Probing the structure of human protein disulfide isomerase by chemical cross-linking combined with mass spectrometry

Li Peng\textsuperscript{a}, Morten Ib Rasmussen\textsuperscript{a}, Anna Chailyan\textsuperscript{a}, Gunnar Houen\textsuperscript{b}, Peter Højrup\textsuperscript{a,⁎}

\textsuperscript{a}Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense, Denmark
\textsuperscript{b}Department of Clinical Biochemistry, Immunology and Genetics, Statens Serum Institut, Copenhagen, Denmark

ARTICLE INFO

Article history:
Received 10 February 2014
Accepted 24 April 2014
Available online 2 May 2014

Keywords:
Cross-linking
BS3
Protein structure
Calnexin
ERP72
Placenta

ABSTRACT

Protein disulfide-isomerase (PDI) is a four-domain flexible protein that catalyzes the formation of disulfide bonds in the endoplasmic reticulum. Here we have analyzed native PDI purified from human placenta by chemical cross-linking followed by mass spectrometry (CXMS). In addition to PDI the sample contained soluble calnexin and ERP72. Extensive cross-linking was observed within the PDI molecule, both intra- and inter-domain, as well as between the different components in the mixture. The high sensitivity of the analysis in the current experiments, combined with a likely promiscuous interaction pattern of the involved proteins, revealed relatively densely populated cross-link heat maps. The established X-ray structure of the monomeric PDI could be confirmed; however, the dimer as presented in the existing models does not seem to be prevalent in solution as modeling on the observed cross-links revealed new models of dimeric PDI. The observed inter-protein cross-links confirmed the existence of a peptide binding area on calnexin that binds strongly both PDI and ERP72. On the other hand, interaction sites on PDI and ERP72 could not be uniquely identified, indicating a more non-specific interaction pattern.

Biological significance
The present work demonstrates the use of chemical cross-linking and mass spectrometry (CXMS) for the determination of a solution structure of natural human PDI and its interaction with the chaperones ERP72 and calnexin. The data shows that the dimeric structure of PDI may be more diverse than indicated by present models. We further observe that the temperature influences the cross-linking pattern of PDI, but this does not influence the overall folding pattern of the molecule.

© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-SA license (http://creativecommons.org/licenses/by-nc-sa/3.0/).

1. Introduction

Elucidating the structure of proteins is an essential part of molecular structure-function studies. However, despite the fact that the amino acid sequence is known for most proteins of many species, the structure of many proteins remains unknown or only partly characterized. This is due mainly to limitations in the availability of the proteins and in the methods used for protein structure elucidation.

X-ray crystallography has for many years been the major method for detailed structure elucidation of proteins [1–6]. This method requires relatively large amounts of pure protein and is limited to proteins, which form suitable crystals. Large
amounts of protein can usually be obtained by recombinant technology; however, this often does not lead to suitable crystals. Obstacles to this include inherent mobility or flexibility in multi-domain proteins, heterogeneity, or a lack of a well-defined structure. In such cases, structural information may be obtained for crystallizable domains or for defined parts of the protein. Even so, X-ray crystallography can only give information about solid state structures of proteins and for many proteins, particularly enzymes, molecular dynamics is vital to the function, e.g. for PDI, which must be able to break improper disulfide bonds and to join the proper ones.

Protein disulfide isomerase (PDI), a 57 kDa protein composed of four thioredoxin domains, two oxido-reductase domains (a and a') with CGHC active sites, two non-catalytic domains (b and b'), a linker region between domains b' and a', and a C-terminal acidic tail [7–9]. The main responsibility of disulfide bonds is to stabilize protein structures, while in some proteins the cysteines perform redox regulation of enzymatic activity. Approximately one third of human proteins contain disulfide bonds [10], and the formation of disulfide bond limits the rate of protein synthesis. Protein disulfide isomerase (PDI), was published in 1963 as the first folding catalyst [11]. It catalyzes disulfide bond formation, breakage and rearrangement, and thus assists proper protein folding [12,13]. PDI is mainly expressed in the endoplasmic reticulum (ER), but also on the cell surfaces of lymphocytes, hepatocytes, and platelets [14,15]. In the ER it also takes part in peptide loading onto major histocompatibility complex class I [16], as well as regulating NAD(P)H oxidase [17].

Although PDI was discovered more than 50 years ago, a complete X-ray structure was not solved until 7 years ago for yeast PDI (yPDI) [18] and has only recently been obtained for human (hPDI) PDI [19]. On the way to solve the full structure, X-ray crystallography models for individual domains were solved and yielded important information about the active sites and substrate binding sites [18,20–22].

Nuclear magnetic resonance (NMR) spectroscopy can yield information about structure and dynamics for proteins in solution and can be used for proteins up to 100 kDa but also requires large amounts of pure protein, which must not aggregate in solution [23–32]. For PDI, NMR spectroscopy has yielded important information about the structure and dynamics of several domains [33–38] but not yet the whole enzyme.

Small angle X-ray scattering (SAXS) can yield low resolution structural information for proteins in solution, but has the same requirements for sample amount and purity as NMR spectroscopy [31,39–43]. For PDI, SAXS has shown an annular arrangement of the four domains [44].

Thus, the above methods each have their advantages and disadvantages, but all are challenged by the requirement for sample purity. Here, we show, using PDI and other endoplasmic reticulum (ER) proteins, that chemical cross-linking in combination with mass spectrometry (CXMS) can yield relevant information about the structure, internal dynamics and intermolecular interactions for low concentrations of mixtures of native proteins in solution. CXMS is particularly useful in combination with known structural models obtained by the methods described above and when handling proteins from natural sources, that have not been obtained in a very high purity.

