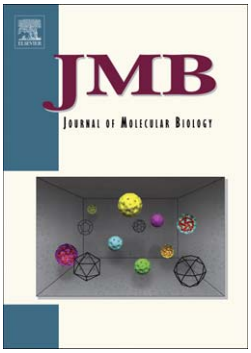



# Alternative conformations of the x region of human protein disulphide-isomerase modulate exposure of the substrate-binding b' domain

Van Dat Nguyen, Katrine Wallis, Mark J. Howard, Antti M. Haapalainen, Kirsi E.H. Salo, Mirva J. Saaranen, Ateesh Sidhu, Rik K. Wierenga, Robert B. Freedman, Lloyd W. Ruddock, Richard A. Williamson

PII: S0022-2836(08)01103-0
DOI: doi: 10.1016/j.jmb.2008.08.085
Reference: YJMBI 60786

To appear in: *Journal of Molecular Biology*

Received date: 23 May 2008
Revised date: 22 August 2008
Accepted date: 26 August 2008

Please cite this article as: Nguyen, V.D., Wallis, K., Howard, M.J., Haapalainen, A.M., Salo, K.E.H., Saaranen, M.J., Sidhu, A., Wierenga, R.K., Freedman, R.B., Ruddock, L.W. & Williamson, R.A., Alternative conformations of the x region of human protein disulphide-isomerase modulate exposure of the substrate-binding b' domain, *Journal of Molecular Biology* (2008), doi: 10.1016/j.jmb.2008.08.085

# Alternative conformations of the **x** region of human protein disulphide-isomerase modulate exposure of the substrate-binding **b'** domain.

**Van Dat Nguyen[1]§, Katrine Wallis[2]§, Mark J. Howard[3], Antti M. Haapalainen[1], Kirsi E.H. Salo[1], Mirva J. Saaranen[1], Ateesh Sidhu[2], Rik K. Wierenga[1], Robert B. Freedman[2], Lloyd W. Ruddock[1]* and Richard A. Williamson[3]**

[1] *Department of Biochemistry, University of Oulu, FIN-90014 Oulu, Finland.*

[2] *Department of Biological Sciences, Warwick University, Coventry CV4 7AL, UK*

[3] *Department of Biosciences, University of Kent, Canterbury CT2 7NJ, UK.*

§ *These authors contributed equally to this work.*

* *Corresponding author*: E-mail address: Lloyd.Ruddock@oulu.fi

*Running title*: Alternative conformations of the **x** region of PDI

*Abbreviations used*: NMR, nuclear magnetic resonance; PDI, protein disulphide isomerase; Pdi1p, yeast PDI (product of the S.cerevisiae *Pdi1* gene); trx, thioredoxin

Protein disulphide-isomerase (PDI) is a key multi-domain protein folding catalyst in the endoplasmic reticulum. The **b'** domain of PDI is essential for the non-covalent binding of incompletely folded protein substrates. We have previously defined the substrate binding site in the **b'** domain of human PDI by modelling and mutagenesis studies. Here we show by fluorescence and NMR that recombinant human PDI **b'x** (comprising the **b'** domain and the subsequent **x** linker region) can assume at least two different conformations in solution. We have screened mutants in the **b'x** region to identify mutations which favour one of these conformers in recombinant **b'x** and have isolated and characterised examples of both types. We have crystallised one mutant of **b'x** (I272A mutation) in which one conformer is stabilized, and determined its crystal structure to a resolution of 2.2Å. This structure shows that the **b'** domain has the typical thioredoxin fold and that the **x** region can interact with the **b'** domain by 'capping' a hydrophobic site on the

**b'** domain. This site is most likely the substrate binding site and hence such 'capping' will inhibit substrate binding. All of the mutations we previously reported to inhibit substrate binding shift the equilibrium towards the 'capped' conformer. Hence these mutations act by altering the natural equilibrium and decreasing the accessibility of the substrate binding site. Furthermore, we have confirmed that the corresponding structural transition occurs in the wild-type full-length PDI. A cross-comparison of our data with that for other PDI-family members, Pdi1p and ERp44, suggests that the **x** region of PDI can adopt alternative conformations during the functional cycle of PDI action and that these are linked to the ability of PDI to interact with folding substrates.

*Keywords*: PDI, crystal structure, **b'** domain, alternative conformations, NMR

## Introduction

The protein disulphide-isomerases (PDIs) are a family of proteins initially defined by their ability to assist the formation of native disulphide bonds in secretory and cell-surface proteins during their biosynthesis, but now characterised structurally by the presence of one or more thioredoxin-fold (trx*)* domains, and by their location within the lumen of the endoplasmic reticulum. The human PDI family is large (> 17 members) and the structures and specific roles of the multiple family members have not yet been defined[1]; the best-characterised is PDI itself, the archetype protein folding catalyst. PDI from a number of mammalian species has been studied for many years[2] and its properties as a catalyst of disulphide-interchange and associated protein folding are well-understood[3-7].

When mammalian PDI was first cloned and sequenced, it was immediately recognised that it comprised multiple homologous domains[8] and subsequently the overall domain organization was defined as **a-b-b'-x-a'-c**, where **a** and **a'** are trx domains containing the active site dithiol sequence WCGHC, **b** and **b'** have the trx conformation but lack the active-site motif, **x** is a linker region and **c** is a C-terminal acidic tail[9-12]. This domain organization was finally confirmed by the x-ray structural analysis of full-length yeast Pdi1p[13-14]. While the **a** and **a'** domains of human PDI can catalyse simple thiol:disulphide interchange reactions, the presence of the **b'** domain is essential for catalysis of protein disulphide isomerizations and for

non-covalent ligand binding[15-16]; the domains operate synergistically in linking catalytic and chaperone activities in protein folding[10].

The **b'** domains of human PDI family members show greater sequence diversity than do the **b**, **a** and **a'** domains, and studies with hybrid and chimaeric species indicate that the specificities of PDIs for ligands and partner proteins are primarily defined by their **b'** domains; this has been confirmed in studies on the pancreas-specific isoform PDIp[17], on the PDI homologue ERp57, which interacts specifically with the lectins calnexin and calreticulin[18-21], and on the newly-described protein ERp27[22] which comprises only **b** and **b'**-like domains. These results make it of particular interest to study -- in molecular detail -- the **b'** domains of PDIs and their interactions with ligands. Frustratingly, the **b'** domain of human PDI is the only domain for which the structure has not been solved and for several members of the family (e.g. PDIp, ERp27), the **b'** domains do not give good yields of well-folded soluble protein when expressed as recombinant constructs in *E.coli.*

We previously reported the expression and characterization of the **b'** and **b'x** domain fragments of human PDI, including defining the substrate binding site by modeling and mutagenesis studies[12]. In this paper we report further characterization of the recombinant **b'x** fragment and show that it exists in two distinct conformations which differ in tertiary structure, specifically in the environment of the unique Trp residue in the **x** region. From a screen of 75 mutants we identify mutants of **b'x** which exclusively adopt one or other of these alternative conformations. In addition, we report the crystal structure of one of these mutants, I272A; this reveals a hydrophobic pocket in the **b'** domain capped by the **x** region. All of the mutations we previously reported to inhibit substrate binding[12] shift the equilibrium in solution towards the capped conformer. Hence these mutations do not define the substrate binding site per se, but rather act indirectly to inhibit substrate binding by decreasing the accessibility of the substrate binding site. We present evidence that the two conformers are also significant in full-length PDI and suggest that these alternative conformations may explain some of the difficulties in characterizing the molecular properties of PDI and its sub-fragments and in understanding the ability of PDI to interact with folding substrates.

## Results

**Preparations of recombinant human PDI b'x contain at least two conformers**

The domain boundaries of human PDI were defined in previous work[11-12]; the constructs studied here correspond to the complete **b'** domain (P218-G332) and subsequent **x** region (K333-P351) with a short N-terminal tag (MHHHHHHMKHNQL-). The **b'** domain of human PDI contains no Trp residues, but there is a single Trp residue located in the **x** region (W347). Despite only having a single tryptophan, preparations of recombinant **b'x** gave a fluorescence emission spectrum with two maxima at approximately 334 nm and 355 nm (Fig 1a and ref 12). Furthermore, in HSQC spectra of $^{15}$N-labelled human PDI **b'x** (Fig 1b) there were two peaks for the characteristic Trp indole sidechain resonance (shown boxed). There was, however, no indication of heterogeneity by mass spectrometry (data not shown). We have previously shown that **b'x** preparations give a CD spectrum consistent with substantial α-helix and β-sheet content and a single denaturation transition curve with a midpoint of 2.32 M guanidine hydrochloride[12], both suggesting that the preparations do not contain any significant proportion of denatured protein. The fluorescence and NMR data in Figs 1a,b therefore indicate the possibility of conformational heterogeneity in the **b'x** preparations and we infer that the **b'x** fragment can exist in at least two alternative structured conformations. In one conformation (which we term the 'capped' form), the Trp residue in the **x** linker is buried in a hydrophobic environment, whereas in the other conformation (which we term the 'uncapped' form) the Trp residue is exposed to solvent. Analysis by fluorescence spectroscopy (Fig 1c,d) and CD (not shown) indicated that the ratio between these conformers could be altered by the applied salt concentration and that in dilute conditions at room temperature this salt-induced change in structure was reversible.

