Structure and Function of *Bacillus subtilis* YphP, a Prokaryotic Disulfide Isomerase with a CXC Catalytic Motif[†,‡]

Urszula Derewenda,[§] Tomasz Boczek,[§,○] Kelly L. Gorres,[∥] Minmin Yu, Li-wei Hung,[#,△] David Cooper,[§] Andrzej Joachimiak,[◇] Ronald T. Raines,*[,∥,⊥] and Zygmunt S. Derewenda*[,§]

[§]*Department of Molecular Physiology and Biological Physics and the ISFI PSI2 Center, University of Virginia School of Medicine, Charlottesville, Virginia 22908-0736,* [∥]*Departments of Biochemistry and* [⊥]*Chemistry, University of Wisconsin−Madison, Madison, Wisconsin 53706,* [#]*Physical Biosciences Division, Lawrence Berkeley National Laboratory, MS4R0230, Berkeley, California 94720,* [△]*Physics Division, MS D454, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, and* [◇]*Biosciences Division, Midwest Center for Structural Genomics and Structural Biology Center, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439* [○]*Present address: Department of Molecular Neurochemistry, Medical University, 6/8 Mazowiecka St., 92-215 Lodz, Poland.*

ABSTRACT: The DUF1094 family contains over 100 bacterial proteins, all containing a conserved CXC motif, with unknown function. We solved the crystal structure of the *Bacillus subtilis* representative, the product of the *yphP* gene. The protein shows remarkable structural similarity to thioredoxins, with a canonical $\alpha\beta\alpha\beta\alpha\beta\beta\alpha$ topology, despite low amino acid sequence identity to thioredoxin. The CXC motif is found in the loop immediately downstream of the first β-strand, in a location equivalent to the CXXC motif of thioredoxins, with the first Cys occupying a position equivalent to the first Cys in canonical thioredoxin. The experimentally determined reduction potential of YphP is $E^{o\prime} = -130$ mV, significantly higher than that of thioredoxin and consistent with disulfide isomerase activity. Functional assays confirmed that the protein displays a level of isomerase activity that might be biologically significant. We propose a mechanism by which the members of this family catalyze isomerization using the CXC catalytic site.

The *Bacillus subtilis yphP* gene codes for a member of a superfamily of over 100 prokaryotic, highly conserved proteins (DUF1094), found predominantly in *Firmicutes* such as *Staphylococcus* and *Bacillus*. No function has been assigned to this family, and to date there is no structure described for any of its representatives. Tertiary structure prediction using 3DJury (*1*, *2*) suggests that YphP has a core domain with high similarity to the canonical thioredoxin (Trx)[1] fold, as exemplified by the *Trypanosoma brucei* Trx (1r26) or *Malassezia sympodialis* Trx (2j23), even though amino acid sequence identity level is very low (~15%). This similarity raises the possibility that members of the DUF1094 proteins constitute a family of oxidoreductases with the ability to catalyze the formation, reduction, and/or isomerization of disulfide bonds. However, they lack the canonical CXXC sequence of the active site and instead contain a CGC motif, atypical of thioredoxins or related oxidoreductases.

The ubiquitous Trx superfamily includes proteins with a unique fold, in which the central, mixed β-sheet is sequestered between several short α-helices (*3*). These oxidoreductases contain a fingerprint CXXC sequence (where X denotes any amino acid), in which the two cysteines alternate between the reduced and oxidized states, allowing them to participate in disulfide-exchange reactions. The location of this active site sequence is highly conserved, at the end of the canonical $\beta2$ strand and into the following α-helix, so that the N-terminal Cys is in an extended conformation, while the remaining three amino acids assume α-helical secondary structure. Interestingly, the sequence of the XX dipeptide varies among the superfamily members and is important for modulating the redox properties of the protein (*4*). For example, in DsbA, which catalyzes the formation of disulfide bonds in proteins secreted into the periplasmic space of *Escherichia coli*, a CPHC motif results in the highest known reduction potential ($E^{o\prime} = -121$ mV) matching the oxidizing environment of the periplasm (*5*). In contrast, cytosolic thioredoxins contain a canonical CGPC motif with an $E^{o\prime}$ value that is 160 mV lower, i.e., -270 mV, which indicates a preference for an oxidized state and matches the reducing environment of the cytoplasm (*6*). Glutaredoxins contain a CPYC motif and their reduction potentials vary depending on structural context, so that the *E. coli* glutaredoxins 1 and 3 have reduction potentials of -233 and -198 mV, respectively (*6*). Finally, the protein disulfide isomerase (PDI) of the yeast endoplasmic reticulum has a CGHC sequence and a reduction potential of -180 mV (*7*).

In spite of this functional diversity, the general architecture of the active site motif is highly conserved, so that no member of the thioredoxin superfamily has been found with fewer or more than

two intervening residues between the cysteines. Consequently, the functional importance of the number of the residues between the cysteines for the reduction potential has been studied in synthetic peptides (*8*). These experiments revealed that for peptides $C(X)_nC$, where $n \leq 5$, those with a single intervening residue have the least stable disulfide. It has also been shown that a CGC-NH_2 tripeptide has a reduction potential of -167 mV, close to that of PDI, and that it exhibits disulfide isomerization activity (*9*). A CXC motif engineered into the active site of an *E. coli* thioredoxin by deleting the Pro residue from the CGPC active site showed likewise a destabilized disulfide with a reduction potential of ≥ -200 mV. In its reduced form, this variant displayed disulfide isomerase activity that was 25-fold greater than that of the synthetic CGC peptide (*9*).