2. Materials and methods

2.1. Materials

NaH2PO4, Na2HPO4, NH4HCO3, CH3CN and TFA were from Sigma Aldrich (St. Louis, MO). Bis[sulfosuccinimidyl] suberate (BS3) was obtained from Thermo Scientific Pierce (Rockford, IL). Porcine trypsin was a gift from Novo Nordisk (Copenhagen, Denmark). Spin filters were from Sertorius Stedim Biotec (Bohemia, NY). Poros Oligo R2 and R3 50 were from Applied Biosystems (Foster City, CA). ReproSil Pur C18 AQ 3 μm was from Dr Maisch GMBH (Ammerbuch-Entringen, Germany). Pure water was obtained from an Elga PureLab Flex (Marlow, UK).

2.2. Purification of PDI

PDI was obtained during purification of human placenta chaperones as described elsewhere [45]. In short, a placenta was homogenized with Triton X-114, the supernatant was separated by temperature-dependent phase separation, and the supernatant from this was precipitated by ammonium sulfate. The new supernatant was ultradialfiltrated, and the retentate was applied to a Q Sepharose Fast Flow column equilibrated with 20 mM Tris, 1 mM CaCl2, pH 7.5. Proteins were eluted with a gradient of NaCl (0-0.5 M). Eluted proteins were monitored by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and enzyme-linked immunosorbent assay (ELISA) using antibodies directed against calreticulin, calnexin, ERp57, ERp72 and PDI. The fraction used for the current analysis corresponds to fractions 132–136 in Fig. 3 of [45]. Two different preparations were used, one for the room temperature experiment, and one for the low/high temperature experiments.

2.3. Cross-linking

50 μg sample (fraction 132–136) was transferred to phosphate buffer (25 mM, pH 7.5) to a concentration of 2 μg/μL using 10 kDa spin filter. Before adding cross-linkers, samples were equilibrated for 30 min under different temperatures, 0 °C on ice water, room temperature and 37 °C separately in a thermostat. Following equilibration, an equal volume of phosphate buffer containing 2500 μg BS3 was added to the samples to a final ratio of 1:50 w/w protein: cross-linker (approximate molar ratio of 1:376), and the cross-linking reactions were carried out under the same temperatures as the equilibration step. After 1 h, the reaction was terminated by the addition of 1 M NH4HCO3 to a final concentration of 20 mM NH4HCO3. After 20 min, the samples were reduced, alkylated and digested by 2% trypsin w/w at 37 °C overnight. Two different batches containing the same proteins in different amounts were used, one for the room temperature assay and another for the 0 °C and 37 °C experiments.
2.4. Strong cation exchange chromatography

The samples were reduced by adding dithiothreitol to a concentration of 5 mM, incubating at 56 °C for 30 min followed by alkylation by iodoacetamide (10 mM) at room temperature for 30 min. Digestion of the protein was performed with 2% trypsin (w/w) overnight at 37 °C, and the resulting peptides were separated by strong cation exchange (SCX) chromatography using a polySULFOETHYL A, 150 × 1 mm, 5 μm particle size, 300 Å pore size column (PolyLC Inc., Columbia, MD). The buffer gradient was: 0–0% B (10 min); 0–40% B (30 min); 40–90% B (5 min); and 100% B (5 min) at a flow rate of 50 μL/min (A buffer, 30% ACN, 0.1% TFA; B buffer, 30% ACN, 0.1% TFA, 500 mM NaCl). The separation was performed on an Agilent 1200 HPLC (Agilent Technologies) monitored at 214/254/280 nm. The entire run was manually collected and divided into 4 fractions. Each fraction was desalted using homemade Poros Oligo R2/R3 (R2:R3 ratio: 4:1) RP microcolumns [45].

2.5. Mass spectrometry

The desalted peptides were loaded onto a 100 μm inner diameter and 18 cm RP capillary column (homepacked with ReproSil Pur C18 AQ 3 μm sorbent) in buffer A (0.1% formic acid, 5% ACN) and run on a Proxeon 1100 easy-nLC system (Thermo Fisher Scientific/Proxeon Biosystems, Odense, Denmark).

The peptides were eluted using either a 50 min gradient (protein identification) or an 80 min one (cross-link identification) from 0 to 34% B-buffer (95% ACN, 0.1% formic acid) at 350 nL/min flow rate and via nanoelectrospray introduced into an LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). A full MS scan in the mass area of 300–1800 m/z was performed in the Orbitrap with a resolution of 30,000 FWHM and a target value of 1 * 10(6) ions. For each full scan the top five ions were selected for either collision induced dissociation (CID) for protein identification or for high energy collision dissociation (HCD) for cross-link identification. CID settings were as follows, activation time, 30 ms; isolation width, 2.5; normalized collision energy, 35; repeat count, 1; repeat duration, 30; and exclusion duration, 45. HCD settings were as follows: threshold for ion selection, 15,000; target value of ions used for HCD 2 * 10(5); activation time, 10 ms; isolation width, 2.5; normalized collision energy, 42; repeat count, 1; repeat duration, 30; and exclusion duration, 45.