**Screening mutations that trap a single conformation of PDI b'x**

To generate homogenous preparations of the 'capped' and 'uncapped' conformers, we decided to screen for mutants in the **b'x** construct which might alter the conformer distribution, stabilising one relative to the other and allowing a single species to be characterised in detail. Since the fluorescence spectrum gave an easily-diagnosable read-out of the proportions of the 'capped' and 'uncapped' forms, this was chosen for the screen. In total 75 mutants of **b'x** were screened, a significant proportion of which were generated previously[12]. These mutants were purified on a micro-scale and their fluorescence spectra recorded in the elution buffer immediately after purification. Many of the mutants -- like the wild type protein -- showed a

mixed fluorescence spectrum with peaks at approximately 334 nm and 355 nm. However, some mutants showed a significant red-shift in fluorescence spectrum towards a single 355 nm peak (e.g. L343A, Fig 2a) while others showed a significant blue-shift towards a single 334nm peak (e.g. I272A, Fig 2b). Since the concentrations of the proteins obtained from the micro-scale purification varied by about three-fold, the parameter chosen to cross-compare spectra was the ratio of the average fluorescence at 331-337 nm to the average fluorescence at 352-358 nm. This ratio in elution buffer was 1.03 for the wild type protein and varied from 0.78 (extremely red-shifted) to 1.29 (extremely blue-shifted) for the mutant proteins (Fig 2c). Many of the mutations which resulted in a significant shift in the fluorescence ratio were either located in the vicinity of the substrate binding site as defined previously[12] or were located in **x**. No significant differences were observed in the fluorescence ratios of the wild type, I272A or L343A mutant **b'x** constructs when excited at 280 or 295nm.

**NMR studies of PDI b'x constructs**

Two red-shifted and three blue-shifted mutants in **b'x** were chosen for purification to homogeneity on the milligram scale. Fluorescence analysis of these mutant proteins confirmed the results obtained from the micro-scale purification (data not shown). The CD spectra were characteristic of globular proteins with defined secondary structure, although the spectra were clearly different in detail from each other and from the wild-type protein (data not shown). The mutant proteins were generated in $^{15}$N-labelled form. HSQC spectra of the five mutants showed significant differences between them in resolved resonances, average line shape and dispersion. The HSQC spectra of the blue-shifted mutants, I272A (Fig 3a), D346A/D348A and I272A/D346A/D348A (data not shown), gave well resolved spectra consistent with a folded protein of the expected molecular size. In contrast, the HSQC spectra of the red-shifted mutants, L343A (Fig 3b) and K349A (not shown), revealed both purified proteins as having poorly defined broad resonances characteristic of conformational exchange over the NMR experimental time scale due to significant structural differences and/or the presence of multimeric interactions. If we assume that the proportion of 'uncapped' and 'capped' forms can be judged from the intensity of the two resonances from the Trp indole HN, then the two red-shifted mutants appear to be entirely in the 'uncapped' form, whereas all blue-shifted mutants had substantially less of the 'uncapped' form than the unfractionated wild-type sample (Fig 1b). The amount of

'uncapped' form was estimated for each blue-shifted mutant and found to decrease in the order D346A/D348A > I272A > the triple mutant I272A/D346A/D348A, with the triple mutant containing essentially none of this conformer. These results suggest that the blue-shift mutations act to stabilise the 'capped' form relative to the 'uncapped' form and that the effect of the blue-shift mutations is additive.

**Crystallography of mutant PDI b'x**

Despite numerous attempts, we have been unable to obtain crystals from wild-type PDI **b'** or **b'x**, a situation we tentatively assigned to the mixed conformers present. To test this hypothesis, we tried to crystallize the I272A and D346A/D348A mutants of human PDI **b'x** which had appeared by NMR to be predominantly in a single conformation. Both mutants gave crystals, but despite extensive optimization we were unable to improve the crystallization of the D346A/D348A mutant to obtain crystals which diffracted well. In contrast, we obtained crystals of the I272A mutant which diffracted well using 3 different conditions. Excluding the high B-factor loop which contains E304, E305 and E306, the crystals obtained under the three conditions gave structures whose backbone traces were effectively super-imposable. We chose to refine the structure of the pH 8.5 crystal, since the diffraction pattern was slightly better than that obtained at pH 7.5 and since it is closer to physiological pH than the crystal obtained at pH 9.5. This structure has been refined at 2.2Å resolution (Table 1). There is one monomer per asymmetric unit. Seventeen residues were not observed in the electron density map, these being the three C-terminal residues (K349-P351) and fourteen N-terminal residues comprising the His-tag and the first residue of the PDI sequence (P218). Therefore the ordered part of the **b'x** construct in the crystal structure extends from L219 to D348. The final model contains 2 *cis*-peptide bonds (C295-P296 and E304-E305).

The crystal structure of the **b'** domain of the I272A **b'x** construct (Fig 4a,b) unsurprisingly exhibits a PDI-type trx fold (ßαßαßαßßα). Part of the **x** region (residues P336-L338) forms an additional β-strand to the mixed β-sheet core of the trx fold and the rest of **x** loops round to remain juxtaposed to **b'** with residues E345-W347 forming a short C-terminal helix.

The **b'** domain of PDI binds peptides and non-native proteins via predominantly hydrophobic interactions[16-17] and hence a hydrophobic surface or solvent exposed hydrophobic pocket might be expected

on **b'**. Surface charge density maps (Fig 4c) of **b'** reveal a hydrophobic pocket on one face of the domain which is capped by the **x** region (Fig 4d). Specifically, the side-chain of M339 (the first residue after the β-strand in the **x** region) binds in a hydrophobic pocket comprising the sidechains of F287, F288, I301, M307 and the alkyl chain of K309. The side-chains of L343 and W347 (further along the **x** region) also make hydrophobic interactions with the **b'** domain and interact with the side-chains of F223, A228, F232, I284, F287, F288, L303, M307 in a broad site immediately adjacent to the binding pocket of M339. Residues F223, I284, F287 and F288 interact most closely with the side-chain of W347, and the site of mutation (A272) is located behind F223 which forms the bottom of this interaction pocket. The indole HN of W347 makes polar interactions with the side-chain carboxyl of E225 which also forms a salt bridge with the guanidino group of R283 on the top rim of the site (Fig 4d). It is noteworthy that the residues M339, L343 and W347 in the **x** region, and those forming the hydrophobic patches on the **b'** domain with which they interact, are well conserved in PDI sequences of metazoan animal species.

This hydrophobic location of the side-chain of W347 is consistent with the single peaked blue-shifted fluorescence spectrum of this mutant. Structure analysis by Internal Coordinates Mechanics (ICM, Molsoft L.L.C., La Jolla, CA) shows vacancies in the structure with a total size of 50 $Å^3$ adjacent to the A272 side chain. This is consistent with the difference in volume between the side chain of the isoleucine found in the wild-type protein at this position and the alanine found in the mutant, suggesting that the structure obtained for the mutant is valid for that of the wild-type **b'x** protein, or at least for the 'capped' conformer of it.

**Evidence for alternative conformations of the x region in full-length PDI**

Previously we showed that limited proteolysis of full length bovine PDI results in the formation of fragments in which the **x** region was associated either with the **b'** domain or with the **a'** domain, but that cleavage at an internal site in the **x** region was never observed[23]. This implies that the **x** region is structured and that this structure can be associated with either of the adjacent domains. While such an observation is consistent with the crystal structure of the capped conformer of **b'x**, care must be taken in interpreting proteolytic data in that conformational exchange of fragments can be different from that in the intact

molecule and in the case of PDI it is likely that proteolytic fragments will be bound by the substrate binding site and will therefore perturb any conformational exchange in the full length protein.

