Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of *Escherichia coli* UDP-*N*-acetylmuramoyl:L-alanine ligase (MurC)

Taru Deva,[a]‡ Edward N. Baker,[a] Christopher J. Squire[a] and Clyde A. Smith[a,b]*

[a]School of Biological Sciences, University of Auckland, Auckland, New Zealand, and [b]Stanford Synchrotron Radiation Laboratory, Menlo Park, CA 94025, USA

‡ Current address: Department of Life Sciences, Aalborg University, DK-9000, Aalborg, Denmark.

Correspondence e-mail: csmith@slac.stanford.edu

Received 24 August 2006
Accepted 19 September 2006

PDB Reference: MurC, 2f00, r2f00sf.

The bacterial cell wall provides essential protection from the external environment and confers strength and rigidity to counteract internal osmotic pressure. Without this layer the cell would be easily ruptured and it is for this reason that biosynthetic pathways leading to the formation of peptidoglycan have for many years been a prime target for effective antibiotics. Central to this pathway are four similar ligase enzymes which add peptide groups to glycan moieties. As part of a program to better understand the structure–function relationships in these four enzymes, the crystal structure of *Escherichia coli* UDP-*N*-acetylmuramoyl:L-alanine ligase (MurC) has been determined to 2.6 Å resolution. The structure was solved by multiwavelength anomalous diffraction methods from a single selenomethionine-substituted crystal and refined to a crystallographic R factor of 0.212 ($R_{\rm free}$ = 0.259). The enzyme has a modular multi-domain structure very similar to those of other members of the *mur* family of ATP-dependent amide-bond ligases. Detailed comparison of these four enzymes shows that considerable conformational changes are possible. These changes, together with the recruitment of two different N-terminal domains, allow this family of enzymes to bind a substrate which is identical at one end and at the other has the growing peptide tail which will ultimately become part of the rigid bacterial cell wall. Comparison of the *E. coli* and *Haemophilus influenzae* structures and analysis of the sequences of known MurC enzymes indicate the presence of a 'dimerization' motif in almost 50% of the MurC enzymes and points to a highly conserved loop in domain 3 that may play a key role in amino-acid ligand specificity.

1. Introduction

The structural integrity of the bacterial cell wall is critical to survival and to this end a rigid scaffold of peptidoglycan, comprising glycan chains and cross-linking peptides, is synthesized and attached outside the cytoplasmic membrane of the cell. The overall shape and strength of the bacterial cell derives primarily from the presence of this layer (Rogers *et al.*, 1980). Both Gram-positive and Gram-negative bacteria have peptidoglycan layers of similar composition, although in Gram-positive bacteria the layer tends to be more cross-linked and consequently thicker. The biosynthesis of peptidoglycan takes place in three stages: (i) formation of UDP-*N*-acetylmuramic acid (UDPMurNAc) from *N*-acetylglucosamine (GlcNAc), (ii) addition of a pentapeptide chain to UDP-MurNAc and (iii) addition of a second GlcNAc to the

Table 1
Data-collection statistics.

Values in parentheses are for the highest resolution shell (2.7–2.6 Å).

	Peak†	Inflection	Remote
Wavelength (Å)	0.9790	0.9792	0.8377
High-resolution limit (Å)	2.6	2.6	2.6
Observed reflections	780423	780563	749744
Unique reflections to $d_{\min}$	38261	38269	38184
$R_{\rm merge}$‡ (%)	0.132 (0.667)	0.136 (0.701)	0.133 (0.698)
$I/\sigma(I)$	8.2 (2.3)	7.6 (2.1)	7.8 (2.15)
Mosaicity§ (°)	0.25	0.23	0.24
Completeness (%)	98.9 (99.8)	99.1 (99.8)	99.3 (99.8)

† The peak data were subsequently used for structure refinement. ‡ $R_{\rm merge} = \sum|I_i - \langle I\rangle|/\sum I_i$, where I_i is the observed intensity and $\langle I\rangle$ is the mean intensity. § The refined mosaicity after post-refinement of the unit-cell parameters.

disaccharide-pentapeptide building block, transport of this unit through the cytoplasmic membrane and incorporation into the growing peptidoglycan layer (van Heijenoort, 1994). In stage (ii), the pentapeptide chain is built sequentially by four enzymes, the *mur* ligases MurC, MurD, MurE and MurF, which in *Escherichia coli* add L-alanine, D-glutamate, *meso*-diaminopimelate (mDAP) and D-alanyl-D-alanine, respectively. These enzymes share limited sequence identity (15–22%), but have several very highly conserved regions (Bouhss *et al.*, 1997; Eveland *et al.*, 1997) that map primarily to the active site. These sequence motifs have also been observed in other members of the *mur* ligase family, including folylpolyglutamate synthetase (FPGS; Sheng *et al.*, 2000), cyanophycin synthetase (Zeigler *et al.*, 1998; Dementin *et al.*, 2001) and the capB enzyme from Bacillales (Candela & Fouet, 2006).

Because peptidoglycan is absent in animal cells, the biosynthetic pathway leading to its synthesis is an important antibiotic target. Drugs have already been developed against some steps in the pathway. These include fosfomycin, which inhibits MurA, the first enzyme in stage (i), cycloserine, which inhibits both alanine racemase and D-Ala-D-Ala ligase, enzymes which produce the precursor dipeptide D-alanyl-D-alanine, and vancomycin, which interferes with the incorporation of the newly synthesized disaccharide-pentapeptide into the existing cell wall. More importantly, the β-lactam antibiotics are potent inhibitors of the steps involved in cross-linking the polypeptide chains at the end of stage (iii). The four cell-wall ligases in stage (ii) have not yet been successfully targeted, although mechanism-based inhibitors have been investigated, the most potent being the phosphinate inhibitors, which are designed to mimic the tetrahedral transition state of phosphoryl transfer (El Zoeiby *et al.*, 2003).

Crystal structures have been determined for the *E. coli* ligases MurD (Bertrand *et al.*, 1997), MurE (Gordon *et al.*, 2001) and MurF (Yan *et al.*, 2000), and for MurC from both *Haemophilus influenzae* (Mol *et al.*, 2003) and *Thermotoga maritima* (Spraggon *et al.*, 2004). Each protein comprises three structural domains: an N-terminal domain primarily responsible for binding the substrate, a large central ATPase domain and a C-terminal domain possibly associated with binding the incoming amino acid (Bertrand *et al.*, 1997). The only other structure of an enzyme from this amide-bond ligase superfamily is FPGS (Sun *et al.*, 1998), which is structurally homologous with the cell-wall ligases (Sheng *et al.*, 2000).