Surprisingly, amino acid sequence alignments showed that YphP, and all other members of the DUF1094 family, contains a CGC motif, involving Cys53 and Cys55, in a location analogous to the active site of thioredoxin. To gain insight into the structure and function of the DUF1094 family, we determined the crystal structure of the *B. subtilis* YphP at 2.1 Å resolution. The refined model confirmed that the protein is structurally closely related to the thioredoxin superfamily and that Cys53 occupies a site corresponding to the canonical N-terminal cysteine in the CXXC motif; Cys55 is located in place of the Pro in Trx. To better understand the function of YphP, we determined its reduction potential and analyzed its isomerase activity by using a continuous assay that measures the rate of isomerization of non-native to native disulfide bonds in tachyplesin I (*10*). We then used two single site variants, C53A and C55A, to investigate the role of each of the two cysteine residues in the active site. Finally, we compared the activity of YphP to that of ΔP34 thioredoxin in the fluorescence assay. The results confirm that YphP represents a new family of oxidoreductases that have a natural CXC motif within its active site. Both the reduction potential and functional assays are consistent with protein disulfide isomerase activity.

MATERIALS AND METHODS

Protein Production and Crystallization. The *B. subtilis* YphP (also denoted APC1446) was originally selected as one of the targets for high-throughput structure determination by the Midwest Center for Structural Genomics (http://www.mcsg.anl. gov). The open reading frame was amplified from genomic DNA with a recombinant KOD HiFi DNA polymerase (Novagen) from *Thermococcus kodakaraensis* using conditions and reagents provided by the vendor (Novagen). The gene was cloned into a pMCSG7 vector (*11*) using a modified ligation-independent cloning protocol (*12*). This construct produced a fusion protein with an N-terminal His_6 tag and a recognition site for the tobacco etch virus (TEV) protease. Unfortunately, further progress was impeded by the recalcitrance of the protein to crystallization. To circumvent this problem, we used the entropy reduction approach (*13–15*). We identified three clusters with high conformational entropy residues: cluster 1, E39, K40, E42; cluster 2, K113, E114; and cluster 3, Q100, E101. Variants were generated in which residues in these clusters were replaced with either Tyr or Ala.

Wild-type protein was expressed with high yield in *E. coli* BL21(DE3)RIPL. The protein was purified using standard methods and a combination of nickel affinity chromatography (Ni-NTA agarose column; Qiagen), followed by rTEV proteolysis to cleave the His tag, and gel filtration in the presence of reducing agents (1 mM DTT, 5 mM β-mercaptoethanol) to circumvent aggregation caused by intermolecular

disulfide formation. The protein was concentrated to $8-15$ mg/mL. Variant proteins were purified in a similar manner.

A custom set of conditions (*16*) was used for crystallization screening against a reservoir of precipitant and an alternative reservoir containing salt (*17*). The cluster 1 Ala variant and cluster 3 Tyr variant yielded poorly reproducible crystals under a variety of conditions. In contrast, the cluster 3 Ala variant (i.e., Q100A, E101A) gave reproducible crystals in two different conditions: 3.2 M $(NH_4)_2SO_4$ and 0.1 M citric acid buffer at pH 5.0 and 2.0 M $(NH_4)_2SO_4$. Following optimization, the best crystals were obtained from 2.6 M $(NH_4)_2SO_4$ and 0.1 M HEPES buffer at pH 7.6 with added 30% trimethylamine *N*-oxide; 5 mM β-mercaptoethanol in the reservoir was necessary for crystals to reach maximal dimensions.

To generate the selenomethionine- (SeMet-) labeled protein samples, the protein was expressed in *E. coli* B834 cells. Seed cultures of 20 mL were grown in Luria broth for 4 h at 37 °C. The seed cultures were spun down at 3000 rpm and used to inoculate autoinduction medium containing SeMet. The cultures were grown for 4.5 h at 37 °C, followed by ~16 h at 20 °C and another 24 h at 10 °C. Harvested cultures yielded ~35 g of wet mass/2 L. Protein was purified using the procedure described above and yielded more than 100 mg of pure SeMet-labeled protein. Protein was concentrated to ~25 mg/ mL in a buffer containing 25 mM NaCl and 25 mM Tris buffer at pH 8.0 and 1 mM DTT. The structure was determined from SeMet-labeled crystals grown from 2 M $(NH_4)_2SO_4$ and 100 mM Tris buffer at pH 6.5 and 5% PEP (pentaerythritol propoxylate) 629.

All single- and multiple-site mutations were introduced using the QuikChange site-directed mutagenesis kit (Stratagene) and confirmed by direct DNA sequencing. The respective protein variants were purified as described above for the wild type.

ΔP34 *E. coli* thioredoxin was produced and purified as we described previously (*9*).

Crystallography. The crystals of the Q100A, E101A variant are tetragonal, $P4_12_12$, with unit cell dimensions $a = b = 68.16$ Å and $c = 302.02$ Å, and contain four molecules in the asymmetric unit. The crystals contain ~55% solvent. For data collection, a crystal was frozen in mother liquor containing additional 20% glycerol. Data were collected at beamline 5.0.2 of the Advanced Light Source (ALS). The multiwavelength anomalous diffraction (MAD) data, used to solve the structure, were collected at three wavelengths: 0.9795 Å (peak), 0.9796 Å (inflection point), and 0.9393 Å (high energy remote). The data extended to $2.3-2.5$ Å with the merging *R* factors ranging between 11.9% and 13.6% (see Table 1). Positions of selenium atoms were identified with SHELXD (*18*). Crystallographic phasing was done with SOLVE (*19*). Phases were improved by iterative model building and refinement in RESOLVE (*19*) followed by ARP-WARP (*20*) and manual rebuilding with O (*21*) and COOT (*22*). Refinement was carried out with REFMAC5 (*23*) and with PHENIX (*24*) using the TLS (translation/libration/ screw) approximation of thermal motion (*25*). Solvent accessibility of the Cys residues was calculated using the PISA server (*26*).