2.6. Data analysis

Raw data were converted to mgf format by Proteome Discover 1.4 (Thermo Fisher Scientific). The data were further filtered by MGF Filter V0.104 to retain only the top 125 most intense peaks per scan (in-house built software, now included in the MassAI search tool). For protein identification the data were searched by the MassAI search engine against the Homo sapiens part of the UNIProt database with the parameters of 10 ppm MS accuracy, 0.1 Da MS/MS accuracy, trypsin as enzyme, three allowed missed cleavages, carbamidomethylation of cysteine as fixed modification, and oxidation of methionine as variable modifications. Based on the proteins identified, a database including only these proteins was generated. Results were further verified by searching the same dataset by the Mascot search engine (Matrix Science, UK).

Searches for crosslinked peptides were conducted against this database using the CrossWork algorithm [46]. The initial searches were carried out using the original CrossWork software (www.massai.dk) with the following parameters, cross-linker: BS3; protease: trypsin with allowance for 3 missed cleavages, fragmentation mode: CID/HCD; MS tolerance: 0.010 Da; MS/MS tolerance: 0.025 Da; cross-link search modes included mass matching, level 1 and level 2 search, as described [46]. During preparation of this paper, the CrossWork algorithm itself was incorporated into the MassAI analytical software package (in-house developed; www.massai.dk). The latter was then used for constructing the cross-link heat maps, using search parameters identical to the CrossWork searches. When constructing the heatmaps, low scoring scans (score < 10) were excluded. Representative scans for the cross-links discussed were annotated and exported (Supplementary Fig. 3).

2.7. Modeling

Two complete structures of human PDI are publicly available, both determined by X-ray crystallography, known as oxidized PDI (4EL1) [19] and reduced PDI (4EKZ) [19]. As our experimental conditions closely match that of the oxidized version, all our further structure analyses are based on the PDI conformation solved in the above-mentioned redox state.

The crystal structure of hPDI, 4EL1, shows two PDI molecules in the asymmetric unit with the a domain of one molecule nestled between the a’ and the a domain of the other. This is likely a result of crystallization, and the dimer of the hPDI molecule is still missing. Whereas for the yeast, Tian et al. have shown that the yPDI has a propensity to dimerize in solution and have deposited the structure in PDB with the code 3BOA.

1. Dimer structure

When aligning yeast and hPDI the sequence identity is 31%. Due to the low sequence identity of yeast and human PDI (not all cross-linked Lys residues are conserved in yeast) it is rather complicated to use it in our analysis, thus a model using the yeast conformation as a template has been constructed. We keep in mind, that the structure of full-length yPDI cannot represent the structure of hPDI, because they are not functionally equivalent and are dissimilar in several aspects, but due to the absence of structural information on the hPDI dimer, we overcome the problem by modeling the hPDI dimer based on yPDI. The following steps have been performed to obtain the above-mentioned dimer structure, in the following called hYeastDimer:

1. As a first step a sequence similarity search on the nr database March 28, 2014 using two iterations of CS-Blast has been performed.
2. We selected the top 100 multiple sequence alignment having an e-value lower than 1e-103.
3. The sequences obtained in the previous step together with the 2 structures i.e. pdb code 4EL1 and 3BOA have
been used to produce the structurally correct sequence alignment of the two above-mentioned structures using PROMALS3D.  

4. The sequence alignment has subsequently been used to build the three-dimensional model of hPDI using Modeller (Fig. 6).  

2. Dimeric structure cross-link list  
In order to exclude the cross-links that were observed to fit the monomer model, all the intra-protein Cα cross-link distances have been calculated using the hPDB structure 4EL1 (as a first step all the coordinates of missing residues in the 4EL1 structure have been modeled using Modeller [47]) and the hYeastDimer model with 30 Å threshold (the sum of the length of a BS3 cross-linker and lysine side chains and 5 Å for flexibility). Out of 261 total cross-links identified at room temperature only 123 fitted the monomeric structure of 4EL1 and 117 the hYeastDimer one. The remaining 138 and 144 cross-links were in discordance thus suggesting the existence of alternative conformations.

3. Cross-link ambiguity  
The 138 cross-links that were in discordance with the 4EL1 structure have been sorted according to the copy number in order to identify the most reliable ones. The total number of cross-links considered for further data analysis was equal to 10 (the number of cross-links seen more than 5 times; this number being a sum of mean (2.6) + sd (2.2) of the data). But since the hPDI is a homodimer, the cross-link list determined experimentally between the residues X and Y and obtained in the previous step can be represented by 2 scenarios:  
1. an inter-protein cross-link between the residue X of chain A and the residue Y of chain B  
2. an inter-protein cross-link between the residue X of chain B and the residue Y of chain A.  
Therefore, the total number of the most reliable 10 cross-links should be multiplied by two resulting in 20 cross-links, but since one of the cross-links is symmetrical (the residue 54 does a cross-link with itself, there is no need to consider 2 scenarios i.e. residue 54 chain A-residue 54 chain B and vice versa as in the remaining 9 cases). 19 cross-links have been selected for the further conformational searches to produce models of the corresponding dimer complexes.  

4. Dimer complex search  
Dimer models of PDI were created based on the monomeric unit of 4EL1 using the RosettaDock modeling suite (http://www.rosettacommons.org). RosettaDock is a well-established docking protocol that incorporates physical and empirical energy functions and fast conformational searching techniques. The algorithm searches the rigid-body and side-chain conformational space of two interacting proteins, making it suitable for modeling protein complexes starting from the free monomers. It is a multi-scale Monte Carlo-based algorithm that combines a low-resolution, a centroid-mode, a coarse-grain stage and a high-resolution all-atom refinement stage that optimizes both rigid-body orientation and side-chain conformation. Scoring in the low-resolution phase includes residue-residue contacts and bumps, knowledge-based terms for residue environment and residue-residue pair propensities. In the high-resolution phase, the energy is dominated by van der Waals energies, orientation-dependent hydrogen bonding, implicit Gaussian solvation, side-chain rotamer probabilities and low-weighted electrostatics energy.