The structural analysis of MurC from *E. coli* was undertaken in order to further investigate the determinants of substrate specificity in the four *mur* ligases and to analyse the similarities and differences that come with biological origin and ligation state. An understanding of these relationships may be critical to the design of broad-spectrum antibiotics able to target a variety of bacterial species. We present the structure of apo *E. coli* MurC (EcMurC), revealing a closed structure with a pre-formed substrate-binding site and structural similarity to the substrate complex of the *H. influenzae* enzyme.

2. Experimental procedures

2.1. Expression, purification, crystallization and diffraction data collection

Native and selenomethionine (SeMet) substituted EcMurC were expressed, purified and crystallized as described previously (Deva *et al.*, 2003). The crystals of both native and SeMet-substituted EcMurC belong to space group $P2_12_12_1$, with unit-cell parameters $a = 73.9$, $b = 93.6$, $c = 176.8$ Å, but it was found that the SeMet crystals gave the best quality diffraction data. Three data sets, each covering 360° of rotation, were collected at the peak of the selenium X-ray absorption edge (0.9790 Å), the inflection point (0.9792 Å) and a point remote from the edge (0.8377 Å). Data-collection statistics are given in Table 1. The high $R_{\rm merge}$ values are primarily a consequence of the high redundancy (around 15-fold on average) and an inherent weakness of the diffraction data [average $I/\sigma(I)$ less than 9 for the three data sets]. There is also evidence of some diffuse scatter on the diffraction images which could also contribute to a larger variance in integrated intensity values (Deva *et al.*, 2003). The $I/\sigma(I)$ values of 2.1–2.3 in the highest resolution shells (2.69–2.6 Å) show that the data at this resolution are still observed.

2.2. Structure solution and refinement

The structure was determined by multiwavelength anomalous diffraction (MAD) techniques. The program *SOLVE* (Terwilliger & Berendzen, 1999) gave the positions of 28 Se atoms (14 per molecule) and *RESOLVE* (Terwilliger, 2000) was used for maximum-likelihood electron-density modification and refinement of the phases. The mean figure of merit (FOM) from *SOLVE* was 0.42 and this increased to 0.65 following density modification. The two molecules in the asymmetric unit were added manually into the $2F_{\rm o} - F_{\rm c}$ density map from *RESOLVE* using the molecular-graphics program *TURBO* (Roussel & Cambillau, 1991) and the model was submitted to refinement with *CNS* (Brünger *et al.*, 1998). The data collected at the peak were used in subsequent structure refinement since high-resolution native data were not available. The initial crystallographic R factor prior to the first round of simulated annealing was 0.462, with an $R_{\rm free}$ of

Table 2
Refinement statistics.

Space group	$P2_12_12_1$
Unit-cell parameters (Å)	a = 73.93, b = 93.13, c = 176.79
Resolution limits (Å)	88.0–2.6
No. of reflections used (reflections used for R_{free})	36340 (1928)
R factor/R_{free}†	0.212 (0.31)/0.259 (0.36)
No. of non-H atoms in the asymmetric unit	
Protein	7196
Solvent	378
Mg^{2+}	2
Average B values (Å^2)	
Molecule A	48.9
Molecule B	35.0
Solvent	39.7
R.m.s. deviations from ideality	
Bonds (Å)	0.008
Angles (°)	1.07
Ramachandran plot	
Residues in most favoured regions (%)	89.9
No. of residues in disallowed regions	4‡

† $R = \sum ||F_o| - |F_c|| / \sum |F_o|$, where F_o are the observed structure-factor amplitudes and F_c are the calculated structure factors. Values in parentheses are for the highest resolution shell (2.7–2.6 Å). ‡ These residues, Ala30*A*, Glu337*A*, Asn340*A* and Ala383*A*, are in poorly defined solvent-exposed loops where the density is ambiguous.

0.477. Subsequent model building with *TURBO* was used to complete the protein structure. Addition of water molecules was carried out with *CNS* by searching an $F_o - F_c$ map using a cutoff of 3.0σ. These water molecules were then inspected with *TURBO* and retained in the model only if they were in good electron density, had reasonable B factors and made chemically sensible interactions with the protein. An Mg^{2+} ion was also located in the active site of each monomer. Refinement with *CNS*, incorporating simulated annealing and energy minimization at 2.6 Å resolution, followed by maximum-likelihood refinement with *REFMAC* (Murshudov *et al.*, 1999) resulted in a final model with an R factor of 0.212 and an R_{free} of 0.258 (see Table 2 for the final refinement statistics). It was found that refinement of individual temperature factors gave a reasonable decrease in R factor and R_{free} compared with refinement of a single overall B factor or residue B factors.

3. Results and discussion

3.1. Structure determination and final model

The crystal structure reported here is for SeMet-substituted EcMurC refined at 2.6 Å resolution (Fig. 1*a*). The structure was solved by MAD and the resulting electron-density map from *RESOLVE*, following solvent flattening, was of very good quality and allowed the unambiguous tracing of the two molecules in the asymmetric unit. The final electron-density map near the ATP-binding site is shown in Fig. 1(*b*). The two molecules are related by an approximate twofold axis and their close association is indicative of a dimer, consistent with the observed activity of EcMurC (Jin *et al.*, 1996) and the solution behaviour as monitored by dynamic light scattering (Deva *et al.*, 2003). In the final model, the two molecules are essentially complete, except for the C-terminal residues 484–491 in both molecules and loop 429–435 in molecule *A*, for which no interpretable density could be found; these residues are assumed to be disordered. Two bound magnesium ions have been identified, one in each monomer. No ligand is present, indicating that this structure represents the apo MurC molecule. The final model conforms well with expected protein geometry, with 89.9% of residues in the most favoured regions of the Ramachandran plot as defined by *PROCHECK* (Laskowski *et al.*, 1993).

3.2. Monomer structure of EcMurC

The two independent EcMurC molecules in the asymmetric unit are essentially identical in structure, with a root-mean-square difference (r.m.s.d.) in atomic positions of 0.73 Å for 475 matching C^α atoms. Each molecule is folded into three contiguous structural domains (Fig. 1*a*): an N-terminal domain (domain 1; residues Met1–Phe118), a central domain (domain 2; residues Arg119–Gly325) and a C-terminal domain (domain 3; residues Arg326–Ala483). The organization of the two other MurC structures, HiMurC (Mol *et al.*, 2003) and TmMurC (Spraggon *et al.*, 2004), is similar, although the domain orientations vary with species and ligation state. The closest match is with HiMurC, consistent with the higher sequence identity: 63% with HiMurC, but only 27% with TmMurC. Accordingly, the r.m.s.d.s in C^α positions for the three structural domains are lower between EcMurC and HiMurC (0.70 Å for 104 matching C^α atoms in domain 1, 0.49 Å for 206 C^α atoms in domain 2 and 0.78 Å for 147 C^α atoms in domain 3) than between EcMurC and TmMurC [1.3 Å (100 C^α atoms), 1.4 Å (185 C^α atoms) and 1.6 Å (123 C^α atoms) for the three domains, respectively]. Both observations are consistent with the closer evolutionary relationship between the first two species (both γ-Proteobacteria) than with *T. maritima* (Thermotogae).