Isomerase Activity. Assays of disulfide isomerase activity were performed with a synthetic fluorescent substrate as described previously (*10*). The final concentration of enzyme was 1 μM. Kinetic parameters were determined by nonlinear

Table 1: Crystallographic Data

	remote	edge	peak
	Data Collection Statistics		
wavelength (Å)	0.9393	0.9796	0.9795
resolution (Å)	75.0−2.30 (2.39−2.30)[a]	75.0−2.50 (2.59−2.50)	75.0−2.30 (2.39−2.30)
total reflections	148455	167160	203269
unique reflections	30085	25434	33433
redundancy	4.93	6.57	6.08
completeness (%)	90.2 (59.1)	94.3 (71.5)	97.2 (82.9)
R_{merge} (%)[b]	13.6 (42.3)	12.4 (48.2)	11.9 (49.1)
average $I/\sigma(I)$	17.4 (2.1)	19.3 (1.9)	19.7 (2.2)
	Refinement Statistics		
resolution (Å)			75.0−2.30 (2.37−2.30)[a]
reflections (working)			31840 (2461)
reflections (test)			1593 (126)
R_{work} (%)[c]			19.0 (31.7)
R_{free} (%)[c]			24.6 (37.3)
no. of protein atoms			4484
no. of water oxygens			249
no. of water sulfate ions			12
average B factors (Å²)[d]			
main chain			50.7
side chain			56.3
waters			50.6
rms deviation from ideal geometry			
bonds (Å)			0.006
angles (deg)			1.02
Ramachandran plot: favored (%)			97.9
Ramachandran plot: outliers (%)			0.0

[a]The numbers in parentheses describe the relevant value for the highest resolution shell. [b]$R_{merge} = \sum|I_i - \langle I \rangle|/\sum I$, where I_i is the intensity of the ith observation and $\langle I \rangle$ is the mean intensity of the reflections. The values are for unmerged Friedel pairs. [c]$R = \sum||F_o| - |F_c||/\sum|F_o|$, crystallographic R factor, and $R_{free} = \sum||F_o| - |F_c||/\sum|F_o|$, where all reflections belong to a test set of randomly selected data; $R_{overall}$ = crystallographic R factor calculated for all data after a final cycle of refinement. [d]B factors were refined using TLS approximation (see Materials and Methods).

least-squares regression analysis using eq 1:

$$F = F_0 + (F_{max} - F_0)(1 - e^{-kt}) \quad (1)$$

Disulfide Bond Reduction Potential. The reduction potential of the disulfide bond in YphP was determined from the thiol−disulfide exchange equilibrium between YphP and glutathione. YphP (10 μM) was incubated in 50 mM sodium phosphate, pH 7.6, containing EDTA (1 mM) and various ratios of reduced/oxidized glutathione, $E^{\circ\prime} = -0.252$ V (*27*). The reaction mixture was blanketed with Ar(g), and the reaction was allowed to proceed for 60 min at 30 °C, after which it was quenched and YphP was precipitated by the addition of trichloroacetic acid to 10%. The precipitated protein was washed with acetone and resuspended in 100 mM sodium phosphate, pH 7.6, containing 5,5′-dithiobis(2-nitrobenzoic acid) (DTNB; 0.10 mg/mL). The concentration of thiols was quantified by the ensuing absorbance at 412 nm. The amount of reduced YphP was compared to that of YphP that had been incubated with DTT (1 mM) to obtain the fraction of reduced YphP (f in eq 2). The standard reduction potential of YphP was calculated by nonlinear least-squares analysis of the data with the Nernst equation (eq 2):

$$f = f_{min} + \frac{f_{max} - f_{min}}{1 + e^{-(E - E^\circ)(nF/RT)}} \quad (2)$$

where n is the number of electrons, F is the Faraday constant (96485 J V^{-1} mol^{-1}), R is the gas constant (8.314 J K^{-1} mol^{-1}), T is the temperature (here, 303 K), E is the reduction potential of the solution, and E° is the reduction potential of YphP.

RESULTS AND DISCUSSION

Crystal Structure. The crystal structure was solved using the multiwavelength anomalous dispersion (MAD) method, and the atomic model was refined to an R factor of 19.4% (Table 1). The final model contains 584 amino acid residues (from four independent protein molecules), 249 water molecules, and 12 SO$_4^{2-}$ ions. Of the residues 93.4% are in the most favored regions of the Ramachandran plot with the remaining residues in additionally allowed regions. Three out of the four molecules show interpretable electron density for all 144 residues with an additional cloning artifact of three residues (SNA) on the N-terminus. The fourth molecule is slightly less well defined, and only residues 1−143 were placed in the electron density. A number of electron density peaks in the solvent region were attributed to sulfate ions. Details of the data collection, refinement, and model quality statistics are listed in Table 1.

The four molecules in the asymmetric unit are arranged into a dimer of dimers in which the major contacts are mediated by an N-terminal helix, unique to YphP and not found in a canonical thioredoxin fold (Figure 1A). A comparison of the tertiary structure of YphP to other known structures using DALI (*28*) shows that, with the exception of the N-terminal helix, the protein is very similar to a typical thioredoxin, with the highest similarity (*Z* score of 11.6) to the human thioredoxin (*29*). The canonical $\alpha\beta\alpha\beta\alpha\beta\beta\alpha$ topology is well conserved, although there are differences in the mutual disposition and lengths of the individual helices and loops (Figure 1B). The tetrameric ensemble appears
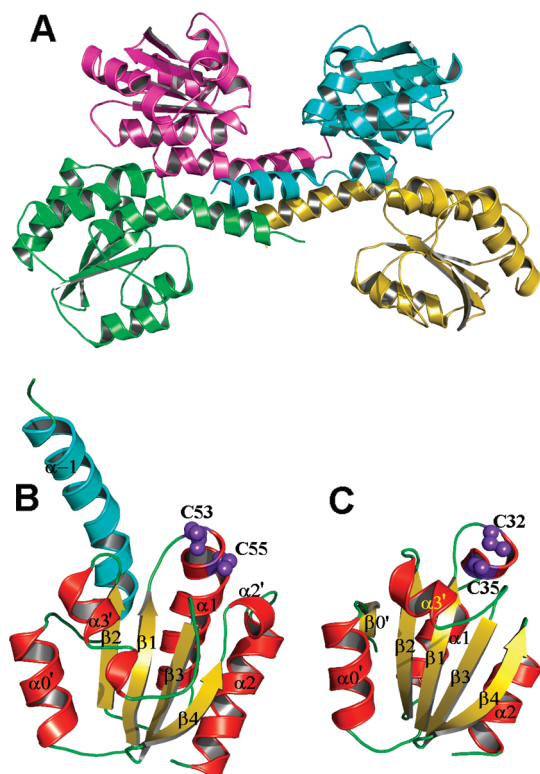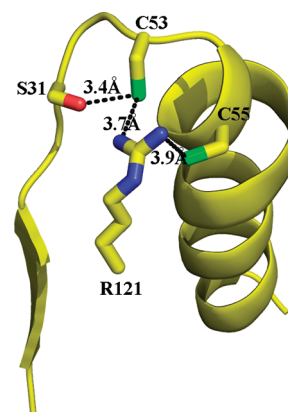
FIGURE 2: Structure of the catalytic loop of YphP. The coordinates used in this diagram are those of chain D but are representative of all four chains in the structure.