3. Results

3.1. Cross-linking

In the purification of ER chaperones, PDI was greatly enriched in some fractions, but was not completely separated from a number of other constituents as displayed by SDS PAGE (Fig. 1). Peptide mapping of the entire fraction revealed the presence of ERp72 (PDI A4), PDI and calnexin as the major constituent proteins with sequence coverages of 93, 92 and 89%, respectively (Supplementary Fig. 1), along with trace amounts of other proteins. The finding of calnexin was unexpected, as it is synthesized as a transmembrane type I protein, and the purification and characterization of this soluble form of the protein have been described elsewhere [45].  

As previous experiments on yPDI have shown great variation in the orientation of the domains [18], the cross-linking experiment was carried out at three different temperatures, 0 °C, room temperature and 37 °C. Each experiment was divided into four fractions and performed in triplicate, thus a total of 45
Fig. 2 – Heat maps showing all observed cross-link scans (spectral observations) in PDI obtained by cross-linking with BS3 and observed with a CrossWork score of at least 10. Total cross-link observations for all three temperatures and repeat experiments combined. Coloring is relative to the total number of observed scans. Hot spots, where at least 10 scans were observed, are shown in dark color. The horizontal and vertical lines are drawn to separate the four domains. The outermost numbering is the lysine numbers according to the nucleotide sequence, the inner numbering is after removal of the signal peptide. Heat maps for individual temperatures and inter-protein interactions are presented in Supplementary Fig. 2.
Table 1 - Total number of intra and inter cross-links observed per domain for PDI with a CrossWork score ≥ 10. Bottom line shows number of amino groups in each domain (N-terminus and lysine residues).

<table>
<thead>
<tr>
<th>Domain</th>
<th>a</th>
<th>b</th>
<th>b'</th>
<th>a'</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1067</td>
<td>336</td>
<td>295</td>
<td>146</td>
</tr>
<tr>
<td>b</td>
<td>36</td>
<td>111</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>b'</td>
<td>325</td>
<td>207</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a'</td>
<td></td>
<td>33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NH₂</td>
<td>13</td>
<td>8</td>
<td>13</td>
<td>15</td>
</tr>
</tbody>
</table>

runs were performed and over 220,000 scans (ms/ms spectra) were recorded.

Running the native and cross-linked proteins on an SDS-PAGE gel revealed some dimerization of PDI and the disappearance of gel bands representing calnexin and ERp72, indicating that there was interaction between the proteins separated in the gel, either as homo- or as hetero-mers.

After chemical cross-linking and strong cation exchange chromatography, followed by mass spectrometry, we identified cross-linked peptides originating from all three proteins. Cross-linking of ERp72 and calnexin individually has been performed previously, and the intra-protein cross-links have been described elsewhere [18,46].

Most of the cross-links in these experiments were observed for PDI. Combining the 15 runs from each temperature triplicate experiment and plotting the observed number of scans (spectral observations) for each cross-link into a grid, we generated a heat-map for all observed cross-links (Fig. 2), along with heat-maps for each individual experiment (Supplementary Fig. 2A–O). Data for the observed interprotein cross-links (PDI/CNX, PDI/ERp72 and ERp72/CNX) are presented in Supplementary Fig. 2D–F. The heat maps were created in MassAI with varying cutoffs in CrossWork score [46] and by checking the related spectra manually. Higher CrossWork scores required for inclusion in the heat map, obviously increased the confidence of the observation, but increasing to more than 10 favored peptides with a particularly good fragmentation rather than increasing the confidence of identification.

3.2. Observed cross-links

Looking at the heat map of all observed cross-links, it is clear that most of the intra domain cross-link scans are present in the a domain (Fig. 2, Table 1). This was not caused by a larger number of amino groups (lysines + N-terminus) present, as the number of amino groups is similar to b' and less than a' (Table 1). The most likely reason could be greater detectability of the peptides, as cross-linked peptides by their nature are larger. Looking at the number of scans per peptide and sorting by mass, a correlation between number of observed scans and mass can be seen (Fig. 3A). Here is a clear trend for large peptides (>2000 Da) to be less represented, with only three peptides (m/z 2010, 2093 and 2639) being highly represented. Both peptides 2093 and 2639 are arginine-terminated, which could play a role due to the higher pKa of arginine. However, when looking at the observed peptides containing a dead-end BS3 linker (Fig. 3B), arginine-containing peptides are not preferentially observed, and the loss of high-mass peptides is only first observed above MH+ 2700. As tryptic cross-linked peptides are typically triply or quadruply charged, these cross-peptides should be well within the mass range of the instrument (m/z 1800), but are clearly disfavored. The three high-occurrence high-mass peptides thus make almost no cross-links with other high-mass peptides, only with small mass peptides and among themselves. Two low mass peptides are not observed, MH+ 303 and 674. The peptide MH+ 303 contains Lys131 and will be seen as a dipeptide (Lys-Arg), which is filtered out by the search program as a peptide with high probability of false identification. Peptide MH+ 674 contains Lys505 and is neither observed as cross-linked nor as a linear peptide (Supplementary Fig. 1). This can be due to the low mass of the peptide, but may also be caused by its absence. The last residue observed with high confidence is residue Lys468, however, a lower confidence peptide spans up to residue Lys502, and it is thus likely that the KDEL sequence or a large part of the acidic C-terminal part is missing. Calnexin purified in the same fraction is missing the C-terminal transmembrane region [45], and for calreticulin, obtained by the same purification scheme, it has been observed that a fraction was missing the C-terminal 6 residues [48]. It thus seems that a part of the ER proteins from the placenta is C-terminally truncated. The low number of observed cross-links relating to the b domain is most likely related to the fact that of the eight lysine residues present, only three peptides with one missed cleavage have a mass between 500 and 2000 Da (Supplementary Table 1).