Each of the three domains in MurC has structural homologues in other proteins. The three individual domains are shown in Fig. 1(*c*). Domain 1 has a Rossmann-type fold comprising a five-stranded parallel β-sheet, β1–β5, flanked by four alternating α-helices, α1–α5, with an additional helix, α0, at the N-terminus. The closest structural match, as indicated by a *DALI* search (Holm & Sander, 1993), is to domain 1 of *E. coli* MurD (PDB code 4uag; Z score = 10.0, r.m.s.d. = 2.6 Å for 101 C^α atoms). Interestingly, the NAD-binding domain of the multifunctional enzyme sirohaem synthase from *Salmonella enterica* (Stroupe *et al.*, 2003; PDB code 1pjq, Z score = 10.2, r.m.s.d. = 2.8 Å for 99 C^α atoms) gives almost as good a match as MurD, as do the Rossmann-fold domains of a number of short-chain dehydrogenase/reductase enzymes.

Domain 2 comprises a seven-stranded parallel β-sheet (β6–β12) surrounded by four α-helices, α6–α9, plus a small separate antiparallel β-sheet (β13-β15); this sheet folds back on the core β-sheet to form an expanded open barrel with strands β10–β12, with an α-helix (α9) sitting in the centre. The core of the domain is typical of the mononucleotide-binding domains in a number of ATP- and GTP-dependent enzymes, including *ras*-P21, adenylate kinase, elongation factors G and Tu, ATP synthase and myosin (Schulz, 1992; Smith &

Rayment, 1996). A *DALI* search gives the top hits as the ATPase domains from MurF (Yan *et al.*, 2000; PDB 1gg4; Z score = 20.6, r.m.s.d. = 2.3 Å for 185 C^{α} atoms), MurE (Gordon *et al.*, 2001; PDB code 1e8c, Z = 19.5, r.m.s.d. = 2.3 Å for 178 C^{α} atoms), MurD (Bertrand *et al.*, 1999; PDB code 4uag, Z = 17.6, r.m.s.d. = 2.4 Å for 163 C^{α} atoms) and FPGS (Sun *et al.*, 1998; PDB code 1fgs; Z = 13.5, r.m.s.d. = 2.5 Å for 154 C^{α} atoms).

Domain 3 also has a Rossmann-type fold, comprising a six-stranded β-sheet, β16–β21, flanked by five helices, α11–α15, although it lacks the typical G*x*G*xx*G fingerprint motif (Schulz, 1992; Bellamacina, 1996). The closest structural matches are to the equivalent domains of MurE (PDB code 1e8c; Z = 13.8, r.m.s.d. = 2.5 Å for 134 C^{α} atoms), MurF (PDB code 1gg4; Z = 12.7, r.m.s.d. = 2.5 Å for 122 C^{α} atoms) and MurD (PDB code 4uag; Z score = 11.9, r.m.s.d. = 2.4 Å for 123 C^{α} atoms) and the C-terminal domain of *E. coli* FPGS (PDB code 1w78; Z = 13.2, r.m.s.d. = 2.5 Å for 127 C^{α} atoms). The short-chain oxidoreductases make up the bulk of the next best matches, including dihydrofolate reductase, the similarity of which has been previously noted (Sun *et al.*, 1998).

