Biochem. J. (2013) **450**, 321–332 (Printed in Great Britain) doi:10.1042/BJ20121635

online data

High-resolution NMR studies of structure and dynamics of human ERp27 indicate extensive interdomain flexibility

Nader T. AMIN*[1], A. Katrine WALLIS*[2], Stephen A. WELLS†[3], Michelle L. ROWE‡, Richard A. WILLIAMSON‡, Mark J. HOWARD‡[4] and Robert B. FREEDMAN*[4]

*School of Life Sciences, University of Warwick, Coventry CV4 7AL, U.K., †Department of Physics, University of Warwick, Coventry CV4 7AL, U.K., and ‡School of Biosciences, University of Kent, Canterbury, Kent CT2 7NJ, U.K.

ERp27 (endoplasmic reticulum protein 27.7 kDa) is a homologue of PDI (protein disulfide-isomerase) localized to the endoplasmic reticulum. ERp27 is predicted to consist of two thioredoxin-fold domains homologous with the non-catalytic **b** and **b′** domains of PDI. The structure in solution of the N-terminal **b**-like domain of ERp27 was solved using high-resolution NMR data. The structure confirms that it has the thioredoxin fold and that ERp27 is a member of the PDI family. ^{15}N-NMR relaxation data were obtained and ModelFree analysis highlighted limited exchange contributions and slow internal motions, and indicated that the domain has an average order parameter S^2 of 0.79. Comparison of the single-domain structure determined in the present study with the equivalent domain within full-length ERp27, determined independently by X-ray diffraction, indicated very close agreement. The domain interface inferred from NMR data in solution was much more extensive than that observed in the X-ray structure, suggesting that the domains flex independently and that crystallization selects one specific interdomain orientation. This led us to apply a new rapid method to simulate the flexibility of the full-length protein, establishing that the domains show considerable freedom to flex (tilt and twist) about the interdomain linker, consistent with the NMR data.

Key words: endoplasmic reticulum, endoplasmic reticulum protein 27.7 kDa (ERp27), NMR, protein disulfide-isomerase (PDI), protein dynamics, thioredoxin-fold.

INTRODUCTION

The lumen of the ER (endoplasmic reticulum) is the subcellular compartment responsible for folding and quality control of proteins exported to the cell surface or extracellular spaces. This compartment contains many resident proteins which carry out these folding and quality assurance functions. One prominent group of such resident proteins is the PDI (protein disulfide-isomerase) family of proteins [1–3]. There are 20 known members of the human PDI family and their best-known role is in facilitating formation of the correct pairing of disulfide bonds, a key aspect of the folding process for most extracellular or cell-surface proteins. Proteins of the PDI family have a range of functions in addition to disulfide bond isomerization, and not all members of the family are catalytically active. All proteins of the PDI family have, or are thought to have, at least one domain with significant structural similarity to Trx (thioredoxin). These Trx-like domains have a mixed β/α fold with the typical secondary structure topology β_1-α_1-β_2-α_2-β_3-α_3-β_4-β_5-α_4 [3]. Those proteins which are catalytically active in disulfide bond isomerization act through a conserved Cys-Xaa-Xaa-Cys active-site motif in a Trx-like domain. This motif is located in the N-terminus of α_2 and incorporates the loop immediately preceding this helix. There is typically also a proline residue in the *cis* conformation in the loop between α_3 and β_4. This loop is adjacent to the active site in the tertiary structure and the *cis*-proline residue is thought to play a role in substrate binding [4]. Those members of the PDI family which are not catalytically active possess at least one Trx-like domain which lacks the active-site motif and may or may not possess the conserved *cis*-proline residue.

PDI is composed of four Trx-like domains, **a**, **b**, **b′** and **a′**, with an acidic α-helical extension on the C-terminus [2,5]. The **a** and **a′** domains contain the catalytic sites involved in thiol-disulfide oxidoreduction and show high sequence similarity to one another and to Trx. The **b** and **b′** domains are non-catalytic; the **b** domain is of unknown function, whereas the **b′** domain plays a major role in substrate binding [6]. This domain contains a hydrophobic pocket which is thought to act as the principal substrate-binding site on the protein [7,8], with the other domains making smaller contributions to substrate binding.

ERp27 is a 27.7 kDa ER-resident protein of unknown function and a putative member of the PDI family [9]. On the basis of sequence analysis, ERp27 is thought to be composed of two non-catalytic PDI-like domains, **b** and **b′**. The protein contains two cysteine residues, one in each domain. These cysteine residues have been shown to exist as free thiols and to be buried within each domain [9]. The protein is thus unlikely to be involved in thiol-disulfide oxidoreduction. The C-terminal domain of ERp27 shows sequence similarity to the **b′** domain of PDI, and the similarity includes residues involved in the substrate-binding site of PDI [7,8], suggesting that ERp27 may possess a similar binding site. Indeed, this domain of ERp27 has been shown to bind the 14-residue peptide Δ-somatostatin [9]. ERp27 has also been shown to interact with ERp57, another member of the PDI family,

Abbreviations used: 2D, two-dimensional; 3D, three-dimensional; ER, endoplasmic reticulum; ERp27, ER protein 27.7 kDa; HSQC, heteronuclear single-quantum coherence; NOE, nuclear Overhauser effect; PDI, protein disulfide-isomerase; RMSD, root mean square deviation; Trx, thioredoxin.

[1] Present address: Astex Therapeutics, 436 Cambridge Science Park, Cambridge CB4 0QA, U.K.
[2] Present address: Faculty of Health and Life Sciences, Conventry University, Richard Crossman Building, Coventry CV1 5FB, U.K.
[3] Present address: Department of Physics, University of Bath, Claverton Down, Bath BA2 7AY, U.K.
[4] Correspondence may be addressed to either of these authors (email m.j.howard@kent.ac.uk or r.b.freedman@warwick.ac.uk).
The structural co-ordinates for the **b** domain of human ERp27 reported will appear in the PDB under accession code 2L4C.

both *in vitro* and *in vivo* [9]. The binding site for ERp57 has been mapped to the C-terminal (**b′**) domain of ERp27.

In the present paper we describe the solution structure and backbone dynamics of the N-terminal (**b**) domain of human ERp27. Structural data have been deposited in the Protein Data Bank (PDB code 2L4C). NMR relaxation data were collected for this domain, indicating regions of increased mobility. A comparison of the HSQC (heteronuclear single-quantum coherence) spectrum for this isolated domain and that for full-length ERp27 identified a large contiguous surface on the **b** domain affected by connection to the **b′** domain. This putative interface region was considerably more extensive than that found in the X-ray structure of full-length ERp27 (PDB code 4F9Z, H. Schindelin and F.-X. Kober) and stimulated an exploration of the flexibility of ERp27 about the interdomain linker. We found that the domains show great interdomain flexibility making extensive transient interdomain contacts, consistent with the NMR data. This implies that crystallization selected a specific interdomain orientation from an ensemble of possible orientations.

EXPERIMENTAL

Protein expression and purification

Mature human ERp27 (Glu^{26}–Leu^{273}) and the individual **b** domain (Glu^{26}–Leu^{141}) were expressed in *Escherichia coli* BL21 (DE3) pLysS cells with an N-terminal His tag (MHHHHHHM) as described previously [9]. ^{15}N- and $^{15}N/^{13}C$-labelled protein was expressed in M9 medium containing [^{15}N]ammonium chloride (1 g/l) or [^{15}N]ammonium chloride (1 g/l) and [^{13}C]glucose (2 g/l). The recombinant protein was purified by immobilized metal-affinity chromatography followed by ion-exchange chromatography. Purified protein was concentrated and buffer-exchanged into 25 mM monosodium phosphate and 100 mM sodium chloride (pH 6.5). Sodium azide (0.01 %) was added as a preservative.

Multidimensional NMR spectroscopy

Protein samples were prepared to a concentration of 1–1.5 mM for structural determination experiments and 0.5 mM for relaxation studies in 25 mM monosodium phosphate and 100 mM sodium chloride (pH 6.5). The samples were placed in 5 mm Shigemi BMS-005V tubes with 10 % ^2H_2O. All NMR experiments were carried out on four-channel Varian Unity INOVA NMR spectrometers with ^1H resonance frequencies of 600 MHz, at the University of Kent, or 800 MHz at the Medical Research Council National Institute for Medical Research, The Ridgeway, Mill Hill, London, U.K. All spectra were acquired at 25 °C. Referencing for the ^1H chemical shift was based on the position of the water ^1H resonance, according to its known relationship with temperature [10] and $^{13}C/^{15}N$ referencing was achieved using gamma ratios. NMR experiments used WATERGATE [11] for solvent suppression to attenuate the ^1H water signal. Indirect dimensions were acquired using the hypercomplex method [12]. Data were processed on a Linux platform (SuSe 10.0) using NMRPipe [13], and analysed using the CcpNmr Analysis package [14].

NMR resonance assignment

^1H-^{15}N HSQC, CBCA(CO)NH and CBCANH standard triple-resonance NMR experiments were used to assign the backbone Cα, Cβ and amide ^{15}N and ^1H resonances. ^{15}N-edited TOCSY-HSQC and HCCH-TOCSY experiments, with mixing times of 60 ms and 18 ms respectively, were used to assign the aliphatic side-chain ^1H and ^{13}C resonances. Where necessary, 3D (three-dimensional) ^{15}N- and ^{13}C-edited NOESY spectra were used to aid in the assignments. Side-chain NH_2 resonances of asparagine and glutamine were assigned using the ^1H-^{15}N HSQC spectrum and a ^{15}N-edited NOESY spectrum. Aromatic side-chain ^1H and ^{13}C resonances were assigned using 2D (two-dimensional) ^1H-^1H TOCSY with 60 ms mixing time, 2D ^1H-^1H NOESY with 100 ms mixing time, 2D aromatic ^1H-^1H NOESY with 80 ms mixing time and a 2D selective aromatic ^1H-^{13}C HSQC spectrum.

^{15}N NMR relaxation measurements

^{15}N T_1 and T_2 experiments were acquired using developed pulse sequences comparable with those described elsewhere [15]. T_1 and T_2 delay times were set as 128, 256 (×2), 384, 512, 640 (×2), 769 and 897 ms, and 20, 40, 60 (×2), 80, 100, 120 (×2), 140 and 160 ms respectively. ^{15}N Heteronuclear NOE (nuclear Overhauser effect) experiments were collected with a relaxation delay of 5 s with and without saturation of the amide protons that was achieved using 120° high-power pulses [16]. Relaxation times were calculated as the exponential fit of single exponential decays to peak intensity values: $I = I_0\exp(-t/T_x)$ where $T_x = T_1$ or T_2, and I = resonance intensity at time t. Heteronuclear NOEs were calculated using the expression $\eta = I/I_0$. ModelFree analysis of the relaxation data was carried out using the ModelFree 4.0 suite of programs [17–19] to obtain the optimal parameter fits of S^2 (order parameter of motion), τ_e (internal correlation time) and R_{ex} (exchange broadening). All ModelFree optimization used the most appropriate model with the lowest number of parameters possible to provide an optimal fit to the relaxation data.

Structure calculations

NOE distance restraints were obtained from 3D ^{15}N-edited NOESY-HSQC and ^{13}C-edited NOESY-HSQC spectra. Both experiments were recorded at 800 MHz with a mixing time of 80 ms. Dihedral angle (ϕ/φ) restraints were generated from backbone chemical-shift data using TALOS [20]. Hydrogen-bond donors were determined from ^{15}N-HSQC experiments involving H/D (hydrogen/deuterium) exchange and hydrogen bond donor–acceptor pairs within secondary structure elements were identified from preliminary structure calculations. Structure calculations initially used only NOE restraints and were refined by subsequent inclusion of the dihedral angle and hydrogen-bond restraints. Structures were calculated using ARIA version 2.1 [21] and CNS version 1.2 [22]. For assignment of ambiguous NOESY cross-peaks with ARIA, the frequency window sizes were left at the default settings of 0.02 p.p.m. for direct and 0.04 p.p.m. for indirect proton dimensions, and 0.5 p.p.m. for heteronuclear dimensions. For each round of structure calculations, ten iterations were performed. For the first iteration, 50 structures were calculated and the 20 lowest-energy structures were used for calibration and violation analysis. For the next seven iterations, 20 structures were calculated, with the seven lowest-energy structures used for calibration and violation analysis. For the ninth iteration, 50 structures were calculated and the 20 lowest-energy structures were used for calibration and violation analysis. For the final (tenth) iteration, 100 structures were calculated. For iterations 0–9, the violation tolerances were set to 1000.0, 5.0, 3.0, 1.0, 1.0, 1.0, 0.1 0.1, 0.1 and 0.1 Å (1 Å = 0.1 nm) respectively and the partial assignment ambiguity cut-off values were set to 1.0, 0.9999, 0.999, 0.99, 0.98, 0.96, 0.93, 0.9, 0.85

and 0.8 respectively. For each iteration, the violation threshold was set to 0.5 Å and the maximum number of contributions for partial assignment was 20. Structure calculations with CNS used torsion angle dynamics. Each iteration used 18000 cooling steps in the simulated annealing protocol fixed in the ratio of 5:4 for the first and second cooling stages respectively.