Full length human PDI contains five tryptophan residues, two in the **a** domain (W35, W111), and two in **a'** (W379, W390), in addition to the one in **x** (W347). Hence its fluorescence emission is complex and it was not possible to use fluorescence to test whether the two conformational states (capped and uncapped) occur in the wild-type protein. The fluorescence of W35 in the WCGHC active site motif is very heavily quenched in both the dithiol and disulphide states[24] and similarly W379 in the WCGHC active site motif of the **a'** domain would be expected to be heavily quenched. Hence in a W111F / W390F mutant PDI, the fluorescence of W347 would be expected to predominate. Micro-scale purification of this mutant revealed a tryptophan fluorescence spectrum that showed two distinct maxima (data not shown) though the spectrum overall was red-shifted compared to wild-type **b'x** (fluorescence ratio = 0.916 ±0.014, n=14). Informed by the **b'x** screen, two 'blue-shift' mutations (I272A and D346A/D348A) were then made in the W111F / W390F mutant background and both resulted in shifts towards the 'capped' form, similar to that observed in the isolated **b'x** construct (fluorescence ratios = 1.068 ±0.015 (n=7) and 1.021 ±0.009 (n=2) respectively). Similarly the 'red-shift' mutation L343A resulted in a shift towards the 'uncapped' form (fluorescence ratio = 0.865 ±0.005, n=3), though with a smaller shift than that observed for wild-type vs L343A **b'x** (see Fig 2). To confirm that the observed two-state nature of the fluorescence spectra was due to alternative environments of W347 in the **x** region, the quadruple-mutant W35F / W111F / W379F / W390F of PDI was made along with one 'red-shift' (L343A) and one 'blue-shift' (I272A) mutant in this background. As for the mutants in the W111F/W390F background, the L343A and I272A mutants in the 'quadruple W-mutant' background resulted in a 'red-shift' and 'blue-shift' respectively. However, proteolysis revealed that the additional mutations in W35 and W379 clearly destabilized the protein relative to wild type (data not shown). Overall the fluorescence data imply that the two conformers seen in the isolated **b'x** domain, corresponding to alternative environments for the **x** region, also occur in full length PDI and that the relative proportions of each conformer are affected in the same way by the same mutations. However, all of these fluorescence data were obtained from mutated forms of PDI.

To examine whether the conformational exchange occurs in wild type full length PDI, NMR based analysis was undertaken. Full length human PDI is 55kDa, too large for a full structural analysis. Instead we

focused on the indole region of the $^{15}$N-$^{1}$H HSQC spectrum as we had already observed a good correlation between shifts in fluorescence spectra and shifts in peak intensities in this region (see Figs. 1-3). Assignments of the individual indole resonances in full length PDI were confirmed from a series of HSQC experiments where tryptophan residues were systematically mutated to phenylalanine in the full length protein and from studies on the isolated **a** domain. The indole resonance of W347 in wild-type **b'x** was seen as two peaks, showing the presence of both 'capped and 'uncapped' species that can be individually defined as distinct single peaks in the mutants **b'x** I272A/D346A/D348A (capped) and **b'x** L343A (uncapped) (Fig. 5). Asymmetric line broadening of the W347 indole resonance in the $^{15}$N-$^{1}$H HSQC spectrum of full length wild type PDI tracks between the observed 'capped' and 'uncapped' indole resonances from both wild-type and mutant **b'x** samples (Fig. 5). This indicates that W347 in full length wild-type PDI also experiences conformational exchange between 'capped' and 'uncapped' forms.

Previously there has been a debate on whether substrate binding by PDI is redox modulated[25-26]. No significant redox dependent changes were seen in the fluorescence spectra of wild type, I272A or L343A mutant **b'x** constructs or in the full length PDI W111F/W390F construct. Furthermore there was no significant redox-dependent difference in the appearance of the W347 resonance in the $^{15}$N-$^{1}$H HSQC spectrum of full length wild type PDI. These observations indicate that there is no redox dependence for the conformational exchange between the 'capped' and 'uncapped' conformers and are consistent with the report that PDI is not a redox-dependent molecular chaperone[26].

## Discussion

### A high resolution structure of the b'x region of human PDI

The biological significance of mammalian PDI in the folding of disulphide-bonded secretory proteins has been clear for many years[2] and throughout that time pure protein has been abundantly available, either from mammalian tissues or recombinant sources[27-28]. No high-resolution structure of a full-length PDI from a multicellular organism has been achieved to date, despite the fact that many groups have attempted to determine the structure of PDIs by x-ray crystallography. This may reflect inter-domain flexibility or the existence of regions within the full-length molecule that can adopt alternative conformations. The structures

of individual **a**, **b** and **a'** domains of human PDI have been determined by NMR, but it is interesting that no structure has been achieved by this approach for the **b'** domain or the combination of **b'** domain and adjacent **x** linker region. The difficulty in previous studies of the **b'** domain or **b'x** combination could have arisen from poor definition of the domain boundary, but it is possible that this domain does show some intrinsic flexibility or conformational heterogeneity.

In the current paper we have analysed recombinant human **b'x** in detail -- using a range of techniques for structural study in solution -- and identified the existence of at least two conformational states which clearly differ in the structural relationship between the **x** region (specifically W347) and the **b'** domain. In order to overcome the complexities introduced by the existence of these states, we sought to generate mutants which might preferentially stabilize one state and eliminate the conformational heterogeneity. Using fluorescence screening we were able to identify mutants which preferentially adopted one or other conformation and we have crystallized and determined the structure of the I272A mutant of **b'x,** which we inferred to be entirely in the 'capped' conformation. In accordance with our inference from solution studies on this mutant and on wild-type **b'x**, the **x** region is intimately associated with the **b'** domain in the crystal structure of **b'x** I272A, and the indole side-chain of W347 is buried in a hydrophobic pocket.

**Comparison with b' and x conformations in other PDI family structures**

In Fig 6a we show a structure-based sequence alignment of the human PDI **b'x** region with the corresponding region of yeast Pdi1p, based on the structure of I272A **b'x** solved here ("capped **b'x**") and on the full-length structure of Pdi1p[13]. This shows that, overall, the two **b'** domains adopt very similar conformations with the exception of the region of Pdi1p (residues 330-344 in mature Pdi1p numbering) which forms a loop and a short additional helix. In capped human PDI **b'x**, this region comprises simply a short loop linking β5 directly to α4. A superposition of the backbone atoms (N, Cα, C) for the core β-strands (42 residues) of yeast and capped human **b'x** using the alignment in Fig 6a gave an RMSD of 0.6Å; this value increased to 1.5Å when both β-strands and helices were included (102 residues). Comparison using DALI[29] of capped human **b'x** to all structures in the protein databank found close similarities with the **b'** domains of other PDI-family members including the **bb'** fragment of human ERp57 (2h81), the **b'** domain of

yeast Pdi1p (2be5), the **bb'** fragment from human PDI (2k18; structure released whilst this manuscript was in review) and the **b'** domain from the thermophilic fungus *Humicola insolens* (2djk) along with other thioredoxin fold containing proteins.

However there is a major difference between the human and yeast PDI structures in the orientation of the **x** region, which in the Pdi1p structure does not cap the hydrophobic surface of **b'** as is observed for capped human PDI **b'x** (Fig 6b). Were human PDI **b'x** to adopt such an arrangement as found in yeast Pdi1p, W347 in the **x** region would be solvent exposed and probably highly mobile, consistent with the fluorescence data for the 'uncapped' conformer of the wild-type human **b'x** protein. We infer that the structure of the 'uncapped' conformer of wild-type human **b'x** has some similarities to that observed in the crystal structure of yeast Pdi1p and that the identified 'red-shifting' mutations act by stabilising this structure in isolated human **b'x**.

In analyzing their structure of yeast Pdi1p, Tian et al.[13] drew attention to a 'highly hydrophobic pocket' on the surface of the **b'** domain (Fig 6b) which – together with hydrophobic patches at homologous positions on the surfaces of the **a**, **b** and **a'** domains – forms a continuous hydrophobic surface. We had previously proposed that the **b'** domain provided the principal binding site on PDI for peptides and unfolded proteins, based on studies on human PDI domains and domain combinations[12,16]. Functional studies on yeast Pdi1p mutants by Tian et al.[13] confirm that this conclusion is also valid for the yeast enzyme. In Fig 6a, the residues which contribute side-chains to the 'highly hydrophobic pocket' identified by Tian et al.[13] are highlighted and the corresponding residues in the structurally-aligned human sequence are also highlighted. Also in Fig 6a, we identify the residues of the **b'** domain of human **b'x** which form hydrophobic contacts with the sidechains of L343 and W347 of the **x** linker. It is clear that the **x** linker is retained in close contact with the **b'** domain by binding in the centre of a wider hydrophobic surface which corresponds to the hydrophobic pocket identified by Tian et al.[13] in the yeast Pdi1p structure and which has been identified as the putative ligand binding site (Fig 6b). Hence the interaction of **x** with **b'** probably mimics substrate binding.

