Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Dimeric structure of the N-terminal domain of PriB protein from *Thermoanaerobacter tengcongensis* solved *ab initio*

Dorothee Liebschner,[a] Krzysztof Brzezinski,[a,b] Miroslawa Dauter,[c] Zbigniew Dauter,[a]* Marta Nowak,[d] Józef Kur[d] and Marcin Olszewski[d]*

[a]Synchrotron Radiation Research Section, MCL, National Cancer Institute, Argonne National Laboratory, Argonne, IL 60439, USA, [b]Institute of Chemistry, University of Bialystok, 15-399 Bialystok, Poland, [c]SAIC-Frederick Inc., Basic Research Program, Argonne National Laboratory, Argonne, IL 60439, USA, and [d]Department of Microbiology, Gdansk University of Technology, 80-952 Gdansk, Poland

Correspondence e-mail: dauter@anl.gov, molszewski@pg.gda.pl

Received 6 September 2012
Accepted 4 October 2012

PDB Reference: N-terminal domain of PriB, 4gs3

PriB is one of the components of the bacterial primosome, which catalyzes the reactivation of stalled replication forks at sites of DNA damage. The N-terminal domain of the PriB protein from the thermophilic bacterium *Thermoanaerobacter tengcongensis* (*Tte*PriB) was expressed and its crystal structure was solved at the atomic resolution of 1.09 Å by direct methods. The protein chain, which encompasses the first 104 residues of the full 220-residue protein, adopts the characteristic oligonucleotide/oligosaccharide-binding (OB) structure consisting of a five-stranded β-barrel filled with hydrophobic residues and equipped with four loops extending from the barrel. In the crystal two protomers dimerize, forming a six-stranded antiparallel β-sheet. The structure of the N-terminal OB domain of *T. tengcongensis* shows significant differences compared with mesophile PriBs. While in all other known structures of PriB a dimer is formed by two identical OB domains in separate chains, *Tte*PriB contains two consecutive OB domains in one chain. However, sequence comparison of both the N-terminal and the C-terminal domains of *Tte*PriB suggests that they have analogous structures and that the natural protein possesses a structure similar to a dimer of two N-terminal domains.

1. Introduction

PriB is one of the essential proteins for the bacterial primosome, a protein complex that plays an important role in DNA replication. At sites of DNA damage, the primosomal protein PriB is one of the components which catalyze the reactivation of stalled replication forks (Cox *et al.*, 2000).

To date, ten PriB structures have been deposited in the Protein Data Bank (PDB; Berman *et al.*, 2003), four of which do not have an accompanying publication. In particular, PriB from *Escherichia coli* has been thoroughly investigated (Lopper *et al.*, 2004, 2007; Liu *et al.*, 2004; Shioi *et al.*, 2005; Huang *et al.*, 2006). The *E. coli* PriB dimer is composed of two identical subunits which contain an oligonucleotide/oligosaccharide-binding (OB) domain. This fold consists of a five-stranded antiparallel β-sheet coiled to form a closed β-barrel and capped by an α-helix located between the third and fourth strands (Murzin, 1993). Three β-hairpin loops protrude from the central core: L_{12}, L_{34} and L_{45} (where the indices indicate the numbering of the β-strands). Loop L_{45} in particular seems to be flexible in the crystal structures, as it is often absent from the PriB models deposited in the PDB. In the PriB dimer, the OB domains are arranged such that the

first β-strands of each chain interact through hydrogen bonds and form a larger six-stranded β-sheet with a slightly convex surface. Loop L_{23} enwraps the other protomer and further stabilizes the dimeric state.

PriBs display high structural similarity to single-stranded DNA-binding proteins (SSBs), which are involved in DNA replication, recombination and repair (Greipel *et al.*, 1989). The arrangement of OB folds and the dimerization interface are similar in SSBs and PriBs. However, the oligomeric state of SSBs is different, as they mostly form tetramers by facing the convex β-sheet surfaces of the dimers against each other. The dimer can consist of two identical protomers or of a long chain composed of two OB domains joined by a linker.

The *Thermoanaerobacter tengcongensis* genome possesses three SSB genes: *ssb*, *ssb2* and *ssb3*. After further analysis, *ssb* was suggested to be the gene encoding the primosomal protein *Tte*PriB (Olszewski *et al.*, 2008), which is the subject of this study. *Tte*PriB possesses two OB folds per monomer, in contrast to the homodimeric PriB from *E. coli* and other bacterial sources. The N-terminal domain of the PriB protein from *T. tengcongensis* (*Tte*PriBN) was identified, expressed and purified. The quality and resolution (1.09 Å) of the diffraction data allowed us to solve the structure of *Tte*PriBN by direct methods. The resulting model is described and compared with other PriB structures.

2. Materials and methods

As crystallization attempts using the full-length sequence of *Tte*PriB were not successful, experiments were carried out with the N-terminal OB domain, which encompasses the first 104 residues of the full 220-residue protein. Unfortunately, isolation and purification of the C-terminal domain was also not successful.

2.1. Cloning, expression and purification

The coding sequence corresponding to the N-terminal domain of the *Tte*PriB protein (*Tte*PriBN; residues 1–104) was amplified by PCR from *T. tengcongensis* genomic DNA. The amplicon was cloned into the pMCSG57 expression vector (available at PSI:Biology-MR) using an LIC reaction. The construct was used to transform the BL21-CodonPlus(DE3)-RIPL strain of *E. coli*.

60 ml LB medium containing 34 µg ml^{-1} chloramphenicol and 100 µg ml^{-1} ampicillin was inoculated with the transformation reaction and grown overnight at 310 K. The overnight culture was used to inoculate 4 l LB medium with appropriate antibiotics and was grown to an OD_{600} of about 1.0. The temperature was decreased to 291 K and protein expression was induced by IPTG at a final concentration of 0.3 m*M*. The cells were harvested 17 h after induction.

The cell pellet was resuspended in buffer *A* (20 m*M* imidazole, 500 m*M* NaCl, 20 m*M* Tris–HCl pH 8.0, 10% glycerol) with the addition of 1 m*M* TCEP, 1 m*M* PMSF, 100 µg ml^{-1} lysozyme and 250 units of benzonase. Cells were disrupted by sonication on ice and centrifuged to remove cell debris. The supernatant was loaded onto Ni Sepharose 6 Fast Flow resin (GE Healthcare) equilibrated with buffer *A*. The protein was eluted with a buffer consisting of 500 m*M* imidazole, 500 m*M* NaCl, 20 m*M* Tris–HCl pH 8.0, 1 m*M* TCEP, 10% glycerol. To remove the His fusion tag, TEV protease was added to a final concentration of 0.5 mg ml^{-1} and the protein solution was dialyzed against buffer *A*. After 12 h of incubation at 277 K, the mixture was loaded onto an Ni Sepharose 6 Fast Flow resin; the protein was eluted with buffer *A* and subsequently exchanged for buffer *B* (800 m*M* NaCl, 30 m*M* HEPES pH 7.5) by dialysis. The protein solution was concentrated by ultracentrifugation using Amicon Ultra-10 filters. The protein solution was loaded onto a Superdex 200 (Pharmacia) gel-filtration column pre-equilibrated with buffer *B*. The protein was eluted as a dimer with buffer *B*. Fractions containing the N-terminal domain of *Tte*PriB were concentrated to 18 mg ml^{-1} by ultracentrifugation and the fresh protein solution was used for crystallization experiments. The size of the expressed protein (about 12.2 kDa) was confirmed by SDS–PAGE. The purified protein contained the N-terminal domain sequence extended at the N-terminus by a short peptide (SNA–) which is an artifact of the cloning and purification procedure.

