Structure of the substrate-binding b’ domain of the Protein disulfide isomerase-like protein of the testis

Sara Bastos-Aristizabal, Guennadi Kozlov & Kalle Gehring

Department of Biochemistry, Groupe de recherche axé sur la structure des protéines, McGill University, 3649 Promenade Sir William Osler, Montréal, Québec, Canada H3G 0B1.

Protein Disulfide Isomerase-Like protein of the Testis (PDILT) is a testis-specific member of the PDI family. PDILT displays similar domain architecture to PDIA1, the founding member of this protein family, but lacks catalytic cysteines needed for oxidoreduction reactions. This suggests special importance of chaperone activity of PDILT, but how it recognizes misfolded protein substrates is unknown. Here, we report the high-resolution crystal structure of the b’ domain of human PDILT. The structure reveals a conserved hydrophobic pocket, which is likely a principal substrate-binding site in PDILT. In the crystal, this pocket is occupied by side chains of tyrosine and tryptophan residues from another PDILT molecule, suggesting a preference for binding exposed aromatic residues in protein substrates. The lack of interaction of the b’ domain with the P-domains of calreticulin-3 and calmegin hints at a novel way of interaction between testis-specific lectin chaperones and PDILT. Further studies of this recently discovered PDI member would help to understand the important role that PDILT plays in the differentiation and maturation of spermatozoids.
in order to be delivered properly onto the sperm surface\textsuperscript{9–11}. This suggests that PDILT plays a role in the folding pathway of certain and specific spermatogenesis-proteins and enzymes required for the proper functioning of the sperm. PDILT possesses the $\beta^9$ domain, which is associated with binding of substrates and protein partners in the PDI family members\textsuperscript{12}, but the molecular details of PDILT interactions with protein substrates remain elusive.

Here, we report the high-resolution crystal structure of the non-catalytic $\beta^9$ domain of human PDILT determined at 2.0 Å. The structure of the PDILT $\beta^9$ domain displays a thioredoxin-like fold comprising a five-stranded $\beta$-sheet with the $\beta^1$-$\beta^5$ topology, flanked by four helices (Fig. 1b). All strands are parallel, except strand $\beta^4$, which is antiparallel. There are four molecules (chains A–D) in the asymmetric unit, which are very similar to each other, as all chains are superimposed using backbone C$_{\alpha}$ atoms with an RMSD oscillating from 0.1 to 0.8 Å. However, there is a feature that divides them into two groups according to the position of the x-linker (residues 369–386 of the construct). Chains A and B have the x-linker pointing away from the thioredoxin domain, while the x-linker folds towards the $\beta$-sheet in chains C and D (Fig. 1b). The x-linker residues Q375–P381 were not observed in the electron density map of chains C and D likely due to its intrinsic mobility.

The conserved hydrophobic pocket in PDILT is primed for binding hydrophobic residues. Analysis of the structure reveals a hydrophobic pocket that is conserved in a number of PDI family members, but the molecular details of PDILT interactions with protein substrates remain elusive.

**Results**

**Structure of the $\beta^9$ domain.** In order to better understand how PDILT interacts with its substrates, we obtained crystals of the $\beta^9$ domain of human PDILT (residues 258–386) that diffraction to 2.0 Å using synchrotron radiation. The three-dimensional structure was solved by molecular replacement using the $\beta^9$ domain of human PDIA1 (PDB accession number 3BJ5), which shares 32% identity with PDILT $\beta^9$ domain sequence, as a search model. A summary of data collection and refinement statistics is given in Table 1.

The structure of the PDILT $\beta^9$ domain displays a thioredoxin-like fold comprising a five-stranded $\beta$-sheet with the $\beta^1$-$\beta^5$ topology, flanked by four helices (Fig. 1b). All strands are parallel, except strand $\beta^4$, which is antiparallel. There are four molecules (chains A–D) in the asymmetric unit, which are very similar to each other, as all chains are superimposed using backbone C$_{\alpha}$ atoms with an RMSD oscillating from 0.1 to 0.8 Å. However, there is a feature that divides them into two groups according to the position of the x-linker (residues 369–386 of the construct). Chains A and B have the x-linker pointing away from the thioredoxin domain, while the x-linker folds towards the $\beta$-sheet in chains C and D (Fig. 1b). The x-linker residues Q375–P381 were not observed in the electron density map of chains C and D likely due to its intrinsic mobility.
A hydrogen bond is also observed between the backbone carbonyl group of W384 and the side chain of K266 in the α1-helix of the b′ domain. Compared with W384, Y383 goes deeper into the hydrophobic pocket and likely plays a more important role in the binding.