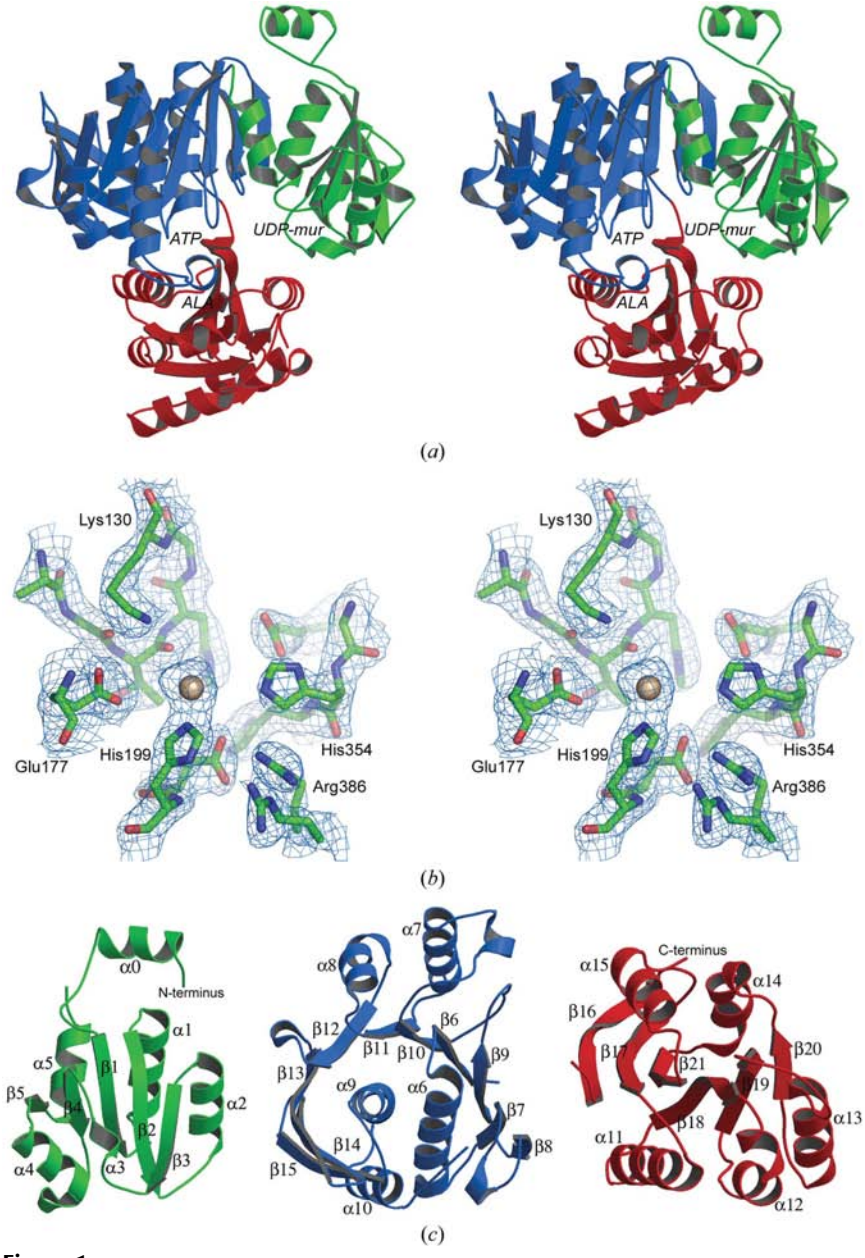

Figure 1
The EcMurC structure. (*a*) Ribbon diagram of *E. coli* MurC, showing domain 1 (green), domain 2 (blue) and domain 3 (red). The ATP-binding site is at the centre of the molecule and is labelled *ATP* and the substrate-binding sites are labelled *UDP-Mur* for the UDPMurNAc site and *ALA* for the L-alanine site. (*b*) The final $2F_o - F_c$ electron density near the ATP-binding site, with the final model shown in ball-and-stick representation. The Mg^{2+} ion found in the active site is shown in the foreground, with the P-loop to the upper left. (*c*) Folding of the three MurC domains. Ribbon diagrams showing domain 1 (green), domain 2 (blue) and domain 3 (red) of EcMurC, along with secondary-structure nomenclature.

3.3. Dimerization

The EcMurC dimer (Fig. 2*a*) is formed by the interaction of the top edge of the molecule (as viewed in Fig. 1*a*) with the equivalent region on the second molecule in a head-to-head fashion. Approximately 1300 $Å^2$ (7%) of surface area is buried per monomer in this interface, which is about average for a homodimer (Jones & Thornton, 1996). Biochemical studies show that *E. coli* MurC exists in an equilibrium between monomeric and dimeric forms, with a K_d of approximately 1 μM, and that it appears to have activity in both forms (Jin *et al.*, 1996).

The EcMurC dimer interface is formed by the interleaving of loops from the top of domain 2 of one molecule with loops from domain 1 of the second molecule (Fig. 2*b*). Several key residues at the edges of the interface play important roles in dimer formation. The most important interactions involve Phe223 and Tyr224 from the loop between helix α7 and strand β11, which are inserted into a highly hydrophobic pocket at the top of domain 1 comprising the side chains of residues Met16, Val19, Val81, Met111 and Ile106. The Phe223-Tyr224 pair is part of a PFYG motif that is conserved in almost 50% of known MurC sequences. Also at the edges of the dimerization interface, two adjacent arginine residues on the loop connecting helix α0 and strand β1 (Arg17 and Arg18) reach towards the second molecule and form salt bridges with two adjacent glutamate residues (Glu306 and Glu307) at the C-terminal end of helix α9. There are additional hydrophobic interactions and two specific intermolecular contacts towards the centre of the interface: the side chain of Arg169 is hydrogen bonded to the C-terminus of helix α1 and the side chains of Gln186 hydrogen bond across the

noncrystallographic twofold axis. This exact same dimer interface is seen in the HiMurC structures, but is completely lacking in TmMurC (and the same head-to-head dimer is not formed).

3.4. Domain movements in MurC

Three apo MurC structures are now available and these, along with the substrate and product complexes of HiMurC, provide interesting insights into domain movements in this enzyme. Superposition of the three apo structures on the basis of domain 2 (Fig. 3*a*) shows that the EcMurC and TmMurC structures are both in closed conformations, whereas an open conformation is observed for apo HiMurC in which domains 1 and 3 have rotated away from domain 2: domain 1 by 20° and domain 3 by 39°. The domain orientations in apo EcMurC much more closely resemble those for the substrate-bound and product-bound forms of HiMurC. Superposition of EcMurC onto the substrate-bound form of HiMurC gives an r.m.s.d. of 0.8 Å for 462 matching C^{α} positions, very similar to the r.m.s.d. values for the individual domains. When the structures are superimposed on the basis of domain 2, the orientations of domains 1 and 3 in the substrate complex of HiMurC only differ by about 2° from the equivalent domains in EcMurC. The occurrence of both open and closed forms in the absence of substrate suggests that there may be very little energy difference between these two states and that a dynamic equilibrium probably exists in solution.

Since the dimer interface spans both domains 1 and 2 where this conformational change takes place, it is possible that dimerization could affect enzyme flexibility and hence activity. Comparison of apo EcMurC and apo HiMurC shows that the movement of domain 1 relative to domain 2 induces a change in the structure of the dimer. Opening of domain 1 rotates the hydrophobic pocket while still maintaining contact with the side chains of Phe223 and Tyr224 from the other molecule. Because the molecules are still locked together through this interaction, the two domains 2 are forced to rotate with respect to each other about an axis running along the interface perpendicular to the noncrystallographic dyad (Fig. 3*b*). This concerted movement opens the rear side of the interface and closes the front side (Fig. 3*c*), breaking the interactions between Arg169 and helix α1 and decreasing the buried surface area to around 1100 $Å^2$ per monomer. Interactions on the front side of the interface involving the glutamine residues are maintained. If the MurC molecule in solution is a state of equilibrium between the open and closed forms, the loss of some of the interface interactions upon going from a closed to an open form may be sufficient to shift the equilibrium in favour of the closed form. This could effect enzyme flexibility and substrate binding and ultimately decrease activity.

If the presence of the PFYG, arginine and glutamate motifs is presumed to be a reasonable predictor of dimerization, alignment of all the known MurC sequences (data not shown) shows some intriguing relationships which may be related to flexibility and activity. Entire genera of bacteria including *Bacillus*, *Clostridium*, *Staphylococcus*, *Streptococcus*, *Mycobacteria* and *Corynebacteria* lack these motifs. Bacilli and clostridia are both spore-forming bacteria and sporulation requires large amounts of peptidoglycan biosynthesis; in *B. subtilis* some of the *mur* genes reside on the same operon as the genes responsible for spor-