Previously we identified the substrate binding site in PDI by a combination of modelling and mutagenesis studies[12]. The binding site identified by those studies was not based on a high resolution structure and does not match that identified here or by Tian et al.[13]. However, all of the mutations previously

identified to inhibit substrate binding (including I272A) are blue-shift mutants (compare data presented in reference 12 with that in fig 2c). Hence these mutations are not in residues directly involved in interaction with substrates but rather they act indirectly by stabilizing the conformer in which the hydrophobic pocket is 'capped' by **x** relative to the 'uncapped' conformer. Hence our previous results showing that these mutations in **b'x** or full length PDI result in a loss of ability to bind substrates suggest that, once capped by **x,** the substrate binding site is unavailable to bind substrates.

**Functional significance of novel structural interaction between b' domain and x linker**

This work has i) demonstrated that there is an intrinsic conformational heterogeneity in isolated human PDI **b'x** between conformers which differ in the structural relationship between the **b'** domain and the **x** region, ii) determined the structure of a mutant **b'x** showing an interaction between parts of the **x** region and sites on the **b'** domain, and iii) confirmed that, in this structure, **x** is interacting with a hydrophobic site homologous to that postulated as the ligand binding site in the structure of full-length yeast Pdi1p. Hence it appears that the **x** region can adopt alternative conformations and can potentially exchange between the 'capped' conformation, in which it occludes a site on the **b'** domain, and an 'uncapped' conformation which reveals this site. Furthermore we have demonstrated that a similar conformational exchange happens in full length human PDI with the **x** region capable of interacting reversibly with the principal substrate binding site. It is likely that this is significant as part of the function of the enzyme in facilitating protein folding. The existence of such alternative conformations for the **x** region in full-length PDI provides one explanation why the structure of full-length mammalian PDI has been so resistant to determination by x-ray crystallography. However it is not the only reason as the full length human PDI I272A mutant appears to be intransigent to crystallization. It is significant that the structure of yeast Pdi1p was generated from a slowly-developing crystal form in which one molecule of the protein occupies the putative ligand binding site of a second, symmetry-related molecule[13]; this may have the effect of trapping the protein in the 'ligand-bound' form and inhibiting conformational exchange. This suggests that the yeast structure represents a conformation in which a protein ligand is bound, displacing the **x** linker, whereas the human **b'x** structure represents an alternative ligand-free state, and that both these conformations may be relevant to the functional cycle of the enzyme. It is well-known that in the functional cycle of GroEL/ES, the

hydrophobic binding site in the apical domain of GroEL that binds unfolded protein substrates alternates between being available for binding such ligands and being occluded by interaction with part of the GroES cap[30]; the interconversion between such states involves relative re-orientations of the domains of the chaperonin molecule. Such re-orientations between domains (and especially of the **x** linker relative to its neighbouring domains) may be a critical feature of the functional cycle of PDI and the process of such re-orientations in preparations of full-length PDI may generate the conformational heterogeneity which has frustrated previous structural studies.

During the preparation of this manuscript data supporting the notion that this conformational exchange may be linked to function has emerged. The crystal structure of human ERp44, a PDI-family member with the domain architecture **a-b-b'**, revealed that the C-terminal tail caps both a hydrophoic pocket in the **b'** domain thought to be the substrate binding site and a region close to the active site in the catalytic **a** domain[31]. The structure of the capped **b'** domain of ER44 is remarkably similar to that of the capped **b'** domain of human PDI we report here (Fig. 7). A superposition of the backbone atoms (N, Cα, C) for 19 residues in structurally analogous positions across all 5 β-strands gave a RMSD of 0.57Å. Likewise the nature of the interaction of the carboxy terminal extension with the hydrophobic pocket in ERp44 is similar to that of the **x** region capping the binding site in the **b'** domain of PDI, with L361 in ERp44 playing a role similar to that of W347 in PDI. Deletion of the C-terminal tail in ERp44 increased the activity of ERp44 and its ability to act as a molecular chaperone[31], both consistent with a regulatory role for the tail. The emergence of two such similar structures suggests that this may be a common regulatory mechanism within the PDI family.

## Materials and Methods

### Mutagenesis, expression and purification

The plasmid containing the gene encoding the human PDI **b'x** domain fragment is a derivative of pET23b (Novagen). The resulting protein is expressed with an N-terminal His tag MHHHHHHM and encodes residues K213 to P351 of mature PDI. Mutagenesis, transformation, expression and purification were done as described previously[12].

Purification on a micro-scale from 5-10 ml of bacterial culture for fluorescence screening of mutants was done using Ni-NTA spin columns (Qiagen) according to the protocol recommended by the manufacturer.

**Spectroscopy**

Fluorescence spectra were measured on a Perkin/Elmer LS50 fluorometer (excitation 280 nm or 295nm for **b'x** and 295nm for the full length protein, emission 300-400 nm or 320-400nm, slit widths 5 nm, scan speed 200 nm/min, average of 4 scans, 25°C). CD spectra were measured on a Jasco J-815 CD (0.1 mm cuvette, 250-190 nm, step resolution 0.2 nm, response 0.25 s, bandwidth 1 nm, scan speed 20 nm/min, average of 8 scans, 25°C)

**NMR spectroscopy**

Cells were grown in minimal medium to allow isotopic labeling using 1 g/l $^{15}N$ ammonium sulphate as sole nitrogen source and labelled protein was expressed and purified as described previously[12]. Samples for NMR analysis were concentrated to between 0.36 mM and 1 mM in 20 mM phosphate buffer (pH 6.5) containing 150 mM NaCl by centrifugation on Vivaspin columns with a 5 kD cut-off and then $D_2O$ added to a final concentration of 10% (v/v). $^{15}N/^{1}H$ HSQC spectra were collected at 25°C as described in Alanen et al.[22].

**Crystallization**

To prepare selenium-labelled protein a fresh colony of *E.coli* B834(DE3) carrying plasmid pAKL73 encoding for human PDI **b'x** I272A was inoculated into 200 ml of sterile M9 minimal medium with an additional 0.5 mM L-methionine at 37°C for 20 hours. The bacteria were collected by centrifugation and washed twice with 200 ml of sterile water. The bacteria were then resuspended in M9 medium with an additional 0.5 mM of selenomethionine to an $OD_{600}$ of 0.3 and grown at 37°C, 200 rpm to an $OD_{600}$ of 0.7 and then induced with 1 mM of IPTG at 25°C, 200 rpm, for 12 hours. The bacteria were then harvested by centrifugation, resuspended in 1/10 volume 20 mM phosphate pH 7.3, DNase added to final concentration of 10 μg/ml and lysozyme to final concentration of 1 mg/ml and then frozen. The protein was purified as for non-labelled protein. 10 mg of purified Se-labeled protein was obtained from 1 litre of culture.

Crystals of the SeMet derivative were obtained after one week by the hanging-drop, vapour diffusion equilibration method[32] using 2 μl of protein solution (10 mg/ml) and 2 μl well solution (0.2 M NaCl, 2.95 M ammonium sulfate, 0.1 M Tris-HCl buffer pH 9.5) at 22°C. Crystals of unlabelled protein were obtained by the same method at pH 7.5, 8.5 and 9.5. Crystals were harvested from drops, put into paraffin oil for a few seconds and then frozen in liquid nitrogen. X-ray data were collected at cryogenic temperature (100 K).

**Data collection, structure determination and refinement**

Multiwavelength anomalous dispersion (MAD) data were collected for the Se-Met crystal on CCD detector at beamline BW7A (DESY, Hamburg, Germany). The x-ray data from native crystals at pH 7.5, 8.5 and 9.5 were collected at X12 (DESY, Hamburg, Germany). Images were processed using the XDS program package[33-34]. Using the Auto-RickShaw protocol[35-36], the MAD-dataset resulted in the initial model that was completed with COOT[37]. Subsequently the structure of the native datasets were refined using REFMAC5[38] with the translation libration screw (TLS) description of the anisotropic rigid body motion[39]. The data collection statistics and refinement statistics of the pH 8.5 structure are given in Table 1. The model includes 45 waters and one sulphate ion.