Omitting the ten cross-links least observed (i.e. two small peptides, 8 large peptides, thus going from 48 amine groups to 38) the number of total observed scans decreases from 2581 to 2571 while the number of different observed cross-links drops from 340 to 333. As the number of potential cross-links drops from 1176 to 741 this means that 45% of all remaining potential cross-links are observed at least once, indicating that a number of short-lived interactions are likely, and only multiply observed cross-links with manually validated scans are thus considered in the following. Cross-links which are observed in at least 10 scans are in the following labeled ‘hot-spots’.

3.3. Intradomain and a–b inter-domain cross-links

Comparing the observed cross-links with the X-ray structure of intact hPDI (PDB 4EL1) [19] shows that all observed intra-domain hot spot cross-links fit within the allowed Co–Co distance of 25 Å for the BS3 linker (Fig. 4; representative...
Fig. 4 – Co intra-protein distance map for N-terminal and lysine residues for PDI structure 4EL1. Distances are color-coded as follows: green < 20 Å; yellow 20–30 Å; orange 30–40 Å; red > 40 Å. The distances for Lys254 are based on modeling of residues 250–254, as these residues are not resolved in structure 4EL1. Numbering is the same as for Fig. 2.
assigned ms/ms scans are shown in Supplementary Fig. 3). This is expected, as each domain has a diameter of 25–30 Å, and most amine groups should thus be reachable by cross-links created by BS3.

Comparing links between the a and b domains, cross-links between Lys200, Lys208 and Lys222 in the b domain and the N-terminus and lysine 71, 81, 103 and 114 are abundant, but also links to Lys31 and 130 are observed. Except for Lys103–Lys222 all Ca distances are more than 30 Å and most exceed 40 Å (Fig. 4A), showing that the domains have to move relative to each other to accommodate the cross-links. As Lys71 and Lys81 are located on the distant face of the a domain, this may even have to rotate in order for the linkage to occur.

3.4. b–b’ inter-domain cross-links

The cross-links between the b and b’ domains are mainly centered on Lys200 in the b domain which cross-links to Lys247, Lys276, and Lys308. The distances to Lys247 and Lys276 are within linking distance in the crystal model, while Lys308 is located on the opposite side of the b’ domain (along with the relatively high copy-linked residues Lys326 and Lys328) indicating inter-protein interaction as no rotation about linker regions can bring these residues into close contact. In the X-ray structure these residues are on the convex ‘outside’ of the dimer, and cannot be accommodated by our dimer models (Supplementary Fig. 5), but necessitate a different form of dimer interaction. Further three 7–9 scan observations with a distance of 30–32 Å indicate some flexibility in the b–b’ interface.

3.5. b’–a’ inter-domain cross-links

Although a number of cross-links can be found between the b’ and the a’ domains, most of these are of low copy number. However, Lys375, which is situated in the interface between the a’ and the b’ domains, links strongly to five lysine residues on the nearby surface of b’: 308, 309, 326, 328, and 352, all of which are within reach in the 4EL1 structure. Lys375 also links to Lys285, which is on the edge of the far side of the b’ domain, which may only be reached if the a’ domain rotates and/or bends around the X-linker region. Lys385 in the b’ domain also cross-links to a lesser degree with the same lysines in the a’–b’ interface, but as these Ca distances are >40 Å, it hints that the a’ domain may be able to rotate and/or shift relative to the plane of the other three domains.

3.6. Longer reaching inter-domain cross-links

10 hot spot cross-links are observed between the a and b’ domains (Fig. 2). Aside from the underrepresented long peptide lysine residues 247, 254, 263, 285, and 350, the linkages are quite evenly distributed with few strong hot spots. Lys114 and Asp18 (N-terminus) both link to Lys308 and Lys328 (distance 38–62 Å), and Lys328 in addition links to Lys31 (43 Å). These linkages are all impossible in the monomeric structure, and even comparing to the potential dimeric structures (distance maps in Supplementary Fig. 5) only a single linkage can be supported.

Cross-links between domains a and a’ domains are somewhat scattered and without clear hot spots, except for Lys467 that links to the N-terminus and to Lys103. In the monomer, these residues are at a distance of more than 50 Å. In the hYeastDimer model there are interacting areas between a and a’ domains (Supplementary Fig. 5), but the cross-links found still have distances larger than 30 Å.

3.7. Influence of temperature on cross-linking

The cross-linking of PDI was performed at three different temperatures, 0 °C, room temperature and 37 °C. When making a cutoff minimum score of 10, 525 scans were obtained in the 0 °C experiment, 1243 scans in the room temperature experiment and 813 scans in the 37 °C experiment. The differences in number of scans obtained are likely explained by the fact that the 22 °C experiment had a different composition than the 0 °C and 37 °C experiments, and for the two temperature experiments, there is a difference in the reactivity of the cross-linking reaction.

