Protein Science (1997), 6:1148–1156. Cambridge University Press. Printed in the USA.

The uncharged surface features surrounding the active site of *Escherichia coli* DsbA are conserved and are implicated in peptide binding

LUKE W. GUDDAT,[1] JAMES C.A. BARDWELL,[2] THOMAS ZANDER,[2] AND JENNIFER L. MARTIN[1]

[1] Centre for Drug Design and Development, University of Queensland, Brisbane, QLD 4072, Australia
[2] Department of Biology, University of Michigan, Ann Arbor, Michigan, 48109-1048

(RECEIVED January 6, 1997; ACCEPTED March 7, 1997)

Abstract

DsbA is a protein-folding catalyst from the periplasm of *Escherichia coli* that interacts with newly translocated polypeptide substrate and catalyzes the formation of disulfide bonds in these secreted proteins. The precise nature of the interaction between DsbA and unfolded substrate is not known. Here, we give a detailed analysis of the DsbA crystal structure, now refined to 1.7 Å, and present a proposal for its interaction with peptide.

The crystal structure of DsbA implies flexibility between the thioredoxin and helical domains that may be an important feature for the disulfide transfer reaction. A hinge point for domain motion is identified—the type IV β-turn Phe 63–Met 64–Gly 65–Gly 66, which connects the two domains.

Three unique features on the active site surface of the DsbA molecule—a groove, hydrophobic pocket, and hydrophobic patch—form an extensive uncharged surface surrounding the active-site disulfide. Residues that contribute to these surface features are shown to be generally conserved in eight DsbA homologues. Furthermore, the residues immediately surrounding the active-site disulfide are uncharged in all nine DsbA proteins.

A model for DsbA–peptide interaction has been derived from the structure of a human thioredoxin:peptide complex. This shows that peptide could interact with DsbA in a manner similar to that with thioredoxin. The active-site disulfide and all three surrounding uncharged surface features of DsbA could, in principle, participate in the binding or stabilization of peptide.

Keywords: DsbA; oxidoreductase; peptide interaction; protein crystallography; protein disulfide isomerase; thioredoxin fold

Disulfides stabilize the tertiary structures of most secreted proteins. The introduction of native disulfides into newly folding proteins is often the critical step in the protein folding pathway of exported proteins. In vivo, the process is catalyzed by specific proteins. In eukaryotes, disulfide bond formation occurs in the endoplasmic reticulum and is catalyzed by protein disulfide isomerase [for a review see Freedman et al. (1994)]. In prokaryotes, the process occurs in the periplasm and is catalyzed by the redox protein DsbA (Bardwell, 1994). DsbA catalyzes disulfide bond formation by transferral of its active-site disulfide to a folding protein substrate, via a mixed disulfide intermediate (Wunderlich & Glockshuber, 1993a, 1993b; Zapun et al., 1994; Zapun & Creighton, 1994). Reduced DsbA, which is produced from this reaction, is re-oxidized through interaction with the integral membrane protein DsbB (Bardwell et al., 1993; Missiakas et al., 1993; Guilhot et al., 1995).

We are investigating the structure and function of DsbA using mutational and crystallographic methods. The crystal structure of oxidized DsbA (refined to a resolution of 2.0 Å) was reported originally in a short letter (Martin et al., 1993a). The structure was shown to incorporate a thioredoxin fold (Martin, 1995) and a helical domain insert and to include several unusual surface features. Here, we give a more detailed description of the DsbA structure, highlighting the possibility of conformational change and defining the unique features of the structure that could be important for interaction with polypeptides and proteins. This description also represents a basis for structural comparison of DsbA mutants and DsbA homologues, some of which are being investigated currently in our laboratories. For our analysis, we use the crystal structure of oxidized wild-type (wt) *Escherichia coli* DsbA, which is now refined to a resolution of 1.7 Å.

Reprint requests to: Jennifer L. Martin, Centre for Drug Design and Development, University of Queensland, Brisbane, QLD 4072, Australia; e-mail: j.martin@mailbox.uq.oz.au.

The crystal structure of DsbA is analyzed for evidence of domain motion by comparing the relative orientations of the two domains in each of the two copies of DsbA in the asymmetric unit. In addition, we examine the characteristic surface features (including a hydrophobic pocket not identified previously) and assess the conservation of contributing residues among the DsbA family. The interaction of peptide with three of these surface features is modeled on the basis of a thioredoxin:peptide complex (Qin et al., 1996b).

Results

The refined 1.7 Å crystal structure of oxidized wt DsbA

Details of data collection statistics for the 1.7 Å wt oxidized DsbA are given in Table 1. DsbA is a monomer, but crystallizes with two molecules in the asymmetric unit. The final 1.7-Å model includes the two DsbA monomers (one of which is shown in Fig. 1) and 268 water molecules in the asymmetric unit. The crystallographic *R*-factor and *R*-free (Brünger, 1992a) for the structure are, respectively, 0.196 and 0.229 (for details see Table 2). For comparison, the DsbA structure refined at 2.0 Å has an *R*-factor of 0.169 (Martin et al., 1993a), but was refined without the use of *R*-free.

Overall, the 1.7-Å and 2.0-Å oxidized wt DsbA structures are very similar, although there are some local differences (described in Materials and methods). The root-mean-square deviation (RMSD) between the two structures for all Cα atoms of both monomers is 0.15 Å. As described previously (Martin et al., 1993a), the structure of DsbA consists of a thioredoxin fold (residues 1–62 and 139–188) with a helical domain insert (residues 63–138).

Comparison of monomers A and B in the asymmetric unit

The high resolution of the wt DsbA structure allows analysis of the structural differences between the two monomers in the asymmetric unit (Fig. 2). Superimposition of all Cα atoms of monomers A and B gives an RMSD of 0.91 Å. This is much higher than the estimated coordinate error (0.2–0.3 Å) of the structure and also higher than the RMSD obtained in comparing the 2.0-Å and 1.7-Å structures (0.15 Å).

Table 1. *Crystallographic data collection statistics*[a]

Resolution (Å)	61.4–1.66
No. of observations	128,237
No. of reflections	48,880
R_{sym}[b]	0.064
R_{sym} (outer shell)[c]	0.328
$I/\sigma(I)$	14.1
$I/\sigma(I)$ (outer shell)	2.35
Completeness	89.0%
Completeness (outer shell)	67.4%

[a]Statistics are for data with $I > 1\ \sigma(I)$.
[b]$R_{sym} = \Sigma|I - \langle I\rangle|/\Sigma\langle I\rangle$.
[c]Outer shell is 1.66–1.72 Å.