Mapping of the b/b′ domain interface on the b domain of ERp27

Residues from the **b** domain of ERp27 potentially forming the interface with the **b′** domain were identified using the minimal shift mapping approach [23,24]. Minimal shift mapping provides a measure of the chemical-shift difference, Δ_{shift}, that corresponds to the minimal, or closest distance, for each **b** peak to any peak in an equivalent full-length ERp27 (**bb′**) spectrum. Therefore a peak in the spectrum of **b** with a corresponding peak in **bb′** in an identical position would score a Δ_{shift} of zero with the value of Δ_{shift} increasing as the **bb′** peak moves further away from being coincident to the **b** peak. These changes in chemical shift are indicative of a change in chemical environment in the full-length protein compared with the single-domain protein. Hence, a comparison of the ^{15}N-HSQC spectra for the **b** domain and the full-length protein provides identification of residues with NH chemical shifts that are different in the **b** construct compared with the **bb′** construct. For each peak in the HSQC spectrum for the **b** domain, the combined ^{1}H and ^{15}N chemical-shift difference was calculated for all peaks in the HSQC spectrum for the full-length protein. The following equation was used:

$$\Delta_{shift}=\sqrt{(\Delta^{1}H_{N})^{2}+(1/6\Delta^{15}N)^{2}}$$

where $\Delta^{1}H_{N}$ is the chemical-shift difference in the ^{1}H dimension and Δ^{15}N is the chemical-shift difference in the ^{15}N dimension. The Δ^{15}N shift difference was reduced by a factor of 6 to account for the difference in ^{15}N-HSQC amide chemical-shift ranges of approximately 30 p.p.m. and 5 p.p.m. for ^{15}N and ^{1}H respectively.

Simulation of the flexibility of full-length ERp27

The atomic co-ordinates of full-length ERp27 (PDB code 4F9Z) were kindly provided by F.-X. Kober and H. Schindelin in advance of publication and were used to simulate and represent the relative motion of the domains. Simulation was by a novel rapid method which analyses the protein structure as a network of rigid clusters and flexible linkers using the program FIRST [25], separately calculates the normal modes of motion of the structure by coarse-grained elastic network modelling using ElNemo [26] and then generates trajectories of motion for the network by geometric simulation using the FRODA algorithm [27] in a calculation in which the normal mode eigenvectors are used to bias the motion of the network [28]. This method of simulation of flexible motion retains an all-atom representation of the protein and uses a simplified physical model [29] to maintain the covalent bond geometry and the network of hydrogen bonds and hydrophobic tethers found in the input structure, and to enforce steric exclusion, while neglecting long-range interactions. This simplification allows for a systematic exploration of motion along multiple normal mode directions, achieving large amplitudes, using modest computational resources. The retention of all-atom detail and of the essential local covalent and noncovalent interactions, meanwhile, makes the generated structures physically realistic and useful for the interpretation of experimental data [30].

Normal modes of motion were generated for both full-length ERp27 (molecule B) and for the central **bb′** domains of full-length yeast PDI (from PDB code 2B5E). For each structure, trajectories of flexible motion were generated using the ten lowest-frequency non-trivial normal modes; these are modes 7–16, as modes 1–6 are combinations of trivial rigid body motions. Biases were applied both parallel and antiparallel to the mode eigenvectors generating two opposed trajectories for each normal mode, indicated as $7-$ and $7+$. The bias step size in each iteration of the simulation was 0.01 Å. The constraint network included all hydrophobic tethers identified by FIRST in the input structures. Hydrogen bonds identified in the input structure are assigned a (negative) energy in FIRST on the basis of their geometry, and the set of bonds retained in the simulations is determined by setting a cut-off value. Simulations were carried out using cut-off values of -1, -2 and -3 kcal/mol (1 kcal = 4.184 kJ), with essentially identical results. Results are presented for the cut-off of -2 kcal/mol. In total 2000 conformations were generated along each trajectory with every 100th conformation being recorded as a PDB file for subsequent analysis. All simulations were carried out overnight on a dual-core 2.4 GHz desktop workstation running Linux.

In this case our principal interest was in the relative orientation of the two domains, and the variation of this orientation in the course of large-scale flexible motion. To represent the relative orientations of the two domains, a plane was calculated as representative of the β-sheet plane in each domain, by selecting the Cα atoms of four 'central' residues, namely alternating central residues in each of the adjacent antiparallel strands β2 and β4. These four atoms in each domain give a quadrilateral, from which the plane normals and interplane axis are calculated. The 'tilt' angle between adjacent domains (θ) is defined using the scalar product of the vectors corresponding to the respective plane normals. The dihedral 'twist' angle (ω) makes use of the vector between the 'average' position of the four selected Cα atoms in each domain. This vector and the plane normal for each domain define a plane; the dihedral angle between these two planes is the 'twist' angle (ω). Tilt and twist values were extracted for the input structures and for the structures generated in the simulations of flexible motion.

RESULTS

Resonance assignment of the ERp27 b domain

Excluding the eight residues of the His tag (MHHHHHHM), chemical-shift assignments were made for 100 % of the backbone amide ^{15}N and ^{1}H resonances, 100 % of the ^{13}Cα and ^{13}Cβ resonances, 96.7 % of the ^{1}Hα resonances, 94.5 % of the ^{1}Hβ resonances, 88.8 % of the other side-chain ^{1}H resonances (^{1}Hγ, ^{1}Hδ, ^{1}Hε, ^{1}Hζ and ^{1}Hη), 70.6 % of the side-chain ^{13}C resonances (^{13}Cγ, ^{13}Cδ, ^{13}Cε, ^{13}Cζ and ^{13}Cη) and 26.1 % of the side-chain ^{15}N resonances (^{15}Nδ, ^{15}Nε, ^{15}Nζ and ^{15}Nη). Six residues had no side-chain assignments other than ^{13}Cβ; Glu^{26}, Glu^{28}, Ser^{30}, Ser^{31}, Glu^{122} and Glu^{125}, of which four (Glu^{26}, Glu^{28}, Ser^{30} and Ser^{31}) are located at the N-terminus and, following structure determination, were found to be in a flexible extension outside of the core Trx fold (see below). The assigned backbone amide resonances are shown in the 2D ^{15}N-^{1}H HSQC spectrum in Figure 1. The well-dispersed pattern of peaks indicates that the protein is folded. A total of 130 HSQC peaks were identified. Of these, 18 were identified as asparagine and glutamine side-chain peaks, one was assigned to the indole amide of the single tryptophan residue, one to the His-tag backbone and the remainder were assigned to the backbone amides of the 110 non-proline residues.

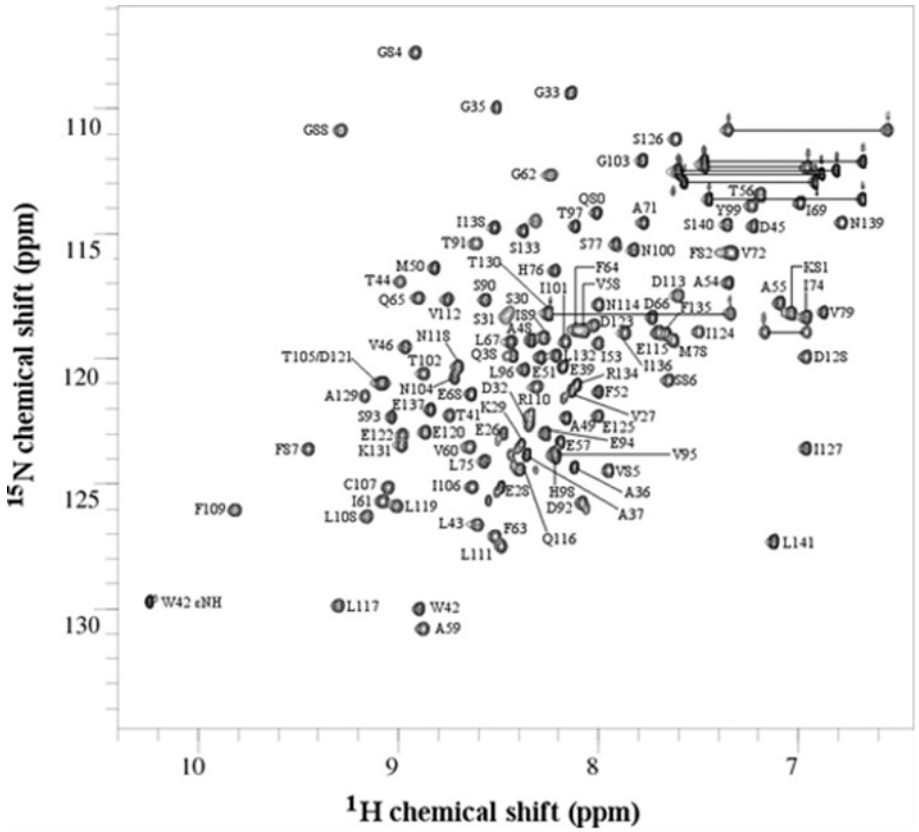

Figure 1 2D ^{15}N-^{1}H HSQC spectrum of the ERp27 b domain

Peaks are annotated to show the backbone amide assignments. All 110 non-proline residues have been assigned. The indole ε1 amide of the single tryptophan residue (W42) has also been assigned. Side-chain amide NH_2 resonances are connected with horizontal lines.

NMR relaxation analysis

^{15}N-NMR relaxation parameters T_1, T_2 and heteronuclear NOE were obtained for the **b** domain of ERp27 at 14.1 T (600 MHz 1H) and 25 °C (Figure 2). The average values for ^{15}N T_1 and T_2 were recorded as 704 ± 46 ms and 122 ± 64 ms respectively; these data were used to estimate the global correlation time for the domain of 7.13 ± 0.71 ns using $\chi 2$ optimization of the theoretical dipolar and practical T_1/T_2 ratios. This estimate of global correlation time confirms that the domain is monomeric under the experimental conditions used. NMR relaxation parameters for each individual ^{15}N were interrogated further using ModelFree analysis which provides motional information in terms of order parameters S^2, internal correlation times τ_e and contributions to chemical exchange R_{ex}, for individual NH vectors across the domain backbone (Supplementary Figure S1 at http://www.biochemj.org/bj/450/bj4500321add.htm). An unstructured N-terminal region was identified from combined low S^2 and high τ_e values for residues 26–38. The average S^2 value across the structured region (residues 39–141) was 0.79 ± 0.04 where S^2 values of 1.0 support a completely rigid domain backbone and 0.01 support an entirely random flexible backbone.

This value of S^2 can be compared with those of oxidized and reduced ERp18 (0.81 and 0.90 respectively) [31] and implies a relatively flexible backbone which may provide capacity for diverse interactions with partner proteins. Average internal correlation times across the structured region were 85 ± 18 ps, a typical value for a well-behaved globular domain. Within the structured region, five residues were estimated to have τ_e values above 300 ps as seen in Supplementary Figure S1(B): Leu[43], Val[72], Val[95], Gly[103] and Leu[141]. These residues are all associated with loops or are the first residues of secondary structure elements and these higher τ_e are due to either depressed T_1 or elevated T_2 values as seen in Figures 2(A) and 2(B). These unusual τ_e values were all obtained with low errors and, according to ModelFree analysis, they indicate significant areas of the domain that have abnormally slow internal motions. Additional ModelFree contributions to R_{ex} were only observed for two residues, Leu[111] and Asp[113] within the loop between strands β4 and β5 (Supplementary Figure S1C), and are indicative of low T_2 observed for the residues in Figure 2(B).