Comparing the extremes of temperatures, 0 °C and 37 °C, only a few differences can be seen (Supplementary Fig. 2A and 2C). Going from zero to 37 °C, most hot spots in the a domain change, but this is mainly the result of the number of observed cross-links in each position, as the pattern is the same. As these residues are well within interaction range in the PDI monomer, the increased number of cross-links at 37 °C is likely a result of increased chemical activity of BS3, but may also be due to an increased flexibility of the PDI molecule, as the distribution of hot spots has changed slightly. For intra-linkages in the other domains the number of linkages is quite small and the changes minimal.

Two residues, Lys81 and Lys208, show a much increased reactivity at increased temperature, both of which are stabilized in the crystal structure by salt bridges to glutamic acid residues (Glu23 and Glu151 respectively). Raising the temperature thus clearly diminishes this interaction, making the lysine residues available for chemical reactions and thus cross-linking.

3.8. Modeling of the dimeric structure

As described above, most cross-links with multiple observations can be assigned to intra-domain distances. However, of the 27 inter-domain hot spots observed, 17 cannot be explained by the monomeric structure. Although the crystal structure 4EL1 shows two PDI molecules, these most likely do not represent a solution dimer, and the distance map (Supplementary Fig. 5A) does not support this either.

Based on these conclusions, we modeled the PDI dimer using monomeric crystallized structure available in PDB using RosettaDock. We generated 20,000 rigid body structure predictions in a blind run of RosettaDock. For each of the 20,000 predicted structures, we checked the distances between cross-linked residues not satisfied in the monomer (Table 2) to find the PDI homodimer model structure that simultaneously satisfies the largest number of cross-links. In this docking we used a cross-linking distance of 30 Å in order to allow slightly more flexibility in the interacting molecules. Following this strategy we expected to identify the best
among the 20,000 models as a result of an extensive exploration of the conformational space.

Fig. 5 shows a heat map of the co-occurrence of all the 19 cross-links (10 cross-links A-B, 10 cross-links B-A, 1 of which is symmetrical). As can be seen, there are 2 large mirror (A-B and B-A) groups of cross-links, that tend to be satisfied together and few times or never with the other cross-links. In this big group we observe 6 cross-links which are satisfied at the same time (the maximum observed number) and are represented by 3 different solutions that are shown in Fig. 6. Table 3 describes each one of them with the list of the satisfied cross-links.

Table 2 - Cross-links used for modeling the PDI dimer. These are the cross-links that do not fit the monomer PDI model and are observed more than five times.

<table>
<thead>
<tr>
<th>Residue number</th>
<th>Cross-link copy number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1–183</td>
<td>6</td>
</tr>
<tr>
<td>1–191</td>
<td>13</td>
</tr>
<tr>
<td>14–183</td>
<td>9</td>
</tr>
<tr>
<td>14–311</td>
<td>8</td>
</tr>
<tr>
<td>14–450</td>
<td>6</td>
</tr>
<tr>
<td>54–54</td>
<td>9</td>
</tr>
<tr>
<td>54–311</td>
<td>7</td>
</tr>
<tr>
<td>86–450</td>
<td>13</td>
</tr>
<tr>
<td>191–291</td>
<td>8</td>
</tr>
<tr>
<td>268–358</td>
<td>7</td>
</tr>
</tbody>
</table>

3.9. Inter-protein cross-links

Cross-links were also observed between both PDI and calnexin, and between PDI and ERP72 as well as all the three major components in the mixture, PDI, ERP72 and calnexin. The observed cross-links are presented as heat-maps in Supplementary Fig. 2A to 2C and the data in Supplementary Tables 3, 4 and 5 respectively.

When cross-linking to calnexin both PDI and ERP72 show a very uneven distribution of crosslinks, with a very high number of observed cross-links to certain lysine residues and none to others. Plotting these lysine residues onto the structure of dog calnexin, which is conserved relative to the human calnexin (e.g. all lysine residues are conserved), a pattern can be seen where all the high copy number linkage positions are located on the same surface close to the peptide binding area as proposed by Chouquet et al. [49] (Fig. 7).

For PDI, the cross-linking to calnexin is clearly dominated by interactions in the α domain (supplementary Fig. 2A), and presenting the high-copy number linkage positions on the structure (Fig. 8) clearly shows that the main interactions are to the α domain and the linking regions of b–b′ and b′–α (the x-linker region). For the interaction of PDI and ERP72, the linkage pattern is much more widely distributed, but the highest copy-number linkage positions are either identical to PDI-calnexin or are neighboring residues.
For the linking to ERp72, the two most linking residues (Lys573 and Lys571) are located in the α′ domain and are the same for interaction with both calnexin and PDI (although most pronounced for calnexin). This is the same for the third most reacting residue, Lys 401 in the b–b′ domain, while the next most linking residues are scattered through the α and α′ domains but with no overlap. The main interaction thus seems to be the α′ domain for the interaction with either protein but with scattered interactions throughout the protein. These interactions are difficult to map onto a structure, as only theoretical models of ERp72 exist [46,50]. Mapping onto the top model reported [46] showed no clear pattern, with several of the lysines being partially buried in the structure leaving the results inconclusive (results not shown).