The structural alignment with the b′ domain of PDIA1 using Dali database gave a Z-score of 17.6 and an RMSD of 1.4 Å over 113 residues, showing that the structures are very similar (Fig. 2c). The major difference is in the orientation of the helix α1, which creates a larger and deeper hydrophobic pocket in PDILT. Also, the overall charge surrounding the hydrophobic pocket is negative for PDIA1, while this area is very positively charged in PDILT. Additionally, K266, S270, L339 and L341 in the substrate-binding pocket of PDILT are not conserved between PDI family members (Fig. 1c). These differences may be the key determinants for PDILT specificity towards in vivo substrates. The size of the hydrophobic pocket in the PDILT structure is more similar to that in the C-terminal domain of ERp27 in the polyethylene glycol-bound conformation with a Z-score of 18.6 and RMSD of 1.7 Å over 111 residues (Fig. 2d). Overall, the structure-based sequence alignment shows a high conservation of residues forming the hydrophobic pocket in PDILT, PDIA1, PDIp and ERp27, explaining why these PDIs are able to directly interact with unfolded proteins (Fig. 1c). Interestingly, H277 of PDILT is positioned at a base of the hydrophobic pocket and is strictly conserved for these PDIs. Our structure demonstrates that the side chain of this histidine forms a hydrogen bond with the hydroxyl group of interacting tyrosine residue and could be employed similarly by other PDIs. The corresponding histidine H256 of PDIA1 was predicted to hydrogen bond with 3-hydroxyl group of 17β-estradiol.

Small amphipathic peptides Δ-somatostatin (AGCKNFVKFTFTSC) and mastoparan (INLKALAAALAKKIL) have been used in several studies of the PDI family proteins in order to mimic unfolded substrates. Interactions of these peptides with PDIA1 b′ domains...
have been characterized previously\textsuperscript{4,16,17}. Interestingly, $\Delta$-somatostatin binds to the hydrophobic pocket of PDIA1 with a higher affinity compared to mastoparan ($K_d$ of 35 $\mu$M and 130 $\mu$M, respectively)\textsuperscript{18}. Since PDIA1 and PDILT display similar substrate-binding pockets, NMR titrations of $^{15}$N-labeled PDILT $b'$ domain with mastoparan are expected to show significant shifts of the NMR signals from PDILT residues of the hydrophobic pocket, similar to those observed for PDIA1 $bb'$ domains. Surprisingly, mastoparan did not interact with the PDILT $b'$ domain, since NMR spectra remained unchanged after the addition of up to 1:10 protein/peptide molar ratio (Supplemental Fig. S2). Interestingly, the binding of somatostatin to PDILT was previously reported\textsuperscript{9} and could be due to the presence of aromatic residues that are absent in mastoparan.

**ERp57-CNIX recognition mechanism is not conserved in PDILT.** Calnexin (CNX) and calreticulin (CRT), which are CMG and CRT3 homologs, use their negatively charged arm-like P-domains to bind to a large positively charged patch in the $b'$ domain of ERp57, another member of the PDI family\textsuperscript{18–20}. In order to determine whether PDILT interacts with CMG and CRT3 in a similar manner, we performed NMR titrations of the $^{15}$N-labeled P-domains of CMG or CRT3 with unlabeled $b'$ domain of PDILT. Addition of the $b'$ domain did not result in any chemical shift changes, demonstrating no interaction (Supplemental Fig. S3). To explain this result, we compared the binding site for CNX and CRT in ERp57 $b'$ domain with the PDILT structure. Both domains could be superimposed using their backbone atoms of 111 residues (out of 129 residues in PDILT $b'$ domain) with an RMSD of 2.6 Å and a Z score of 13.4 using Dali database (Fig. 3). In ERp57, the CNX-binding surface is centered on helix $\alpha_2$ and is characterized by a pronounced positive charge (Fig. 3a). Interestingly, PDILT does not display a positively charged patch in the corresponding surface. This is consistent with the fact that PDILT and ERp57 $b'$-domains share only 11% identity, and the amino acids implicated in the ERp57/CNX interaction are not conserved in PDILT (Fig. 1c).

**Discussion**

Despite many years of studies, molecular details of substrate recognition by protein disulfide isomerasers remain elusive. In particular, structural characterization of substrate-binding sites provides important information about this process. Here, we determined a high-resolution structure of the substrate-binding domain of PDILT, which is also the first structure of any PDILT domain. Importantly, the structure reveals a large hydrophobic pocket that seems to be primed for binding aromatic residues. In PDIA1, this hydrophobic pocket is a binding site for protein substrates and stabilizes hydrophobic residues exposed in unfolded regions of protein substrates\textsuperscript{21–23}. The higher affinity of the $b'$ domain for aromatic tryptophan and tyrosine residues was previously observed for the other tissue-specific PDI, PDIp. Using a seven amino acid peptide phage-display library, Ruddock and co-workers showed that tyrosine or tryptophan residues within a folding polypeptide trigger the recognition and binding to PDIp\textsuperscript{24}. It is possible that PDILT uses similar preference for aromatic residues reflecting the selectivity for spermatogenesis-specific proteins. However, further studies are required to determine how PDILT recognizes its substrates in vivo.