**Protein Data Bank accession code**

The atomic coordinates and structure factors of the **b'x** domain of human PDI have been deposited in the Protein Data Bank with accession code 3BJ5.

## Acknowledgements

This work was supported by the Academy of Finland, Sigrid Juselius Foundation, the Biocenter Oulu, the University of Oulu, the Biotechnology and Biological Sciences Research Council, UK (Grant BB/D017807), and the Vice-Chancellor's Fund, Warwick University. We thank Lee Byrne (University of Kent) for helpful discussions.

## References

1. Ellgaard, L. & Ruddock, L.W. (2005) The human protein disulphide isomerase family: substrate interactions and functional properties. *EMBO Reports* **6**, 28-32
2. Freedman, R. B. (1984) Native disulphide bond formation in protein biosynthesis: evidence for the role of protein disulphide isomerase**.** *Trends Biochem. Sci.* **9**, 438-441
3. Creighton, T.E., Hillson, D.A. & Freedman, R.B. (1980) Catalysis by protein-disulphide isomerase of the unfolding and refolding of proteins with disulphide bonds. *J. Mol .Biol.* **142**, 43-62
4. Creighton, T.E., Bagley, C.J., Cooper, L., Darby, N.J., Freedman, R.B., Kemmink, J. & Sheikh, A. (1993) On the biosynthesis of bovine pancreatic trypsin inhibitor (BPTI): structure, processing, folding and disulphide bond formation of the precursor *in vitro* and in microsomes. *J. Mol. Biol.* **232**, 1176-1196.
5. Hawkins, H.C., Blackburn, E.C. & Freedman, R.B. (1991) Comparison of the activities of protein disulphide-isomerase and thioredoxin in catalysing disulphide isomerization in a protein substrate. *Biochem. J.* **275**, 349-353.
6. Darby, N.J., Freedman, R.B. & Creighton, T.E. (1994) Dissecting the Mechanism of Protein Disulfide Isomerase: Catalysis of Disulfide Bond Formation in a Model Peptide. *Biochemistry* **33**, 7937-7947
7. Ruoppolo, M., Freedman, R.B., Pucci, P. & Marino, G. (1996) Glutathione-dependent pathways of refolding of RNase $T_1$ by oxidation and disulfide isomerization: catalysis by protein disulfide isomerase. *Biochemistry* **35**, 13636-13646
8. Edman, J.C., Ellis, L., Blacher, R.W., Roth, R.A. & Rutter, W.J. (1985) Sequence of protein disulphide isomerase and implications of its relationship to thioredoxin. *Nature* **317**, 267-270
9. Kemmink, J., Darby, N.J., Dijkstra, K., Nilges, M. & Creighton, T.E. (1997) The folding catalyst protein disulfide isomerase is constructed of active and inactive thioredoxin modules. *Curr. Biol.* **7**, 239-245
10. Freedman, R.B., Klappa, P. & Ruddock, L.W. (2002) Protein disulfide isomerases exploit synergy between catalytic and specific binding domains. *EMBO Reports* **3**, 136-140
11. Alanen, H.I., Salo, K.E.H., Pekkala, M., Siekkinen, H.M., Pirneskoski, A. & Ruddock, L.W. (2003) Defining the domain boundaries of the human protein disulfide isomerases. *Antioxid. Redox. Signal.* **5**, 367-374

12. Pirneskoski, A., Klappa, P., Lobell, M., Williamson, R.A., Byrne, L., Alanen, H.I., Salo, K.E.H., Kivirikko, K.I., Freedman, R.B. & Ruddock, L.W. (2004) Molecular characterization of the principal substrate binding site of the ubiquitous folding catalyst protein disulfide isomerase. *J. Biol. Chem*. **279**, 10374-10381

13. Tian, G., Xiang, S., Noiva, R., Lennarz, W.J. & Schindelin, H. (2006) The crystal structure of yeast protein disulfide isomerase suggests cooperativity between its active sites. *Cell* **124**, 61-73

14. Gruber, C.W., Cemazar, M., Heras, B., Martin, J.L. & Craik, D.J. (2006) Protein disulfide isomerase: the structure of oxidative folding. *Trends. Biochem. Sci.* **31**, 455-464

15. Darby, N.J., Penka, E. & Vincentelli, R. (1998) The multi-domain structure of protein disulfide isomerase is essential for high catalytic efficiency. *J. Mol.Biol.* **276**, 239-247

16. Klappa, P., Ruddock, L.W. Darby, N.J. & Freedman, R.B. (1998) The **b'** domain provides the principal peptide-binding site of protein disulfide isomerase but all domains contribute to binding of misfolded proteins. *EMBO J.* **17**, 927-935.

17. Klappa, P., Freedman, R.B., Langenbuch, M., Robinson, G.K. & Ruddock, L.W. (2001) The pancreas-specific protein disulphide-isomerase PDIp interacts with a hydroxyaryl group in ligands. *Biochem. J.* **354**, 553-559

18. Pollock, S., Kozlov, G., Pelletier, M.-F., Trempe, J.-F., Jansen, G., Sitnikov, D., Bergeron, J.J.M., Gehring, K., Ekiel, I., & Thomas, D.Y. (2004) Specific interaction of ERp57 and calnexin determined by NMR spectroscopy and an ER two-hybrid system. *EMBO J.* **23**, 1020-1029

19. Russell, S.J., Ruddock, L.W., Salo, K.E.H., Oliver, J.D., Roebuck, Q.P., Llewellyn, D.H., Roderick, H.L., Koivunen, P., Myllyharju, J. & High, S. (2004) The primary substrate binding site in the b' domain of ERp57 is adapted for endoplasmic reticulum lectin association. *J. Biol. Chem*. **279**, 18861-18869

20. Silvennoinen, L., Myllyharju, J., Ruoppolo, M., Orru, S., Caterino, M., Kivirikko, K.I., & Koivunen, P. (2004) Identification and characterization of structural domains of human ERp57: association with calreticulin requires several domains. *J. Biol. Chem.* **279**, 13607-13615

21. Urade, R., Okudo, H., Kato, H., Moriyama, T. & Arakaki, Y. (2004) ER-60 domains responsible for interaction with calnexin and calreticulin. *Biochemistry* **43**, 8858-8868

22. Alanen, H.I., Williamson, R.A., Howard, M.J., Hatahet, F.S., Salo, K.E.H., Kauppila, A., Kellokumpu, S. & Ruddock, L.W. (2006) ERp27, a new non-catalytic endoplasmic reticulum-located human protein disulfide isomerase family member, interacts with ERp57. *J. Biol. Chem*. **281**, 33727-33737

23. Freedman, R.B., Gane, P.J., Hawkins, H.C., Hlodan, R., McLaughlin, S.H. & Parry, J.W.L. (1998) Experimental and theoretical analyses of the domain architecture of mammalian protein disulphide-isomerase *Biol.Chem.* **379** 321-328

24. Lappi A.K., Lensink, M.F., Alanen, H.I., Salo, K.E.H., Lobell, M., Juffer, A.H. & Ruddock, L.W. (2004) A conserved arginine plays a role in the catalytic cycle of the protein disulphide isomerases. *J. Mol. Biol.* **335**, 283-295

25. Tsai, B., Rodighiero, C., Lencer, W.I. & Rapoport, T. (2001) Protein disulphide isomerase acts as a redox-dependent chaperone to unfold cholera toxin. *Cell* **104**, 937-948

26. Lumb, R.A. & Bulleid, N.J. (2002) Is protein disulfide isomerase a redox-dependent molecular chaperone? *EMBO J.* **21**, 6763-6770

27. Lambert, N. & Freedman, R.B. (1983) Structural properties of homogeneous protein disulphide-isomerase from bovine liver purified by a rapid high-yielding procedure. *Biochem. J.* **213**, 225-234

28. Vuori, K., Myllyla, R., Pihlajaniemi, T. & Kivirikko, K.I. (1992) Expression and site-directed mutagenesis of human protein disulfide isomerase in *Escherichia coli*. This multifunctional polypeptide has two independently acting catalytic sites for the isomerase activity. *J. Biol. Chem.* **267**, 7211-7214

29. Holm, L. & Sander, C. (1996) Mapping the protein universe. *Science* **273**, 595-602

30. Horwich, A.L., Farr, G.W. & Fenton, W.A. (2006) GroEL-GroES-Mediated Protein Folding. *Chem. Rev.* **106**, 1917-1930.