The structural differences between the two monomers appear to arise from different crystal packing environments and the inherent mobility of solvent-exposed regions. The most striking example is at the N-terminus (Fig. 2); Gln 2 has a backbone α-helical conformation in monomer A and a β-strand conformation in monomer B. Accompanying this change in main-chain conformation is a 9-Å shift in the mean coordinate position of Ala 1. If just these two N-terminal residues are excluded from the superimposition, the Cα RMSD for monomers A and B drops to 0.57 Å. The B conformation of residues 1 and 2 cannot be adopted by monomer A in the crystal lattice because a steric clash would occur with Lys 70 and Ala 1 side chains from a symmetry-related molecule. In the reverse case, where the conformation of residues 1 and 2 from monomer A is overlaid onto monomer B in the crystal, there are no symmetry clashes. Consequently, the N-terminal conformation of monomer B is probably more representative of the unhindered solution structure.

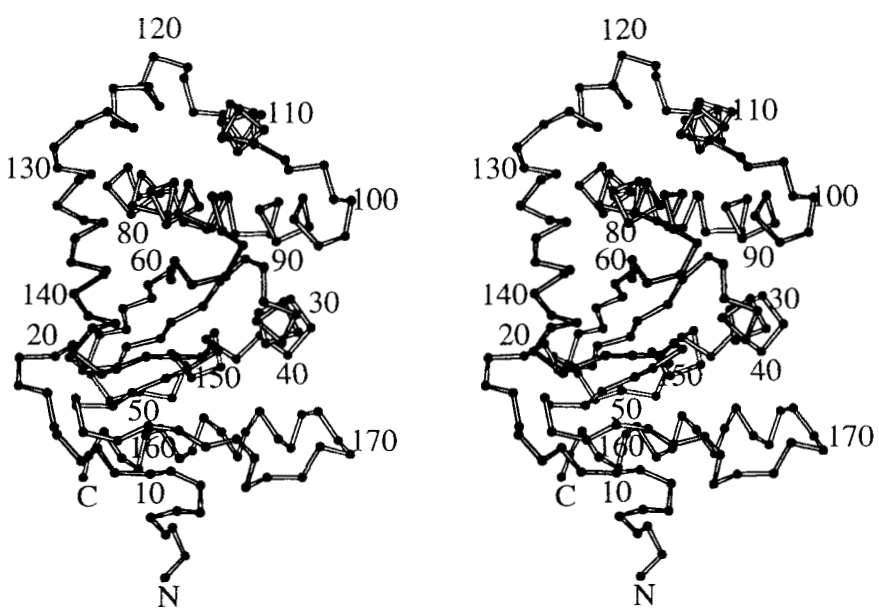

Fig. 1. Stereo view of the structure of oxidized wt *E. coli* DsbA (from monomer A in the asymmetric unit). This figure was prepared with the program MOLSCRIPT (Kraulis, 1991).

Table 2. *Crystallographic refinement statistics*

	wt 1.7 Å[a]	wt 2.0 Å[b]
Resolution range (Å)	50.0–1.7	6–2.0
No. of reflections	46,502	25,426
R-factor[c]	0.196	0.169
R-free[d]	0.229	—
RMSDs from ideal		
Bond length (Å)	0.006	0.011
Bond angle (°)	1.39	1.43
Dihedral angle (°)	21.86	21.96
Improper angle (°)	1.06	1.19
Average *B*-factor[e] (Å^2)	33.1	37.8
No. of water molecules	268	195
Ramachandran plot statistics		
Residues in most favored regions (%)	94.3	93.7
Residues in additionally allowed regions (%)	5.7	6.3
Residues in disallowed regions (%)	0.0	0.0

[a]For data with ($F > 1\sigma(F)$).
[b]For data with ($F > 2\sigma(F)$), see Martin et al. (1993a).
[c]R-factor = $\sum|F_{obs} - F_{calc}|/\sum F_{obs}$.
[d]R-free as defined by Brünger (1992a), using 10% of the data.
[e]Including waters.

The other large difference between the two monomers is in the stretch of residues 163–174 (Fig. 2), located directly after strand β5 in the polypeptide sequence. These residues form a mobile loop (residues 163–169, average B-factor 47 Å^2) and the first turn of helix α7 (residues 170–174, average B-factor 33 Å^2). As with the N-terminus, the conformational changes in this region are a result of different crystal contacts and intrinsic flexibility.

Domain motion in DsbA

If residues 1–2 and 163–174 are removed from the Cα superimposition of monomers A and B, the RMSD falls to 0.43 Å for the 174 residues. This is still somewhat higher than the estimated coordinate error of the structure. We therefore investigated the possibility of domain mobility in DsbA by superimposing, separately, only the secondary structural features of each domain (helix α6 was not included because it straddles both domains). For the thioredoxin domain, Cα atoms from 62 residues (8–12, 21–37, 41–50, 55–61, 152–156, 158–162, 175–187) superimpose to give an RMSD of 0.25 Å. Similarly, superimposition of Cα atoms from 47 residues of the helical domain (67–81, 85–97, 105–114, 119–127) gives an RMSD of 0.27 Å. When the Cα atoms of these residues from *both* domains are superimposed, the resulting RMSD is 0.34 Å for the 109 residues.

To analyze further the possibility of domain motion, the angular and translational differences between the domains in monomers A and B were calculated. First, the two monomers were superimposed using the same Cα atoms from secondary structural elements of the helical domain defined above. Then, using this helical domain superimposition as the starting point, the monomers were again overlaid—this time using only thioredoxin domain elements of the structure. The movement required to achieve this second superimposition corresponds to the difference in domain position between the two monomers. This movement in DsbA involves a translation of 0.66 Å along the major axis of the domain and three angular modifications, a "roll" of −2.4° along the long axis of the domain, plus a "tilt" of 1.8° and a "swivel" of −2.0° [for a description of these terms, see Huber et al. (1971) and Herron et al. (1991)]. Compared with the range of angular and translational differences that occur at the V_L-V_H domain interface on binding of antigen (Herron et al., 1991; Stanfield et al., 1993), the "motion" in the DsbA domains is small but significant. The difference observed between the two molecules in the crystal structure could represent the lower bound, and by no means the upper bound, of any possible domain movement in DsbA.