Structure calculations

In the final round of structure calculations, 1939 NOE-derived distance restraints, 117 dihedral angle restraints and 42 hydrogen-bond restraints were used. In total 100 structures were generated, of which 92 had converged satisfactorily. The converged structures had no distance violations greater than 0.5 Å and no dihedral angle restraint violations greater than 5°. Of the 92 converged structures, the representative structure ensemble was chosen as the 50 lowest-energy structures; superimposed backbone tracings of these structures are shown in Figure 3(A). None of the structures in the ensemble had NOE or hydrogen-bond restraint violations greater than 0.3 Å. The NMR restraints and structural statistics are summarized in Table 1. NOE distribution, structural calculations and NMR relaxation data analysis confirmed that the structured

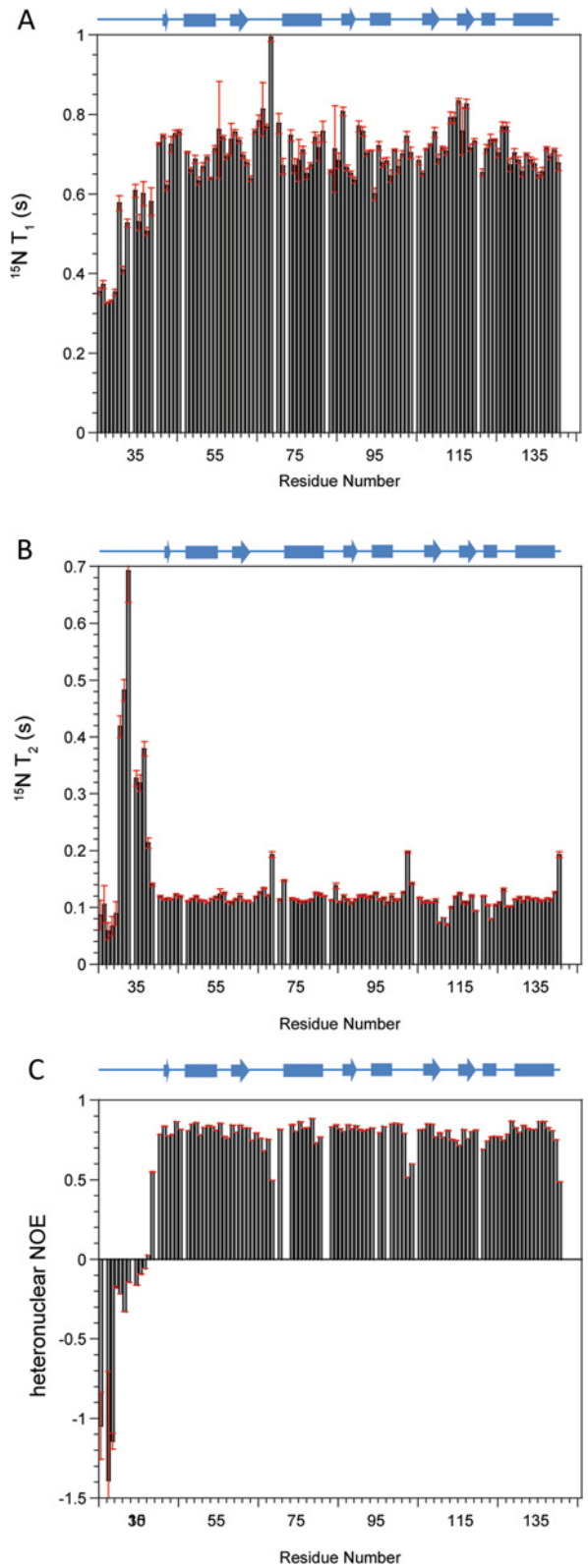

Figure 2 ^{15}N NMR relaxation parameters of the ERp27 b domain

Spin–lattice relaxation times T_1 (**A**), spin–spin relaxation times T_2 (**B**) and ^{1}H-^{15}N heteronuclear NOE data (**C**) were collected at 14.1 T and 25 °C and are plotted as a function of residue number. A schematic diagram of the domain secondary structure as solved by NMR is shown above each plot. Cylinders represent α-helices and arrows represent β-strands.

region of the **b** domain extends from Glu^{39} to Leu^{141}. Figure 3(A) illustrates the diversity of conformations of the N-terminal region in the individual structures and the structural limits of the domain are also confirmed by comparing the RMSD (root mean square deviation) values between residues 26 and 141 and those between residues 39 and 141 (Table 1). Description and analysis of the structure below is restricted to this structured region (Glu^{39}–Leu^{141}) and is based on the individual structure in the ensemble closest to the mean structure; this is illustrated in Figures 3(B) and 3(C) showing the assigned secondary structure.

The ordered region of the protein (residues 39–141) consists of a five-stranded mixed β-sheet surrounded by five α-helices (Figure 3). The secondary structure topology is β_1-α_1-β_2-α_2-β_3-α_3-β_4-β_5-α_4-α_5. Strands 1, 2 and 3 are parallel with each other, whereas strands 2, 4 and 5 are antiparallel. This corresponds to the variant of the canonical Trx fold (SCOP reference number 52832) commonly found in the PDI family [3], but with an α-helical insert (α4) before the C-terminal α-helix of the fold (α5). The β_1 strand is short (two residues), but this is common in the family (e.g. in the various domains of yeast PDI [5] this strand comprises two, three, zero and four residues). This structural analysis is supported by chemical-shift data (Supplementary Figure S2 at http://www.biochemj.org/bj/450/bj4500321add.htm).

The structure ensemble was analysed using PROCHECK-NMR [32,33]. Over residues 39–141 of the ensemble, 74 % of the non-glycine dihedral angles were in the most-favoured regions and 26 % were in the additionally allowed regions. Gly^{103} was the only residue with dihedral angles in a generously allowed region of the plot. This residue is located on the loop between α3 and β4. This corresponds to the *cis*-proline loop found in many domains in the PDI family. Although there is no conserved proline residue in this loop in ERp27, Gly^{103} is in the equivalent position as the residue preceding the *cis*-proline in the catalytic PDI domains.

Structural homology

Structural homologues of the **b** domain of ERp27 were identified by a search of the PDB database using DALI [34]. In total 591 structure matches with a *Z*-score greater than 2.0 were found. Matches were detected to each of the five classes of Trx-fold proteins: Trx, glutaredoxin, glutathione transferase, DsbA and glutathione peroxidase. The highest-scoring matches were the structures of the **b** domains of ERp57 [35], human PDI [36], calsequestrin [37], ERp72 [38], yeast PDI [5], ERp29 [39] and ERp44 [40]. ERp27 **b** also showed significant structural similarity to Trx despite lacking the active-site residues and the conserved *cis*-proline residue. Ten out of the 20 highest-scoring matches were to Trxs.

The structure of ERp27 **b** domain is notable in having an α-helical insert between β5 and α4 of the standard PDI fold. This helix is short, consisting of only four residues, and is situated adjacent in the tertiary structure to the loop which contains the active site (in **a**-type PDI domains). A similar non-canonical helix is found in the yeast PDI **b** domain [5]. However, the structure of yeast PDI **b** differs significantly from that of ERp27 **b** in that it does not contain α3 of the canonical fold. Both ERp57 **b** and ERp72 **b** differ from the structure of ERp27 **b** most significantly in the loop region between β4 and β5, which is extended in these two proteins. Structure superimpositions between ERp27 **b** and selected homologous proteins are shown in Figure 4. A structure-based multiple sequence alignment between ERp27 **b** and selected homologues is shown in Supplementary Figure S3 (at http://www.biochemj.org/bj/450/bj4500321add.htm).

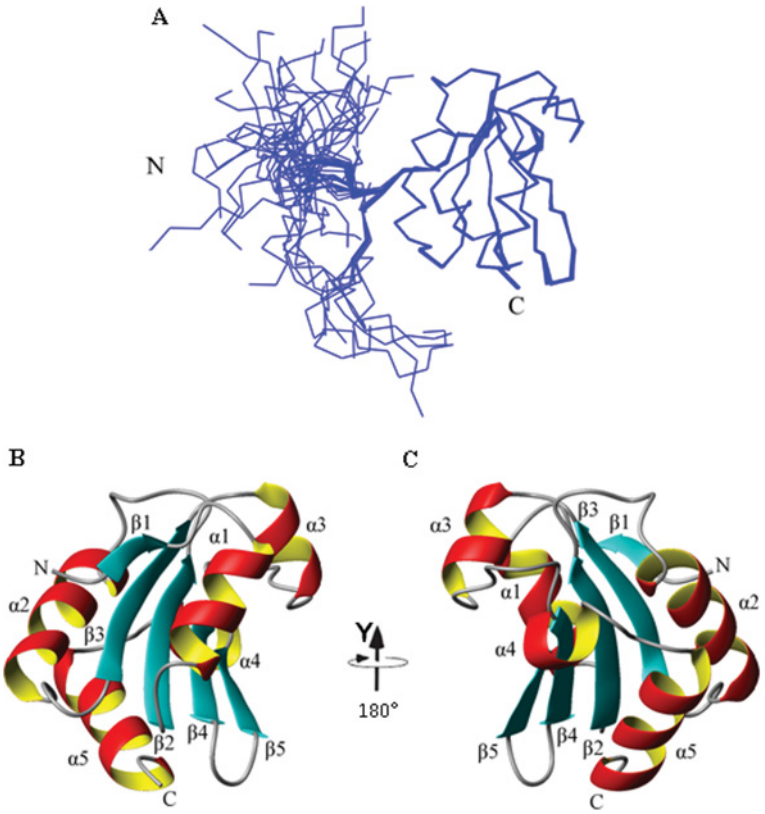

Figure 3 Solution structure of the ERp27 b domain

(**A**) Superimposition of the ensemble of 50 representative structures, residues 26–141. Each structural model in the ensemble was superimposed over the mean structure for the ordered region 39–141. The mean structure was calculated using MOLMOL [46]. The Cα traces are shown. The positions of the N- and C-termini are labelled N and C respectively. (**B**) and (**C**) show the single representative structure of ERp27 **b**, residues 39–141 and are related by a 180° rotation about the *y*-axis. Secondary structure elements are labelled.

Sequence conservation

The sequence of ERp27 has only been found in vertebrate genomes. The sequence alignment in Supplementary Figure S4 (at http://www.biochemj.org/bj/450/bj4500321add.htm) shows a range of ERp27 **b** sequences from mammalian, bird, reptile and fish genomes. The sequence of ERp27 **b** is well conserved across species. The majority of conserved residues are located in the hydrophobic core of the protein, predominantly on β-strands 2, 3 and 4 and on the interior faces of the surrounding α-helices. In the alignment in Supplementary Figure S4 there are 13 positions that are invariant between ERp27 sequences. These correspond to Leu^{43}, Glu^{57}, Val^{58}, Gly^{62}, Phe^{63}, Glu^{68}, Val^{85}, Val^{95}, Leu^{108}, Phe^{109}, Arg^{110}, Asp^{113} and Phe^{135} of human ERp27. Of these residues, nine (Leu^{43}, Glu^{58}, Val^{58}, Glu^{68}, Val^{85}, Phe^{109}, Arg^{110}, Asp^{113} and Phe^{135}) are surface-exposed. Gly^{62} and Phe^{63} are buried within the hydrophobic core of the protein and are situated at the C-terminal end of β2. These two residues are highly conserved in all **b** domains of the human PDI family (Supplementary Figure S3), with the exception of ERp29. Glu^{68} is located on the loop between β2 and α2. This is the location of the active site in the catalytically active domains of the PDI family. Glu^{57} and Val^{58} are situated at the N-terminus of β2, Arg^{110} is situated at the C-terminus of β4 and Asp^{113} is situated on the loop between β4 and β5. These residues are surface-exposed on a common face of the protein. The side chains of Arg^{110} and Asp^{113} are positioned so as to form a salt bridge. These residues are also present in the **b** domains of human PDI, ERp57 and ERp29, although in human PDI the arginine residue is replaced by a lysine (Supplementary Figure S3).