2.2. Crystallization

Protein solution (18 mg ml^{-1}) in buffer *B* was used for initial screening of crystallization conditions using a sparse-matrix screen (Jancarik & Kim, 1991) from Hampton Research (Crystal Screen and Crystal Screen 2) and a Mosquito robot. Initial crystals were obtained from condition No. 10 of Crystal Screen (0.2 *M* ammonium acetate, 0.1 *M* acetate buffer pH 4.6, 30% PEG 4K). The crystallization conditions were optimized by adjusting the concentration of the precipitation agent. The best crystals were obtained with a well solution consisting of 0.14 *M* ammonium acetate, 0.07 *M* acetate buffer pH 4.6, 21% PEG 4K and were grown using the hanging-drop vapor-diffusion method at 293 K by mixing 1 µl protein solution and 1 µl precipitant solution on a coverslide and equilibrating the drop against 1.0 ml precipitant solution. Single crystals suitable for X-ray diffraction experiments appeared within 3 d. Crystals were cryocooled using a solution consisting of 0.14 *M* ammonium acetate, 0.07 *M* acetate buffer pH 4.6, 21% PEG 4K, 27% glycerol.

2.3. Data collection

Diffraction data were collected to 1.09 Å resolution from a cryocooled single crystal on the SER-CAT beamline 22-ID at APS, Argonne National Laboratory using a MAR 300 CCD detector and a wavelength of 1.000 Å. Two passes of data were collected with different exposures in order to avoid overloaded and overlapping reflections for low-resolution and high-resolution reflections, respectively, choosing suitable exposure times, oscillation angles and crystal-to-detector distances for each pass. The data were processed, scaled and merged with *HKL*-2000 (Otwinowski & Minor, 1997). The intensities of the two passes were merged together to produce

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

Data collection	
Wavelength (Å)	1.000
Space group	*C*2
Unit-cell parameters	
a (Å)	55.56
b (Å)	41.18
c (Å)	38.78
β (°)	94.8
Resolution (Å)	30–1.09 (1.11–1.09)
Multiplicity	4.1 (2.2)
Completeness (%)	99.0 (88.3)
$R_{\rm merge}$ (%)	4.7 (22.9)
$\langle I/\sigma(I)\rangle$	29.3 (1.9)
Refinement	
Total reflections	36546
Reflections, free set	1830
$R_{\rm cryst}$ [$F > 4\sigma(F)$] (%)	12.92
$R_{\rm free}$ [$F > 4\sigma(F)$] (%)	15.44
Wilson *B* factor (Å^2)	12.4
Average *B* factor (Å^2)	21.6
R.m.s.d. bond lengths (Å)	0.016
R.m.s.d. bond angles (°)	2.80
Water molecules	96
Ramachandran analysis	
Favored regions (%)	94.2
Allowed regions (%)	5.8
PDB code	4gs3

an overall completeness of 99.0% in space group *C*2, with unit-cell parameters $a = 55.56$, $b = 41.18$, $c = 38.78$ Å, $\beta = 94.8^\circ$. The resulting diffraction data statistics are summarized in Table 1.

2.4. Phasing by direct methods

The reflection file obtained after merging was converted to *SHELX* (Sheldrick, 2008) format using *SHELXPRO*. For structure solution using dual-space direct methods with *SHELXD* (Sheldrick, 2008), the instruction file requires information about the cell and its content, an estimated number of atoms (keyword FIND) and/or the number of peaks to start with during each cycle of the peaklist-optimization step (Sheldrick & Gould, 1995; keyword PLOP). According to the sequence, *Tte*PriBN is composed of 107 residues (104 residues corresponding to the N-terminal domain of the *Tte*PriB protein plus three residues remaining from the cloning and purification steps), *i.e.* 552 C atoms, 144 N atoms, 163 O atoms and two S atoms, yielding a total of 861 non-H atoms. As the known SSB structures contain several loops, which are usually disordered and therefore contribute only weakly to diffraction, the estimated number of atoms was presumed to be smaller than 861. Several *SHELXD* runs with a varying estimated number of sites and appropriate changes of the number of cycles for peaklist optimization were carried out. The attempt with FIND = 650 and PLOP = 700 800 900 provided the best result, yielding a final correlation coefficient of 69.21. The obtained model was composed of 616 C and O atoms and was split into several fragments. This model was input to *ARP*/*wARP* (Langer *et al.*, 2008), which connected the fragments of the initial model and built 96 amino-acid residues into one chain extending from residues 11 to 106. The number of water molecules was 63.

2.5. Refinement

The model obtained from *ARP*/*wARP* was used for further refinement with *SHELXL*. After a first round of 15 conjugate-gradient least-squares (CGLS) cycles, the model was inspected with the graphics program *Coot* (Emsley & Cowtan, 2004) and compared with the initial electron-density map from phasing in order to identify potentially erroneous parts. The loop between residues 83 and 91 was obviously not correct, as it overlapped with a symmetry-related monomer. Residues 54–61, which correspond to a small helix in other SSB structures, only had poor electron density and were deleted. Subsequent refinement was carried out initially with isotropic and subsequently with anisotropic treatment of atomic displacement parameters (ADPs). Every 15 cycles of refinement, the model was inspected and if necessary corrected with *Coot*. Water molecules were added with occupancies of 1.0 or 0.5. Special attention was paid to the missing parts of the sequence, which were added if significant electron-density peaks and stereochemistry justified their positioning. H atoms were added as 'riding' on their parent atoms, with the exception of those of O—H groups with a rotational degree of freedom (Ser, Tyr and Thr). Throughout refinement, a randomly selected 5% of the reflections were used for cross-validation (Brünger, 1992). In the last refinement stage, the model was refined against all reflections and one cycle of blocked full-matrix least-squares refinement was carried out without application of parameter shifts (DAMP 0 0) in order to estimate the standard uncertainties (SUs) of all refined and derived parameters. This refinement was accepted as final and the resulting model was deposited in the PDB as entry 4gs3.

Figures displaying electron density were created with *PyMOL* (DeLano, 2002). The contact surface in the dimer was calculated with *PISA* (Krissinel & Henrick, 2007). Structural superposition was achieved with *LSQKAB* (Kabsch, 1976) by aligning the C^α atoms of the β-barrel.