molecules range from 3.6 to 5.9 Å, ruling out intramolecular disulfide bonds in the crystal structure; thus, the crystal structure represents the reduced form of the enzyme, as expected from the presence of reducing agents. The sulfhydryl of Cys53 is involved in two hydrogen bonds: with the hydroxyl O^γ of Ser51 and with $N^{\eta 1}$ of Arg121 (Figure 2), although specific distances vary depending on the side chain conformation. Both of these residues are completely conserved within the family, suggesting a functional role. Interactions of arginines with thiolates are known in other redox-regulated proteins. For example, in the hydroperoxide resistance protein Ohr, the active site cysteine Cys60, which is directly involved in peroxide reduction, accepts a hydrogen bond from $N^{\eta 1}$ of Arg18 at 3.4 Å (*30*). In the *B. subtilis* chaperone Hsp33, the Zn^{2+}-binding Cys235 is within 3.4 Å of $N^{\eta 1}$ of Arg86 (*31*). By contrast, Cys55 has no similar interactions. These different environments suggest that the pK_a value of Cys53 is lower than that of Cys55 and that Cys53 could be the key nucleophile in the active site.

In spite of the obvious similarity of the CGC motifs in YphP and the ΔP34 variant of Trx (*9*), the active sites of these two proteins need not be structurally equivalent. In Trx, the deletion of Pro34 might cause a compensating rotation of the α-helix, although there is no X-ray structure available to prove this notion. Also, given the amphipathic character of this helix, it is difficult to see how a deletion of Pro34 could be accomplished without disturbing the protein's stability. In fact, the thermal stability of the ΔP34 variant has been shown to be lower than that of the wild-type protein by ∼10 °C (*9*). In YphP, Cys55 simply occupies the place of Pro34 in Trx, while the position immediately downstream is occupied by an Ala, thus avoiding the strain that might characterize the ΔP34 Trx variant.

DUF1094 Family. Figure 3 shows an alignment of representative amino acid sequences of proteins in the DUF1094 family, as listed in the Pfam database (*32*). This family comprises 102 sequences in 62 species. The majority (80 sequences) are found among the *Firmicutes*, mostly in the *Bacillus* and *Staphylococcus* genera. Interestingly, in these genera there are nearly always two homologous genes of this family in any species. The remaining sequences are in *Acidobacteria* and *Bacterioidetes*, and there is only a single gene in each species. The family shows a relatively high level of amino acid conservation with several completely conserved motifs in addition to the CGCA active site. Topologically, all of these conserved motifs are clustered primarily in loops around the active site, suggesting a possible role in

FIGURE 1: (A) Packing of four crystallographically independent molecules of YphP. (B) A single molecule of YphP with the secondary structure elements identified (α-helices red and β-strands yellow) and the two Cys side chains of the CXC motif shown as spheres and labeled; the additional N-terminal α-helix unique to the DUF1094 family is shown in cyan. (C) Analogous view of the human thioredoxin (PDB code 1AUC) shown for comparison. The annotations used here for the α-helices and β-sheets follow the standard convention for thioredoxins (*45*).

to constitute a crystallographic artifact, because gel filtration data (not shown) show that the protein is monomeric in solution.

The four crystallographically independent molecules display a degree of conformational variability with respect to the N-terminal α-helix, suggesting that the latter is flexible with respect to the core of the molecule. This may account, at least in part, for the recalcitrance of the wild-type protein to crystallization. Several intermolecular contacts define the molecular packing. Importantly, three out of four YphP molecules in the asymmetric unit form crystal contacts directly via the low-entropy patches generated by the Q100A, E101A substitutions. This pattern suggests that an oligomeric ensemble might first form transiently in solution and then be incorporated into the crystal via surface patches introduced by mutagenesis. A similar mechanism is observed for other proteins crystallized by this approach (Z. S. Derewenda, unpublished).