PDILT interacts in vivo with testsis-specific lectin-type chaperones, calmegin (CMG) and calreticulin 3 (CRT3)\textsuperscript{3,15}. Our results suggest that the mechanism of binding of PDILT to CMG and CRT3 is different from the one that occurs in ERp57/CNX, and other domains are necessary to mediate binding. For instance, the C-terminal tail of ERp57 is an additional site of interaction with CNX\textsuperscript{9}. While the C-terminus alone does not interact with CNX, the basic stretch (residues $^{258}$KPKKKKK$^{262}$ of ERp57) is needed for high-affinity binding. Interestingly, the C-terminal tail of PDILT also contains a positively charged region (residues KKKKSEEEVVVAKKPK-GPPVQK); therefore, it is possible that the C-terminus of PDILT in combination with other domains is required for binding to CMG and CRT3. More studies are needed to determine the mechanism underlying these interactions. Noteworthy, ERp57 is not able to bind to its substrates directly and requires CRT or CNX for substrate recognition\textsuperscript{25}, while the PDILT $b'$ domain is more PDIA1-like and should be able to directly bind the non-native substrates through the hydrophobic pocket.

The specificity of substrate binding by protein disulfide isomerasers is still poorly understood. PDIA1 acts on very broad range of substrates\textsuperscript{26,27}, while PDILT is expected to interact with spermatogenesis-specific protein substrates unique to this cell type. This structure is another step towards understanding of how substrate binding is mediated by PDI family members and PDILT in particular. The study of ER protein folding in the testis will lead to better understanding of pathways associated with male infertility.

**Methods**

**Protein expression and purification.** The PDILT $b'$ domain (V258-Q386) was amplified by PCR and sub-cloned into pGEX-6P-1 (Amersham Pharmacia) using BamHI and XhoI restriction sites. Two modules of the CMG P-domain comprising the repeats 3 and 4 (residues 300–369) were cloned in pET15b (Novagen) containing BamHI and XhoI restriction sites. Two modules of the CMG P-domain comprising the repeats 3 and 4 (residues 300–369) were cloned in pET15b (Novagen) containing BamHI and XhoI restriction sites. Two modules of the CMG P-domain comprising the repeats 3 and 4 (residues 198–291, 3-modules) were also cloned into pET15b using NdeI and BamHI restriction sites. Clones were confirmed by sequencing.

Protein was expressed in *E. coli* BL21 (DE3) cells. Expression of PDILT $b'$ domain with a N-terminal GST-tag were accomplished by growing cells in LB medium at 37 °C and shaking until the optical density (OD) at 600 nm reached 0.8; then protein production was induced by adding isopropyl-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. Cells were incubated with shaking at 30 °C for another 4 hours. The GST fusion protein was purified by affinity chromatography on glutathione-Sepharose resin (GE Healthcare). Once eluted from the resin, the tag was removed by thrombin.

**Protein expression and purification.** The PDILT $b'$ domain (V258-Q386) was amplified by PCR and sub-cloned into pGEX-6P-1 (Amersham Pharmacia) using BamHI and XhoI restriction sites. Two modules of the CMG P-domain comprising the repeats 3 and 4 (residues 300–369) were cloned in pET15b (Novagen) containing BamHI and XhoI restriction sites. Two modules of the CMG P-domain comprising the repeats 3 and 4 (residues 300–369) were cloned in pET15b (Novagen) containing BamHI and XhoI restriction sites. Two modules of the CMG P-domain comprising the repeats 3 and 4 (residues 198–291, 3-modules) were also cloned into pET15b using NdeI and BamHI restriction sites. Clones were confirmed by sequencing.

Protein was expressed in *E. coli* BL21 (DE3) cells. Expression of PDILT $b'$ domain with a N-terminal GST-tag were accomplished by growing cells in LB medium at 37 °C and shaking until the optical density (OD) at 600 nm reached 0.8; then protein production was induced by adding isopropyl-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. Cells were incubated with shaking at 30 °C for another 4 hours. The GST fusion protein was purified by affinity chromatography on glutathione-Sepharose resin (GE Healthcare). Once eluted from the resin, the tag was removed by thrombin.
removed by cleavage with PreScission protease leaving a N-terminal GPLGS sequence. The resulting PDILT b’ domain protein with a molecular weight of 15.5 kDa was additionally purified using size-exclusion (Superdex-75) chromatography using 20 mM HEPES pH 7.5, 0.15 M NaCl buffer, and then concentrated using a Centricon tube (Millipore) with the 3-kDa cut-off. His-tagged proteins were purified by affinity chromatography on Ni-NTA (nickel-nitrilotriacetic acid) resin (QIAGEN). Eluted proteins were then cleaved with thrombin protease to remove the His-tag, leaving a N-terminal GSHM sequence. Proteins were further purified as described above.