2.6. Calculation of the electrostatic potential

The electrostatic potential of a protein can be calculated by assigning point charges to atoms and applying the Coulomb equation

$$\varphi(\mathbf{r}') = \frac{Q}{4\pi\varepsilon_0|\mathbf{r} - \mathbf{r}'|}, \qquad (1)$$

which gives the potential φ at $\mathbf{r}'$ owing to a charge Q at $\mathbf{r}$, where ε_0 represents the vacuum permittivity.

A more elaborate way of calculating the electrostatic potential is achieved by using a continuous charge distribution instead of point charges. The potential $\varphi(\mathbf{r}')$ is then given by

$$\varphi(\mathbf{r}') = \int \frac{\rho_{\rm total}(\mathbf{r})}{4\pi\varepsilon_0|\mathbf{r} - \mathbf{r}'|}\, {\rm d}\mathbf{r} \qquad (2)$$

(Coppens, 1997), where $\varphi_{\rm total}(\mathbf{r})$ represents the nuclear and the electronic charges.

For protein crystals, the resolution of the diffraction data is usually not sufficient to conduct the refinement of parameters describing the continuous distribution of the electron density. However, these parameters can be transferred from a database such as ELMAM2 (Domagała *et al.*, 2012) or other pseudoatom data banks (Dittrich *et al.*, 2006; Dominiak *et al.*, 2006), allowing the subsequent calculation of the electrostatic potential. In the case of *Tte*PriB, the final structural model containing all H atoms was used for the potential calculations. Atomic charges and multipole parameters (up to octupoles for C, O and N atoms and to dipoles for H atoms) were transferred from the ELMAM2 database onto the protein model. The covalent X—H bonds were elongated to standard neutron distances (Allen, 1986) and the protein was neutralized electrically so that the sum of all charges was zero (the charge shift was 4×10^{-3} electrons per atom). The total electrostatic potential was calculated with the *VMoPro* software (Jelsch *et al.*, 2005); water molecules were not included so as to estimate only the protein potential.

3. Results and discussion

The calculated Matthews coefficient suggests the presence of one molecule in the asymmetric unit with a solvent content of 32%. The highest oligomer the protein can form is therefore a homodimer located at the twofold axis of space group *C*2.

3.1. Phasing by direct methods

The successful *SHELXD* trial yielded a set of atoms which was suitable for subsequent interpretation and further refinement since, in contrast to other direct-methods programs, *SHELXD* presents results in the form of a compact model that represents as closely as possible one molecule with all atoms connected by bonds. This model already showed the closed β-barrel that is a typical feature of OB domains. A close-up of the unbiased electron-density map from direct-methods phasing, superposed on the model from *SHELXD*, is displayed in Fig. 1(*a*). Many details, such as a phenylalanine ring (center) or a valine side chain (top left) as well as the trace of the main chain (center), are clearly visible in the electron density.

According to the PDB, only 18 novel protein structures and those of a number of smaller oligopeptides and oligonucleotides have previously been solved by *ab initio* direct methods. The PDB codes and some characteristics of these models are summarized in Table 2. The structure of *Tte*PriBN, with a 107-residue sequence, is one of the largest protein structures to be solved by direct methods; only lysozyme T4 (Mooers & Matthews, 2004) and phospholipase A_2 (Liu *et al.*, 2003), which possess 164 and 134 residues, respectively, are larger (Table 2). Several test structures, such as hen egg-white lysozyme or bovine trypsin, have been solved by direct methods in the course of various methodological calculations but have not been deposited in the PDB.

It is known that the presence of atoms heavier than C, N and O makes the direct-methods solution of macromolecules easier. The crystal of lysozyme T4 was deliberately soaked in RbCl solution and several Rb^+ and Cl^- ion sites were identified in its structure. Phospholipase A_2 contains 14 cysteines and three methionines, so that there are 17 S atoms present in the structure. In contrast, only two methionines and no cysteines are present in the *Tte*PriBN sequence, with one methionine residing in the completely disordered N-terminal part and thus not contributing to the recorded diffraction. It may be concluded that the *Tte*PriBN model is closest to the theoretical 'equal-atom structure' among the largest macromolecules solved by direct methods.

The main limit for *ab initio* phasing with direct methods is the resolution of the diffraction data, which has to be at least 1.2 Å (Morris & Bricogne, 2003; Sheldrick, 1990). It is therefore interesting to note that ten of the 18 structures have been solved using diffraction data between 1.0 and 1.2 Å resolution. The number of high-resolution protein structures is steadily increasing thanks to methodological and technical improvements in macromolecular crystallography (Petrova & Podjarny, 2004; Dauter, 2003). At present, the PDB contains about 1600 protein structures that have been determined at a resolution better than 1.2 Å; about half of these have a

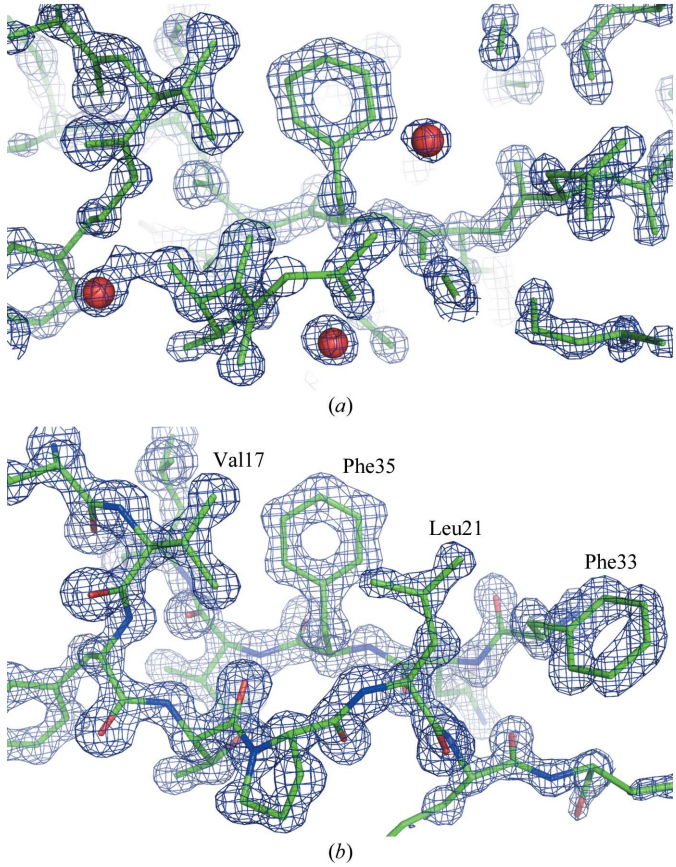

Figure 1
(*a*) $2F_o - F_c$ electron-density map calculated from the model obtained by direct methods contoured at the 2.5σ level and superposed on the *SHELXD* model of *Tte*PriBN. (*b*) The same view of *Tte*PriBN as in (*a*) using the final model and the corresponding $2F_o - F_c$ electron-density map contoured at 2σ. For a clearer view, only residues 16–23 and 33–47 are displayed.

Table 2
Overview of structures in the PDB solved *ab initio* by direct methods.