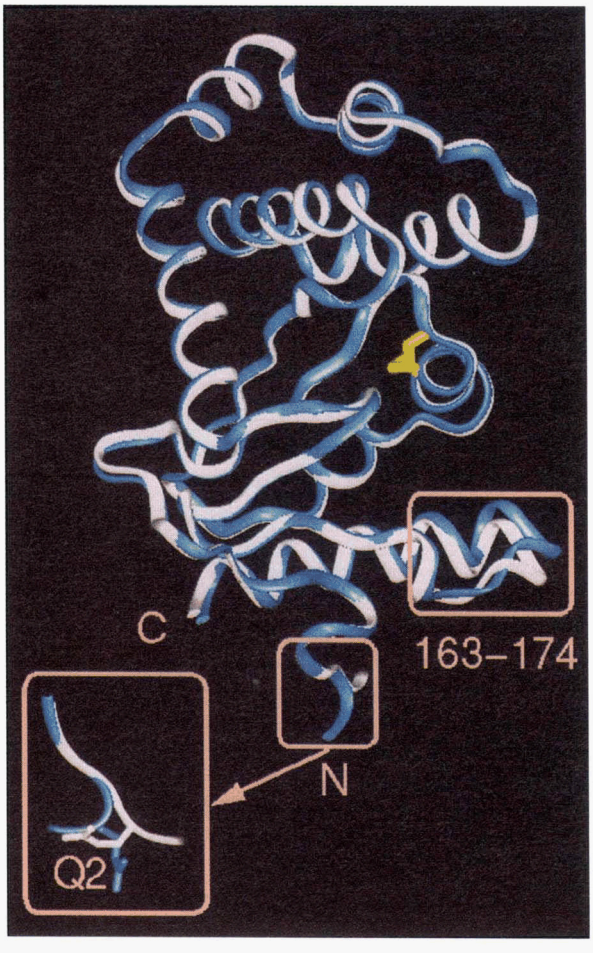

Fig. 2. Comparison of the two monomers of DsbA. Superimposition of the two monomers (monomer A in blue, monomer B in white), indicating the regions of largest main-chain difference at the N terminus (N) and the flexible loop region (residues 163–174). The N terminus is magnified to show the two alternate conformations, with residue Gln 2 labeled (Q2).

Furthermore, the connection between the two DsbA domains—between strand β3 and helix α2—is a type IV β-turn contributed from residues Phe 63–Met 64–Gly 65–Gly 66. The difference in relative domain orientation observed for the two DsbA monomers in this crystal form could be accounted for by differences of 10–20° in the main-chain ϕ,ψ torsion angles of Gly 65 and Gly 66. In addition, the first residue of this turn, Phe 63, is highly strained—having an unusually large main-chain angle for N-Cα-C of ~120° (ideal value ~110°) in both monomers. We suggest that this β-turn therefore represents a hinge point for domain motion.

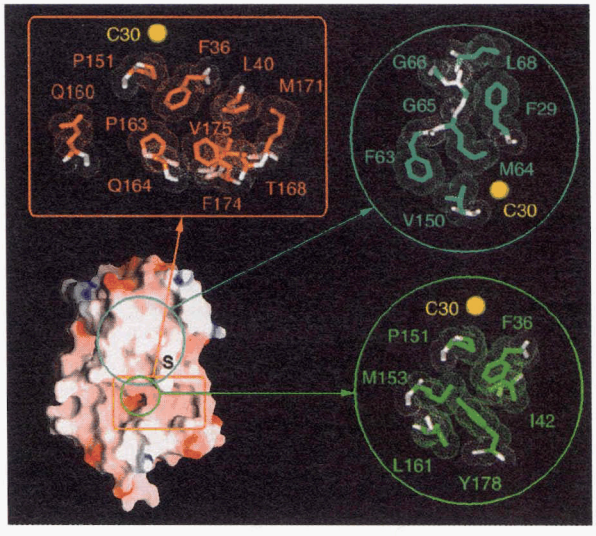

Fig. 3. Characteristic surface features of *E. coli* DsbA. The electrostatic surface of DsbA [generated using GRASP (Nicholls et al., 1993)], colored according to charge (red for negative, blue for positive, white for uncharged regions) is shown in the same orientation as Figure 1. The positions of the disulfide active site (S), the groove (orange), the pocket (green), and the hydrophobic patch (aqua-blue) are indicated. Residues Phe 36 and Pro 151 contribute to both the groove and the pocket. Magnified views show residues contributing to each of these features, with the position of the Sγ atom of Cys 30 shown as a yellow sphere. Note that residues Leu 40, Thr 168, and Met 171 lining the groove in *E. coli* DsbA are located in parts of the sequence that are not strictly conserved in all nine homologues.

The proposed substrate binding surface of DsbA

Several studies have shown that DsbA interacts with other parts of the folding polypeptide substrate in addition to the cysteine thiol (Wunderlich et al., 1993; Darby & Creighton, 1995; Frech et al., 1996). The details of the interaction between DsbA and polypeptide are not known. We propose that the surface of the DsbA structure surrounding the active-site disulfide is critical for interaction with unfolded substrate. Figure 3 depicts the electrostatic surface of one face of the DsbA molecule. This face includes the accessible sulfur of the active-site disulfide, an uncharged groove below, and a hydrophobic patch above the accessible sulfur [identified in Martin et al. (1993a)]. Within the groove is a hydrophobic pocket, not identified previously. Each of these characteristic surface features is described in detail below, and the residues that contribute to each feature are identified. The sequence alignment of nine DsbA proteins (Fig. 4) gives a useful guide to the conservation of these residues (and hence conservation of the surface characteristics) within the DsbA family.

A peptide binding groove . . .

The groove below the active-site disulfide is approximately 20-Å long, 10-Å wide, and 7-Å deep. It incorporates an indentation, or pocket, directly below the active-site disulfide that could accommodate a small hydrophobic side-chain (see below). The groove is formed exclusively from residues of the thioredoxin domain and is lined with several solvent-exposed and uncharged (hydrophobic and polar) residues. These residues are from the α1–loop–α1′ helix (Phe 36, Leu 40), the active site *cis*-Pro 151, strand β5 (Gln 160) and several residues from the flexible loop region between β5 and the first turn of helix α7 (Pro 163, Gln 164 and Thr 168, Met 171, Phe 174, Val 175).

Of these 10 residues, three (Phe 36, Leu 40, Pro 151) are conserved as hydrophobic, two (Pro 163, Thr 168) are conserved as uncharged residues, and two (Phe 174, Val 175) are uncharged in eight of the nine DsbA sequences. The remaining three residues (Gln 160, Gln 164, and Met 171) are variable. Some explanation is required regarding Leu 40, which is not conserved, in the strictest sense of the term, in all nine proteins. This residue is located in a three-residue loop that kinks the α1/α1′ helix of *E. coli* DsbA and is found in only five of the nine DsbA proteins. However, in all five cases, the residue aligned with Leu 40 is hydrophobic. In the other four cases, the residue following the three-residue loop deletion is hydrophobic.

. . . hydrophobic pocket within the groove . . .