31. Wang, L., Wang, L., Vavassori, S., Li, S., Ke, H., Anelli, T., Degano, M., Ronzoni, R., Sitia, R., Sun, F. & Wang, C.C. (2008) Crystal structure of human ERp44 shows a dynamic functional modulation by its carboxy-terminal tail. *EMBO Rep*. **9**, 642-7

32. Jancarik, J. & Kim, S,H, (1991) Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl.Crystallogr.* **24**, 409-411

33. Kabsch, W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795-800.

34. Kursula, P. (2004) *XDSi*: a graphical interface for the data processing program *XDS. J. Appl. Crystallogr.* **37**, 347-348

35. Panjikar S., Parthasarathy V., Lamzin V.S., Weiss M.S. & Tucker P.A.(2005) *Auto-Rickshaw*: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Crystallogr. D* **61**, 449-457.

36. Schneider TR & Sheldrick GM (2002) Substructure solution with SHELXD. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1772-1779

37. Emsley, P. & Cowtan, K. (2004) *Coot*: model-building tools for molecular graphics. *Acta Crystallogr D* **60**, 2126-2132

38. Murshudov, A., Vagin, A. & Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240-255

39. Winn, M., Isupov, M. & Murshudov, G.N. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D* **57**, 122-133

40. Koradi, R., Billeter, M. & Wüthrich, K. (1996) MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graphics* **14**, 51-55

## Figure legends

**Figure 1.** Spectroscopic properties of human PDI **b'x**

(a) Representative relative fluorescence emission spectrum, excitation at 280nm, and (b) $^{15}N/^{1}H$ HSQC NMR spectrum for purified wt protein. Emission maxima at 334nm and 355nm are indicated by the dotted and dashed lines respectively and the position of the two Trp indole cross-peaks in (b) are shown by the dashed box. (c and d) Shift in fluorescence spectra upon changing the ionic strength of the solution. (c) **b'x** (4.4 mg/ml) pre-incubated in 20 mM phosphate buffer, pH 7.3, and (d) **b'x** (4.4 mg/ml) pre-incubated in 20mM phosphate buffer, pH 7.3, containing 1.0M sodium chloride. Both (c) and (d) were pre-incubated for 30 minutes at room temperature and then fluorescence emission spectra collected at 25$^{o}$C after diluting 100-fold into 20mM phosphate buffer (solid line) or 20mM phosphate buffer plus 1.0M sodium chloride (dashed line). Separate kinetic measurements showed that the interconversion between the two forms occurred within the

manual mixing time. Similar shifts in fluorescence spectra were observed using a range of salt solutions indicating that the effects were dependent on ionic strength and not specific to the salt used. (AU = arbitrary units)

**Figure 2.** Screening of **b'x** mutants to identify mutations stabilising a single conformer

Representative relative fluorescence emission spectra, excitation at 280nm, of micro-scale preparations of (a) L343A and (b) I272A mutants of human PDI **b'x**. Emission maxima at 334nm and 355nm are indicated by the dotted and dashed lines respectively. (c) Shifts in fluorescence ratios (331-337nm/352-358nm) for all 75 mutants of human PDI **b'x** relative to the wild type protein (wild type fluorescence ratio = 1.03). The data shows the mean from at least 2 independent measurements and the horizontal dashed lines correspond to a ratio of +/- 0.15 from the mean value for the wild-type. The mutants L343A and I272A are shown with open boxes.

**Figure 3.** NMR studies of b'x mutants

$^{15}N/^{1}H$ HSQC NMR spectra of (a) I272A and (b) L343A mutants. The position of the Trp indole cross-peaks is shown by the dashed box.

**Figure 4.** Crystal structure of **b'x** I272A

(a) Ribbon diagram showing the trx fold of **b'x** with the **x** region coloured in green. (b) Rotation of the above structure through approx. 90 degrees along both the x- and z-axes to bring the **x** region to the front of the molecule. The side-chains of M339, L343 and W347 are shown as green sticks. (c) Surface charge density map for the **b'** domain in the same orientation as shown in (b) with the **x**-region shown in ribbon format. Blue = positive potential, red = negative potential, white = neutral. (d) Expanded view of the **x**-region binding site. The sidechains of residues F223, A228, F232, I284, F287, F288, L303 and M307 are coloured magenta and form a broad hydrophobic site in which residues L343 and W347 are bound. The sidechains of I301 and the alkyl chain of K309 are coloured light blue and together with F287, F288, M307 (coloured magenta) form a binding pocket for M339. The sidechains of E225 and R283 (charged atoms coloured grey)

form a salt bridge on the top rim of the binding site close to the indole HN of W347. This figure was produced using MOLMOL[40].

**Figure 5.** Conformational exchange in the environment of W347.

Overlay of the tryptophan indole regions of 600 MHz $^{15}$N-$^{1}$H HSQC NMR spectra of PDI **b'-x** L343A mutant (red), **b'x** I272A/D346A/D348A mutant (blue), **b'x** wild-type (green) and full length PDI (black). Assignments of all full length PDI indole resonances are indicated.

**Figure 6.** Structural comparison of human and yeast **b'x** domains

(a) Structure-based alignment of the human and yeast **b'x** sequences. A cyan background denotes residues identified as part of the hydrophobic binding site in the yeast structure[13] and the corresponding residues in the human protein. Two further residues (F287 and M307) appear to contribute to the human hydrophobic site (shown underlined in red) whereas two residues (N281 and R299) appear to fall outside the site (shown underlined in black). Residues in the human sequence coloured red comprise the binding site for L343 and W347 as judged from the **b'x** structure (at least one side-chain to side-chain atom distance within 5Å). The **x** region for both sequences is shown on a grey background with the hydrophobic residues in **x** important for interaction with **b'** shown in green. The residue numbering is for the mature protein in both cases and the secondary structure assignments (as defined in the pdb header for yeast (2b5e) and using the Kabsch and Sander algorithm as implemented in MOLMOL[39] for human) are shown above or below the sequence as appropriate. (b and c) Surface representation of the hydrophobic binding sites as defined in the sequence alignment above (those residues highlighted in cyan and underlined in red) for yeast and human **b'x** respectively. The side-chains of the hydrophobic sites are shown in white whilst the backbone and all other **b'** residues are coloured blue. The **x** regions are shown in ribbon format and the side-chains of M339, L343 and W347 in the human structure are shown in green. The two views are in the same orientation (achieved by superposition of the core β-sheet) corresponding to that in Fig 4b and c. This figure was generated using MOLMOL[40].

**Figure 7** Structural comparison of human PDI **b'x** and ERp44

The two structures are shown in the same orientation as determined by superposition of their β-strands. The **x** region is shown in green for human PDI and magenta for ERp44. The sidechains of the residues important for hydrophobic interaction of **x** with **b'** (L343 and W347 in PDI, and F358 and L361in ERp44) are also shown. The sidechain of L361 in ERp44 is in an analogous position to W347 in PDI.

**Table 1.** Crystallographic statistics of the structure of the non-labelled human PDI **b'x** domain I272A mutant grown at pH 8.5 as processed by XDS. The values in parentheses are for the highest resolution shell.

| **Data collection statistics** | |
|---|---|
| Space group | P3(1)21 |
| Unit cell parameters. | |
| a, b, c (Å) | 57.37, 57.37, 68.31 |
| Temperature (K) | 100 |
| Wavelength (Å) | 0.89997 |
| Resolution (Å) | 50-2.2 |
| Rmerge (%) | 5.5 (29.3) |
| Completeness (%) | 99.3 (99.6) |
| $I/\sigma_I$ | 17.84 (5.0) |
| Unique reflections | 6879 (849) |
| Redundancy | 4.3 (4.4) |
| Mosaicity (°) | 0.17 |
| B-factor from Wilson plot ($Å^2$) | 41 |
| **Refinement statistics** | |
| Resolution (Å) | 50-2.2 |
| Total number of reflections | 6877 |

| | |
|---|---|
| Working set: number of reflections | 6533 |
| $R_{factor}$ (%) | 19.4 |
| Test set: number of reflections | 344 |
| $R_{free}$ (%) | 25.5 |
| Protein atoms | 1063 |
| Water atoms | 45 |
| $SO_4^{2-}$ atoms | 5 |
| **Geometry statistics** | |
| Rmsd (bond distance) (Å) | 0.012 |
| Rmsd (bond angle) (°) | 1.4 |
| Rmsd B | |
| Main chain bonded atoms ($Å^2$) | 0.8 |
| Side chain bonded atoms ($Å^2$) | 1.8 |
| Main chain angle ($Å^2$) | 1.3 |
| Side chain angle ($Å^2$) | 2.8 |
| Average B | |
| Main chain atoms ($Å^2$) | 36.4 |
| Side chain atoms ($Å^2$) | 37.7 |
| Water molecules ($Å^2$) | 39.5 |
| Sulfate atoms ($Å^2$) | 80.7 |
| Ramachandran plot | |
| Most favored region (%) | 94 |
| Additionally allowed regions (%) | 6 |
| Generously allowed regions (%) | 0 |
| Disallowed regions (%) | 0 |