PDB code	Sample	No. of residues	No. of atoms	Space group	Resolution (Å)	Program used	Reference
Proteins and peptides							
1sx2	Lysozyme T4	164	1449	$P3_221$	1.06	*ACORN*	Mooers & Matthews (2004)
1swz	Lysozyme T4	164	1410	$P3_221$	1.06	*SIR*2002	Mooers & Matthews (2004)
1mc2	Phospholipase A_2	134	1018	$C2$	0.85	*SnB*	Liu *et al.* (2003)
4gs3	*Tte*PriBN	107	860	$C2$	1.09	*SHELXD*	Present work
1ctj	Cytochrome c_6	90	767	$R3$	1.10	*SHELXS*	Frazão *et al.* (1995)
1tuk	Lipid protein 2G	67	498	$C2$	1.12	*SnB*	Hoh *et al.* (2005)
1yk4	Rubredoxin Pa	52	919	$P2_12_12_1$	0.69	*SHELXD*	Bönisch *et al.* (2005)
2pya	Rubredoxin Pa	52	851	$P2_12_12_1$	0.86	*SHELXD*	Bönisch *et al.* (2007)
1bx7	Hirustasin	51	387	$P4_32_12$	1.20	*SHELXD*	Usón *et al.* (1999)
1bx8	Hirustasin	49	355	$P4_32_12$	1.40	*SHELXD*	Usón *et al.* (1999)
3psm	Defensin SPE10	2 × 47	758	$P2_1$	0.98	*SnB*	H. Zhou, X. Song & W. Gong (unpublished work)
1p9g	EAFP-2	41	305	$P2_1$	0.84	*SnB*	Xiang *et al.* (2004)
1lu0	CMTI inhibitor	2 × 29	444	$P2_12_12$	1.03	*SHELXD*	Thaimattam *et al.* (2002)
1vrz	21-Residue peptide	22	286	$C2$	1.05	*SnB*	Rudresh *et al.* (2004)
3njw	Lasso peptide	19	144	$P2_12_12_1$	0.86	*SHELXD*	Nar *et al.* (2010)
1a7z	Actinomycin Z3	3 × 11	330	$P2_12_12_1$	0.95	*SHELXD*	Schäfer *et al.* (1998)
1a7y	Actinomycin D	3 × 11	314	$P1$	0.94	*SHELXD*	Schäfer *et al.* (1998)
1t7h	Endothelin-1	2 × 18	307	$P2_12_12$	1.13	*SnB*	Hoh *et al.* (2004)
1sho	Vancomycin	2 × 9	210	$P4_32_12$	1.09	*SHELXD*	Schäfer *et al.* (1996)
Oligonucleotides							
3mbu	PNA	4 × 10	740	$P2$	1.05	*SHELXD*	Yeh *et al.* (2010)
3mbs	PNA	9 + 8 + 9 + 8	624	$P1$	1.27	*SHELXD*	Yeh *et al.* (2010)
1g4q	RNA/DNA hybrid	20	414	$P2_12_12_1$	1.15	*SnB*	Han (2001)
3omj	Py-im polyamide	2 × 10	406	$P4_12_12$	0.95	*SHELXD*	Chenoweth & Dervan (2010)
3i5l	DNA oligomer	2 × 10	441	$P1$	1.18	*SHELXD*	Chenoweth & Dervan (2009)
3i5e	DNA oligomer	10	297	$C2$	0.98	*SHELXD*	Chenoweth & Dervan (2009)
3ngg	Omwaprin A	2 × 50	714	$P2_1/c$	1.33	*SHELXD*	Banigan *et al.* (2010)
3odv	α-Ktx toxin	2 × 38	566	$P\bar{1}$	0.95	*SHELXD*	Pentelute *et al.* (2010)
3e7r	Plectasin	40	309	$P\bar{1}$	1.00	*SHELXS*	Mandal, Pentelute, Tereshko, Thammavongsa *et al.* (2009)
3e8y	BMBKTX1 toxin	30	240	$I4_1/a$	1.10	*SHELXS*	Mandal, Pentelute, Tereshko, Kossiakoff *et al.* (2009)

sequence length smaller than 200 (survey performed in May 2012). With such high-resolution diffraction data available, direct-methods phasing is no longer only restricted to small-molecule structures and constitutes a good way of obtaining unbiased phases.

3.2. Refinement and structural description of the final model

The final model consists of 90 amino acids in two fragments composed of residues 7–80 and 89–104. The loop located between the fourth and fifth β-strands (L_{45}) and extending from residue 81 to 88 as well as the first six residues of the N-terminus (and the three residues from the purification step) were not visible in electron-density maps and were therefore not modeled. The lack of electron density is likely to reflect the dynamic nature of these elements in the crystal structure. 96 water molecules were placed with occupancies equal to 1.0 or 0.5. The model refined to an R_{cryst} of 12.92% for 29 367 reflections stronger than $4\sigma(F_o)$. No backbone torsion angles fall into disallowed regions of the Ramachandran plot. A close-up of the $2F_o - F_c$ electron density superposed on the final model is displayed in Fig. 1(*b*). The view is the same as for the unbiased electron-density map from direct-methods phasing and illustrates the changes between the initial model from *SHELXD* and the refined structure. Some isolated water molecules are now part of the main chain or side chains, fragments (bottom right) are linked together and the electron density of some side chains is better defined (Phe33, middle right; Leu21, front center).

*Tte*PriBN adopts the typical OB-domain folding: a closed β-barrel formed by a coiled five-stranded β-sheet. Usually, the

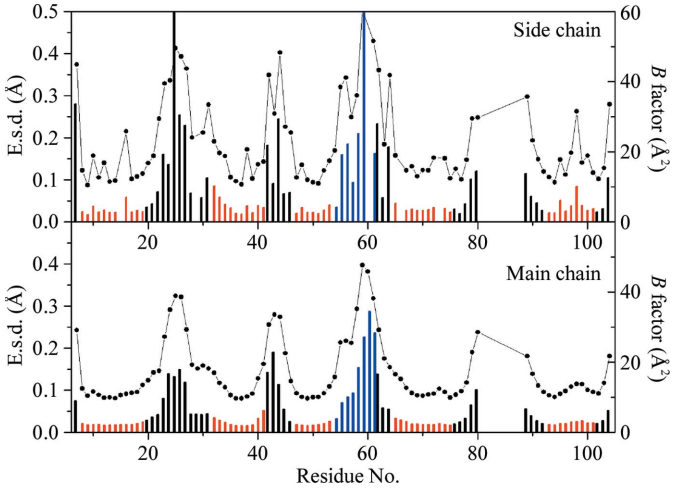

Figure 2
Averaged standard uncertainties of atomic coordinates and average isotropic *B* factors for main-chain (bottom) and side-chain (top) atoms, represented by bars and connected black circles, respectively. Coiled conformation and β-sheets are colored black and red, respectively. Residues which adopt a helical conformation in other PriB structures are represented in blue.

barrel is capped by an α-helix located between the third and fourth strands; however, in *Tte*PriBN the main chain does not adopt a helical conformation.

