found to be important in C-P4H assembly (Koivunen et al. 2005) map to this region of the ERp57 a domain. It has been shown that an ERp57 a domain mutation H103W, when introduced into an ERp57aPDIbb’a’c hybrid polypeptide, increases C-P4H assembly (Koivunen et al. 2005), indicating that the residue at this site is involved in protein-protein interactions. The potential role of the large hydrophobic region in interactions with other parts of the ERp57 molecule or its substrates remains to be studied. The four thioredoxin-like domains of yeast PDI form the shape of a twisted U, its inside containing a continuous hydrophobic surface and all four domains contributing to its formation (Tian et al. 2006). As clear differences were found in the distribution of hydrophobic regions in the ERp57 and PDI a domain surfaces, it is very likely that there are differences in the hydrophobic surfaces of the full-length polypeptides.
ERp57, together with calnexin, calreticulin, the accessory molecule tapasin and the transporter associated with antigen processing, is involved in the folding of major histocompatibility complex (MHC) class I molecules and in the formation of a peptide loading complex, which plays a critical role in immune responses by presenting foreign antigenic peptides to cytotoxic T-cells (Morrice & Powis 1998, Hughes & Cresswell 1998). Tapasin forms a mixed disulfide with the catalytic site of the ERp57 α domain and is an essential structural component in the assembly of a stable peptide loading complex (Dick et al. 2002, Garbi et al. 2006). Interestingly, noncovalent interactions between the proteins were recently shown to maintain this interaction by inhibiting the reductase activity of the catalytic site of this domain (Peaper et al. 2005). It is possible that such interactions involving ERp57 are provided by the α domain residues in the hydrophobic region.
References


Gong QH, Fukuda T, Parkison C & Cheng SY (1988) Nucleotide sequence of a full-length cDNA clone encoding a mouse cellular thyroid hormone binding protein (p55) that is homologous to protein disulfide isomerase and the β-subunit of prolyl-4-hydroxylase. Nucleic Acids Res 16: 1203.


Wang CC & Tsou CL (1993) Protein disulfide isomerase is both an enzyme and a chaperone. FASEB J 7: 1515-1517.


Yao Y, Zhou Y & Wang C (1997) Both the isomerase and chaperone activities of protein disulfide isomerase are required for the reactivation of reduced and denatured acidic phospholipase A2. EMBO J 16: 651-658.

Original articles


The permission of the following copyright owners to reproduce the original articles has been granted

(I) Reprinted with permission of the American Society for Biochemistry and Molecular Biology

(II and IV) Reprinted with kind permission of Springer Science and Business Media

Original papers are not included in the electronic version of the dissertation.
859. Härkönen, Päivi (2005) 17β-Hydroxysteroid dehydrogenases/17-ketosteroid reductases (17HSD/KSRs) in prostate cancer. The role of 17HSD/KSR types 2, 5, and 7 in steroid hormone action and loss of heterozygosity at chromosome region 16q

860. Yrjänä, Sanna (2005) Implementation of 0.23 T magnetic resonance scanner to perioperative imaging in neurosurgery


862. Ylönen, Riikka (2005) Characterization of the function of type XIII collagen in mice; specific roles during cardiovascular development and postnatally in bone modeling

863. Karihtala, Peeter (2005) Oxidative damage and counteracting mechanisms in breast carcinoma


870. Hinkula, Marianne (2006) Incidence of gynaecological cancers and overall and cause specific mortality of grand multiparous women in Finland

Laura Silvennoinen

ERP57—CHARACTERIZATION OF ITS DOMAINS AND DETERMINATION OF SOLUTION STRUCTURES OF THE CATALYTIC DOMAINS