The CXC motif (Figure 2) is located within a loop following the first β-strand, as predicted by 3D-Jury (*1*) based on the amino acid sequence. Cys53 is in an extended conformation ($\phi = \sim-70$, $\psi = \sim165$; the dihedral angles are similar in all four molecules), and its position within the active site loop is analogous to the first Cys of the canonical CXXC motif. Residue Gly54 ($\phi = \sim-45$, $\psi = \sim-40$) is the first amino acid of the α1 helix, while Cys55 occupies the position of the second X (Pro in thioredoxin). Of the two cysteines, Cys53 is more solvent exposed (∼42 Å^2) than Cys55 (∼17 Å^2), although there is some variability among the four molecules, and there is evidence of static disorder of the side chain of Cys55. The distances between the two S$^\gamma$ atoms in the four
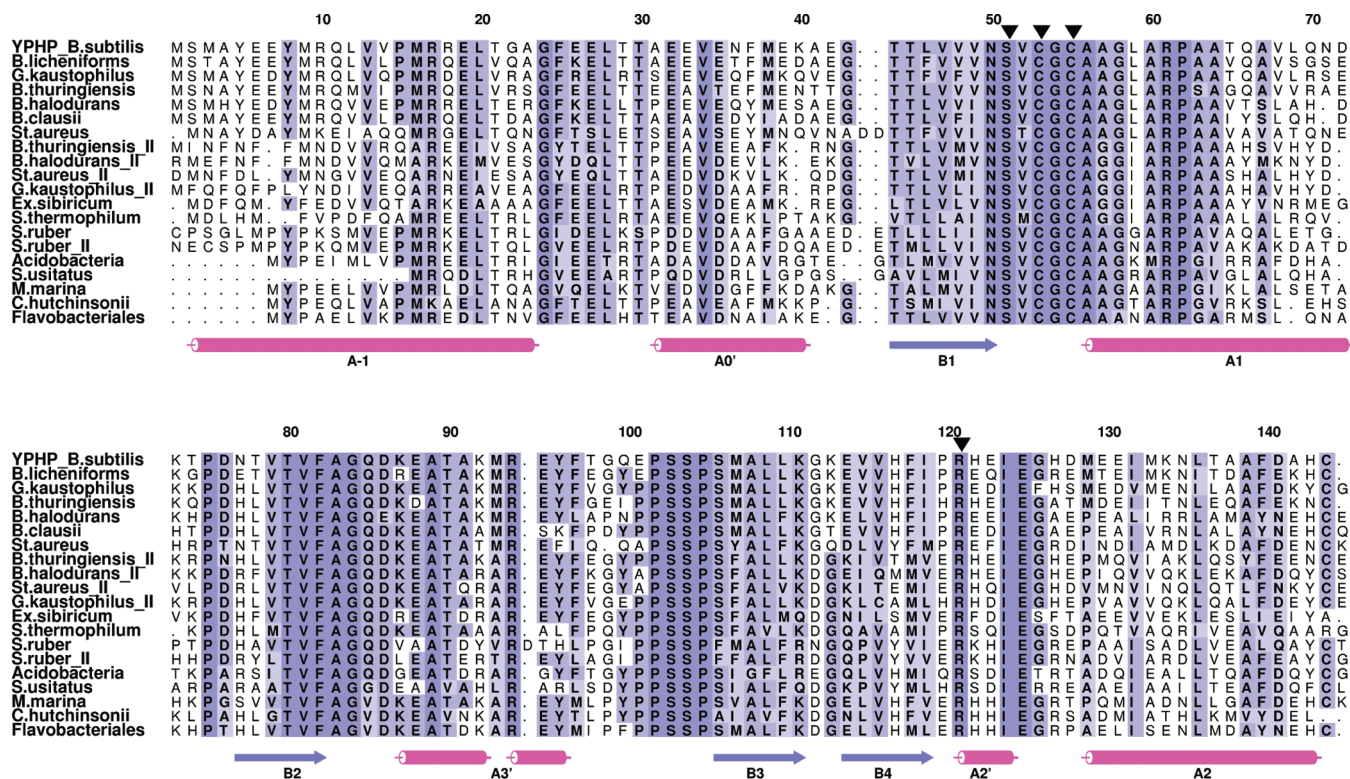
FIGURE 3: Sequence alignment of representative members of the DUF1094 family. Black triangles show the four residues that we identify as the catalytic amino acids (further details in the text). Semiconserved and fully conserved residues are indicated by darkening blue color. The secondary structure elements are shown below the alignment as cylinders (α-helices) and arrows (β-sheets).

substrate recognition. The atypical, N-terminal α-helix is the least conserved fragment in the DUF1094 family. We were unable to identify any homologues of these proteins in eukaryotic genomes.

Activity Assays. The overall structural similarity of YphP to thioredoxins and the presence of the CGC motif in the putative active site suggest that the protein, as well as its homologues in the DUF1094 family, constitutes a family of oxidoreductases. To confirm this hypothesis, we carried out functional assays.

First, we determined the reduction potential, $E^{o'}$, of YphP to be -130 ± 5 mV (Figure 4). As pointed out above, for the known oxidoreductases with CXXC motifs in their active sites, the reduction potentials range from -270 mV in the *E. coli* thioredoxin (6), a cytosolic reductant, to -122 mV in DsbA, which catalyzes oxidation of proteins secreted into the periplasm (5). The stability of the active site disulfide bond is tuned, in part, by the amino acid residues between the active site cysteines, i.e., the XX residues in the CXXC motif. The reduction potential of YphP was determined to be $E^{o'} = -130 \pm 5$ mV, a notably high value, consistent with YphP being an isomerase/oxidase rather than a reductase. The only other protein containing a CXC motif within an intact protein, for which $E^{o'}$ has been determined, is the ΔP34 variant of Trx; in that case $E^{o'} \geq -200$ mV, at least 70 mV higher than in wild-type Trx (9). It seems that decreasing the number of intervening residues between the cysteines from two (CXXC) to one (CXC) provides at least one mechanism to increase the reduction potential of a thiol−disulfide oxidoreductase. It is noteworthy, however, that for a synthetic CysGlyCysNH$_2$ peptide the reduction potential $E^{o'} = -167$ mV (9). Thus, the structural scaffold of YphP is also responsible for a further increase of $E^{o'}$ by about 40 mV.

We also showed that YphP was able to catalyze the isomerization of non-native to native disulfide bonds using tachyplesin I peptide as a substrate, yielding a k_{cat}/K_M value of $(4.7 \pm 0.03) \times 10^3$ M^{-1} s^{-1}. This activity is ~3% that of protein disulfide
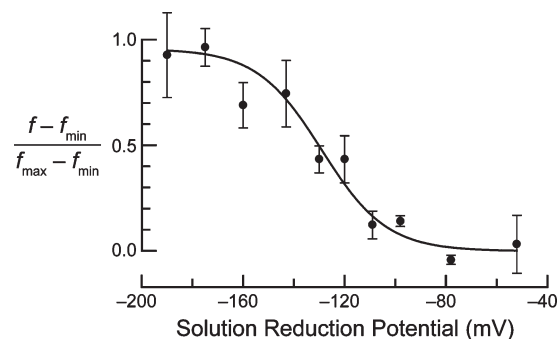

FIGURE 4: Reduction potential of YphP. The fraction of reduced YphP (*f*) is plotted as a function of the solution reduction potential, which was established at 30 °C by using reduced and oxidized glutathione. Values are the mean (±SE) from three experiments. The data were fitted to eq 2 to give $E^{o'} = -130 \pm 5$ mV.