The b to b′ domain interface on the b domain

^{1}H-^{15}N HSQC spectra have been collected for both full-length ERp27 and the individual **b** domain (Supplementary Figure S5 at http://www.biochemj.org/bj/450/bj4500321add.htm). Signals in these spectra are highly sensitive to the local chemical environment of the backbone amide groups. A comparison of these spectra may therefore allow identification of those residues in **b** which are in contact with **b′** in full-length ERp27. Residues in **b** potentially forming the interface between **b** and **b′** in full-length ERp27 were therefore identified using the minimal shift approach [23,24]. The distance between each assigned cross-peak in the spectrum for the **b** domain to the nearest cross-peak in the spectrum for the full-length protein was measured. The minimal shift values per residue are plotted in the histogram in Figure 5(A). Residues 136–141 at the C-terminus of the **b** domain are inevitably close to the subsequent **b′** domain; omitting these from further consideration, the average minimal shift of the remaining residues was 0.0333 ± 0.0292. Figure 5(A) shows that the residues with large minimal shifts are clustered in three regions of sequence, namely residues 53–65, 80–87 and 110–119; within these regions, 19 out of 31 residues have minimal shifts >1 S.D. above the mean and most of the others have shifts above the mean. To interpret the shifts in relation to structure, all residues

Table 1 Structural statistics and RMSD values for the ERp27 b domain ensemble (50 structures)

For (A), (B), (D) and (E), values are reported ± S.D. (A) The energy was obtained using GROMOS96 within DEEPVIEW. (B) The RMSD to the mean structure was calculated using MOLMOL [46]. The mean structure was calculated independently for residues 26–141 and residues 39–141. The RMSD over the secondary structure regions used the mean structure calculated for residues 39–141. The secondary structure was as assigned using DSSP-CONT. (C) The total number of restraints was calculated using the total number of restrained hydrogen bonds, rather than the total number of hydrogen-bond restraints. This is because each restrained hydrogen bond had a set of two distance restraints. (D) The average number of all restraint violations (NOE, hydrogen bond and dihedral restraint violations) was calculated for all NOE and hydrogen bond violations greater than 0.1 Å and all dihedral angle violations greater than 5°. The data in (D) and (E) were generated using CNS. (F) Procheck Ramachandran plot statistics calculated for non-glycine residues from 39 to 141 for the ensemble of 50 structures.

Parameter	Measurement
(A) Average energy (kJ/mol)	-2165 ± 121
(B) Average RMSD to mean structure (Å)	
Residues 26–141	
Backbone	3.27 ± 0.82
Heavy	3.48 ± 0.81
Residues 39–141	
Backbone	0.10 ± 0.02
Heavy	0.54 ± 0.03
Secondary structure	
Backbone	0.09 ± 0.02
Heavy	0.52 ± 0.03
(C) Number of restraints	
Total	2098
Average per residue	18.1
NOE	
Total	1939
Ambiguous	1059
Unambiguous	880
Intra ($i-j=0$)	296 (33.6 %)
Sequential ($\|i-j\|=1$)	232 (26.4 %)
Medium-range ($1<\|i-j\|<5$)	146 (16.6 %)
Long-range ($\|i-j\|>5$)	206 (23.4 %)
Hydrogen bond	
Total constrained	42
Total constraints	84
Long-range constraints ($\|i-j\|>5$)	38
Dihedral angle (φ/ψ)	
Total	117
(D) Restraint violations (mean values per structure)	
All restraints	16.0 ± 1.7
NOE	
Violations>0.5 Å	0
Violations>0.3 Å	0
Violations>0.1 Å	13.1 ± 1.7
Hydrogen bond	
Violations>0.5 Å	0
Violations>0.3 Å	0
Violations>0.1 Å	2.9 ± 0.3
Dihedral angle	
Violations>5°	0
(E) RMSD from experimental restraints	
NOEs (Å)	0.020
Unambiguous (Å)	0.018 ± 0.001
Ambiguous (Å)	0.022
Hydrogen bonds (Å)	0.031 ± 0.001
Dihedral angles (°)	0.110 ± 0.015
(F) Procheck statistics (residues 39–141, 50 structures)	
Most favoured regions	3403 (74 %)
Additional allowed regions	1197 (26 %)
Generously allowed regions	0
Disallowed regions	0

within these regions were included in a map of the possible interdomain surface; these were plotted on to the domain structure in Figure 5(B), which shows clearly that the regions identified in this way are located on one face of the protein. Similar regions of the fold make up the **b** to **b′** domain interfaces in PDI, ERp57, ERp44, calsequestrin and ERp72 [5,35,37,38,40]. Calculation of the electrostatic potential shows the postulated interface region to be predominantly hydrophobic with charged residues around the edge (Figure 5C). There are patches of negative charge created, for example, by the surface-exposed side chains of Glu^{57}, Asp^{113} and Glu^{137}. There is considerable sequence conservation on the surface of the postulated **b** to **b′** interface (Figure 5D). In particular, Glu^{57}, Val^{58}, Val^{85}, Arg^{110}, Asp^{113} and Phe^{135} are strictly conserved in ERp27 (Supplementary Figure S4), whereas Ala^{59}, Val^{112}, Asn^{139} and Leu^{141} in the interface are conserved in over 80 % of ERp27 sequences.

We were provided with the atomic co-ordinates of an X-ray structure of full-length ERp27 in advance of publication (PDB code 4F9Z) allowing a direct comparison of our inferred interface with that determined by X-ray diffraction.

The NMR minimal shift data in the present paper predict that the interdomain interface is homologous with that between the **b** and **b′** domains in other multidomain members of the PDI family and this is broadly confirmed by the X-ray structure of the full-length protein, although the X-ray structure indicates an unusual interdomain angle. For a precise comparison, we selected 'domain contact residues' on the **b** domain by choosing the residues showing the greatest 'minimal shifts' in ^{1}H-^{15}N resonances in the NMR data and the closest interdomain atom contacts in the crystal structure data, excluding the residues at the C-terminus of the **b** domain (residues 136–141) which are inevitably in contact with or close to the **b′** domain. Both approaches highlighted the same three regions on the **b** domain: $I^{53}AATEVAVIGFFQ^{65}$, $Q^{80}KFPGVSF^{87}$ and $R^{110}LVDNEQLNL^{119}$. Within these regions, eight residues were identified as contact residues by both criteria (Ile^{53}, Ala^{54}, Ala^{55}, Glu^{57}, Gly^{84}, Leu^{111}, Val^{112} and Asn^{114}), 11 were identified by the NMR criterion only (Ala^{59}, Ile^{61}, Gln^{65}, Gln^{80}, Val^{85}, Ser^{86}, Phe^{87}, Arg^{110}, Gln^{116}, Leu^{117} and Leu^{119}), and four by the X-ray criterion only (Thr^{56}, Val^{58}, Asp^{113} and Glu^{115}). Hence the contact surface highlighted by NMR is more extensive than that indicated in the crystal structure.

The X-ray data represent a static structure whereas shifts in NMR resonances could represent indirect effects, but will certainly also reflect perturbations that arise from solution dynamics. We therefore attempted to explore the potential flexibility of the molecule using a new rapid method [28].

Simulation of the flexibility of full-length ERp27 reconciles NMR and X-ray analyses of the interdomain interface

Examination of the trajectories of flexible motion showed that, for both ERp27 and PDI **bb′**, the largest amplitudes of flexible motion are achieved by the three lowest-frequency non-trivial normal modes (modes 7, 8 and 9), which are combinations of rotation about the interdomain axis, and tilting around the interdomain interface. Further discussion of results is focussed on these modes. To analyse these motions, we have defined the planes represented by the core β-sheet in each domain and extracted the 'tilt' and 'twist' angles between these planes (see the Materials and methods section). Figure 6(A) is a 'tilt/twist' plot of the interdomain orientations and confirms that the relative orientation of the domains in ERp27 is distinct from that in yeast PDI (arrows); it shows that the two structures differ in twist angle as well as in tilt angle (ERp27, tilt $+34^{\circ}$ and twist -33°; compared with PDI **bb′**, tilt $+51^{\circ}$ and twist $+53^{\circ}$).

Trajectories of flexible motion are shown by calculating and plotting tilt and twist angles for every 100th structure in a

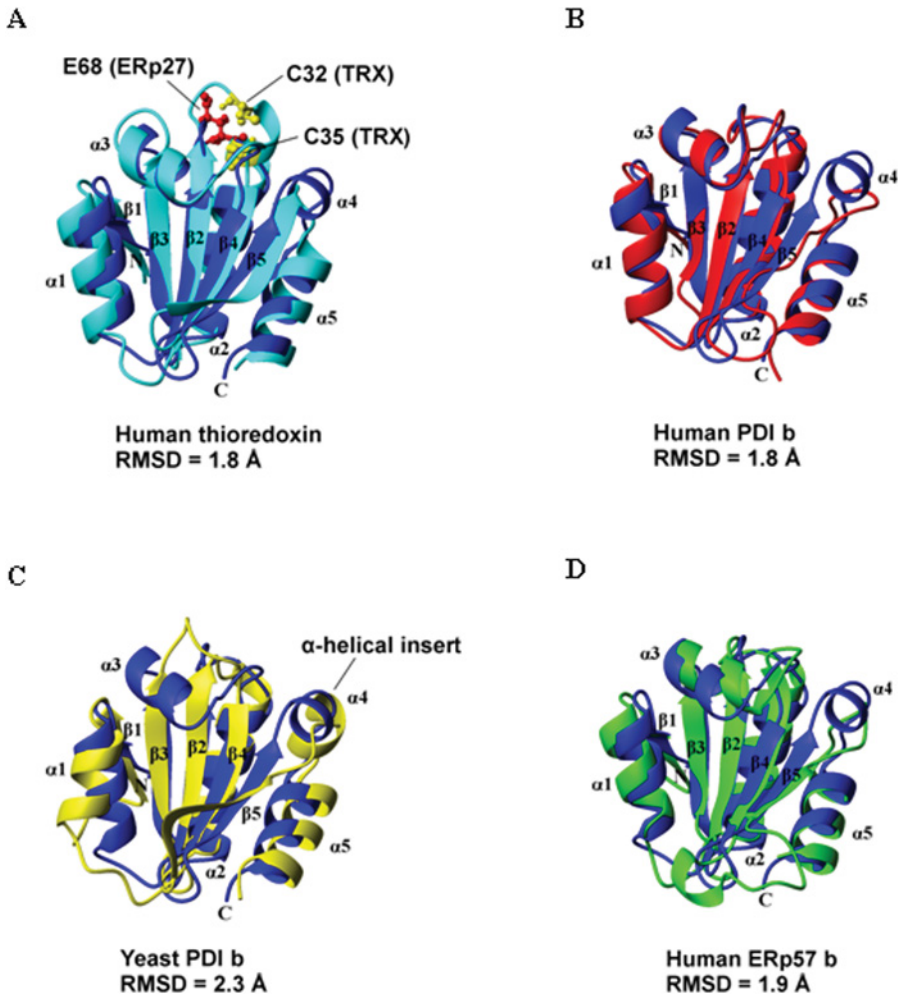

Figure 4 Structure superimpositions between ERp27 b and homologous proteins

Structures were aligned using DALI [34]. The single structure of ERp27 **b** with the lowest backbone RMSD to the mean structure over residues 39–141 was used. The structure of ERp27 **b** is shown in blue. The secondary structure elements of ERp27 **b** are labelled. The N- and C-termini of ERp27 **b** are labelled N and C. ERp27 **b** is aligned with (**A**) human Trx (PDB code 1ERT, residues 1–105, RMSD = 1.8 Å over 95 residues), (**B**) human PDI **b** domain (PDB code 2BJX, residues 119–216, RMSD = 1.8 Å over 98 residues), (**C**) yeast PDI **b** domain (PDB code 2B5E, residues 142–237, RMSD = 2.3 Å over 90 residues) and (**D**) human ERp57 **b′** domain (PDB code 2H8L, residues 135–240, RMSD = 1.9 Å over 97 residues). In (**A**) the catalytic cysteine residues of Trx are displayed in ball-and-stick representation. Glu^{68} of ERp27 **b** is also shown. This residue is located in the active-site loop and is strictly conserved in ERp27. In (**C**) the α-helical insertion in ERp27 **b** (α4) and yeast PDI **b** is indicated.