4. Discussion

The combined use of high-resolution methods for protein structure elucidation (e.g. X-ray crystallography, NMR spectroscopy), low-resolution methods (SAXS, CXMS, electron microscopy, and others) and molecular biological methods is needed to gain a detailed understanding of protein structure and function. Each method has its advantages and limitations.
and they complement each other in several ways. An excellent example of this is PDI. This enzyme plays an important role in the proper folding and quality control of ER proteins and collaborates with many other ER chaperones and folding catalysts.

Due to the importance of PDI, it has been suggested to be involved in the development of several diseases like AIDS [51,52], ovarian cancer [53], thrombosis [54–58], and neurodegenerative diseases [59–62] mainly due to its important chaperone functions.

The combined use of all the methods mentioned above has shown PDI to contain a large hydrophobic pocket suggesting domain b’ as the major substrate binding site [38,63–65].

Tian G et al. [18] revealed that at 4 °C the four thioredoxin-like domains of yPDI are arranged in a twisted ‘U’ shape with domains a and a’ facing each other and the inside surface of the ‘U’ shape containing a high abundance of hydrophobic residues, thus providing binding sites for substrate proteins, while at 22 °C the domain a rotates 123° around the junction between domains a and b, and hence the four domains form a ‘boat’ shape. It was demonstrated that the distance between domains a and b is even more flexible than the one between domains b’ and a’, which was critical for the enzymatic activity both in vivo and in vitro. Wang C et al. [19] revealed that the four domains of hPDI were arranged as a horseshoe shape, with two CGHC active sites, located in domains a and a’ respectively, facing each other, but hPDI exists in different conformations under oxidized and reduced states. The oxidized human PDI exists in an open state with many other ER chaperones and folding catalysts.

Here we have analyzed hPDI co-purified with calnexin and ERp72 by chemical cross-linking and mass spectrometry. We have used the lysine-reacting compound bis [sulfosuccinimidyl] suberate (BS3) which cross-links amino groups (i.e. N-terminus and lysine residues). A feature of cross-linking with BS3 is that the reaction is relatively slow with reaction times of more than 1/2 h (in this experiment BS3 was allowed to react for one hour). If the protein has a very stable structure, we will observe specific interactions, i.e. only relatively few cross-links. However, if the structure is flexible, intra-domain cross-links are expected to dominate, as the linking points have a relatively stable position relative to each other, while inter-domain cross-links are less frequent if the domains are flexible and have active movements relative to each other while the reaction takes place. Another bias that has to be considered, is that high-mass peptides are disfavored in cross-linked peptides (Fig. 3A), which however cannot be observed for the peptides derivatized by a hydrolysed BS3 molecule (dead-end linkers).

The present results have been compared to the full-length structure of oxidized hPDI structure, PDB code 4EL1 [7,19]. This structure is based on recombinant PDI expressed without the signal peptide, but with an N-terminal purification tag. However, as the PDI used in the current analysis was purified from human placenta, there may be slight structural differences. We also chose to compare to the oxidized form, as this has been analyzed in a tris–HCl, pH 8.0 buffer, while the reduced PDI (PDB 4EKZ) has been analyzed in the presence of 50 mM DTT. When compared to the reduced form (distance map in Supplementary Fig. 4), only minor differences can be seen, and no additional cross-link mapping can be assigned.

Comparing the combined cross-links (Fig. 2) to the lysine Cα distances in PDI (Fig. 4A) it can be observed that all observed intra-domain hot-spot linkages fit with the 4EL1 structure. However, when comparing interactions between domains, only 10 out of 27 cross-links are still supported by the monomer model. The non-explained cross-links are most

<table>
<thead>
<tr>
<th>Model number</th>
<th>54_515</th>
<th>1_644</th>
<th>1_652</th>
<th>14_644</th>
<th>14_772</th>
<th>14_911</th>
<th>54_772</th>
<th>86_911</th>
<th>191_752</th>
<th>268_819</th>
</tr>
</thead>
<tbody>
<tr>
<td>4EL1_01827</td>
<td>46</td>
<td>9</td>
<td>28</td>
<td>28</td>
<td>47</td>
<td>78</td>
<td>59</td>
<td>89</td>
<td>84</td>
<td>75</td>
</tr>
<tr>
<td>4EL1_08502</td>
<td>50</td>
<td>26</td>
<td>49</td>
<td>15</td>
<td>56</td>
<td>91</td>
<td>77</td>
<td>89</td>
<td>71</td>
<td>49</td>
</tr>
<tr>
<td>4EL1_09304</td>
<td>23</td>
<td>16</td>
<td>30</td>
<td>27</td>
<td>74</td>
<td>107</td>
<td>62</td>
<td>102</td>
<td>81</td>
<td>93</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model number</th>
<th>462_183</th>
<th>462_191</th>
<th>475_183</th>
<th>475_311</th>
<th>475_450</th>
<th>457_311</th>
<th>547_450</th>
<th>572_291</th>
<th>729_358</th>
</tr>
</thead>
<tbody>
<tr>
<td>4EL1_01827</td>
<td>10</td>
<td>28</td>
<td>28</td>
<td>47</td>
<td>78</td>
<td>59</td>
<td>89</td>
<td>84</td>
<td>75</td>
</tr>
<tr>
<td>4EL1_08502</td>
<td>9</td>
<td>23</td>
<td>27</td>
<td>28</td>
<td>68</td>
<td>49</td>
<td>69</td>
<td>62</td>
<td>32</td>
</tr>
<tr>
<td>4EL1_09304</td>
<td>29</td>
<td>25</td>
<td>24</td>
<td>65</td>
<td>100</td>
<td>64</td>
<td>103</td>
<td>88</td>
<td>95</td>
</tr>
</tbody>
</table>
likely explained by the occurrence of a dimer, which is supported by the SDS PAGE gel (Fig. 1). In the X-ray crystal of hPDI (4EL1), two PDI molecules in the asymmetric units are oriented in a concave fashion. A dimerized yPDI was also suggested by Tian and co-workers [21] which also showed a dimer with the a domain nestled in the concave cavity of the opposite PDI. We subsequently modeled the hPDI sequence into the yeast dimer and calculated lysine–lysine distances. However, as neither model could satisfy our cross-linking results, it led us to model alternative PDI dimer models using RosettaDock. We identified 10 cross-links that could not be accommodated by the monomer structure and were determined to be reliable and stable (Table 2). We generated 20,000 models and sorted them according to their fit to the cross-links (Fig. 5). This resulted in three modalities (Fig. 6) showing that instead of the dimer models, where the a domain of one PDI molecule is nestled inside the U form of another PDI molecule (Fig. 6A), we have interaction between the outside of the a–b domains. Either domain a against b in an antiparallel fashion (Fig. 6B and D), or slightly cross-wise centered on the a–b linker region (Fig. 6C). None of the dimer models satisfied all of the ‘missing’ cross-links (Table 3), indicating that we have a number of different interactions. Furthermore, we are likely to have alternative conformations of PDI, as the present modeling was performed based on a rigid-body docking that did not take intra-chain flexibility into consideration. As a large number of the observed cross-links indicate that we have considerable flexibility, it is likely that this could satisfy the missing cross-links.