For NMR experiments, the 15N-labeled proteins were expressed by growing in M9 minimal medium with 15N-ammonium chloride as the unique source of nitrogen. Expression and purification conditions were the same as for unlabelled proteins.

### Protein crystallization

Crystallization conditions for PDILT b’ domain were identified using commercial PACT suite (QIAGEN). The best PDILT b’ domain crystals were obtained by equilibrating a 0.6 µl drop of the protein solution (7.2 mg/ml) in 20 mM HEPES pH 7.5, 150 mM NaCl, mixed with 0.6 µl of reservoir solution containing 0.2 M sodium formate, 0.1 M Bicine pH 9.0, 18% PEG3350, 0.1 M cobalt chloride and suspended over 1 ml of reservoir solution. Crystals grew in 4 days at 22 °C. For data collection, crystals were cryoprotected by addition of 20% (v/v) of ethylene glycol and flash cooled in a N2 cold stream. Crystals of PDILT b’ domain belong to the primitive monoclinic system, space group P21, with four protein molecules per asymmetric unit corresponding to a solvent content of 36.9%.

### Structure solution and refinement

Diffraction data from native PDILT b’ domain crystals were collected using a single wavelength of 0.977 Å regime with an ADSC Quantum-210 CCD detector (Area Detector Systems Corp.) on beamline A1 at the Cornell High-Energy Synchrotron Source (CHESS). Data processing and scaling were obtained with HKL-2000. The structure of human PDIA1 b’ domain (residues 230–368, PDB accession code 3BJ5) was used as the search model for protein crystals. The protein crystals grew in 4 days at 22 °C. For data collection, crystals were cryoprotected by addition of 20% (v/v) of ethylene glycol and flash cooled in a N2 cold stream. Crystals of PDILT b’ domain belong to the primitive monoclinic system, space group P21, with four protein molecules per asymmetric unit corresponding to a solvent content of 36.9%.

### NMR titrations

All NMR experiments were performed at 25 °C on a Bruker DRX 600 MHz spectrometer. NMR samples were prepared with 90% of NMR buffer (20 mM HEPES pH 7.5 and 150 mM NaCl) and 10% of D2O. NMR titrations were carried out by acquiring H-15N heteronuclear single quantum correlation (HSQC) spectra of 0.2 mM of 15N-labeled protein. Subsequent spectra were taken 600 MHz spectrometer. NMR samples were prepared with 90% of NMR buffer (20 mM HEPES pH 7.5 and 150 mM NaCl) and 10% of D2O. NMR titrations were carried out by acquiring H-15N heteronuclear single quantum correlation (HSQC) spectra of 0.2 mM of 15N-labeled protein. Subsequent spectra were taken of 1H-15N heteronuclear single quantum correlation (HSQC) spectra of 0.2 mM of 15N-labeled protein. Subsequent spectra were taken manually with the program Coot 30. The resulting model was improved by several cycles of refinement, using the program REFMAC 31 and model refitting, followed by manual refinement with the program Coot 30. The resulting model was improved by several cycles of refinement, using the program REFMAC 31 and model refitting, followed by manual refinement with the program Coot 30.

### Additional information

**Acknowledgments**

Data acquisition at the Macromolecular Diffraction (MacCHESS) facility of the Cornell High Energy Synchrotron Source (CHESS) was supported by the National Science Foundation award DMR 0225180 and the National Institutes of Health award RR-01646. This study was funded by an NSERC Discovery grant RGPIN-338873 to K.G.

**Author contributions**

S.B.A. and G.K. designed the experiments; S.B.A. performed the experiments; S.B.A. and G.K. wrote the main manuscript; G.K. and S.B.A. designed the experiments; S.B.A. performed the experiments; S.B.A. and G.K. wrote the main manuscript; G.K. and S.B.A. analyzed the data; S.B.A., G.K. and K.G. wrote the main manuscript; G.K. and S.B.A. prepared the figures. All authors reviewed the manuscript.

**Additional information**

**Data**

Coordinates have been deposited into the Protein Data Bank database under the accession number 4NWY.

**Supplementary information**

accompanies this paper at http://www.nature.com/scireports

**Competing financial interests**

The authors declare no competing financial interests.