Table 2: Values of k_{cat}/K_M for Catalysis of Disulfide Bond Isomerization

enzyme	variant	k_{cat}/K_M (10^3 M^{-1} s^{-1})
YphP	wild type	4.7 ± 0.1
YphP	C53A	2.6 ± 0.2
YphP	C55A	3.3 ± 0.1
YphP	C53A/C55A	0.3 ± 0.2
Trx	ΔP34	11.2 ± 0.8
PDI	wild type	170 ± 50^{a}
PDI	CGHA/CGHA	86 ± 35^{a}

[a]Data from ref *10*.

isomerase (PDI) (*10*) and 42% that of ΔP34 thioredoxin (Table 2). To investigate the roles of the cysteine thiols of YphP, we replaced each of the active site cysteine residues with alanine, thus generating single-site C53A and C55A variants.
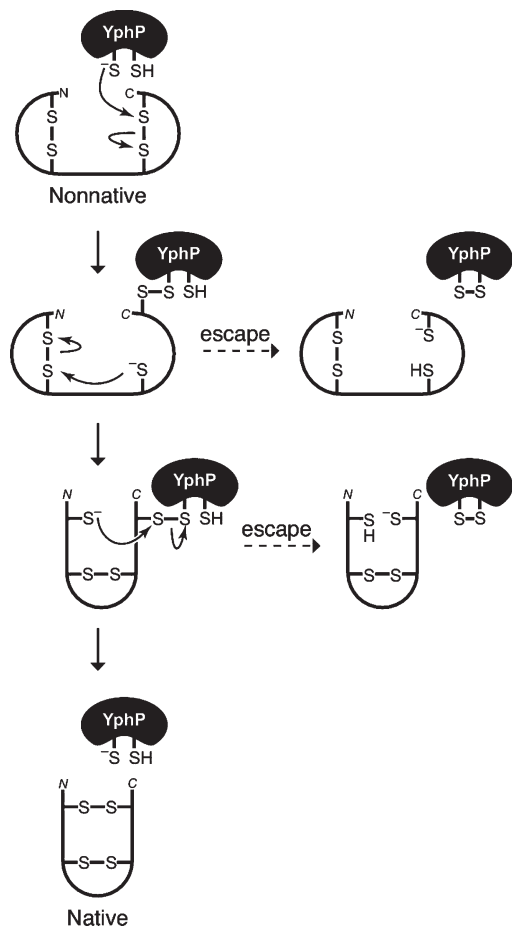
Figure 5: Diagram of the proposed reaction mechanism catalyzed by YphP and its homologues.

Figure 6: Structures of known oxidized CXC motifs and the reduced CXC motif of YphP; for details see text. The colors reflect atom type: yellow, carbon; blue, nitrogen; red, oxygen; green, sulfur.

Unexpectedly, both of these variants catalyzed the isomerization of disulfide bonds in the scrambled tachyplesin substrate. The C53A and C55A variants had k_{cat}/K_M values of $(2.6 \pm 0.2) \times 10^3$ and $(3.3 \pm 0.1) \times 10^3$ M^{-1} s^{-1}, respectively, only slightly lower than that of the wild-type protein. Yet, when *both* active site cysteine residues were replaced with alanines, resulting in the C53A/C55A variant, isomerase activity was abolished completely (Table 2).

The Catalytic Mechanism. The canonical mechanism of disulfide bond isomerization by PDI involves an attack on the substrate disulfide bond by the thiolate of the N-terminal cysteine of the CXXC motif (*33−35*). A mixed disulfide is formed between PDI and the substrate, enabling the thiol and disulfides of the substrate to react until native disulfide bonds are formed in the substrate, so that PDI is no longer tethered to it by a mixed disulfide. Replacing the N-terminal Cys in PDI with a Ser eliminates enzymatic activity (*36*). In contrast, replacing the C-terminal Cys eliminates the ability of the enzyme to catalyze substrate reduction and oxidation but not isomerization (*7, 37*). During the rearrangement of disulfides in the substrate protein, a stable mixed disulfide between PDI and the substrate could prevent the release of PDI. The C-terminal cysteine of PDI is thought to provide an "escape" mechanism in which a disulfide is formed within the active site of PDI to allow for substrate release (*8−11, 37*). A similar role has been ascribed to the C-terminal cysteine residue of thioredoxin (*38*).

The crystal structure of YphP suggests that the protein functions as a disulfide isomerase using a mechanism similar to that of protein disulfide isomerase (Figure 5). In the reduced
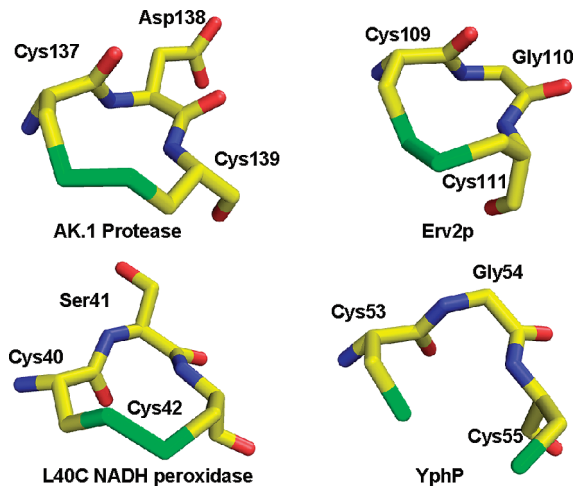
form, Cys53 is most likely the attacking nucleophile, activated by Arg121. When a stable intermolecular disulfide forms, an escape route is provided by Cys55 which, we suggest, also becomes activated by Arg121 swinging into a new position, so that an intramolecular disulfide can form. This putative mechanism involves a crucial role assigned to Arg121, which is absolutely conserved in the DUF1094 family (Figure 3). We suggest that the escape route involves transient formation of a Cys53−Cys55 disulfide, although we have not confirmed this experimentally.