The hydrophobic pocket is also located within the thioredoxin domain, directly below the active-site disulfide at the midpoint of the groove described above, with dimensions approximately 6 Å by 7 Å, and 5-Å deep. The pocket is large enough to accommodate a small hydrophobic group—the side-chain of an alanine or valine residue, for instance. Residues that form this pocket are: Phe 36 and Ile 42 from the α1–loop–α1′ helix, *cis*-Pro 151 from the active site, Met 153 from β4, Leu 161 from β5, and Tyr 178 from α7. The buried acidic residue Glu 24 of strand β2 (see below) lies below the pocket and contributes to the negative electrostatic charge within the pocket (red tint of the hydrophobic pocket in Fig. 3), although this buried residue may not be charged at physiological pH [the equivalent residue in thioredoxin has a pK_a above 7 (Jeng et al., 1995; Wilson et al., 1995; Jeng & Dyson, 1996; Qin et al., 1996a)].

Residues Phe 36 and *cis*-Pro 151, which are identified above as forming part of the putative peptide binding groove, are highly conserved in the DsbA family. The hydrophobicity of the other four residues is completely conserved in the nine DsbA sequences except in two cases: Ile 42 is a serine in *Azotobacter vinelandii* and Leu 161 is a threonine in *Legionella pneumophila*.

. . . and the hydrophobic patch

The hydrophobic patch, an area of roughly 10 Å × 15 Å located above the active-site disulfide in the orientation shown in Figure 3, is composed of exposed hydrophobic residues. The residues contributing to this feature are donated from both the thioredoxin and helical domains. These are: Phe 29 from a type I β-turn preceding the active-site disulfide; all four residues in the type IV "hinge-point" β-turn (Phe 63–Met 64–Gly 65–Gly 66) that connects the thioredoxin and helical domains; Leu 68 from helix α2; and Val 150 from a loop located between α6 and β4. Two of these residues—Phe 29 and Phe 63—are conserved as hydrophobic residues (Phe, Tyr, Val, Ile, Leu, Met, or Gly) in the nine sequences. However, the Phe 63–Gly 66 stretch of residues is variable in length. Residue 64 is present in only six sequences (but all are hydrophobic) and residue 65 is present in only three (all three are Gly). Of the remaining three residues, Leu 68 and Val 150 are conserved as uncharged (either hydrophobic or polar) and Gly 66 is uncharged in eight of the nine sequences.

Model of a DsbA–peptide complex

The conservation of uncharged/hydrophobic residues from the surface features surrounding the DsbA active site implies that these

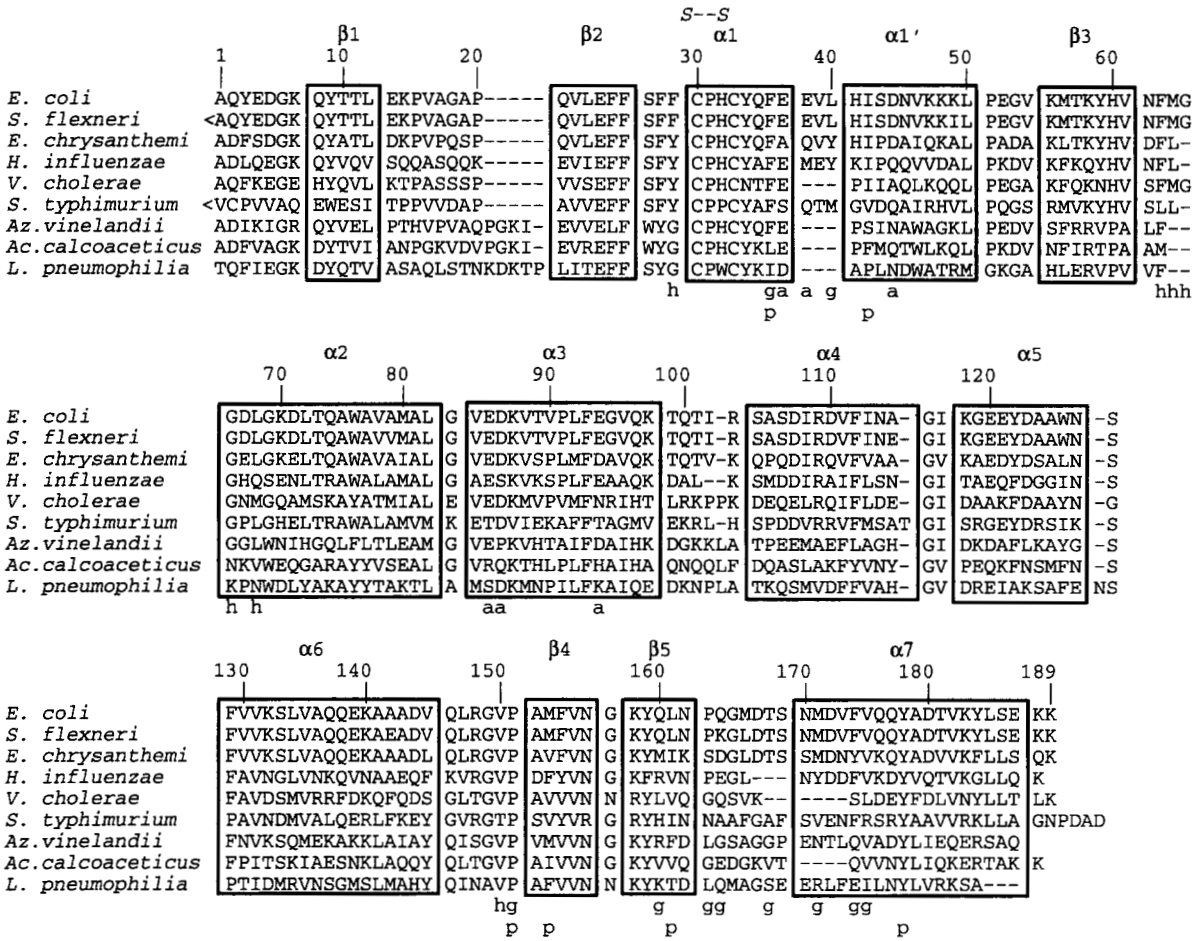

Fig. 4. Sequence alignment of DsbA homologues. Sequences are arranged in descending sequence identity with *E. coli* DsbA (*S. flexnerii* 96%, *E. chrysanthemi* 66%, *H. influenzae* 45%, *V. cholerae* 38%, *S. typhimurium* 34%, *Az. vinelandii* 31%, *Ac. calcoaceticus* 29%, *L. pneumophila*, 21%). The secondary structure identified from *E. coli* DsbA is boxed and labeled, and the active-site disulfide is labeled S–S. The N-terminal residues are those predicted after signal sequence cleavage, except in the case of *S. typhimurium* (where no reasonable cleavage site was predicted) and *Az. vinelandii* (where the predicted cleavage site has extra residues at the N terminus). For both these sequences, the N-terminus is shown flush with the other DsbA homologues, although this may not be accurate. The *S. typhimurium* sequence is not a chromosomal sequence, but is from a virulence plasmid, so it could have a somewhat different function from the chromosomal copy. Residues from characteristic structural features of DsbA are denoted "g" for those lining the groove, "p" for those forming the pocket, "h" for those in the hydrophobic patch, and "a" for those in the acidic patch. (See text and Figs. 3, 6.)

uncharged regions are also present on other DsbA proteins. Furthermore, the placement of these uncharged regions around the active-site disulfide suggests that they may be important for binding and/or stabilization of hydrophobic residues in polypeptide substrates.