The standard deviation (e.s.d.) of atomic coordinates and the *B* factor, both averaged over main-chain and side-chain atoms, are represented in Fig. 2. The thermal motion is particularly low in the β-sheets, where the *B* factor has values of between 15 and 20 Å^2 for main-chain atoms, with slightly higher values in the side chains. As expected, the loops connecting the β-strands are more disordered and the *B* factor reaches values of up to 40 and 50 Å^2 in the main chain and side chains, respectively. The part of the sequence that adopts a helical conformation in other PriB structures is particularly flexible, with *B* factors of up to 50 Å^2 in main-chain atoms. The e.s.d.s, which are represented as bars in Fig. 2, follow the behavior of the thermal motion: the average e.s.d. is low in the β-sheets (below 0.05 Å for main-chain atoms and below 0.1 Å in the side chains), increases in the disordered loop regions (up to 0.02 and 0.5 Å for main chain and side chains, respectively) and has even higher values in the part that corresponds to the helix in other PriB structures. This behavior reflects the well known correlation between thermal disorder and positional uncertainty.

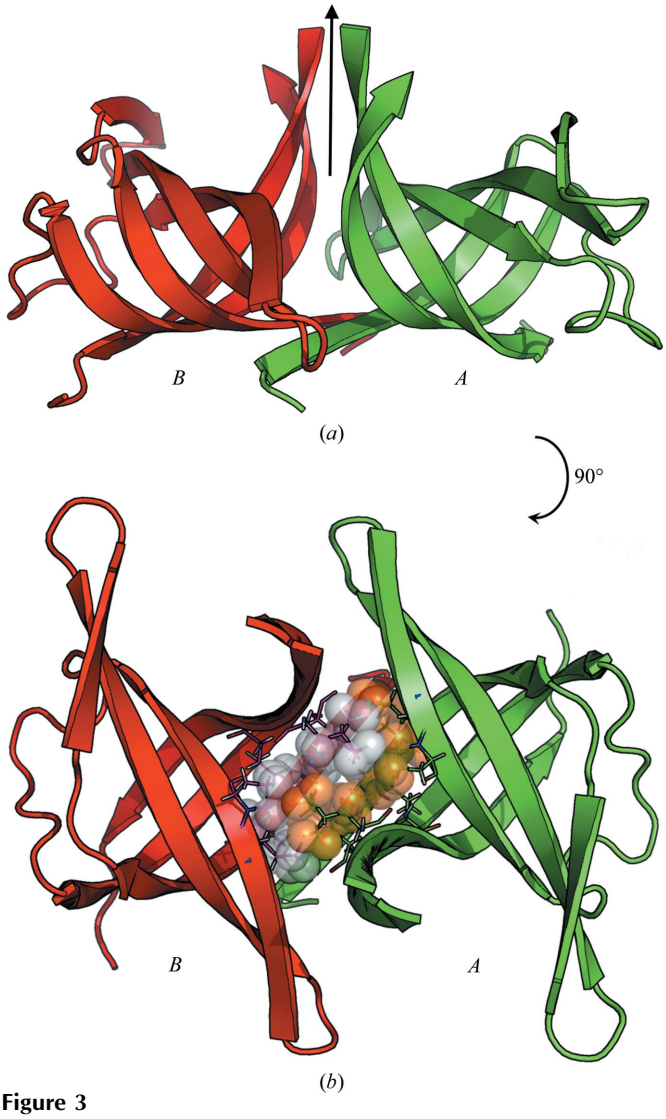

Figure 3
(*a*) Cartoon representation of the *Tte*PriBN homodimer created by the crystallographic twofold axis (black arrow). Monomers *A* and *B* are colored green and red, respectively. (*b*) A 90° rotated view of the dimer illustrating the hydrophobic interactions between the monomers. The H atoms of hydrophobic residues are represented as orange (monomer *A*) and white (monomer *B*) spheres.

3.3. Oligomerization

3.3.1. Dimerization. *Tte*PriBN forms a homodimer which is located at the twofold axis of space group *C*2. Fig. 3(*a*) shows the dimer in a cartoon representation. The protomers are intertwined and exhibit an extensive dimerization interface, which amounts to 920 Å^2 per molecule. Contacts are made at the open end of the β-barrel and *via* loop L_{23}, which wraps around the β-strands of the other chain. The monomers are joined together by six hydrogen bonds between peptide units of the first β-strand of each chain. The β-sheet hydrogen-bond pattern spans from residues 8 to 14; the linked OB domains therefore form a large ten-stranded β-barrel. Furthermore, hydrophobic interactions between side chains in the core of the molecule, two hydrogen bonds involving main chains and side chains between the second and third strands of different monomers and one hydrogen bond between Asn7 and Pro40 in L_{23} stabilize the dimer. Fig. 3(*b*) illustrates the hydrophobic interactions between the monomers.

3.3.2. Crystal contacts. The solvent content of the crystal is 32%, which is rather low for a protein. Indeed, the crystal packing is dense: the closest interactions are found within the molecules forming the dimer. Furthermore, one monomer interacts *via* 12 hydrogen bonds and four salt bridges with five

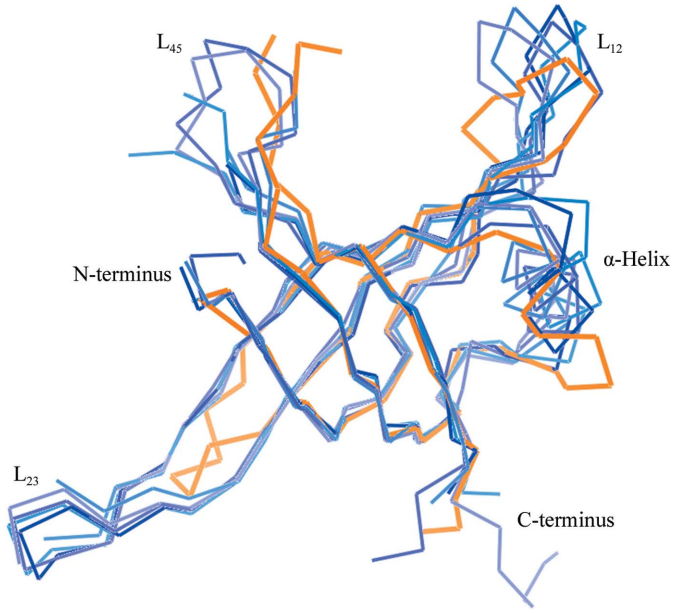

Figure 4
Superposition of *Tte*PriBN (orange line) and the five other PriB structures (shades of blue) determined to date. The following PriB models from different bacterial sources were used: PDB entries 1txy, 3k8a, 3en2, 3fhw and 4apv.

Table 3
PriB structures available in the PDB.

Each structure is characterized by the PDB code, protein source, crystal symmetry, data resolution, the oligomeric state and the number of molecules in the asymmetric unit.