Oxidized CXC sequences, which are unusual 11-member ring structures, are extremely rare in nature. The only examples in the PDB of high-resolution structures containing such motifs are the thermophilic Ak.1 protease from *B. subtilis*, 1DBI (*39*), the L40C variant of the NADH peroxidase, 1NHR (*40*), and Erv2p, 1JR8 (*41*) (Figure 6). A similar disulfide is also reported in the *Mengo encephalomyelitis* virus coat protein (*42*), although low resolution precludes detailed stereochemical analysis. Sep15 has a redox-active CX(C/U) motif, where "U" refers to selenocysteine and the use of C or U varies between organisms (*43*). Finally, there is evidence that a CXC disulfide plays a functional role in the chaperone Hsp33 (*44*), but no structure of the oxidized protein is available. In all three cases where high-resolution data are available, the torsion angles of the disulfide-containing ring suggest significant strain. It is interesting to note that the backbone conformations in the 11-member rings are different: in the Ak.1 protease and in Erv2P the residue in the X position is in the left-handed helical conformation, while in the NADH peroxidase variant a serine occupying the equivalent position is in the right-handed helical conformation, very similar to Gly in YphP (Figure 6). Based on these structural comparisons, it appears that a disulfide bond in YphP is stereochemically possible, with a slight conformational adjustment to Cys55.

The observed activity of the two variants with only one of either of the active site cysteines is easily rationalized in view of the fact that the escape mechanism is not necessary for the enzyme to function in assays with peptides, as described elsewhere (*10*). In the wild-type protein, Cys53 is likely to be the key nucleophile that attacks the substrate. It is more solvent exposed than Cys55, and the proximity of Arg121 makes its thiolate form more stable. However, in the absence of Cys53, e.g., in the C53A variant, Arg121 could easily swing toward Cys55, which is still sufficiently exposed to be reactive. This arrangement is different

from that in *Saccharomyces cerevisiae* PDI, where the C-terminal Cys is buried and unable to catalyze isomerization in the absence of the N-terminal Cys. Thus, the C53A and C55A variants can both catalyze isomerization.

CONCLUSION

The *B. subtilis* YphP is a member of an unusual oxidoreductase superfamily of enzymes (DUF1094) with a natural CXC motif within the active site and a tertiary fold closely reminiscent of the thioredoxins. The experimentally determined reduction potential, $E^{o\prime}$, is -130 ± 5 mV. Functional assays show that YphP has a level of disulfide isomerase activity that might be biologically significant with natural substrates. The enzymes in this family are highly conserved and occur primarily in the *Staphylococcus* and *Bacillus* species. All of these organisms contain canonical genes that encode thioredoxins, which are cellular reducing agents that have only been crystallized in an oxidized state. It is possible that, unlike thioredoxins, the DUF1094 enzymes are cellular oxidizing agents or *in vivo* catalysts of oxidative protein folding. Further biological studies will be needed to answer this question.