simulated trajectory. In the case of ERp27, modes 8 and 9 correspond to tilting motions with the interdomain link acting as a hinge (as shown by a clear change in tilt angle but a very limited change in twist angle); PDI **bb′** shows similar behaviour in modes 7 and 9 although the change in twist angle is slightly greater. For ERp27, mode 7, the lowest frequency non-trivial normal mode, shows the most striking results; this mode generates a pronounced twist, effectively a counter-rotation of the two domains around the axis joining the two domains. A similar motion in PDI **bb′** is generated by its normal mode 8. What is notable is that this flexibility is much greater for ERp27 (a total amplitude of twist from -120^{o} to $+60^{o}$) than for PDI **bb′**, and that the extent of this motion for ERp27 allows it to take up a relative orientation very close to that found initially in PDI **bb′**. This reorientation of ERp27 to a more PDI-like structure is illustrated in Figures 6(B) and 6(C). Furthermore, in the course of facile motion along modes 7, 8 and 9, many residues in the loops identified by NMR as 'contact loops' move to make new contacts with the **b′** domain (Figures 6D–6F). Using as the contact criterion that any backbone atom of a residue in the **b** domain is at a distance <7 Å from any non-hydrogen atom in the **b′** domain, we find that Ile^{53}–Val^{58}, Gly^{84}–Val^{85} and Leu^{111}–Glu^{115} initially make contact with the **b′** domain in at least one of the five molecules of ERp27 in the crystal unit cell. Flexible motion along the trajectories of modes 7–9 brings about new contacts, and almost all of the residues identified as being 'contact' residues by the NMR criterion are now seen to make contact with the **b′** domain. Flexible twisting motion along the mode 7− trajectory (towards the PDI **bb′** conformation) generates contacts between the **b′** domain residues Ala^{59}, Phe^{82}, Pro^{83}, Ser^{86} and Arg^{110}, whereas motion along trajectory 7+ generates new contacts with Pro^{83}, Phe^{87}, Arg^{110}, Gln^{116}, Leu^{117} and Asn^{118}. Similarly, tilting motion along mode 8− generates new contacts with Ala^{59}, Val^{60}, Gln^{80}, Lys^{81}, Phe^{82}, Pro^{83}, Ser^{86}, Phe^{87} and Arg^{110} and that along mode 8+ generates new contacts with Arg^{110}, Gln^{116}, Leu^{117} and Asn^{118}. New contacts made by motion along trajectories 9− and 9+ are all drawn from the same set of residues. Of the 11 non-proline residues in this set, four (Ala^{59}, Gln^{80}, Ser^{86} and Leu^{117}) are the residues showing the greatest resonance shifts in the NMR analysis (Figure 5A) and these are four out of five residues showing shifts of greater

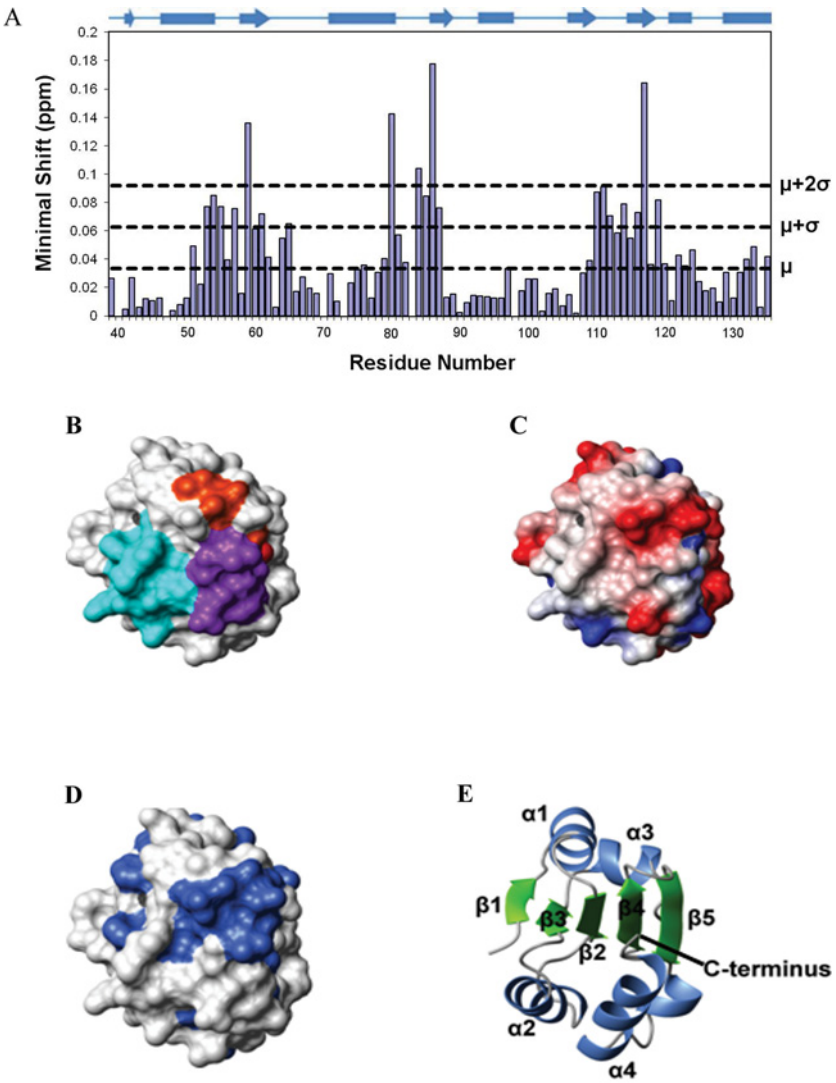

Figure 5 Mapping of the b to b′ domain interface on the b domain of ERp27

(**A**) Histogram showing the minimal chemical shifts per residue for ERp27 **b** (residues 39–135). The secondary structure is indicated above the histogram. Cylinders represent α-helices and arrows represent β-strands. The minimal shift values corresponding to the mean (μ), mean + 1 S.D. ($\mu + \sigma$) and mean + 2 S.D. ($\mu + 2\sigma$) are marked by horizontal lines. (**B**–**D**) The contact molecular surface generated using MOLMOL with default settings [46]. In (**B**), residues are coloured according to the regions of structure showing the highest minimal shift values; residues 53–65 are coloured cyan, residues 80–87 are coloured orange and residues 110–119 are coloured purple. In (**C**), the electrostatic potential generated using MOLMOL is shown; blue represents positive charge, red represents negative charge and white represents neutral charge. In (**D**), sequence conservation is mapped on to the structure; residues with >80 % identity (in the alignment in Supplementary Figure S3 at http://www.biochemj.org/bj/450/bj4500321add.htm) are coloured. (**E**) Cartoon representation of the protein backbone showing the orientation of the protein used for (**B**–**D**).

than the mean + 2 S.D., whereas a further three of the 'new contact' set (Phe[87], Arg[110] and Gln[116]) have shifts greater than the mean + S.D. Hence our simulations of flexible motion in ERp27 reconcile the apparent discrepancy between the interdomain contact region identified by NMR and that identified in the crystal structure

DISCUSSION

The N-terminal (**b**) domain of ERp27 was expressed in high yield, was highly soluble and very stable in solution, allowing the collection of high-quality 2D and 3D NMR data. By contrast, the C-terminal domain (**b′**) could not be expressed in isolation as a soluble protein and we did not succeed in strategies aimed at deriving it from cleavage of a fusion construct. Full-length ERp27 was readily expressed and gave a well-dispersed HSQC spectrum [9], but did not remain soluble through long-term NMR data collection at temperatures >30 °C, preventing the collection of usable 3D data sets for full assignment and structure calculations. It is now well-established that the ligand-binding **b′**-type domains of the PDI family are the most difficult to generate and study in isolation; high-resolution structures of these domains have only been determined in the presence either of an adjacent domain or of a C-terminal extension sequence which is capable of occupying the hydrophobic cleft which constitutes the substrate-binding site [41]. Without one or other of these extensions in place, the isolated **b′** domains have a tendency to oligomerize and aggregate [42], presumably as a result of the exposure of their hydrophobic ligand-binding sites.

The solution structure of the **b** domain reported in the present paper is highly congruent with the structure of this domain in

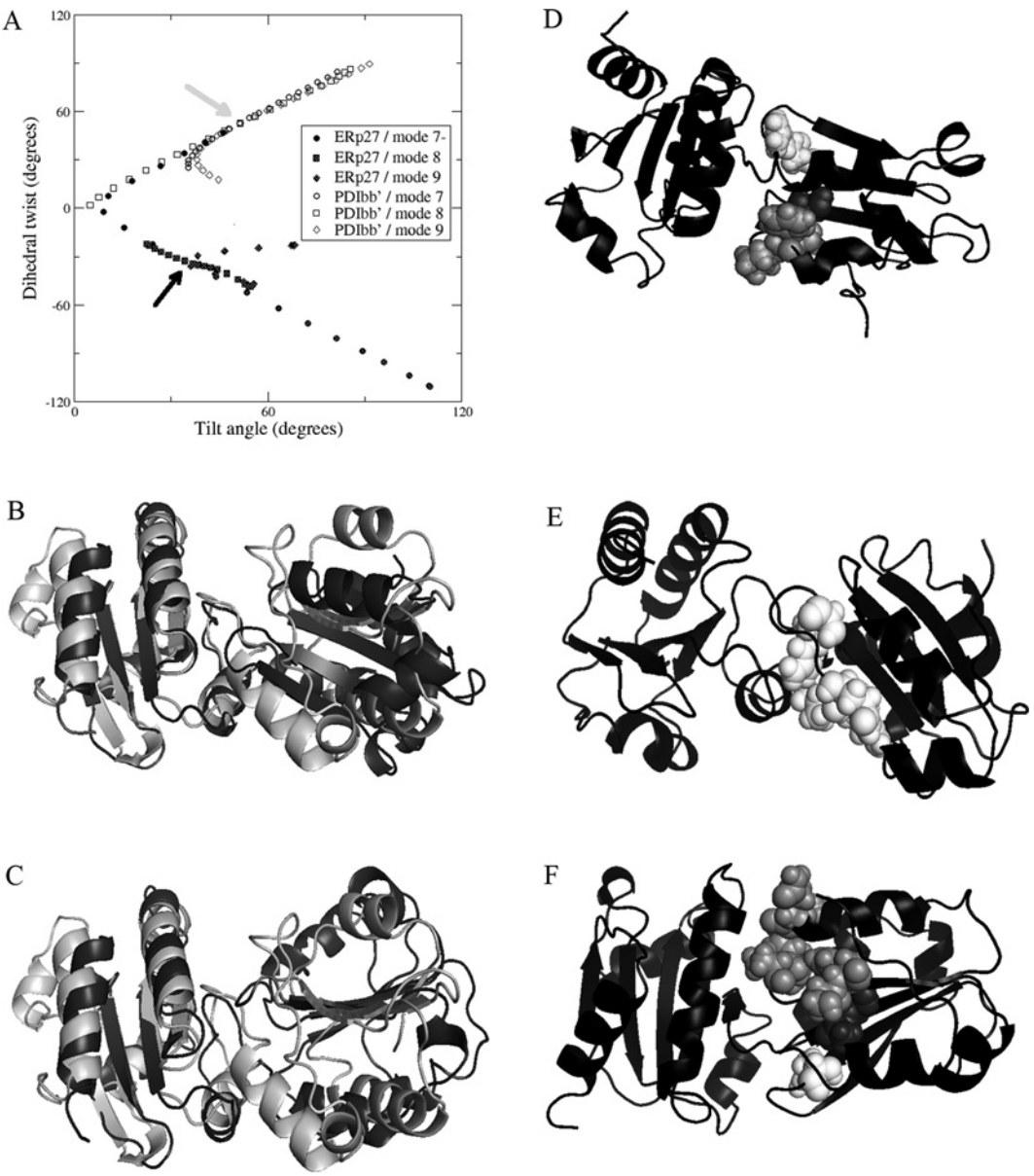

Figure 6 Domain orientation, flexibility and interdomain contacts in ERp27

(**A**) Relative orientation of the β-sheet planes in adjacent Trx-fold domains for ERp27 and for the **b-b′** moiety of yeast PDI. The planes are defined by the positions of four Cα atoms in each domain, as follows: for ERp27 domain **b**, Val^{60}, Gly^{62}, Ile^{106} and Leu^{108}; for ERp27 domain **b′**, Leu^{164}, Leu^{166}, Leu^{222} and Ile^{224}; for PDI **b** domain, Ile^{163}, Gln^{165}, Leu^{202} and Ile^{204}; and for PDI **b′** domain, Gly^{260}, Leu^{262}, Phe^{314} and Ile^{316}. Orientation is described by a tilt angle between the plane normals, and by a dihedral twist angle formed by the plane normals and the interplane vector. The orientation in the PDB code 2B5E crystal structure of PDI is indicated by a grey arrow and that in the ERp27 crystal structure (PDB code 4F9Z [43]) by a black arrow. Symbols (open for PDI, closed for ERp27) show flexible motion along the three lowest-frequency non-trivial elastic network modes (modes 7, 8 and 9), in positive and negative directions, for each structure. Motion of ERp27 along mode 7 − leads to PDI-like interdomain orientations. (**B**) Backbone cartoon overlay of ERp27 (black) with PDI **b-b′** (white) crystal structures, aligned on the **b′** domain (left-hand side) only, showing very different **b-b′** orientations. (**C**) Corresponding backbone cartoon overlay of ERp27 after projection along mode 7 − (black) with PDI **b-b′** (white) crystal structure; the **b** domain β-sheets of ERp27 and PDI are now coplanar. (**D**) Backbone cartoon of ERp27 after projection along mode 7 − (black). Residues 59 (dark grey), 82, 83 and 86 (all light grey), and 110 (white) are shown as spheres. These residues make new contacts with domain **b′**, not found in the input crystal structure. (**E**) Backbone cartoon of ERp27 after projection along mode 7 + (black), showing residues 110, 116, 117 and 118 as white spheres, forming new contacts with domain **b′**. (**F**) Backbone cartoon of ERp27 after projection along mode 8 − (black), showing residues 59 and 60 (dark grey spheres), 80–83, 86 and 87 (light-grey spheres), and 110 (white spheres) forming new contacts with domain **b′**.

full-length ERp27, determined by X-ray diffraction [43]. The two structures superimpose extremely closely with RMSD values of 0.96 Å over 102 Cα atoms and of 1.02 Å over the corresponding 408 backbone atoms (taking molecule D of the crystal structure as representative). The residues showing the greatest RMSD are in loops at either end of β3 (Gly^{103} and Val^{112}) and in the loop containing the 'additional' helix (Glu^{122}). The structural context of Gly^{103} has been noted above as corresponding to the residue preceding the conserved *cis*-proline residue in catalytic redox-active Trx domains, and this residue also shows up in the relaxation data as having an unusually slow internal correlation time.