A model of a DsbA–peptide complex was derived by comparison of the structure of unliganded DsbA with the NMR structure of a complex of human thioredoxin with its target peptide from Ref-1 [(Qin et al., 1996b), PDB accession code 1CQG]. The thioredoxin domain of DsbA was superimposed onto the human thioredoxin structure, and the thioredoxin-bound peptide was examined in terms of the DsbA structure. In this model, the peptide is quite clearly oriented in a manner that would favor interaction with all three of the uncharged regions surrounding the active-site disulfide of DsbA (Fig. 5). The DsbA–peptide model provides further confirmation that the uncharged features may be involved in substrate binding, but does not give precise detail of these interactions.

Also shown in Figure 5 is a depiction of the active-site surface of DsbA, highlighting all residues that are uncharged in all nine DsbA sequences. Notably, the active-site disulfide is encompassed by a broad patch of conserved uncharged residues contributed from all three surface features—the groove, pocket, and patch. Comparison of the DsbA–peptide model with the conserved uncharged residues in DsbA shows that the model peptide binding site is contained entirely within the conserved uncharged surface. We submit that these observations lend strong support to the hypothesis that the conserved surface features of DsbA are required for interaction with regions of the unfolded substrate that ultimately form the hydrophobic core of the folded protein product.

The acidic patch

The acidic patch (Fig. 6) is located on a groove between the thioredoxin and helical domains on the opposite face of the mol-

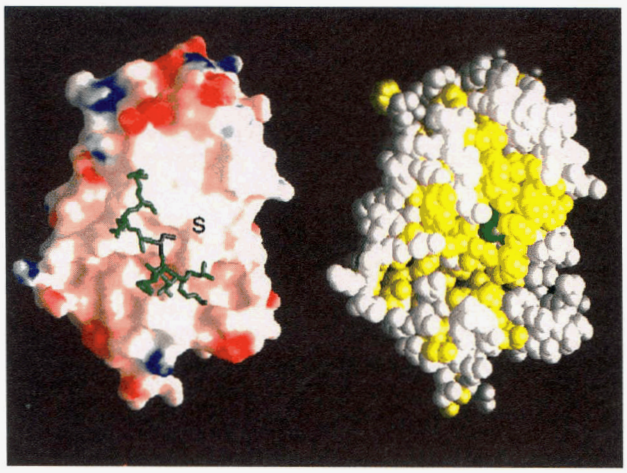

Fig. 5. The proposed peptide interaction surface of DsbA. Left: A model of the interaction between peptide substrate and DsbA, based on the human thioredoxin:Ref-1 peptide complex [(Qin et al., 1996b) PDB accession code 1CQG]. The electrostatic surface of DsbA is shown [generated using GRASP (Nicholls et al., 1993)] with the peptide in the thioredoxin-bound conformation superimposed. The position of the accessible sulfur of the active-site disulfide is denoted by "S." Right: CPK model of DsbA, showing conserved uncharged residues in yellow. Residues with at least one occurrence of a charged amino acid (Asp, Glu, Lys, or Arg) in the nine aligned sequences are shown in white. The conserved active-site cysteines are shown in green.

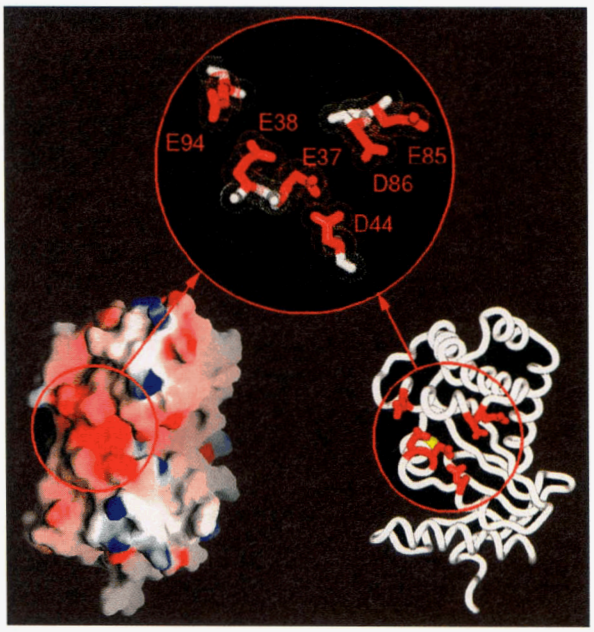

Fig. 6. Acidic patch of *E. coli* DsbA. The acidic patch is located on the opposite face of the molecule to that shown in Figure 3. Lower left: Electrostatic surface [generated using GRASP (Nicholls et al., 1993)]. Lower right: Ribbon representation of *E. coli* DsbA. Side chains of acidic residues are shown in red. The magnified view identifies each of these residues and includes their van der Waals surface.

ecule to that presenting the active-site disulfide. The groove is wedge-shaped, measuring approximately 14-Å long and varying in width (8–11 Å) and depth (5–12 Å). The sides of the wedge are formed from the α1–loop–α1′ helix of the thioredoxin domain and α3 of the helical domain, with the floor composed of residues from β2 of the thioredoxin domain.

The seven acidic residues in this region of *E. coli* DsbA are Glu 24 from strand β2, Glu 37, Glu 38, and Asp 44 from the α1–loop–α1′ region, and Glu 85, Asp 86, and Glu 94 from helix α3. Although it is conserved among the DsbA sequences, Glu 24 is probably not involved in the acidic patch because it is completely buried. Of the remaining six acidic residues, some are reasonably well-conserved among the DsbA sequences (Glu or Asp at position 37 is present in 8/9 sequences, Glu 85 and Asp 86 in 6/9), but others are not conserved at all (Glu 38 is present in only 2/9 and Asp 44 in only 3/9).