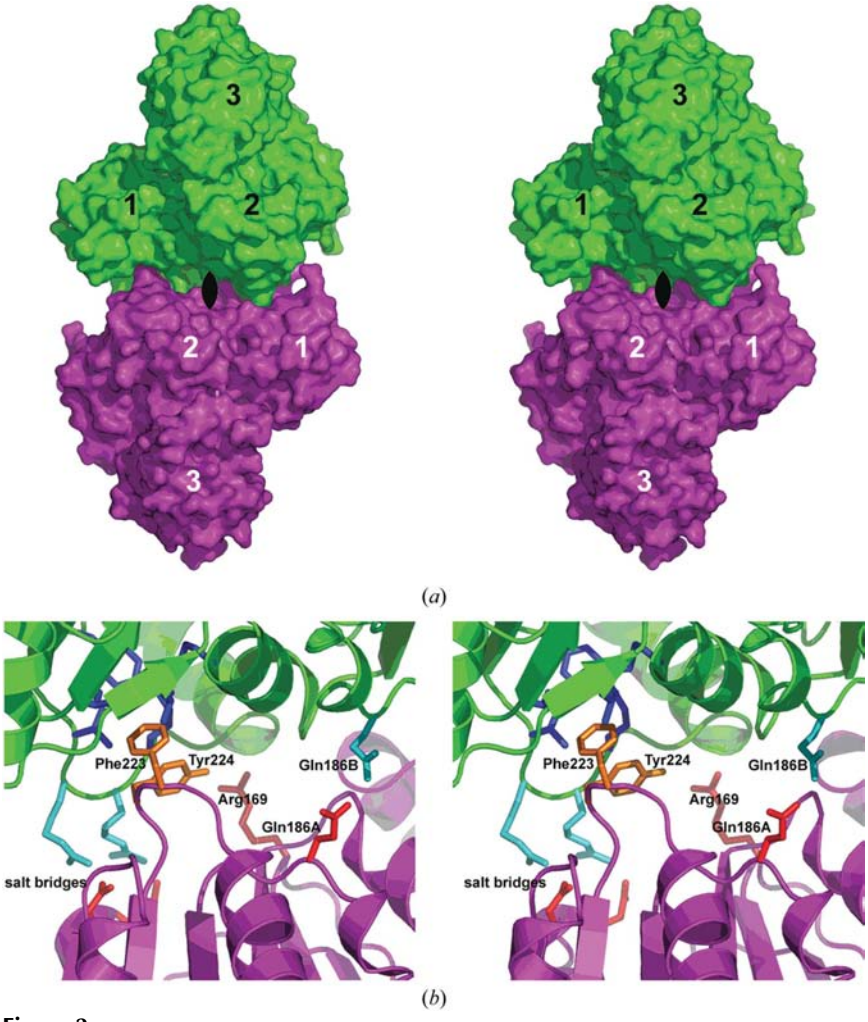

Figure 2
The EcMurC dimer. (*a*) Stereoview of the surface representation of the dimer, showing the two molecules in green and magenta. The locations of the three domains are indicated. The noncrystallographic dyad is indicated at the centre of the intermolecular interface. (*b*) Stereoview of part of the dimer interface, where loops from domain 2 of molecule *A* (magenta) interlock with loops from domain 1 of molecule *B* (green). The salt bridges formed by two arginine residues from one molecule (cyan sticks) with two glutamate residues from the other (red sticks, partially obscured) are indicated.

ulation (Daniel & Errington, 1993). In these species, it could be that the requirement for increased peptidoglycan synthesis has selected a monomeric MurC, since it is less constrained and possibly more active than a dimer. However, there is no single common factor which would explain why a MurC monomer might be preferred over a dimer in the bacteria where the dimerization motifs are missing; for example, the above argument does not explain the lack of the motifs in a number of extremophiles, including *Methanopyrus kandleri*, *Thermoanaerobacter tengcongensis* and *Thermotoga maritima* (although, surprisingly, not *Aquifex aeolicus*), since this seems contrary to the idea that increased oligomerization is one of the key features of thermophile stability (Charlier & Droogmans, 2005). The functional relevance, if any, of MurC dimerization remains an open question.

3.5. Active site

Comparison of apo EcMurC with the substrate complex of HiMurC (Fig. 4*a*) shows that the residues that form the UDPMurNAc-binding site are all in perfect position for substrate binding even though no ligand is bound and implies that the binding mode is the same in the two enzymes. The substrate binds in a cleft formed by the $\beta2$–$\alpha2$ and $\beta4$–$\alpha4$ loops, with the uracil ring sandwiched between hydrophobic residues from the two loops (Leu51 and Ile88 in EcMurC and Ile50 and Ile87 in HiMurC). A hydrogen bond with the conserved His70 further holds the uracil ring in place. The ribose moiety interacts through both hydroxyl groups with the side chain of Asp49 from the C-terminus of strand $\beta2$ and the diphosphate is hydrogen bonded to the side chain of Ser84 and the main-chain amide N atoms of a glycine-rich loop (residues 25–31). This glycine-rich loop, which has the consensus sequence GIGG*x*GM in the MurC enzymes, is comparable to the canonical G*x*G*xx*G dinucleotide-binding motif found in oxidoreductases (Schulz, 1992; Bellamacina, 1996).

The ATP-binding site in MurC is located in domain 2. The key binding determinant is a second glycine-rich loop linking strand $\beta6$ and helix $\alpha6$ and equivalent in location and sequence to the classical mononucleotide binding P-loop observed in many kinases and ATPases (Schulz *et al.*, 1974; Smith & Rayment, 1996). In nucleotide complexes of HiMurC (PDB codes 1p3d and 1gqy), the triphosphate is anchored by hydrogen bonds from main-chain amide N atoms of the P-loop and from the side chain of Lys129 (invariant in all *mur* enzymes) at the N-terminus of helix $\alpha6$. Despite the absence of bound ATP in EcMurC, the P-loop and the side chain of this lysine, Lys130, have identical conformations to those observed in the HiMurC–AMPPNP complex (Fig. 4*b*). Two other residues that help bind an essential Mg^{2+} ion, Glu172 and Thr130 in HiMurC, are also identically positioned in EcMurC, although this Mg^{2+} is missing in the absence of nucleotide.

The *mur* ligases also have a second Mg^{2+} ion in their active site which plays a role in stabilizing the interaction between the γ-phosphate and the free carboxylate during phosphoryl transfer (Sun *et al.*, 2001). Although no ATP is bound in EcMurC, this second Mg^{2+} ion is present, bound by the side