PDB code	Protein source	Symmetry	Resolution (Å)	Oligomeric state	Molecules per asymmetric unit	Reference
1txy	*E. coli*	$P2_12_12_1$	2.00	Dimer	2	Lopper *et al.* (2004)
1v1q	*E. coli*	$P2_12_12_1$	2.10	Dimer	2	Liu *et al.* (2004)
1woc	*E. coli*	$C2$	2.00	Dimer × 2	4	Shioi *et al.* (2005)
2ccz	*E. coli* + DNA	$P2_12_12_1$	2.70	Dimer	2	Huang *et al.* (2006)
2pnh	*E. coli*	$P2_12_12_1$	2.25	Dimer	2	Lopper *et al.* (2007)
3k8a	*N. gonorrhoeae*	$I4_1$	2.70	Dimer	2	Dong *et al.* (2010)
3en2	*R. solanacaerum*	$I4_122$	2.30	Dimer	1	Northeast Structural Genomics Consortium (unpublished work)
3fhw	*B. parapertussis*	$P2_12_12$	1.90	Dimer	2	Northeast Structural Genomics Consortium (unpublished work)
3klw	*B. parapertussis*	$P6_3$	2.00	Dimer	2	Northeast Structural Genomics Consortium (unpublished work)
4apv	*K. pneumoniae*	$C2$	2.10	dimer	1	Y.-H. Lo, Y.-H. Huang, C.-D. Hsiao & C.-Y. Huang (unpublished work)

neighboring PriB molecules. It forms also hydrophobic interactions with two more symmetry-related molecules.

3.4. Comparison of *Tte*PriBN with other PriB proteins

3.4.1. Structure. To date, ten PriB structures from five different bacterial sources have been determined: *E. coli* (*Eco*PriB), *Neisseria gonorrhoeae* (*Ngo*PriB), *Ralstonia solanacaerum* (*Rso*PriB), *Bacillus parapertussis* (*Bpa*PriB) and *Klebsiella pneumoniae* (*Kpn*PriB). Their PDB codes, references and some of their characteristics are summarized in Table 3. All of the structures are dimers formed by two identical protein chains; however, the sequence of full-length *Tte*PriB contains an N-terminal and C-terminal OB domain, comparable to the thermostable SSBs from *Thermus aquaticus* (Jędrzejczak *et al.*, 2006) and the thermophilic *Deinococcus* (Eggington *et al.*, 2004) and *Thermus* genera (Dąbrowski *et al.*, 2002).

Fig. 4 shows a superposition of the PriB structures from the five different bacterial sources on the model of *Tte*PriBN. The β-barrel is very well conserved in all structures: the maximum $C^{\alpha}-C^{\alpha}$ distance between *Tte*PriBN and *Ngo*PriB amounts to 1.7 Å. For mesophile PriBs, other structural features also superpose quite well, such as loop L_{23}, which participates in dimerization contacts; even the more flexible loops L_{12} and L_{45} are similar. On the other hand, *Tte*PriBN shows several differences: loop L_{23} is seven residues shorter than those in mesophile PriBs and L_{45} is five amino acids longer than in *Eco*PriB and *Kpn*PriB and is seven residues longer than in the other structures. As a consequence, the intertwining of L_{23} with the other protomer is less pronounced. Loop L_{12} has a similar length, but does not superpose well on those of the

Figure 5
Sequence alignment of the N- and C-termini of *Tte*PriB and five other PriB structures. The same PriB models as in Fig. 4 were used. Amino acids in and near β-strands are highlighted in gray; these residues are the most structurally conserved in all PriB models (with an r.m.s.d. of <2 Å). Residues with left-handed helix dihedral angles and conserved residues participating in ssDNA binding are highlighted in yellow and magenta, respectively.

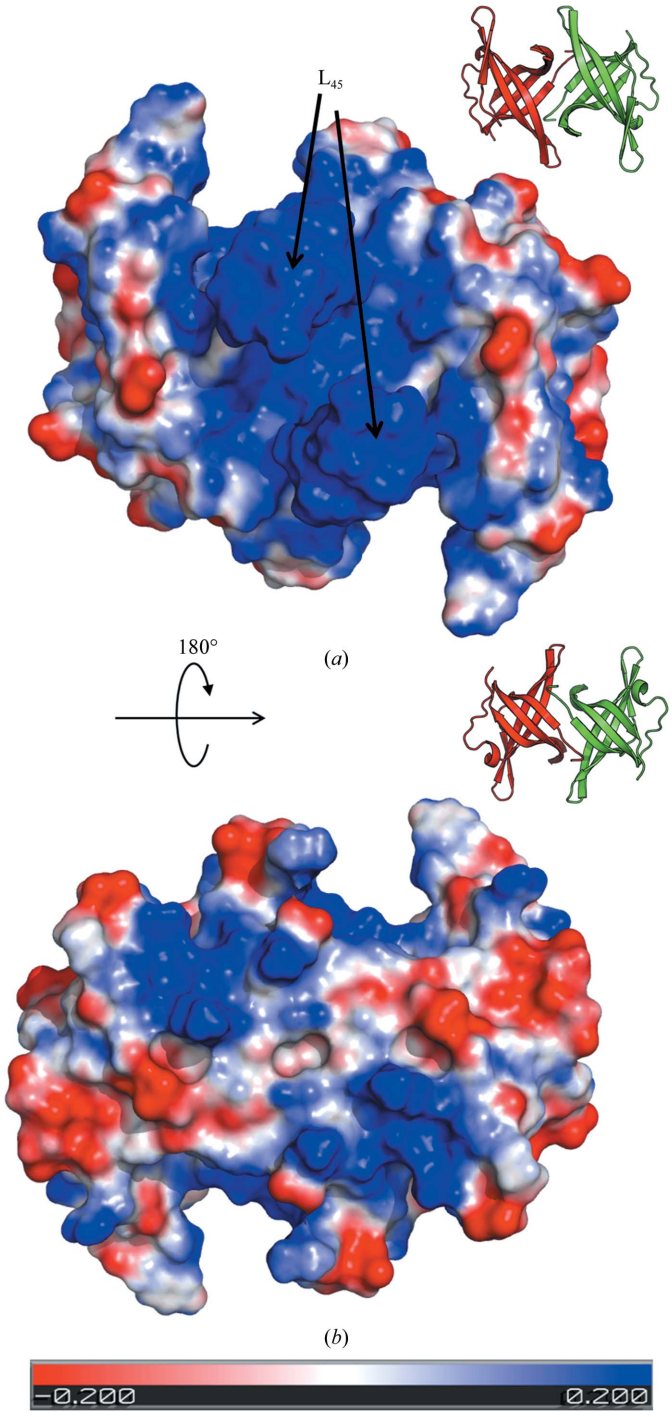

Figure 6
Total electrostatic potential of *Tte*PriBN shown as a color-coded map displayed on the surface of the protein. Negative, zero and positive potentials are displayed in red, white and blue, respectively. The scale goes from −0.2 to +0.2 |e| Å^{-1}. The L_{45} loops point out of the plane in (*a*) and into the plane in (*b*). The same orientation of the dimer is represented as cartoon image.

other PriBs. However, as this region seems flexible in all PriB structures, this feature is not necessarily unique to *Tte*PriBN.