The lack of precise conservation of acidic residues prompted us to examine the *E. coli* DsbA structure to identify solvent-accessible residues in the vicinity of the acidic patch. We reasoned that a precise alignment of charged residues of DsbA homologues with those in *E. coli* DsbA is not required to produce a net negative charge in this region. Presumably, nearby residues could also contribute to the overall charge in the acidic patch of DsbA homologues. We found that residues 37, 38, 41, 43, 44, 47, 48 from the α1–loop–α1′, residue 80 from the last turn of helix α2, residue 83 from the turn between α2 and α3, residues 85, 86, 87, 89, 90, 94, 98 from helix α3, and residue 99 from the turn following α3 are solvent accessible and could contribute to the overall charge in the acidic patch. In *E. coli* DsbA, 6 of these 17 residues are acidic and 4 are basic, giving a net charge of -2 for this region. For the other eight sequences, the ratio of acidic/basic residues is 6/4 for *Shigella flexneri*, 4/3 for *Erwinia chrysanthemi*, 5/4 for *Haemophilus influenzae*, 4/2 for *Vibrio cholerae*, 4/3 for *Salmonella typhimurium*, 5/1 for *Azotobacter vinelandii*, 2/2 for *Acinetobacter calcoaceticus*, and 4/2 for *Legionella pneumophila*. Thus, it appears that eight of the nine DsbA proteins have a net negative charge in the region of the acidic patch of *E. coli* DsbA.

Discussion

The structure of DsbA combines several unique features that may be critical for function. First, there is the intriguing possibility of domain flexibility. The movements noted upon comparison of the domains of the two monomers in the asymmetric unit are small and could, on their own, appear circumstantial. The evidence for domain motion is more compelling, given the additional observation of a strained residue (in both monomers) at a position placed perfectly for hinge motion between the two domains. We intend to investigate the possibility of larger domain motions in DsbA by studying the structures of DsbA–peptide complexes, reduced DsbA, and oxidized DsbA from alternative crystal forms.

The four striking surface features of the DsbA structure (the peptide binding groove, hydrophobic pocket, hydrophobic patch, and acidic patch) appear to be conserved throughout the DsbA family. Conservation of these characteristic traits further supports the hypothesis that these features are required for function. The precise role each surface feature plays must remain speculative at this stage, but there are some clues. Thus, the dimensions and location of the groove near to the active-site disulfide is suggestive of a peptide binding groove. The flexibility of residues 163–174, lining one edge of the groove, could be significant in the context of the nonspecific nature of the interaction between DsbA and substrate. The hydrophobic pocket within the groove could interact with the side chain of a small hydrophobic residue. Residues from the groove, the pocket and the hydrophobic patch form a con-

served, uncharged surface surrounding the accessible sulfur of the DsbA active site. All three of these surface features are implicated in substrate binding based on a model of the DsbA–peptide complex. However, the precise details of the binding interaction between DsbA and substrate peptide must await definitive experimental evidence.

The location of the acidic patch on the molecular face opposite to that with the active-site disulfide implies that this surface feature is not important in catalysis of disulfide bond interchange. In addition, the lack of precise alignment of the acidic residues in the nine DsbA sequences suggests that protein–protein interactions that might target the acidic patch may be nonspecific and mediated by electrostatic attraction. A candidate for interaction at this site is DsbB, an integral membrane protein that assists the function of DsbA by re-oxidizing its disulfide bond. The interaction between DsbA and DsbB would necessitate molecular recognition over large distances, a requirement that is met by electrostatic attraction. Nothing is known about the three-dimensional structure of DsbB, although the transmembrane topology (Jander et al., 1994) does not predict highly charged periplasmic domains.

In conclusion, our detailed analysis of the DsbA structure suggests the possibility of domain motion for function, and identifies a potential hinge point. The characteristic surface features on the active-site face of the DsbA structure are shown to be conserved among the family of DsbA proteins. These features—a peptide groove, hydrophobic pocket, and hydrophobic patch—form a distinctive uncharged surface around the active site that is conserved and likely to be involved in substrate binding. The significance of the acidic patch on the opposite face of DsbA is unclear, but its conservation suggests a function that we propose could be related to the binding of DsbB.

Materials and methods

Protein crystallization and X-ray data collection

Recombinant wt DsbA was expressed and purified in *E. coli* as described previously (Bardwell et al., 1991). Samples for crystallization were oxidized using 1.5 mM copper [II] phenanthroline, dialyzed against 20 mM Hepes/NaOH, pH 7.5, and purified by FPLC on a HR 5/5 Mono Q column followed by concentration to 20 mg/mL.

Crystals of wt DsbA (space group $C2$, unit cell dimensions a = 117.7 Å, b = 65.1 Å, c = 76.4 Å, β = 126.3°) were grown using conditions described previously (Martin et al., 1993b). The calculated solvent content with two DsbA molecules per asymmetric unit is 55.9%. A triangular-shaped crystal of dimensions 1.1 × 0.8 × 0.5 mm was used for data collection. Crystallographic data were measured at 16 °C using an RAXIS-IIC imaging plate area detector with RU-200 rotating anode generator. The X-ray generator was operated at 50 kV and 100 mA and the crystal to detector distance was 65 mm. An oscillation range of 2° was used with an exposure time of 30 min. The frames were integrated, scaled, and merged (with data cutoff $I > 1\ \sigma(I)$) using the data processing package provided with the RAXIS-IIC system (Higashi, 1990). Crystallographic statistical results for wt DsbA crystals are presented in Table 1.

Crystallographic refinement of wt DsbA structure

The previously refined 2.0-Å structure of DsbA (PDB accession code 1DSB) (Martin et al., 1993a), with solvent molecules removed, was used as the starting model for the high-resolution structure. Although data were measured to 1.66 Å, a cutoff of 1.7 Å was used for refinement: the R-factor and R-free for the 1.7–1.66-Å shell are both ~40% and data completeness in this resolution shell is less than 50%, with multiple measurements for only 48 of the 2,295 reflections. Structural refinement was guided by the use of R-free (Brünger, 1992a; Kleywegt & Jones, 1996) at all stages. Refinement was performed in X-PLOR v. 3.1 (Brünger, 1992b) using simulated annealing, positional refinement, individual B-factor refinement, and bulk solvent correction, according to standard protocols and using standard parameters (Engh & Huber, 1991). Models were built and examined using the graphics program O (Jones et al., 1991).