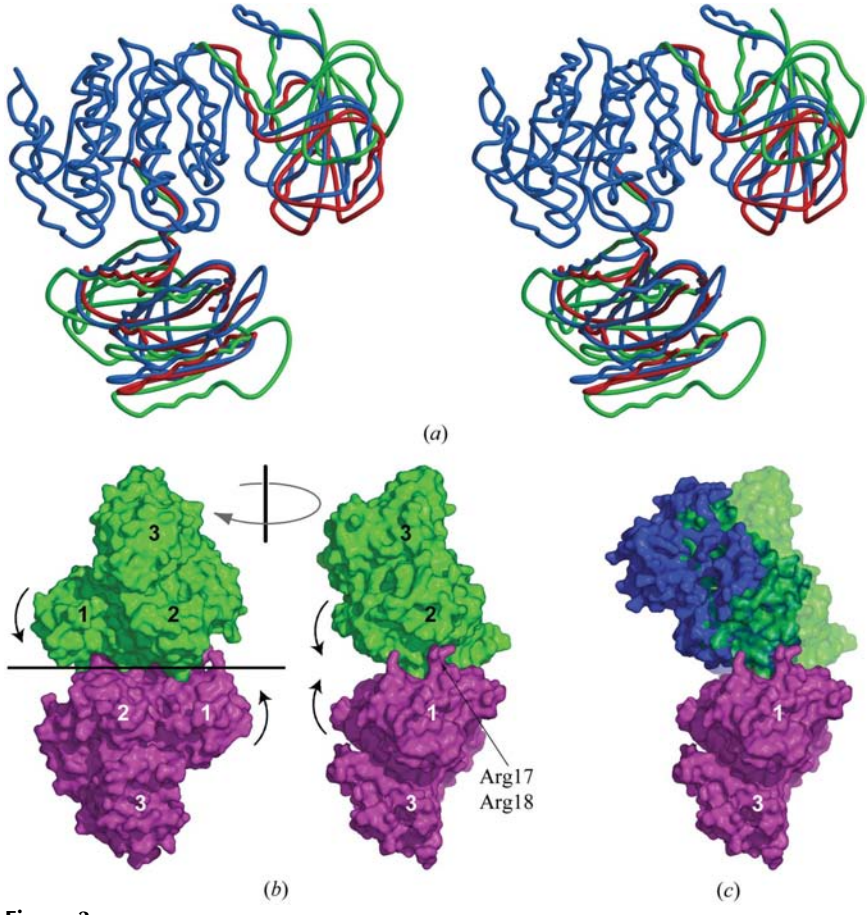

Figure 3
Conformational change in MurC. (*a*) Stereoview of the superposition of apo HiMurC (green) and apo TmMurC (red) onto apo EcMurC (blue) based on domain 2, showing the extent of movement of domains 1 and 3. Relative to EcMurC, domain 1 of apo HiMurC has rotated about an axis normal to the plane of the paper, whereas domain 3 has twisted about an axis parallel to the paper. (*b*) Surface representation of the EcMurC dimer in two orientations; left, similar to Fig. 2(*a*) and right, rotated 90° about a vertical axis as indicated by the grey arrow. The direction of rotation of domain 1 going from the closed to open form is indicated by the curved black arrows on the left figure. One monomer rotates relative to the other about a horizontal axis through the interface indicated by the solid black line, leading to an opening of the right side of the interface and a closure of the left side (indicated by the two curved black arrows on the right figure). The location of the two arginine residues (Arg17 and Arg18) from domain 1 which form salt bridges with two domain 2 glutamate residues (Glu306 and Glu307) can clearly be seen in this orientation. (*c*) The same molecular-surface representation with the surface of one monomer of apo HiMurC (dark blue) superimposed, showing the approximate extent of closure of the left side of the interface. The second monomer of HiMurC has been omitted for clarity.

chain of His199 (Fig. 4*b*) and surrounded by water molecules, which are in turn stabilized by hydrogen bonding to several conserved residues, namely Asp198, His354 and Glu177 (see Fig. 1*b*). The latter residue occupies the same spatial position as the carbamoylated lysine in MurD, MurE and FPGS, which similarly hydrogen bonds to the water molecules coordinating the second Mg^{2+} ion (Bertrand *et al.*, 1999; Gordon *et al.*, 2001; Sun *et al.*, 2001).

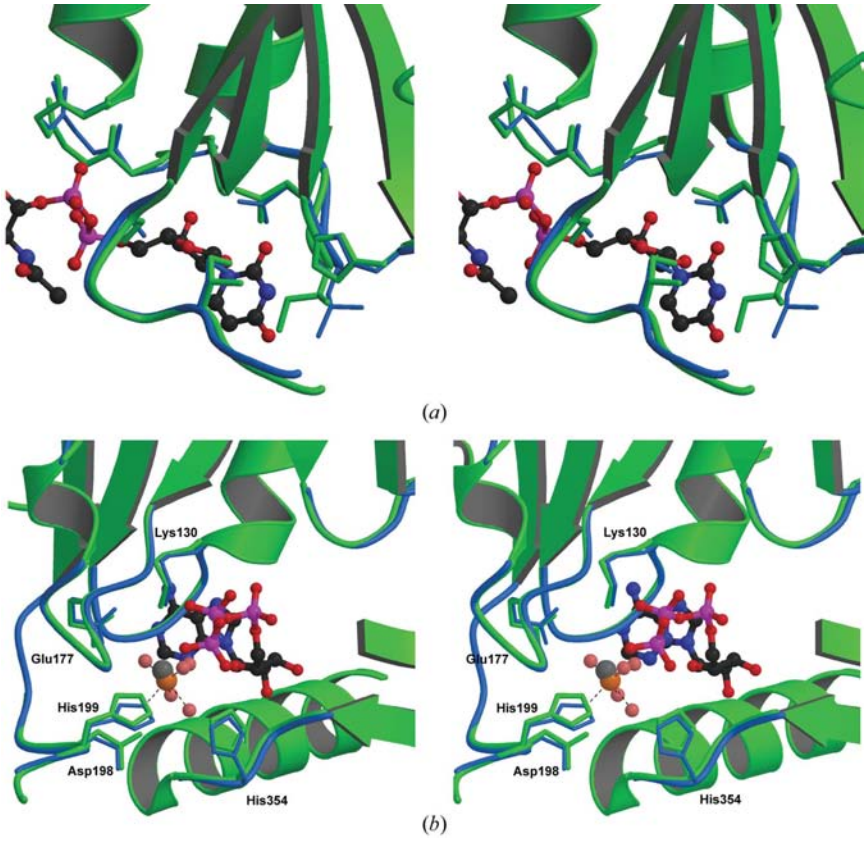

Figure 4
Active site of MurC. (*a*) Stereoview of the UDPMurNAc-binding site in MurC. The EcMurC (blue) and HiMurC (green) structures are superimposed, with the bound UDPMurNAc from the HiMurC structure shown in ball-and-stick representation. Side chains involved in substrate binding are shown in stick mode and are in green for HiMurC and in blue for EcMurC. (*b*) Stereoview of the ATP-binding site of EcMurC (blue) superimposed onto the HiMurC–ANPPNP complex. The ANPNP is shown in ball-and-stick representation. The second Mg^{2+} ion is shown as an orange sphere for EcMurC and a grey sphere for HiMurC. Red spheres indicate the water molecules which are coordinated to the Mg^{2+} ion in EcMurC. Residues involved in stabilizing the Mg^{2+} are indicated (blue bonds for EcMurC and green bonds for HiMurC).