3.4.2. Sequence comparison. Based on structural superposition, the sequence of *Tte*PriBN was aligned with those of the other PriB structures. Furthermore, the C-terminus of *Tte*PriB was aligned by identifying similar motifs with respect to the N-terminal sequence. The alignment is displayed in Fig. 5. There are numerous hydrophobic residues located in the β-strands (highlighted in gray) which agglutinate in the core of the protein. Interestingly, there are several residues in a left-handed helical conformation in all PriB structures. The first occurrence is located before the second β-strand β_2, the second is in turn L_{23} and the third is in β_4. This conformation is a consequence of the turn of the main chain which is necessary for formation of the β-sheet.

In *Tte*PriBN, the distance between the C-terminal C^{α} of Glu107 and the N-terminal C^{α} of Glu7 of the symmetry-related molecule in the dimer amounts to 16.5 Å. According to the sequence alignment, there are five residues available to bridge this distance. In one of the β-strands, the distance over six $C^{\alpha}-C^{\alpha}$ bonds (corresponding to five C^{α} atoms between the end points) is 20.1 Å, which is larger than the length of the domain linker necessary to bridge the N- and C-terminal domains. Therefore, it is physically possible that the N- and C-termini of the full-length *Tte*PriB are arranged like the monomers in the *Tte*PriBN homodimer.

3.4.3. ssDNA binding. To date, only one structure of PriB complexed with ssDNA has been reported in the literature: the structure of *Eco*PriB bound to an oligonucleotide of 15 bases in length (dT_{15}; Huang *et al.*, 2006). In the crystal structure, the ssDNA is enclosed by monomer *A* from one dimer and the adjacent monomer *B*′ from a symmetry-related molecule. The 5′ and 3′ termini interact mainly with monomer *A*, while the central part of dT_{15} is coordinated to monomer

Table 4
List of *Eco*PriB–ssDNA interactions and equivalent residues in *Tte*PriB.

T and P stand for the base and phosphate groups of ssDNA, respectively. Conserved residues of the same or a similar type are highlighted in bold; the fourth column gives the positional shift in the sequence alignment.

	*Eco*PriB	*Tte*PriB	Shift
T2	His64 (*A*″)	—	—
T3	**Arg34 (*B*)**	**Arg40 (C-ter)**	***i***
	Trp47 (*B*)	**Trp47 (C-ter)**	***i* + 2**
T4	Ser22 (*A*)	Tyr28	*i*
$P_{4/5}$	**Lys18 (*B*′)**	**Arg23 (C-ter)**	***i***
T5	Asn85 (*A*)	—	—
	Ser88 (*A*)	—	—
	Lys18 (*B*′)	**Arg23 (C-ter)**	***i***
T6	Lys18 (*A*)	Phe23	*i* − 1
		His25	*i* + 1
$P_{6/7}$	**Lys89 (*A*)**	**Lys89**	***i* − 2**
T7	**Arg13 (*B*′)**	**Lys18 (C-ter)**	***i***
	Leu16 (*B*′)	Val21 (C-ter)	*i*
$P_{8/9}$	**Lys84 (*B*′)**	**Lys82 (C-ter)**	***i* + 1**
	Lys89 (*B*′)	**Lys91 (C-ter)**	***i***
T9	Asn85 (*B*′)	Lys82 (C-ter)	*i*
$P_{11/12}$	Arg13 (*A*)	—	—
T11	Gln49 (*B*′)	Tyr47 (C-ter)	*i*
T12	**Glu32 (*A*)**	**Glu38**	***i***
T13	**Glu32 (*A*)**	**Glu38**	***i***
	Arg34 (*A*)	**Lys46**	**Next strand**
	Trp47 (*A*)	—	—
$P_{13/14}$	**Lys82 (*B*)**	**Arg78 (C-ter)**	***i* − 1**
T14	His57 (*A*‴)	—	—
	Arg34 (*B*′)	**Arg40 (C-ter)**	***i* + 1**

B'. The oligonucleotide is therefore in an Ω-shaped conformation. *Eco*PriB binds to ssDNA essentially through electrostatic interactions; the binding surface potential is highly positive (Lopper *et al.*, 2004). It was shown by mutational studies that this region, especially that within loop L_{45} (residues Lys82, Lys84 and Lys89), plays a critical role in ssDNA binding.

By comparing the superposed *Eco*PriB and *Tte*PriB structures as well as their sequence alignment, it was investigated whether the residues that contribute to ssDNA binding in *Eco*PriB are conserved in *Tte*PriB. Monomer *A* and its symmetry mate B' were compared with the N- and C-termini of *Tte*PriB, respectively, by inspecting the aligned sequences (Fig. 5) and, for the former, by analyzing the superposed models of *Eco*PriB–ssDNA (monomer *A*) and *Tte*PriBN. The *Eco*PriB–ssDNA interactions and equivalent residues in *Tte*PriB are listed in Table 4. A large proportion of the residues that participate in ssDNA binding are conserved, *i.e.* are of the same or a similar type, in *Tte*PriB (highlighted in bold), especially the basic residues such as lysine and arginine. For example, the three lysine residues in loop L_{45} (residues Lys82, Lys84 and Lys89) have two and three equivalents in the N- and C-termini of *Tte*PriB (Lys81 and Lys88 as well as Lys82, Lys83 and Lys91), respectively. This strongly suggests that the DNA-binding mode of *Tte*PriB is similar to that of *Eco*PriB.

3.4.4. Electrostatic potential. The electrostatic potential of the *Eco*PriB dimer is negative on the surface corresponding to the continuous six-bladed β-barrel (Lopper *et al.*, 2004), whereas it is positive on the opposite surface where the L_{45} loops protrude from the core and where the ssDNA binds to the protein. The electrostatic potential of *Tte*PriB represented in two orientations is displayed in Fig. 6. As in *Eco*PriB, the potential in the region around the two L_{45} loops is positive, as expected from the presence of numerous positively charged residues. On the opposite surface and the sides of the β-barrel there are several positive, negative and neutral patches. As previously indicated by the sequence comparison (§3.4.3), *Tte*PriBN and *Eco*PriB have a similar ssDNA-binding mode which is governed by electrostatic interactions. This is in contrast to SSB proteins, which share structural similarity and bind ssDNA mainly *via* base-stacking interactions.