Water molecules were included in the structure using these criteria: peak above 3σ in $F_o - F_c$ maps and peak above 1σ in $2F_o - F_c$ maps and at least one hydrogen bonding partner within a 3.4-Å radius. Residues Ser A133, Gln A176, Glu A187, and Ser B133 were assigned two conformations of half occupancy each. Residues Glu A13, Lys A14, Glu A52, Lys A55, Arg A148, Gln A164, Glu B13, Lys B14, Lys B47, Glu B52, Lys B132, and Arg B148 were modeled as alanine. Refinement statistics for the 1.7 Å wt DsbA structure are presented in Table 2, with those from the 2.0 Å structure given for comparison. In both monomers A and B, the electron density is continuous at the 1σ level for all main-chain atoms, an improvement on the original 2.0-Å structure, which had discontinuous density for residues 13–18 and 51–55 in both monomers. However, no density is observed for the C-terminal Lys 189 in either monomer A or B (as was the case for the 2.0-Å structure).

Validation and analysis of crystal structure

The program packages PROCHECK (Laskowski et al., 1993), PROMOTIF (Hutchinson & Thornton, 1996), GRASP (Nicholls et al., 1993), and INSIGHT (Biosym Corporation) implemented on a Silicon Graphics Indigo R4400 were used to analyze, characterize, and superimpose the DsbA structure and to produce many of the figures for this paper. The angular and translational differences between the domains in monomers A and B were calculated using the program ROTMOL (Huber et al., 1971). Residues contributing to surface features were identified using O. Side chains accessible to solvent were identified through solvent-accessibility calculations in X-PLOR 3.1 (using a probe radius of 1.4 Å).

Although there are no major structural changes between the two, the 1.7-Å DsbA structure corrects some errors in the 2.0-Å structure. For example, the peptide linkage between Ala B17 and Gly B18 is flipped compared with that in the 2.0-Å structure and, as a result, has an improved peptide orientation. Five lysine residues (A7, A47, A118, B140, B158) in the 1.7-Å structure are shifted compared with the 2.0-Å structure (Nζ changes of 2.1–4.4 Å). In addition, the side-chain orientation of Gln B21 is significantly different from that in the 2.0-Å structure, with the Nϵ2 atom shifted by 5.4 Å. Other smaller side-chain shifts, of the order of 1.0–2.0 Å, were observed for residues Glu A38, Asn A114, Gln A146, Asp B5, Met B56, Glu B94, and Gln B137. In general, these changes result in improved side-chain conformations and a better match to electron density.

Coordinates of DsbA (1.7 Å) have been deposited with the Brookhaven Protein Data Bank (Bernstein et al., 1977) (accession code 1FVK). Structure factors have also been deposited with the Protein Data Bank (R1FVKSF), but these will be held for release for two years.

Sequence alignment

Sequences from the following DsbA homologues were used: *E. coli* (Bardwell et al., 1991), *S. flexneri* (Watarai et al., 1995), *E. chrysanthemi* (Shevchik et al., 1995), *H. influenzae* (Tomb, 1992), *V. cholerae* (Peek & Taylor, 1992; Yu et al., 1992), *S. typhimurium* (Friedrich et al., 1993), *Az. vinelandii* (Ng et al., 1996, Genbank accession number L76098), *Ac. calcoaceticus* (P. Rauch & K.J. Hellingwerf, pers. comm.), *L. pneumophila* (A.B. Sadosky & H.A. Shuman, unpubl., Genbank accession number U15278).

The N-terminal residue for each of the sequences is that predicted by PSORT (Nakai & Kanehisa, 1991) after signal sequence cleavage, except in the case of *S. typhimurium*, where the predicted cleavage point is past the active site, and *Az. vinelandii*, where the predicted cleavage point leaves a much larger N-terminal region. In both cases, the sequences were trimmed to the same length as the other sequences. The initial sequence alignment was generated by Clustal analysis [Clustal W (Thompson et al., 1994)] using the full dynamic programming algorithm. Minor manual adjustments were made to optimize the alignment of secondary structural features. The region of residues 162–175 was the most difficult to align, especially for proteins from *V. cholerae* and *Ac. calcoaceticus*. We used the preliminary crystal structure of *V. cholerae* DsbA (Hu et al., 1997) to check this region of the alignment.

Acknowledgments

We thank Rudi Glockshuber and Martina Huber-Wunderlich for helpful advice and many stimulating discussions. We are grateful to Matthew Wilce for advice on the use of bulk solvent correction in crystallographic refinement. We thank Dr. Rauch and Prof. Hellingwerf for providing us with the sequence of *Acinetobacter calcoaceticus* DsbA prior to publication. We acknowledge the assistance of Mr. Trent Goodwin and Mr. Aaron Schindeler in the preparation of figures and bibliography, and Mr. Alun Jones for mass spectrometric analyses. This work is supported by a grant from the Australian Research Council (A09531886) and a Queen Elizabeth II Fellowship to J.L.M.

References

Bardwell JCA. 1994. Building bridges: Disulfide bond formation in the cell. *Mol Microbiol 14*:199–205.

Bardwell JCA, Lee JO, Jander G, Martin N, Belin D, Beckwith J. 1993. A pathway for disulfide bond formation in vivo. *Proc Natl Acad Sci USA 90*:1038–1042.

Bardwell JCA, McGovern K, Beckwith J. 1991. Identification of a protein required for disulfide bond formation in vivo. *Cell 67*:581–589.

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol 112*:535–542.

Brünger AT. 1992a. Free *R* value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature 355*:472–475.

Brünger AT. 1992b. *X-PLOR (version 3.1) manual*. New Haven, Connecticut: Yale University.

Darby NJ, Creighton TE. 1995. Catalytic mechanism of DsbA and its comparison with that of protein disulfide isomerase. *Biochemistry 34*:3576–3587.

Engh RA, Huber R. 1991. Accurate bond lengths and angle parameters for X-ray protein structure refinement. *Acta Crystallogr A 47*:392–400.

Frech C, Wunderlich M, Glockshuber R, Schmid FX. 1996. Preferential binding of unfolded protein to DsbA. *EMBO J 15*:392–398.

Freedman RB, Hirst TR, Tuite MF. 1994. Protein disulphide isomerase—Building bridges in protein folding. *Trends Biochem Sci 19*:331–336.

Friedrich MJ, Kinsey NE, Vila J, Kadner RJ. 1993. Nucleotide sequence of a 13.9 kb segment of the 90 kb virulence plasmid of *Salmonella typhimurium*: The presence of fimbral biosynthetic genes. *Mol Microbiol 8*:543–558.

Guilhot C, Jander G, Martin NL, Beckwith J. 1995. Evidence that the pathway of disulfide bond formation in *Escherichia coli* involves interactions between the cysteines of DsbB and DsbA. *Proc Natl Acad Sci USA 92*:9895–9899.