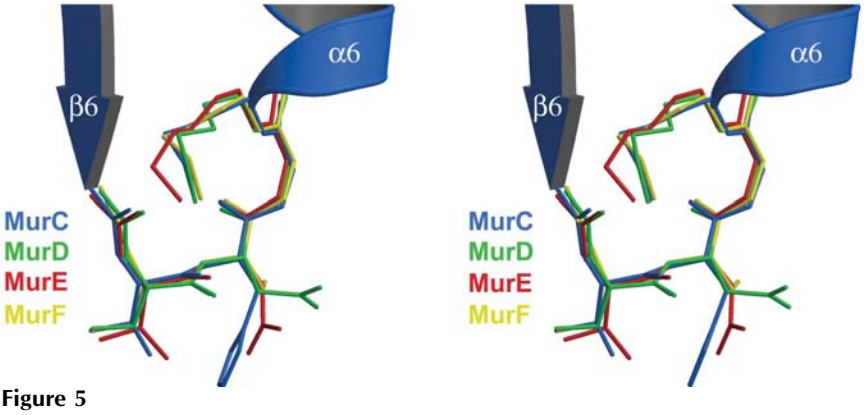

Figure 5
Stereoview of the superposition of the P-loop in the four *mur* ligases. For clarity, only EcMurC strand $\beta6$ and helix $\alpha6$ (blue) are shown. The residues comprising the P-loop are shown as coloured bonds for MurC (blue), MurD (green), MurE (red) and MurF (yellow).

The binding site for the third substrate, L-alanine, has not yet been determined experimentally, although circumstantial evidence points to two loops on the upper surface of domain 3, adjacent to the active site. The product complexes of HiMurC (Mol *et al.*, 2003), MurD (Bertrand *et al.*, 1999) and MurE (Gordon *et al.*, 2001) all have the terminal residue lying in pockets in this region of domain 3, although the structural details and the residues involved vary across the enzymes. Comparison of the three MurC structures shows that there are a number of residues which, based upon the HiMurC–product complex, could be involved in amino-acid binding. In HiMurC, these residues include His376, Arg377 and Arg380 in the $\beta18$–$\alpha12$ loop (the two arginine residues interact with the L-alanine carboxylate), along with Tyr346 and His348 in the adjacent $\beta17$–$\alpha11$ loop. The first three residues are part of a highly conserved motif with consensus sequence FQP**HR**-F/Y-T/S-**R**. Tyr346 and His348 are in a conserved DP**Y**-G/A-**H**HP motif (the aspartate residue is hydrogen bonded to the ATP ribose). This latter motif is also partially conserved in the other *mur* ligases (Sheng *et al.*, 2000) including FPGS, but the arginine-rich $\beta18$–$\alpha12$ loop appears to be MurC-specific. This loop shows wide variation in sequence and structure in the other *mur* enzymes, but adopts an identical conformation in both HiMurC and EcMurC (it is disordered in TmMurC); this raises the possibility that the specificity for a particular amino-acid ligand in the *mur* ligase family may be provided, at least in part, by this variable loop.

3.6. The *mur* amide-bond ligase family

The *mur* ligases all share a common three-domain architecture. They are organized around a common central ATP-binding domain, flanked by N- and C-terminal domains which move to accommodate the growing substrate. The structural and functional relationships between the four *mur* ligases have recently been reviewed (Smith, 2006). Domain 2 is central to this fold and is structurally conserved in all enzymes of the *mur* family. That this domain is essentially invariant makes perfect mechanistic sense,

given that it binds ATP, the one substrate that is common to all the *mur* ligases. The conservation of the ATP-binding domain is accompanied by conservation of the ATP-binding apparatus, the P-loop. The EcMurC consensus sequence, AGTHGKTTTT, closely matches the sequences of the other *mur* ligase P-loops (Smith, 2006) and superposition of the P-loop residues in the four *mur* ligases gives r.m.s.d. values of 0.30–0.35 Å for the 20 matching C^{α} positions (Fig. 5). This P-loop is a truncated form of the canonical P-loop observed in other mononucleotide-binding enzymes (Smith & Rayment, 1996) and when enzymes with more conventional P-loops are compared with the *mur* ligases, the P-loops in these enzymes are longer and have a more ledge-like structure on which the triphosphate sits (Sheng *et al.*, 2000). Despite this difference, the P-loop in the *mur* ligases still binds the triphosphate in a similar way as the enzymes with longer P-loops (Bertrand *et al.*, 1999; Sun *et al.*, 2001; Mol *et al.*, 2003).

Domain 1 shows the greatest structural and sequence diversity amongst the *mur* ligases. The first two enzymes in the pathway, MurC and MurD, share a common Rossmann-type α/β fold, whereas the last two enzymes in the pathway, MurE and MurF, have an α/β-fold with a completely different topology (Smith, 2006). The variations in this domain lead to different UDP-binding modes that can be related to the differing sizes of substrate that each enzyme binds. The UDP-binding pocket in MurD (Bertrand *et al.*, 1999) is essentially identical to that for MurC; however, domain 1 of MurD has rotated away from domain 2 by approximately 50°. In MurE (and presumably MurF, although there is no substrate complex for this enzyme), the UDP is bound by a long loop between the second strand and the second helix of domain 1. When the first three ligases are superimposed based upon domain 2, the distance from the ATP-binding site to the UDP-binding sites progressively increases from MurC to MurE (Fig. 6), consistent with the increasing length of the substrate and the mechanistic constraint that the terminal carboxylate must be near the γ-phosphate of the ATP. MurF presumably moves the substrate even further away, perhaps by rotation of domain 1 relative to domain 2, as is seen for MurD.

Progressing from one enzyme to the next, the growing length of the polypeptide tail is accommodated by either a reorientation of the substrate-binding domain (in the case of MurD) or by the recruitment of a completely different domain (in the case of MurE). All the while, the substrate is being moved further from the active site in such a way as to place the terminal carboxylate in roughly the same location. Clearly, structural flexibility, conformational changes and the use of two widely different substrate-binding domains are what allow these four enzymes to undertake the same identical reaction on a substrate which is progressively lengthened as it passes from one enzyme to the next.