REFERENCES

1. Ginalski, K., Elofsson, A., Fischer, D., and Rychlewski, L. (2003) 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics 19*, 1015–1018.
2. von Grotthuss, M., Pas, J., Wyrwicz, L., Ginalski, K., and Rychlewski, L. (2003) Application of 3D-Jury, GRDB, and Verify3D in fold recognition. *Proteins 53* (Suppl. 6), 418–423.
3. Pan, J. L., and Bardwell, J. C. (2006) The origami of thioredoxin-like folds. *Protein Sci. 15*, 2217–2227.
4. Quan, S., Schneider, I., Pan, J., Von Hacht, A., and Bardwell, J. C. (2007) The CXXC motif is more than a redox rheostat. *J. Biol. Chem. 282*, 28823–28833.
5. Grauschopf, U., Winther, J. R., Korber, P., Zander, T., Dallinger, P., and Bardwell, J. C. (1995) Why is DsbA such an oxidizing disulfide catalyst? *Cell 83*, 947–955.
6. Aslund, F., Berndt, K. D., and Holmgren, A. (1997) Reduction potentials of glutaredoxins and other thiol-disulfide oxidoreductases of the thioredoxin superfamily determined by direct protein-protein redox equilibria. *J. Biol. Chem. 272*, 30780–30786.
7. Laboissière, M. C., Sturley, S. L., and Raines, R. T. (1995) The essential function of protein-disulfide isomerase is to unscramble non-native disulfide bonds. *J. Biol. Chem. 270*, 28006–28009.
8. Zhang, R. M., and Snyder, G. H. (1989) Dependence of formation of small disulfide loops in two-cysteine peptides on the number and types of intervening amino acids. *J. Biol. Chem. 264*, 18472–18479.
9. Woycechowsky, K. J., and Raines, R. T. (2003) The CXC motif: a functional mimic of protein disulfide isomerase. *Biochemistry 42*, 5387–5394.
10. Kersteen, E. A., Barrows, S. R., and Raines, R. T. (2005) Catalysis of protein disulfide bond isomerization in a homogeneous substrate. *Biochemistry 44*, 12168–12178.
11. Stols, L., Gu, M., Dieckman, L., Raffen, R., Collart, F. R., and Donnelley, M. I. (2002) A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. *Protein Expression Purif. 25*, 1–7.
12. Dieckman, L., Gu, M., Stols, L., Donnelley, M. I., and Collart, F. R. (2002) High throughput methods for gene cloning and expression. *Protein Expression Purif. 25*, 8–15.
13. Goldschmidt, L., Cooper, D. R., Derewenda, Z. S., and Eisenberg, D. (2007) Toward rational protein crystallization: a Web server for the design of crystallizable protein variants. *Protein Sci. 16*, 1569–1576.
14. Derewenda, Z. S., and Vekilov, P. G. (2006) Entropy and surface engineering in protein crystallization. *Acta Crystallogr. D62*, 116–124.
15. Derewenda, Z. S. (2004) Rational protein crystallization by mutational surface engineering. *Structure 12*, 529–535.
16. Cooper, D. R., Boczek, T., Grelewska, K., Pinkowska, M., Sikorska, M., Zawadzki, M., and Derewenda, Z. (2007) Protein crystallization by surface entropy reduction: optimization of the SER strategy. *Acta Crystallogr. A63*, 636–645.
17. Newman, J. (2005) Expanding screening space through the use of alternative reservoirs in vapor-diffusion experiments. *Acta Crystallogr. D61*, 490–493.
18. Schneider, T. R., and Sheldrick, G. M. (2002) Substructure solution with SHELXD. *Acta Crystallogr. D58*, 1772–1779.
19. Terwilliger, T. C. (2003) SOLVE and RESOLVE: automated structure solution and density modification. *Methods Enzymol. 374*, 22–37.
20. Morris, R. J., Perrakis, A., and Lamzin, V. S. (2003) ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol. 374*, 229–244.
21. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A47*, 110–119.
22. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D60*, 2126–2132.
23. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D53*, 240–255.
24. Zwart, P. H., Afonine, P. V., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., McKee, E., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., Storoni, L. C., Terwilliger, T. C., and Adams, P. D. (2008) Automated structure solution with the PHENIX suite. *Methods Mol. Biol. 426*, 419–435.
25. Winn, M. D., Isupov, M. N., and Murshudov, G. N. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D57*, 122–133.
26. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol. 372*, 774–797.
27. Lees, W. J., and Whitesides, G. M. (1993) Equilibrium constants for thiol-disulfide interchange reactions: a coherent, corrected set. *J. Org. Chem. 58*, 642–647.
28. Holm, L., Kaariainen, S., Wilton, C., and Plewczynski, D. (2006) Using Dali for structural comparison of proteins, Curr. Protoc. Bioinf., Chapter 5, Unit 5 5.
29. Andersen, J. F., Sanders, D. A., Gasdaska, J. R., Weichsel, A., Powis, G., and Montfort, W. R. (1997) Human thioredoxin homodimers: regulation by pH, role of aspartate 60, and crystal structure of the aspartate 60 → asparagine mutant. *Biochemistry 36*, 13979–13988.
30. Lesniak, J., Barton, W. A., and Nikolov, D. B. (2002) Structural and functional characterization of the *Pseudomonas* hydroperoxide resistance protein Ohr. *EMBO J. 21*, 6649–6659.
31. Janda, I., Devedjiev, Y., Derewenda, U., Dauter, Z., Bielnicki, J., Cooper, D. R., Graf, P. C., Joachimiak, A., Jakob, U., and Derewenda, Z. S. (2004) The crystal structure of the reduced, Zn^{2+}-bound form of the *B. subtilis* Hsp33 chaperone and its implications for the activation mechanism. *Structure 12*, 1901–1907.
32. Finn, R. D., Tate, J., Mistry, J., Coggill, P. C., Sammut, S. J., Hotz, H. R., Ceric, G., Forslund, K., Eddy, S. R., Sonnhammer, E. L., and Bateman, A. (2008) The Pfam protein families database. *Nucleic Acids Res. 36*, D281–D288.
33. Woycechowsky, K. J., and Raines, R. T. (2000) Native disulfide bond formation in proteins. *Curr. Opin. Chem. Biol. 4*, 533–539.
34. Wilkinson, B., and Gilbert, H. F. (2004) Protein disulfide isomerase. *Biochim. Biophys. Acta 1699*, 35–44.
35. Kersteen, E. A., and Raines, R. T. (2003) Catalysis of protein folding by protein disulfide isomerase and small-molecule mimics. *Antioxid. Redox Signaling 5*, 413–424.
36. Vuori, K., Myllylä, R., Pihlajaniemi, T., and Kivirikko, K. I. (1992) Expression and site-directed mutagenesis of human protein disulfide isomerase in *Escherichia coli*. *J. Biol. Chem. 267*, 7211–7214.
37. Walker, K. W., Lyles, M. M., and Gilbert, H. F. (1996) Catalysis of oxidative protein folding by mutants of protein disulfide

isomerase with a single active-site cysteine. *Biochemistry 35*, 1972–1980.

38. Chivers, P. T., Laboissière, M. C. A., and Raines, R. T. (1996) The CXXC motif: imperatives for the formation of native disulfide bonds in the cell. *EMBO J. 16*, 2659–2667.

39. Smith, C. A., Toogood, H. S., Baker, H. M., Daniel, R. M., and Baker, E. N. (1999) Calcium-mediated thermostability in the subtilisin superfamily: the crystal structure of *Bacillus* Ak.1 protease at 1.8 Å resolution. *J. Mol. Biol. 294*, 1027–1040.

40. Miller, H., Mande, S. S., Parsonage, D., Sarfaty, S. H., Hol, W. G., and Claiborne, A. (1995) An L40C mutation converts the cysteine-sulfenic acid redox center in enterococcal NADH peroxidase to a disulfide. *Biochemistry 34*, 5180–5190.

41. Gross, E., Sevier, C. S., Vala, A., Kaiser, C. A., and Fass, D. (2002) A new FAD-binding fold and intersubunit disulfide shuttle in the thiol oxidase Erv2p. *Nat. Struct. Biol. 9*, 61–67.

42. Krishnaswamy, S., and Rossmann, M. G. (1990) Structural refinement and analysis of Mengo virus. *J. Mol. Biol. 211*, 803–844.

43. Ferguson, A. D., Labunskyy, V. M., Fomenko, D. E., Arac, D., Chelliah, Y., Amezcua, C. A., Rizo, J., Gladyshev, V. N., and Deisenhofer, J. (2006) NMR structures of the selenoproteins Sep15 and SelM reveal redox activity of a new thioredoxin-like family. *J. Biol. Chem. 281*, 3536–3543.

44. Jakob, U., Muse, W., Eser, M., and Bardwell, J. C. (1999) Chaperone activity with a redox switch. *Cell 96*, 341–352.

45. Qi, Y., and Grishin, N. V. (2005) Structural classification of thioredoxin-like proteins. *Proteins*: *Struct. Funct. Bioinf. 58*, 376–388.