It was noted (above) that ERp27 **b** has an α-helical insert between β5 and α4 of the conventional PDI domain fold. The equivalent region in the catalytic **a** and **a′** domains of PDI, the

loop between β5 and α4, contains a conserved arginine residue that has been implicated in the catalytic mechanism [44]. Short α-helical inserts also exist in analogous positions in yeast PDI **b** [4] and ERp44 **b′** [40]. There is some sequence similarity between these inserts, with the C-terminus of each helix ending in the sequence Asp-X, where X is a hydrophobic branched-chain aliphatic residue. The side chain of the aspartate residue is surface-exposed and that of the hydrophobic residue is directed inwards, towards the core of the protein. The **b′** domain of yeast PDI also contains an α-helical insert at this position, but this is part of a larger insert between β5 and α4 of the fold. There is no structural information available for the **b** domains of PDIp or PDILT, but these proteins appear to have insertions in the protein sequence at a similar position to the helical inserts in ERp27 **b** and yeast PDI **b** (Supplementary Figure S3). It is therefore possible that PDIp and PDILT also possess the structural feature of an additional α-helix at this location.

We found that ERp27 has a large capacity for flexible motion owing to its two-domain structure with relatively few interdomain constraints. The lowest-frequency non-trivial normal modes of motion for ERp27 represent a twisting motion around the interdomain axis (mode 7) and tilting motions where the interdomain interface acts as a hinge (modes 8 and 9). The **b-b′** moiety of PDI has a similar structure with slightly more interdomain constraints and a similar (although slightly smaller) range of flexible variation. Twist motion of ERp27 along mode 7 − carries its interdomain orientation from the distinctive state identified in the crystal structure to a PDI-like state. The inference that ERp27 is a highly flexible molecule is corroborated by interdomain contact information from NMR, which also suggests considerable interdomain motion (including motion towards a more PDI-like orientational state). Simulations of flexible motion along low-frequency modes generate multiple new interdomain contacts which correlate well with the interdomain contacts suggested by NMR. A further implication is that both ERp27 and PDI have considerable conformational flexibility in solution and that the crystal structures have captured particular orientations from a much wider ensemble of flexible variation. The crystallization of full-length ERp27 in a novel interdomain orientation has revealed the extent of this variation, not previously revealed by other PDI family crystal structures.

The paucity of information on the function of the **b** domain of ERp27 makes it difficult to interpret these results in functional terms. If the function of the **b** domain is purely passive (e.g. to confer solubility on the functional ligand-binding **b′** domain), then the flexibility of the interdomain linker could arise negatively from the lack of any functional constraints determining relative orientation; this would contrast with PDI and other multidomain family members, where the **b-b′** domain pair must partially constrain the positions and orientations of the adjacent functional **a** and **a′** domains [4,45]. Alternatively, if the **b** domain plays a role in interactions between ERp27 and other chaperones or folding factors within the ER lumen, then the interdomain flexibility may be a positive factor in enabling alternative orientations for the domains in interactions with different partners. A recent publication confirmed that many members of the PDI family, including PDI, PDIr, ERp57, ERp72, P5, ERdj5 and ERp29, make multiple interactions with partner proteins within the ER ([45] see, for example, Figure 2A in this reference). Unfortunately, ERp27 was not included as a subject in that study, and in any case we lack structural information to underpin the majority of these inferred protein–protein interactions. In the absence of more information on the cellular role of ERp27, further comments on the significance of the flexible interdomain hinge between **b** and **b′** domains are speculative.

AUTHOR CONTRIBUTION

The project was devised and supervised by Robert Freedman, Mark Howard and Richard Williamson. Nader Amin performed the experimental work and most of the data analysis, with guidance and hands-on support from Katrine Wallis and Michelle Rowe. Stephen Wells carried out the simulations of flexible motion. Nader Amin, Katrine Wallis and Robert Freedman drafted the paper, and all authors contributed to generating the final version of the paper.

ACKNOWLEDGEMENTS

We thank Franz-Xaver Kober and Hermann Schindelin (Rudolf Virchow Center, University of Würzburg, Würzburg, Germany) for communicating their structure of full-length ERp27 and for very helpful discussions. The majority of NMR data were collected at the University of Kent Biomolecular NMR facility which is supported by the Wellcome Trust. We are also grateful to Dr Geoff Kelly and Dr Tom Frenkiel (MRC National Institute for Medical Research, Mill Hill, London, U.K.) for assistance in running 800 MHz NMR datasets.

FUNDING

This work was supported by the Leverhulme Trust (Early Career Research Fellowship to S.A.W.) and the Wellcome Trust ('Value in People' Fellowship to A.K.W.).

REFERENCES

1 Ellgaard, L. and Ruddock, L. W. (2005) The human protein disulphide isomerase family: substrate interactions and functional properties. EMBO Rep. **6**, 28–32
2 Freedman, R. B., Klappa, P. and Ruddock, L. W. (2002) Protein disulfide isomerases exploit synergy between catalytic and specific binding domains. EMBO Rep. **3**, 136–140
3 Kozlov, G., Määttänen, P., Thomas, D. Y. and Gehring, K. (2010) A structural overview of the PDI family of proteins. FEBS J. **277**, 3924–39364
4 Guddat, L. W., Bardwell, J. C., Zander, T. and Martin, J. L. (1997) The uncharged surface features surrounding the active site of *Escherichia coli* DsbA are conserved and are implicated in peptide binding. Protein Sci. **6**, 1148–11565
5 Tian, G., Xiang, S., Noiva, R., Lennarz, W. J. and Schindelin, H. (2006) The crystal structure of yeast protein disulfide isomerase suggests cooperativity between its active sites. Cell **124**, 61–736
6 Klappa, P., Ruddock, L. W., Darby, N. J. and 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
7 Byrne, L. J., Sidhu, A., Wallis, A. K., Ruddock, L. W., Freedman, R. B., Howard, M. J. and Williamson, R. A. (2009) Mapping of the ligand-binding site on the b′ domain of human PDI: interaction with peptide ligands and the x-linker region. Biochem. J. **423**, 209–217
8 Denisov, A. Y., Määttänen, P., Dabrowski, C., Kozlov, G., Thomas, D. Y. and Gehring, K. (2009) Solution structure of the bb' domains of human protein disulfide isomerase. FEBS J. **276**, 1440–1449
9 Alanen, H. I., Williamson, R. A., Howard, M. J., Hatahet, F. S., Salo, K. E., Kauppila, A., Kellokumpu, S. and 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**, 33727–33738
10 Wishart, D. S. and Sykes, B. D. (1994) Chemical shifts as a tool for structure determination. Methods Enzymol. **239**, 363–392
11 Piotto, M., Saudek, V. and Sklenář, V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. J. Biomol. NMR **2**, 661–665
12 States, D. J., Haberkorn, R. A. and Ruben, D. J. (1982) A two-dimensional nuclear Overhauser experiment with pure absorption phase in 4 quadrants. J. Magn. Reson. **48**, 286–292
13 Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR **6**, 277–293
14 Wim, F. V., Wayne, B., Tim, J. S., Rasmus, H. F., Anne, P., Miguel, L., Eldon, L. U., John, L. M., John, I. and Ernest, D. L. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins: Struct., Funct., Bioinf. **59**, 687–696