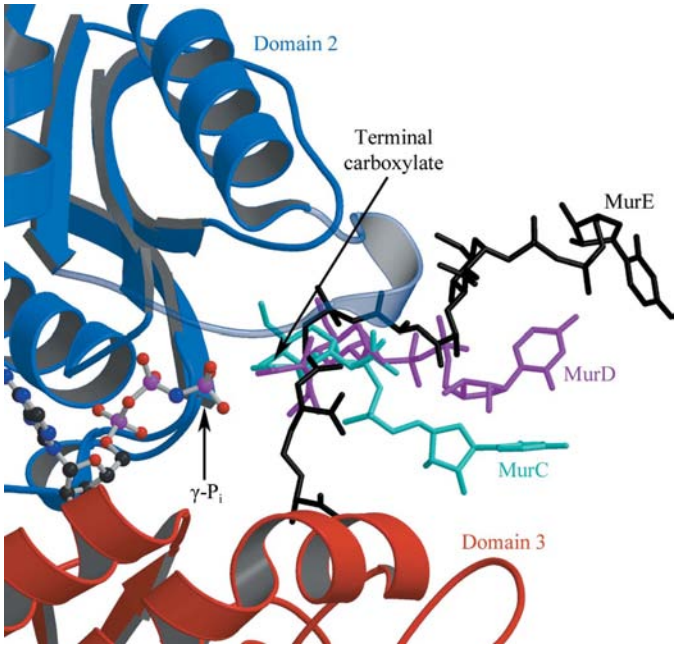

Figure 6
Relative positions of the substrates for MurC (UDPMurNAc; cyan), MurD (UDPMurNAc:L-Ala; magenta) and MurE (UDPMurNAc–tripeptide product; black) after superposition of the three structures based upon domain 2. Domains 2 and 3 for MurC are shown, but domain 1 is omitted for clarity. The bound nucleotide (AMPPNP) is shown in ball-and-stick representation and the approximate location of the terminal carboxylate group is indicated.

This paper is dedicated to the memory of Barbara Leiting, who supplied us with the *E. coli murc* gene for the initial crystallization work and who was a very enthusiastic collaborator on our *mur* ligase structural work. This work was supported by grants from the Health Research Council of New Zealand (CAS and ENB) and the Wellcome Trust (UK). The X-ray data collection was conducted at the Stanford Synchrotron Radiation Laboratory (SSRL), which is funded by the Department of Energy (BES, BER) and the National Institutes of Health (NCRR, NIGMS).

References

Bellamacina, C. R. (1996). *FASEB J.* **10**, 1257–1269.
Bertrand, J. A., Auger, G., Fanchon, E., Martin, L., Blanot, D., van Heijenoort, J. & Dideberg, O. (1997). *EMBO J.* **16**, 3416–3425.
Bertrand, J. A., Auger, G., Martin, L., Fanchon, E., Blanot, D., Le Beller, D., van Heijenoort, J. & Dideberg, O. (1999). *J. Mol. Biol.* **289**, 579–590.
Bouhss, A., Mengin-Lecreulx, D., Blanot, D., van Heijenoort, J. & Parquet, C. (1997). *Biochemistry*, **36**, 11556–11563.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Candela, T. & Fouet, A. (2006). *Mol. Microbiol.* **60**, 1091–1098.
Charlier, D. & Droogmans, L. (2005). *Cell. Mol. Life Sci.* **62**, 2974–2984.
Daniel, R. A. & Errington, J. (1993). *J. Gen. Microbiol.* **139**, 361–370.
Dementin, S., Bouhss, A., Auger, G., Parquet, C., Mengin-Lecreulx, D., Dideberg, O., van Heijenoort, J. & Blanot, D. (2001). *Eur. J. Biochem.* **268**, 5800–5807.
Deva, T., Pryor, K. D., Leiting, B., Baker, E. N. & Smith, C. A. (2003). *Acta Cryst.* D**59**, 1510–1513.
El Zoeiby, A., Sanschagrin, F. & Levesque, R. C. (2003). *Mol. Microbiol.* **47**, 1–12.

Eveland, S. S., Pompliano, D. L. & Anderson, M. S. (1997). *Biochemistry*, **36**, 6223–6229.
Gordon, E., Flouret, B., Chantalat, L., van Heijenoort, J., Mengin-Lecreulx, D. & Dideberg, O. (2001). *J. Biol. Chem.* **276**, 10999–11006.
Heijenoort, J. van (1994). *Bacterial Cell Wall*, edited by J. M. Ghuysen & R. Hakenbeck, pp. 39–54. Amsterdam: Elsevier.
Holm, L. & Sander, C. (1993). *J. Mol. Biol.* **233**, 123–138.
Jin, H., Emanuele, J. J., Fairman, R., Robertson, J. G., Hail, M. E., Ho, H. T., Falk, P. J. & Villafranca, J. J. (1996). *Biochemistry*, **35**, 1423–1431.
Jones, S. & Thornton, J. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 13–20.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Mol, C. D., Brooun, A., Dougan, D. R., Hilgers, M. T., Tari, L. W., Wijnands, R. A., Knuth, M. W., McRee, D. E. & Swanson, R. V. (2003). *J. Bacteriol.* **185**, 4152–4162.
Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). *Acta Cryst.* D**55**, 247–255.
Rogers, H. J., Perkins, H. R. & Ward, J. B. (1980). *Microbial Cell Walls and Membranes*, edited by H. J. Rogers, pp. 239–297. London: Chapman & Hall.
Roussel, A. & Cambillau, C. (1991). *TURBO-FRODO*. http://www.afmb.univ-mrs.fr/-TURBO-.
Schulz, G. E. (1992). *Curr. Opin. Struct. Biol.* **2**, 61–67.
Schulz, G. E., Elzinga, M., Marx, F. & Schirmer, R. H. (1974). *Nature (London)*, **250**, 120–123.
Sheng, Y., Sun, X., Shen, Y., Bognar, A. L. & Smith, C. A. (2000). *J. Mol. Biol.* **302**, 427–440.
Smith, C. A. (2006). *J. Mol. Biol.* **362**, 640–655.
Smith, C. A. & Rayment, I. (1996). *Biophys. J.* **70**, 1590–1602.
Spraggon, G. *et al.* (2004). *Proteins*, **55**, 1078–1081.
Stroupe, M. E., Leech, H. K., Daniels, D. S., Warren, M. J. & Getzoff, E. D. (2003). *Nature Struct. Biol.* **10**, 1064–1073.
Sun, X., Bognar, A. L., Baker, E. N. & Smith, C. A. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 6647–6652.
Sun, X., Cross, J. A., Bognar, A. L., Baker, E. N. & Smith, C. A. (2001). *J. Mol. Biol.* **310**, 1067–1078.
Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Yan, Y., Munshi, S., Leiting, B., Anderson, M. S., Chrzas, J. & Chen, Z. (2000). *J. Mol. Biol.* **304**, 435–445.
Zeigler, K., Diener, A., Herpin, C., Richter, R., Deutzmann, R. & Lockau, W. (1998). *Eur. J. Biochem.* **254**, 154–159.