Herron JN, He XM, Ballard DW, Blier PR, Pace PE, Bothwell ALM, Voss EW Jr, Edmundson AB. 1991. An autoantibody to single-stranded DNA: Comparison of the three-dimensional structures of the unliganded Fab and a deoxynucleotide–Fab complex. *Proteins Struct Funct Genet 11*:159–175.

Higashi T. 1990. R-AXIS-IIC, a program for indexing and processing R-AXIS IIC imaging plate data. Danvers, Massachusetts: Rigaku.

Hu SH, Peek JA, Rattigan E, Taylor RK, Martin JL. 1997. Structure of TcpG, the DsbA protein folding catalyst from *Vibrio cholerae*. *J Mol Biol 268*:137–146.

Huber R, Epp O, Steigemann W, Formanek H. 1971. The atomic structure of erythrocruorin in the light of the chemical sequence and its comparison with myoglobin. *Eur J Biochem 19*:42–50.

Hutchinson EG, Thornton JM. 1996. PROMOTIF—A program to identify and analyze structural motifs in proteins. *Protein Sci 5*:212–220.

Jander G, Martin NL, Beckwith J. 1994. Two cysteines in each periplasmic domain of the membrane protein DsbB are required for its function in protein disulfide bond formation. *EMBO J 13*:5121–5127.

Jeng MF, Dyson HJ. 1996. Direct measurement of the aspartic acid 26 p*K*a for reduced *Escherichia coli* thioredoxin by 13C NMR. *Biochemistry 35*:1–6.

Jeng MF, Holmgren A, Dyson HJ. 1995. Proton sharing between cysteine thiols in *Escherichia coli* thioredoxin: Implications for the mechanism of protein disulfide reduction. *Biochemistry 34*:10101–10105.

Jones TA, Zou JY, Cowan SW, Kjeldgaard M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A 47*:110–119.

Kleywegt GJ, Jones TA. 1996. Good model-building and refinement practice. *Methods Enzymol*. Forthcoming.

Kraulis PJ. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr 24*:946–950.

Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Crystallogr 26*:283–291.

Martin JL. 1995. Thioredoxin—A fold for all reasons. *Structure 3*:245–250.

Martin JL, Bardwell JCA, Kuriyan J. 1993a. Crystal structure of the DsbA protein required for disulphide bond formation in vivo. *Nature 365*:464–468.

Martin JL, Waksman G, Bardwell JCA, Beckwith J, Kuriyan J. 1993b. Crystallization of DsbA, an *Escherichia coli* protein required for disulphide bond formation in vivo. *J Mol Biol 230*:1097–1100.

Missiakas D, Georgopoulos C, Raina S. 1993. Identification and characterization of the *Escherichia coli* gene *dsbB*, whose product is involved in the formation of disulfide bonds in vivo. *Proc Natl Acad Sci USA 90*:7084–7088.

Nakai K, Kanehisa M. 1991. Expert system for predicting protein localization sites in Gram-negative bacteria. *Proteins Struct Funct Genet 11*:95–110.

Ng TCN, Kwik JF, Maier RJ. 1996. Cloning and expression of the gene for a protein disulfide oxidoreductase from *Azotobacter vinelandii*: Complementation of an *E. coli* dsba mutant strain. *Gene*. Forthcoming.

Nicholls A, Bharadwaj R, Honig B. 1993. GRASP: Graphical representation and analysis of surface properties. *Biophysical J 64*:A116.

Peek JA, Taylor RK. 1992. Characterization of a periplasmic thiol:disulfide interchange protein required for the functional maturation of secreted virulence factors of *Vibrio cholerae*. *Proc Natl Acad Sci USA 89*:6210–6214.

Qin J, Clore GM, Gronenborn AM. 1996a. Ionization equilibria for side-chain carboxyl groups in oxidized and reduced human thioredoxin and in the complex with its target peptide from the transcription factor NFκB. *Biochemistry 35*:7–13.

Qin J, Clore GM, Kennedy WP, Kuszewski J, Gronenborn AM. 1996b. The solution structure of human thioredoxin complexed with its target from Ref-1 reveals peptide chain reversal. *Structure 4*:613–620.

Shevchik VE, Bortoli-German I, Robert-Baudouy J, Robinet S, Barras F, Condemine G. 1995. Differential effect of *dsbA* and *dsbC* mutations on extracellular enzyme secretion in *Erwinia chrysanthemi*. *Mol Microbiol 16*:745–753.

Stanfield RL, Takimoto-Kamimura M, Rini JM, Profy AT, Wilson IA. 1993. Major antigen-induced domain rearrangements in an antibody. *Structure 1*:83–93.

Thompson JD, Higgins DG, Gibson TJ. 1994. Clustal W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res 22*:4673–4680.

Tomb JF. 1992. A periplasmic protein disulfide oxidoreductase is required for transformation of *Haemophilus influenzae* Rd. *Proc Natl Acad Sci USA 89*:10252–10256.

Watarai M, Tobe T, Yoshikawa M, Sasakawa C. 1995. Disulfide oxidoreductase activity of *Shigella flexneri* is required for release of Ipa proteins and invasion of epithelial cells. *Proc Natl Acad Sci USA 92*:4927–4931.

Wilson NA, Barbar E, Fuchs JA, Woodward C. 1995. Aspartic acid 26 in *Escherichia coli* thioredoxin has a $pK_a > 9$. *Biochemistry 34*:8931–8939.

Wunderlich M, Glockshuber R. 1993a. In vivo control of redox potential during protein folding catalyzed by bacterial protein disulfide isomerase (DsbA). *J Biol Chem 268*:24547–24550.

Wunderlich M, Glockshuber R. 1993b. Redox properties of protein disulfide isomerase (DsbA) from *Escherichia coli*. *Protein Sci 2*:717–726.

Wunderlich M, Otto A, Seckler R, Glockshuber R. 1993. Bacterial protein disulfide isomerase (DsbA): Efficient catalysis of oxidative protein folding at acidic pH. *Biochemistry 32*:12251–12256.

Yu J, Webb H, Hurst TR. 1992. A homologue of the *Escherichia coli* DsbA protein involved in disulphide bond formation is required for the enterotoxin biogenesis in *Vibrio cholerae*. *Mol Microbiol 6*:1949–1958.

Zapun A, Cooper L, Creighton TE. 1994. Replacement of the active-site cysteine residues of DsbA, a protein required for disulfide bond formation in vivo. *Biochemistry 33*:1907–1914.

Zapun A, Creighton TE. 1994. Effects of DsbA on the disulfide folding of bovine pancreatic trypsin inhibitor and α-lactalbumin. *Biochemistry 33*:5202–5211.