15 Barbato, G., Ikura, M., Kay, L. E., Pastor, R. W. and Bax, A. (1992) Backbone dynamics of calmodulin studied by ^{15}N relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible. Biochemistry **31**, 5269–5278
16 Neuhaus, D. and Vanmierlo, C. P. M. (1992) Measurement of heteronuclear NOE enhancements in biological macromolecules: a convenient pulse sequence for use with aqueous solutions. J. Magn. Reson. **100**, 221–228
17 Lipari, G. and Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules. 1. Theory and range of validity. J. Am. Chem. Soc. **104**, 4546–4559
18 Lipari, G. and Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules. 2. Analysis of experimental results. J. Am. Chem. Soc. **104**, 4559–4570
19 Palmer, A. G., Rance, M. and Wright, P. E. (1991) Intramolecular motions of a zinc finger DNA-binding domain from Xfin characterized by proton-detected natural abundance C-12 heteronuclear NMR spectroscopy. J. Am. Chem. Soc. **113**, 4371–4380
20 Cornilescu, G., Delaglio, F. and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J. Biomol. NMR **13**, 289–302
21 Rieping, W., Habeck, M., Bardiaux, B., Bernard, A., Malliavin, T. E. and Nilges, M. (2007) ARIA2: automated NOE assignment and data integration in NMR structure calculation. Bioinformatics **23**, 381–382
22 Brünger, A., Adams, P., Clore, G., DeLano, W., Gros, P., Grosse-Kunstleve, R., Jiang, J., Kuszewski, J., Nilges, M., Pannu, N. et al. (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr., Sect. D: Biol. Crystallogr. **1**, 905–921
23 Waters, L. C., Veverka, V., Bohm, M., Schmedt, T., Choong, P. T., Muskett, F. W., Klempnauer, K. H. and Carr, M. D. (2007) Structure of the C-terminal MA-3 domain of the tumour suppressor protein Pdcd4 and characterization of its interaction with eIF4A. Oncogene **26**, 4941–4950
24 Williamson, R. A., Carr, M. D., Frenkiel, T. A., Feeney, J. and Freedman, R. B. (1997) Mapping the binding site for matrix metalloproteinase on the N-terminal domain of the tissue inhibitor of metalloproteinases-2 by NMR chemical shift perturbation. Biochemistry **36**, 13882–13889
25 Jacobs, D. J., Rader, A. J., Kuhn, L. A. and Thorpe, M. F. (2001) Protein flexibility predictions using graph theory. Proteins **44**, 150–165
26 Suhre, K. and Sanejouand, Y. H. (2004) ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. Nucleic Acids Res. **32**, W610–W614
27 Wells, S., Menor, S., Hespenheide, B. and Thorpe, M. F. (2005) Constrained geometric simulation of diffusive motion in proteins. Phys. Biol. **2**, S127–S136
28 Jimenez-Roldan, J. E., Freedman, R. B., Römer, R. A. and Wells, S. A. (2012) Rapid simulation of protein motion: merging flexibility, rigidity and normal mode analyses. Phys. Biol. **9**, 016008
29 Wells, S. A. and Sartbaeva, A. (2012) Template-based geometric simulation of flexible frameworks. Materials **5**, 415–431
30 Li, H., Wells, S. A., Jimenez-Roldan, J. E., Römer, R. A., Zhao, Y., Sadler, P. J. and O'Connor, P. B. (2012) Protein flexibility is key to cisplatin crosslinking in calmodulin. Protein Sci. **21**, 1269–1279
31 Rowe, M. L., Ruddock, L. W., Kelly, G., Schmidt, J. M., Williamson, R. A. and Howard, M. J. (2009) Solution structure and dynamics of ERp18, a small endoplasmic reticulum resident oxidoreductase. Biochemistry **48**, 4596–4606
32 Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. **26**, 283–291
33 Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R. and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR **8**, 477–486
34 Holm, L., Kaariainen, S., Rosenstrom, P. and Schenkel, A. (2008) Searching protein structure databases with DaliLite v.3. Bioinformatics **24**, 2780–2781
35 Kozlov, G., Maattanen, P., Schrag, J. D., Pollock, S., Cygler, M., Nagar, B., Thomas, D. Y. and Gehring, K. (2006) Crystal structure of the bb' domains of the protein disulfide isomerase ERp57. Structure **14**, 1331–1339
36 Kemmink, J., Dijkstra, K., Mariani, M., Scheek, R. M., Penka, E., Nilges, M. and Darby, N. J. (1999) The structure in solution of the b domain of protein disulfide isomerase. J. Biomol. NMR **13**, 357–368
37 Wang, S., Trumble, W. R., Liao, H., Wesson, C. R., Dunker, A. K. and Kang, C. (1998) Crystal structure of calsequestrin from rabbit skeletal muscle sarcoplasmic reticulum. Nat. Struct. Biol. **5**, 476–483
38 Kozlov, G., Määttänen, P., Schrag, J. D., Hura, G. L., Gabrielli, L., Cygler, M., Thomas, D. Y. and Gehring, K. (2009) Structure of the noncatalytic domains and global fold of the protein disulfide isomerase ERp72. Structure **17**, 651–659
39 Barak, N. N., Neumann, P., Sevvana, M., Schutkowski, M., Naumann, K., Malesevic, M., Reichardt, H., Fischer, G., Stubbs, M. T. and Ferrari, D. M. (2009) Crystal structure and functional analysis of the protein disulfide isomerase-related protein ERp29. J. Mol. Biol. **385**, 1630–1642
40 Wang, L., Wang, L., Vavassori, S., Li, S., Ke, H., Anelli, T., Degano, M., Ronzoni, R., Sitia, R., Sun, F. and Wang, C.-c. (2008) Crystal structure of human ERp44 shows a dynamic functional modulation by its carboxy-terminal tail. EMBO Rep. **9**, 642–647
41 Nguyen, V. D., Wallis, K., Howard, M. J., Haapalainen, A. M., Salo, K. E., Saaranen, M. J., Sidhu, A., Wierenga, R. K., Freedman, R. B., Ruddock, L. W. and Williamson, R. A. (2008) Alternative conformations of the x region of human protein disulphide-isomerase modulate exposure of the substrate binding b' domain. J. Mol. Biol. **383**, 1144–1155
42 Wallis, A. K., Sidhu, A., Byrne, L. J., Howard, M. J., Ruddock, L. W., Williamson, R. A. and Freedman, R. B. (2009) The ligand-binding b' domain of human protein disulphide-isomerase mediates homodimerization. Protein Sci. **18**, 2569–2577
43 Kober, F. -X., Koelmel, W., Kuper, J., Drechsler, J., Mais, C., Hermanns, H. H. and Schindelin, H. (2012) The crystal structure of the protein disulfide isomerase family members ERp27 provide insights into its substrate-binding capabilities. J. Mol. Chem., doi:10.1074/jbc.M112.410522
44 Lappi, A. K., Lensink, M. F., Alanen, H. I., Salo, K. E., Lobell, M., Juffer, A. H. and 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
45 Jansen, G., Määttänen, P., Denisov, A. Y., Scarffe, L., Schade, B., Balghi, H., Dejgaard, K., Chen, L. Y., Muller, W. J., Gehring, K. and Thomas, D. Y. (2012) An interaction map of ER chaperones and foldases. Mol. Cell. Proteomics **11**, 710–723
46 Koradi, R., Billeter, M. and Wüthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. J. Mol. Graphics **14**, 51–55

Received 25 October 2012/28 November 2012; accepted 12 December 2012
Published as BJ Immediate Publication 12 December 2012, doi:10.1042/BJ20121635

Biochem. J. (2013) **450**, 321–332 (Printed in Great Britain) doi:10.1042/BJ20121635

online data

SUPPLEMENTARY ONLINE DATA

High-resolution NMR studies of structure and dynamics of human ERp27 indicate extensive interdomain flexibility

Nader T. AMIN*[1], A. Katrine WALLIS*[2], Stephen A. WELLS†[3], Michelle L. ROWE‡, Richard A. WILLIAMSON‡, Mark J. HOWARD‡[4] and Robert B. FREEDMAN*[4]

*School of Life Sciences, University of Warwick, Coventry CV4 7AL, U.K., †Department of Physics, University of Warwick, Coventry CV4 7AL, U.K., and ‡School of Biosciences, University of Kent, Canterbury, Kent CT2 7NJ, U.K.

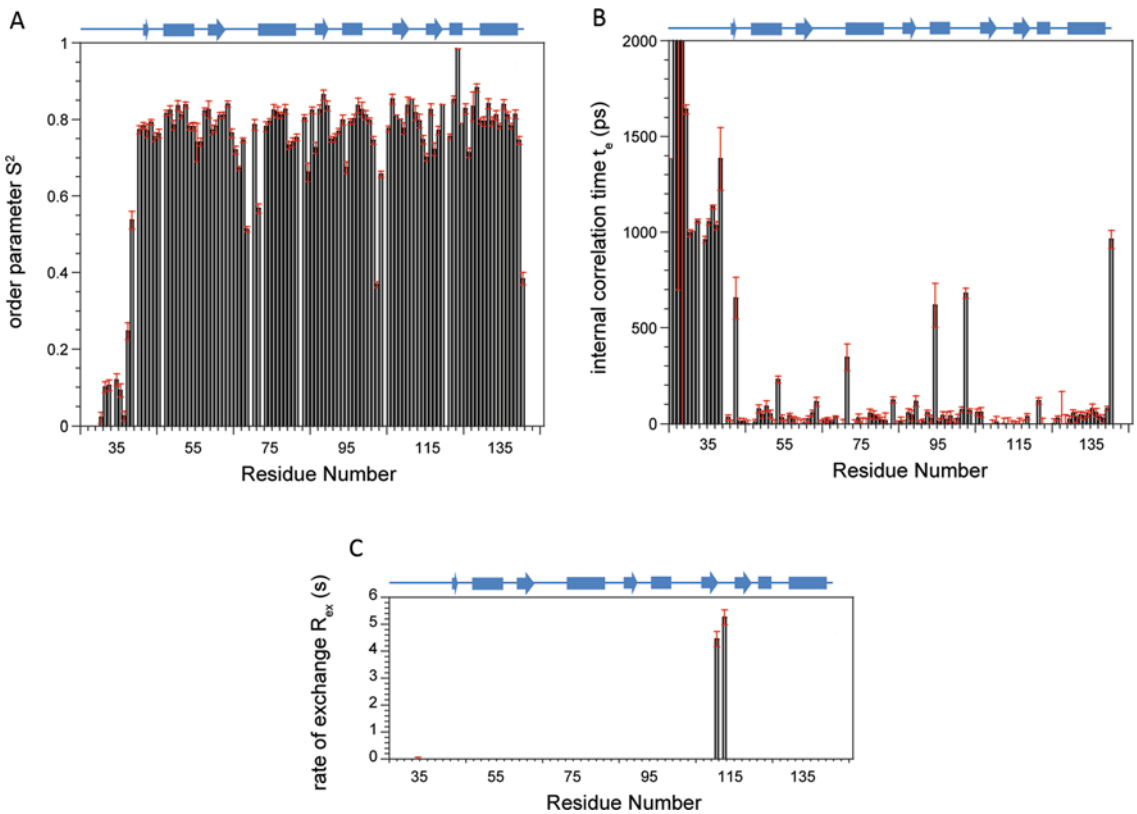

Figure S1 ModelFree analysis of ^{15}N-NMR relaxation data of the ERp27 b domain

Data were collected at 25 °C and 14.1 T. Histograms illustrate the changes across the sequence of (**A**) order parameter S^2, (**B**) internal correlation time τ_e and (**C**) rate of chemical exchange broadening R_{ex}. A schematic diagram of the domain secondary structure as solved by NMR is shown above each plot. Cylinders represent α-halices and arrows represent β-strands.

[1] Present address: Astex Therapeutics, 436 Cambridge Science Park, Cambridge CB4 0QA, U.K.
[2] Present address: Faculty of Health and Life Sciences, Coventry University, Richard Crossman Building, Coventry CV1 5FB, U.K.
[3] Present address: Department of Physics, University of Bath, Claverton Down, Bath BA2 7AY, U.K.
[4] Correspondence may be addressed to either of these authors (email m.j.howard@kent.ac.uk or r.b.freedman@warwick.ac.uk).
The structural co-ordinates for the **b** domain of human ERp27 reported will appear in the PDB under accession code 2L4C.

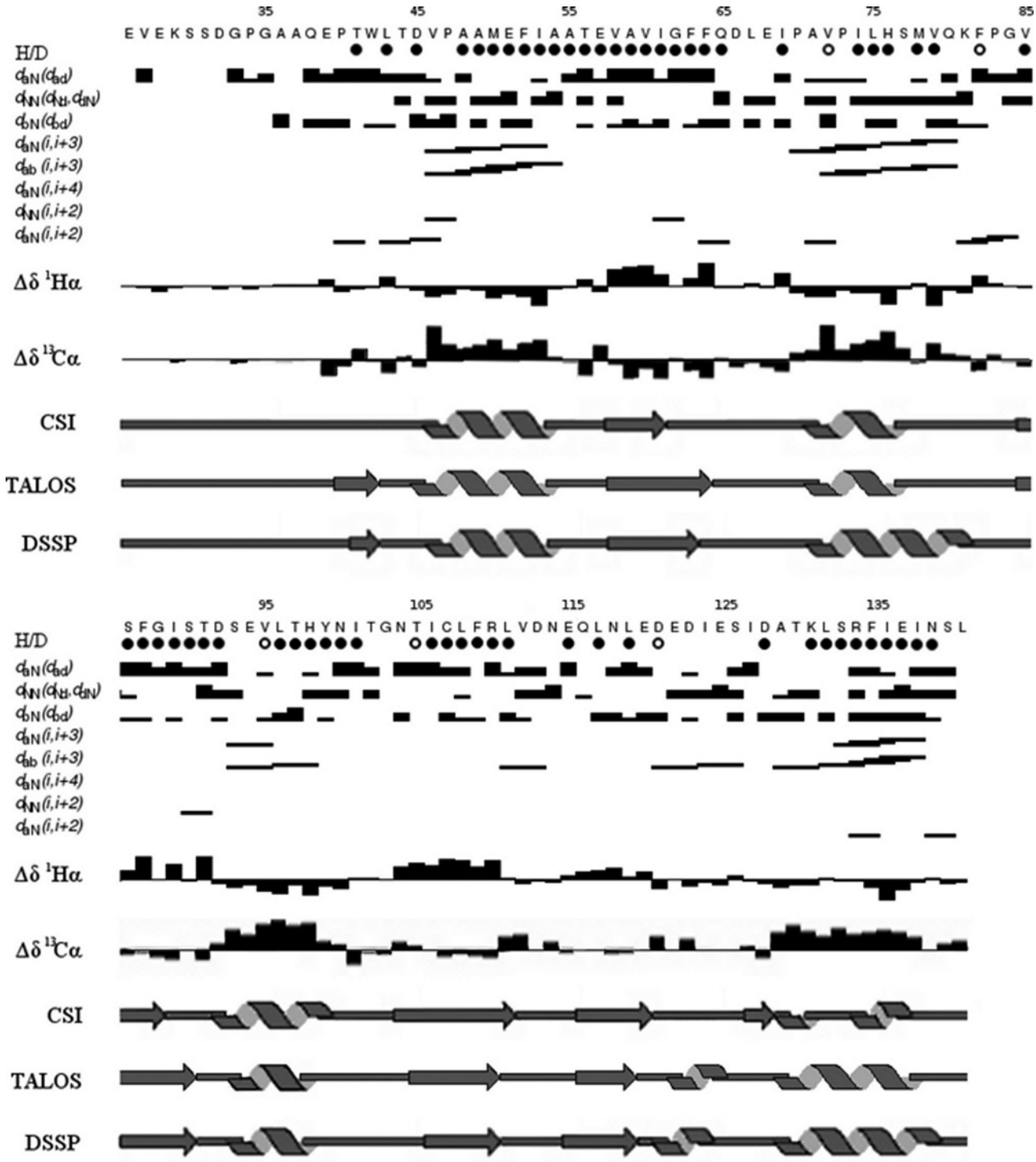

Figure S2 Chemical-shift data supporting the identification of the secondary structure in the ERp27 b domain

A closed circle represents a residue with an amide proton which is still present in the 2D ^{15}N-^{1}H spectrum after incubation in $^{2}H_2O$ for 300 s. An open circle represents an amide proton still present in the 2D ^{15}N-^{1}H spectrum after exchange to $^{2}H_2O$ for which there is ambiguity over the assignment. The sequential proton–proton NOE connectivities ($d_{\alpha N}$, d_{NN} and $d_{\beta N}$) are represented by thick and thin bars. These correspond to strong and weak NOE intensities respectively. The medium-range proton–proton NOE connectivities [$d_{\alpha N}(i,i+3)$, $d_{\alpha\beta}(i,i+3)$, $d_{\alpha N}(i,i+4)$, $d_{NN}(i,i+2)$ and $d_{\alpha N}(i,i+2)$] are represented by lines connecting the residues whose protons are correlated. $^{1}H\alpha$ and $^{13}C\alpha$ chemical-shift deviations from random coil values, as determined by CSI are plotted and are labelled as $\Delta\delta$ $^{1}H\alpha$ and $\Delta\delta$ $^{13}C\alpha$ respectively, with units on the *y*-axis of p.p.m. Positive values represent shifts to lower field. The CSI and TALOS secondary structure predictions and the DSSP-CONT secondary structure assignment are represented by arrows and ribbons. Arrows represent β-sheets and ribbons represent α-helices. The Figure was generated using CCPN Analysis.