Figure 1

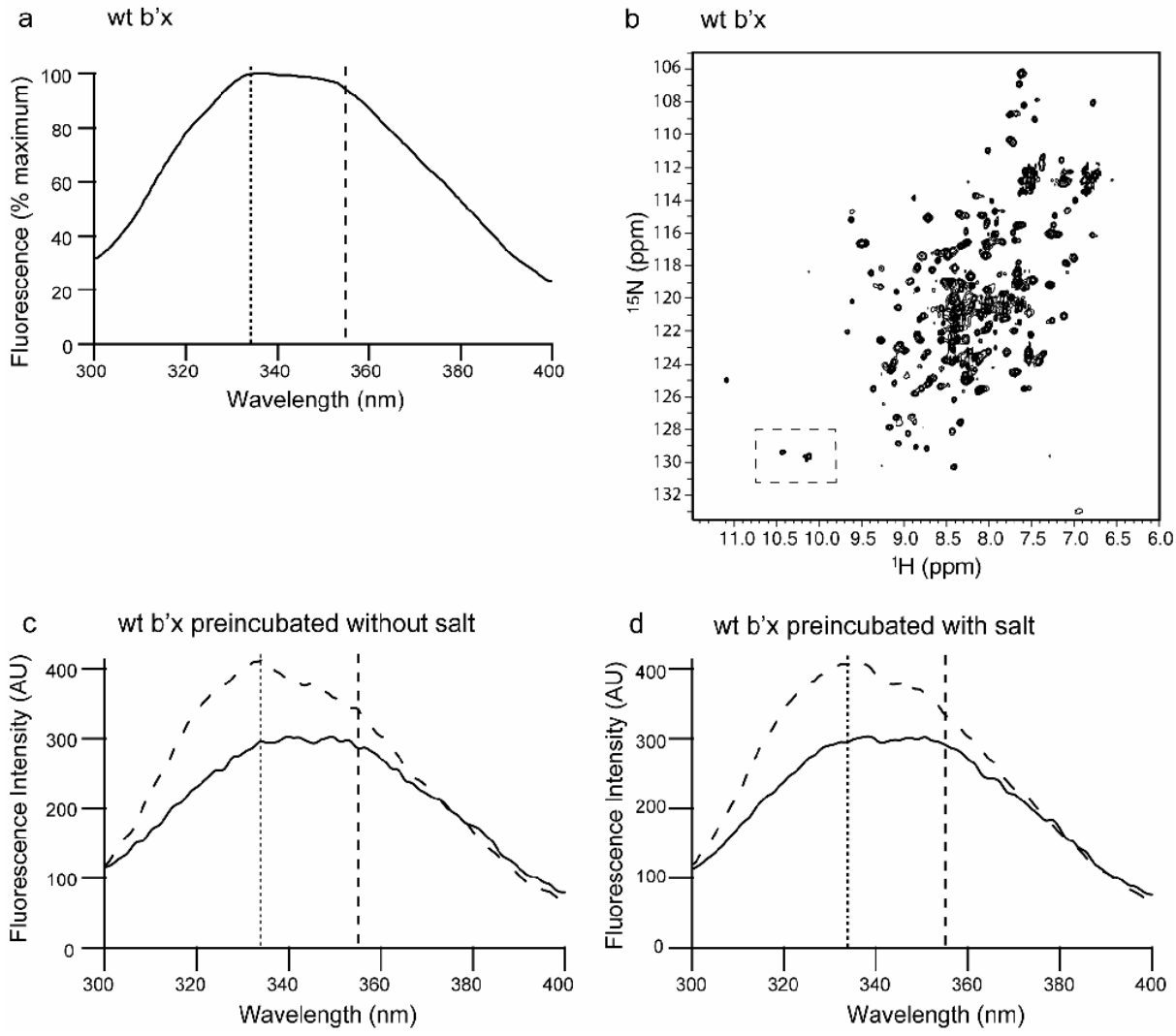


Figure 2

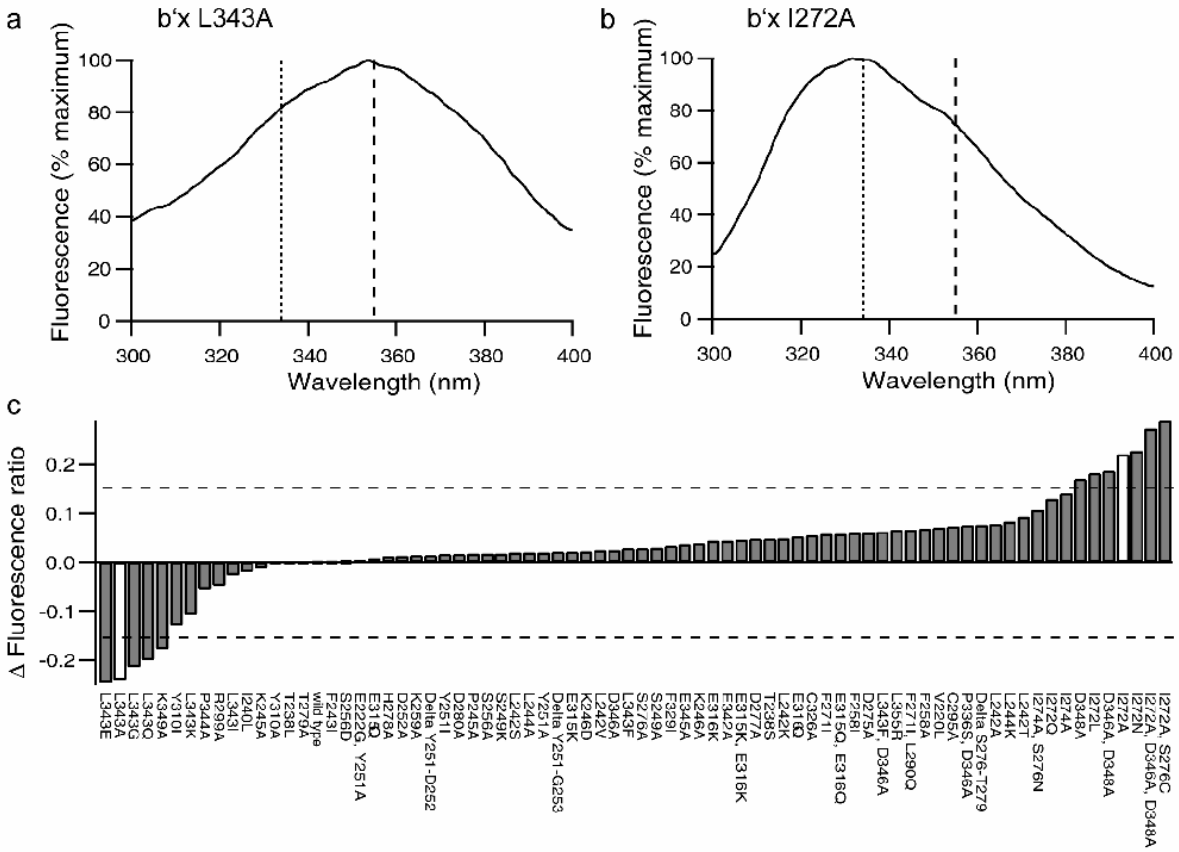


Figure 3

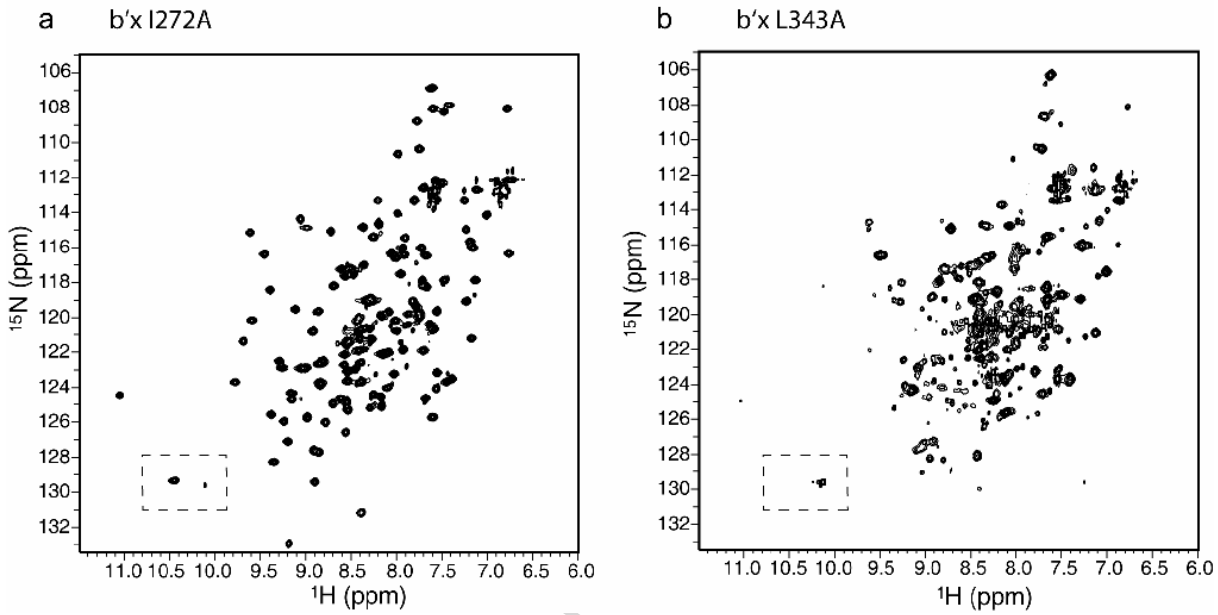


Figure 4

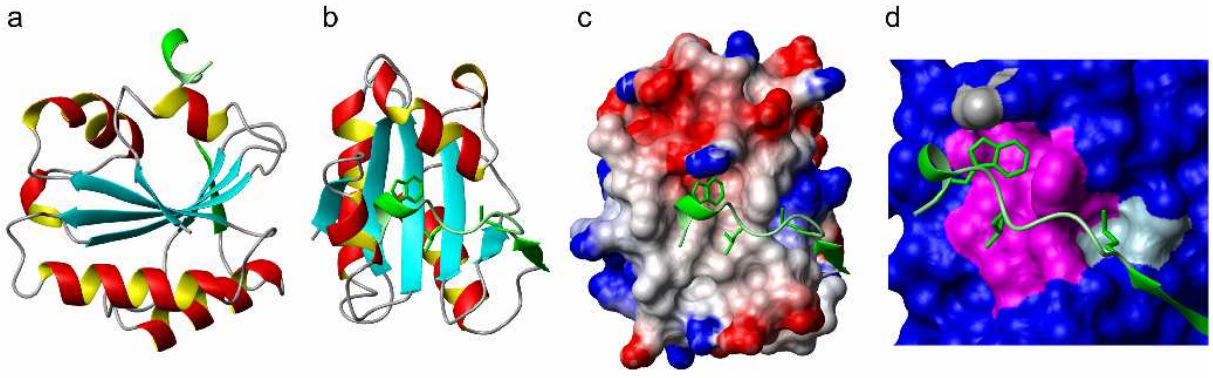


Figure 5

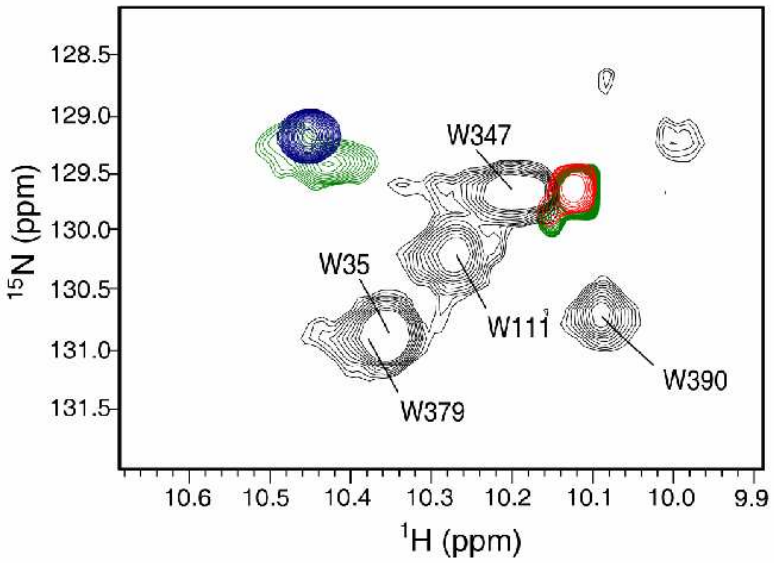


Figure 6

a

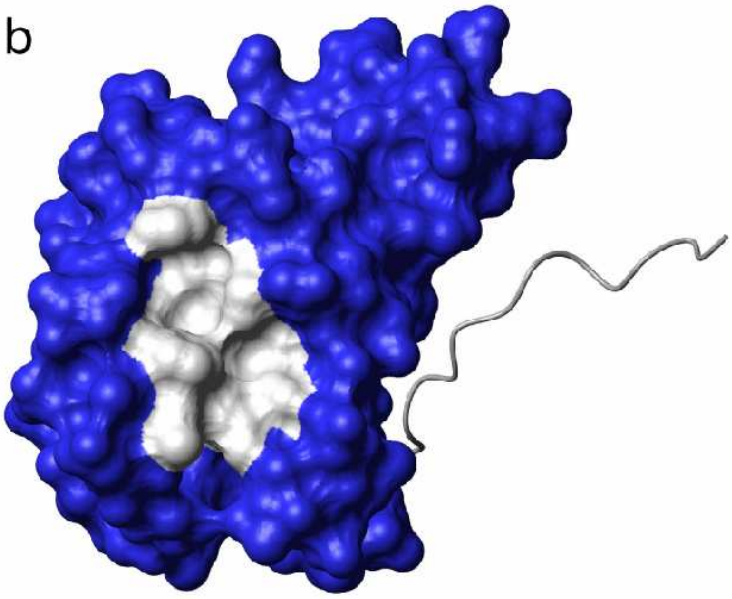
```
                  s  ss      hhh  hh      sss  ssss
Human    219  LV  IEFTEQTAPK  IFGGEIKTHI  LLFLPKSVSD
Yeast    241   F  GEIDGSVFAQ  YVESGL-PLG  YLFY-NDEEE
                   hhhhhhh  hhhh     ss  ssss  hhhh

     hhhhhhhhhh  hhh       s  sssss   hh  hhhhhhhh