As the crystal structure of the yPDI showed distinct differences in structure when crystallized at 0 °C and at room temperature, we performed the cross-linking experiment at 0 and 37 °C. Based on the heat maps (Supplementary Fig. 2A and 2C) it can be concluded, that there are no great differences in the solution structures at either temperature. The main difference observed is a markedly lower cross-linking ability of residues Lys81 and Lys208 at 0 °C, which is likely caused by a stronger stabilization by salt bridges to glutamic acid residues at low temperature.

A large number of inter-protein cross-links were also observed between the major components of the fractions (PDI, calnexin and ERp72). For all interactions involving calnexin, the cross-linking was clear to the surface of the molecule containing the peptide and carbohydrate binding area [49,67] (Fig. 7), thus confirming this part as the main peptide interaction region. For PDI the interactions with ERp72 and calnexin were mostly similar, with the a domain being by far the strongest cross-linker (Fig. 8). For ERp72 the situation was different, as the main interaction to PDI shared only part

Fig. 7 – Inter-cross-linking sites on calnexin model 1JHN. The general structure is shown as backbone, lysine residues are displayed in space filling mode. Low scan number cross-linking lysines are shown in blue. High scan number lysine residues unique for cross-links to PDI are shown in green, unique for ERp72 in cyan, common for both PDI and ERp72 are shown in orange.
of the interaction with calreticulin. Although most of the participating lysines were cross-linking for both proteins, the distribution of the intensities was quite different with the interaction to calreticulin being more localized, in agreement with calreticulin being a smaller molecule with a single globular domain. The binding residues were the same, particularly in the a domain, but the distribution of linked residues was slightly different as observed for calreticulin, probably reflecting the size of the binding proteins. Although there seems to be no reports on a biological interaction of PDI, ERp72 and calnexin, the fact that they co-purify through a lengthy purification process, indicates that the interaction is both specific and stable. In our purification of placental chaperones, we have observed other protein mixtures, which may represent similar complexes. Another question is whether it is possible to analyze the binding PDI to substrate proteins by CXMS. In this case the interaction is expected to be rather short-lived, and the use of a fast reacting cross-linker is thus necessary. The most likely choice would be photoreactive cross-linkers, however, they are quite unspecific, which will lead to a lower yield of the individual cross-link and a much larger search space for the search program. Combined with the large size of PDI, this seems to be out of reach by the current technology.

5. Conclusion

The current results show that it is possible to generate a wealth of data using standard cross-linking techniques when combined with sensitive mass spectrometric analysis and detection software. For the analyzed hPDI protein an X-ray structure of the intact protein exists, and the current experiments are for a large part consistent with the model. PDI is a very flexible molecule as the a and a' domains can move through large angles relative to the b and b' domains. X-ray experiments have shown a large variation in the orientation of the a and a' domains relative to the b and b' domains, both regarding temperature (for yPDI) and oxidation state (for hPDI). Based on the current results, we hypothesize that the different conformations observed in X-ray structures may be ‘frozen’ conformations that may be entered under most/all conditions, and which happen to be stabilized under the specific crystallization conditions used. Furthermore, the structure of the yeast dimer is unlikely to be representative of the hPDI dimer. Although the substrate interaction generally has been located to the b' domain, the present results indicate that the in-solution homo-dimer does not bind in this area, but in the domain a-b interface. Our data thus suggest a convex dimer interaction rather than a concave.
Interaction was observed between all the major components in the mixture. While the interaction to calnexin was clearly localized, the interaction to PDI was more diffuse over the a' and part of the b' domain. Interaction to Erp72 was difficult to determine, as a complete structural model is not available yet.

In conclusion, cross-linking in combination with mass spectrometry is a valuable supplement to other structural methods and can be used to obtain information about protein structure, flexibility and interaction, as exemplified here for PDI, calnexin and Erp72.

**References**