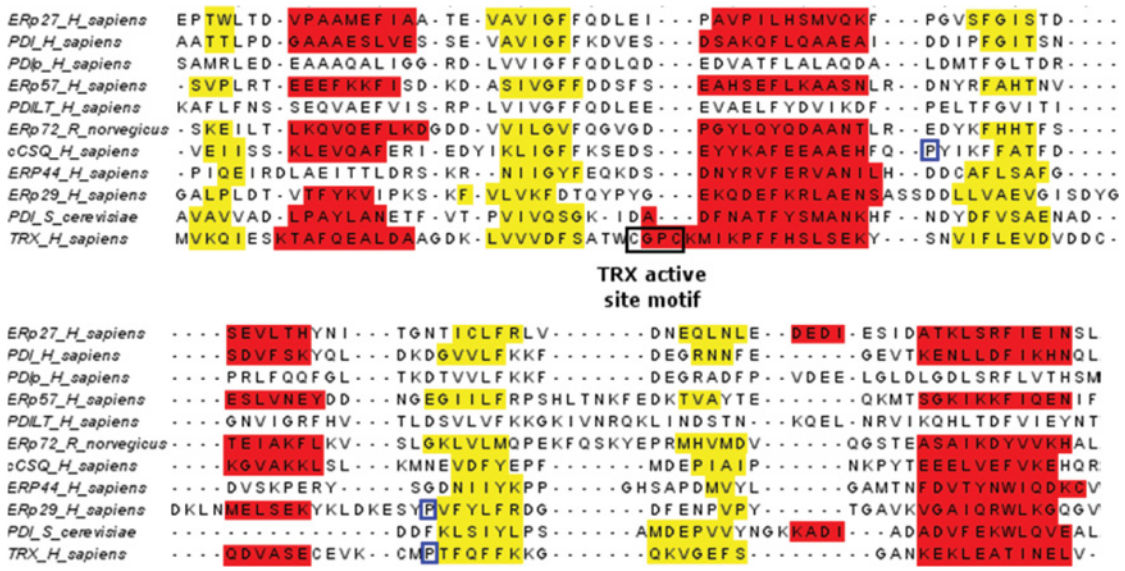

Figure S3 Multiple sequence alignment between ERp27 b and selected homologous proteins

Sequences of human Trx and various **b** domains of the PDI family are shown. The alignment was generated using ClustalW and modified to take into account known secondary structure. Secondary structure elements were defined using DSSP. Strands are shown in yellow and helices are shown in red. The conserved *cis*-proline residues in the *cis*-proline loop of Trx and ERp29 are boxed in blue. The conserved *cis*-proline in the loop between $\alpha 2$ to $\beta 3$ of calsequestrin is also boxed in blue. The active site CGPC motif of Trx is boxed in black. *H_sapiens*, *Homo sapiens*; *R_norvegicus*, *Rattus norvegicus*; *S_cerevisiae*, *Saccharomyces cerevisiae*.

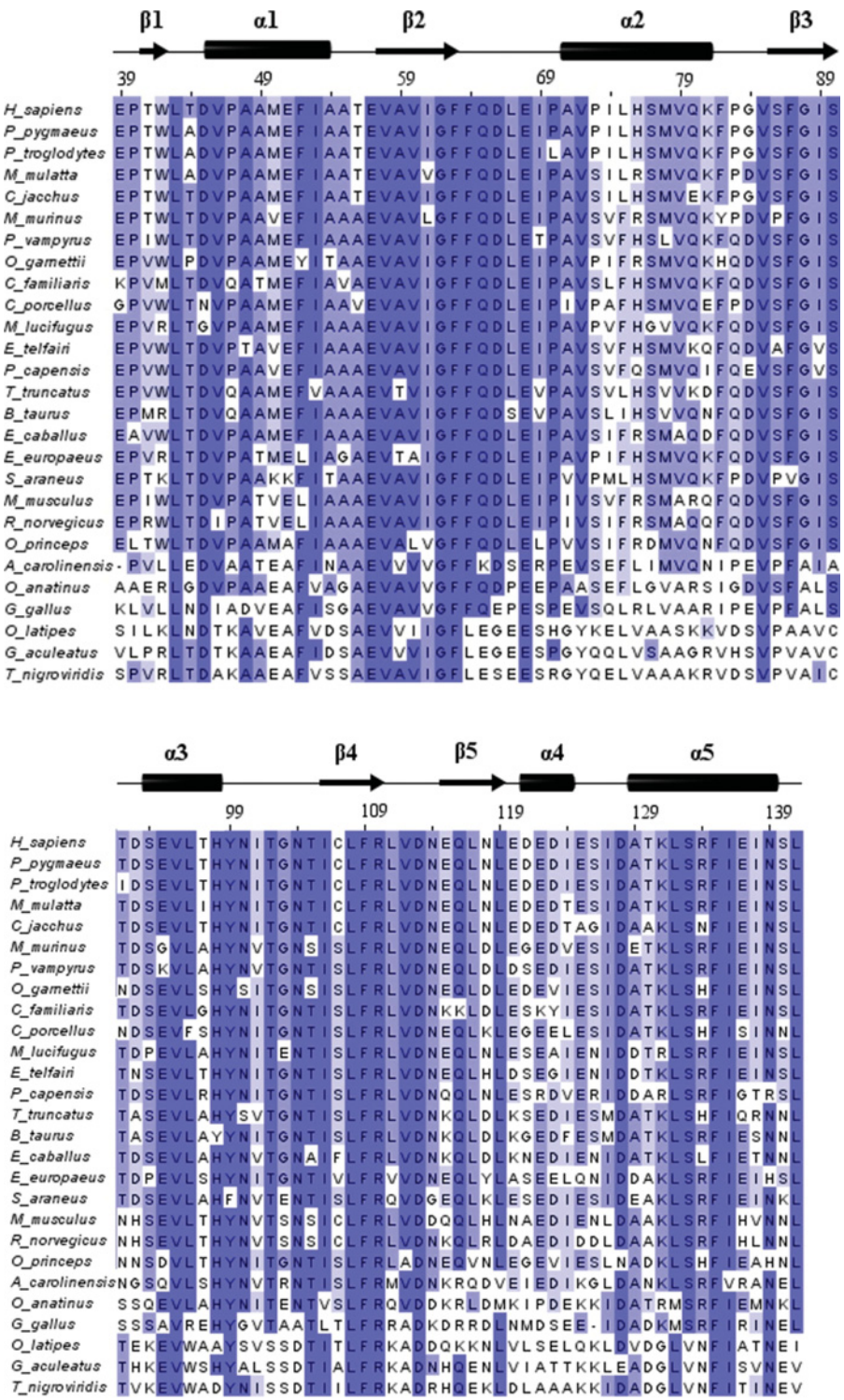

Figure S4 Sequence conservation between species in the b domain of ERp27

ERp27 **b** sequences from 27 different species were aligned using ClustalW. Secondary structure is indicated by cylinders for α-helices and arrows for β-strands. Sequence conservation is denoted by colour coding: light blue for >40 % identity, medium blue for >60 % identity and dark blue for >80 % identity. Residue numbering above the alignment is for human ERp27. *A_carolinensis*, *Anolis carolinensis*; *B_taurus*, *Bos taurus*; *C_familiaris*, *Canis familiaris*; *C_jacchus*, *Callithrix jacchus*; *C_porcellus*, *Cavia porcellus*; *E_caballus*, *Equus caballus*; *E_europaeus*, *Euonymus europaeus*; *E_telfairi*, *Echinops telfairi*; *G_aculeatus*, *Gasterosteus aculeatus*; *G_gallus*, *Gallus gallus*; *H_sapiens*, *Homo sapiens*; *M_lucifugus*, *Myotis lucifugus*; *M_mulatta*, *Macaca mulatta*; *M_murinus*, *Microcebus murinus*; *M_musculus*, *Mus musculus*; *O_anatinus*, *Ornithorhynchus anatinus*; *O_gamettii*, *Otolemur gamettii*; *O_latipes*, *Orizyas latipes*; *O_princeps*, *Ochotona princeps*; *P_capensis*, *Procavia capensis*; *P_pygmaeus*, *Pongo pygmaeus*; *P_troglodytes*, *Pan troglodytes*; *P_vampyrus*, *Pteropus vampyrus*; *R_norvegicus*, *Rattus norvegicus*; *S_araneus*, *Sorex araneus*; *T_nigroviridis*, *Tetraodon nigroviridis*; *T_truncatus*, *Tupsiops truncatus*.

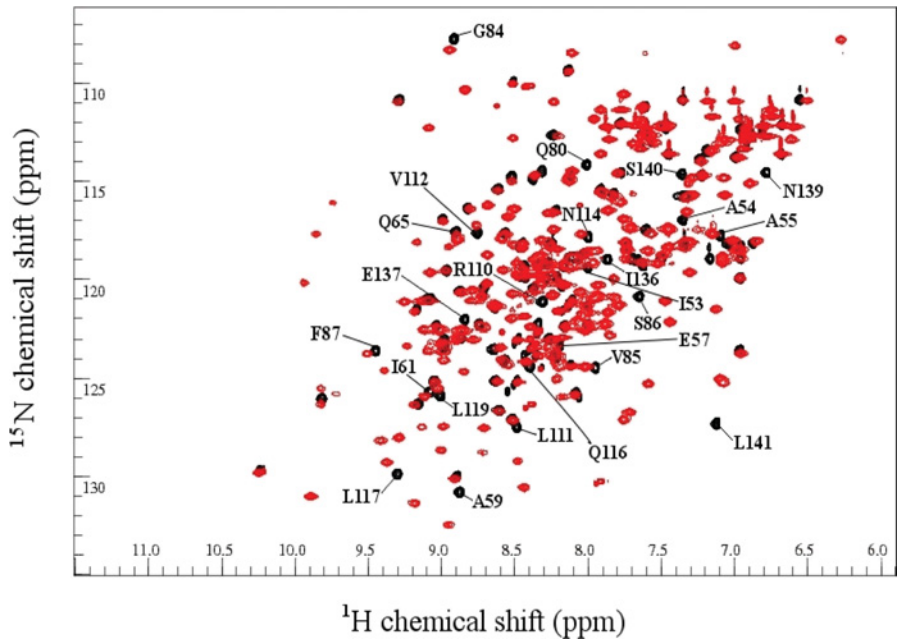

Figure S5 Comparison of spectra of b domain and full-length ERp27

Overlay of the ^{15}N-^{1}H HSQC spectra for the **b** domain of ERp27 (black) and full-length ERp27 (red). Residues with minimal shifts greater than 1 S.D. from the mean for residues 39–135 (0.0333 + 0.0292 p.p.m.) are labelled.

Received 25 October 2012/28 November 2012; accepted 12 December 2012
Published as BJ Immediate Publication 12 December 2012, doi:10.1042/BJ20121635