251  YDGKLSNFKT  AAESFKGKIL  FAFIDSDHTD  NQRILEFFGL
270  LEEYKPLFTE  LAKKNRGLMN  FVSIDA--RK  FGRHAGNLNM
     hhhhhhhhhh  hhhhh     s  sssssh   hh  hhhhhhhh

      hhh   ssss  ss      sss                      hh
291  KKEECPAVRL  ITLEEEMTKY  KPES------EEL------TAE
308  KEQ-FPLFAI  HDMT-EDLKY  GLPQLSEEAFDELSDKIVLESK
           ssss  ss      sss  s    hhhhhhh        hhh

     hhhhhh  hhh  h     sss        hhh
321  RITEFCHRFL  EGKIKPHLMS  QELPEDWD  348
348  AIESLVKDFL  KGDASPIVKS  QEIFENQD  375
     hhhhhhhhhh  hh
```

b



c

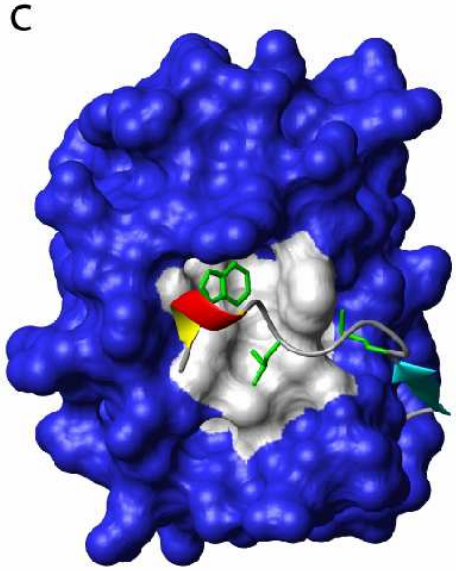


Figure 7