4. Conclusion

The N-terminal domain of the thermostable *Tte*PriB was expressed, purified and crystallized. The resolution and quality of the diffraction data allowed *ab initio* solution of the structure without using prior knowledge such as heavy-atom positions. The resulting *Tte*PriBN model is one of the largest protein structures among macromolecules solved by direct methods. *Tte*PriBN adopts the OB-domain fold and is in a dimeric oligomerization state. In the crystal lattice, one loop fragment is disordered and is not visible in electron-density maps. The model displays significant differences to PriB structures from other bacterial sources; namely, the loops L_{23} and L_{45} are shorter and longer, respectively. Furthermore, the part which adopts a helical conformation in other PriB structures is in a coiled arrangement in *Tte*PriBN. Comparison of the structure and sequence with an *E. coli* ssDNA–PriB complex suggests that the binding mode of ssDNA is similar, *i.e.* it is governed by electrostatic interactions between the ssDNA and positively charged residues. The positive electrostatic potential around loop L_{45}, which contains several positively charged residues crucial for ssDNA binding in *Eco*PriB, also implies the electrostatic nature of the binding mechanism.

This work was supported in part with Federal funds from the National Cancer Institute, National Institutes of Health contract No. NO1-CO-12400, the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research and Polish State Committee for Scientific Research grants No. N302 116636 and N302 439439. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does the mention of trade names, commercial products or organizations imply endorsement by the US Government. Diffraction data were collected on the SER-CAT beamline 22-ID at the Advanced Photon Source, Argonne National Laboratory. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. W-31-109-Eng-38.

References

Allen, F. H. (1986). *Acta Cryst.* B**42**, 515–522.
Banigan, J. R., Mandal, K., Sawaya, M. R., Thammavongsa, V., Hendrickx, A. P., Schneewind, O., Yeates, T. O. & Kent, S. B. (2010). *Protein Sci.* **19**, 1840–1849.
Berman, H., Henrick, K. & Nakamura, H. (2003). *Nature Struct. Biol.* **10**, 980.
Bönisch, H., Schmidt, C. L., Bianco, P. & Ladenstein, R. (2005). *Acta Cryst.* D**61**, 990–1004.
Bönisch, H., Schmidt, C. L., Bianco, P. & Ladenstein, R. (2007). *J. Biol. Inorg. Chem.* **12**, 1163–1171.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
Chenoweth, D. M. & Dervan, P. B. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 13175–13179.
Chenoweth, D. M. & Dervan, P. B. (2010). *J. Am. Chem. Soc.* **132**, 14521–14529.
Coppens, P. (1997). *X-ray Charge Densities and Chemical Bonding*. Oxford University Press.
Cox, M. M., Goodman, M. F., Kreuzer, K. N., Sherratt, D. J., Sandler, S. J. & Marians, K. J. (2000). *Nature (London)*, **404**, 37–41.
Dąbrowski, S., Olszewski, M., Piątek, R., Brillowska-Dąbrowska, A., Konopa, G. & Kur, J. (2002). *Microbiology*, **148**, 3307–3315.
Dauter, Z. (2003). *Methods Enzymol.* **368**, 288–337.
DeLano, W. L. (2002). *PyMOL*. http://www.pymol.org.
Dittrich, B., Hübschle, C. B., Luger, P. & Spackman, M. A. (2006). *Acta Cryst.* D**62**, 1325–1335.
Domagała, S., Fournier, B., Liebschner, D., Guillot, B. & Jelsch, C. (2012). *Acta Cryst.* A**68**, 337–351.
Dominiak, P. M., Volkov, A., Li, X., Messerschmidt, M. & Coppens, P. (2006). *J. Chem. Theory Comput.* **3**, 232–247.
Dong, J., George, N. P., Duckett, K. L., DeBeer, M. A. & Lopper, M. E. (2010). *Nucleic Acids Res.* **38**, 499–509.
Eggington, J. M., Haruta, N., Wood, E. A. & Cox, M. M. (2004). *BMC Microbiol.* **4**, 2.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Frazão, C., Soares, C. M., Carrondo, M. A., Pohl, E., Dauter, Z., Wilson, K. S., Hervás, M., Navarro, J. A., De la Rosa, M. A. & Sheldrick, G. M. (1995). *Structure*, **3**, 1159–1169.
Greipel, J., Urbanke, C. & Maass, G. (1989). *Protein–Nucleic Acid Interaction*, edited by W. Saenger & U. Heinemann, pp. 61–86. London: Macmillan.
Han, G. W. (2001). *Acta Cryst.* D**57**, 213–218.
Hoh, F., Cerdan, R., Kaas, Q., Nishi, Y., Chiche, L., Kubo, S., Chino, N., Kobayashi, Y., Dumas, C. & Aumelas, A. (2004). *Biochemistry*, **43**, 15154–15168.
Hoh, F., Pons, J.-L., Gautier, M.-F., de Lamotte, F. & Dumas, C. (2005). *Acta Cryst.* D**61**, 397–406.
Huang, C.-Y., Hsu, C.-H., Sun, Y.-J., Wu, H.-N. & Hsiao, C.-D. (2006). *Nucleic Acids Res.* **34**, 3878–3886.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Jędrzejczak, R., Dauter, M., Dauter, Z., Olszewski, M., Długołęcka, A. & Kur, J. (2006). *Acta Cryst.* D**62**, 1407–1412.
Jelsch, C., Guillot, B., Lagoutte, A. & Lecomte, C. (2005). *J. Appl. Cryst.* **38**, 38–54.
Kabsch, W. (1976). *Acta Cryst.* A**32**, 922–923.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nature Protoc.* **3**, 1171–1179.
Liu, J.-H., Chang, T.-W., Huang, C.-Y., Chen, S.-U., Wu, H.-N., Chang, M.-C. & Hsiao, C.-D. (2004). *J. Biol. Chem.* **279**, 50465–50471.
Liu, Q., Huang, Q., Teng, M., Weeks, C. M., Jelsch, C., Zhang, R. & Niu, L. (2003). *J. Biol. Chem.* **278**, 41400–41408.
Lopper, M., Boonsombat, R., Sandler, S. J. & Keck, J. L. (2007). *Mol. Cell*, **26**, 781–793.
Lopper, M., Holton, J. M. & Keck, J. L. (2004). *Structure*, **12**, 1967–1975.
Mandal, K., Pentelute, B. L., Tereshko, V., Kossiakoff, A. A. & Kent, S. B. (2009). *J. Am. Chem. Soc.* **131**, 1362–1363.
Mandal, K., Pentelute, B. L., Tereshko, V., Thammavongsa, V., Schneewind, O., Kossiakoff, A. A. & Kent, S. B. (2009). *Protein Sci.* **18**, 1146–1154.
Mooers, B. H. M. & Matthews, B. W. (2004). *Acta Cryst.* D**60**, 1726–1737.
Morris, R. J. & Bricogne, G. (2003). *Acta Cryst.* D**59**, 615–617.
Murzin, A. G. (1993). *EMBO J.* **12**, 861–867.
Nar, H., Schmid, A., Puder, C. & Potterat, O. (2010). *ChemMedChem*, **5**, 1689–1692.
Olszewski, M., Mickiewicz, M. & Kur, J. (2008). *Arch. Microbiol.* **190**, 79–87.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pentelute, B. L., Mandal, K., Gates, Z. P., Sawaya, M. R., Yeates, T. O. & Kent, S. B. H. (2010). *Chem. Commun.* **46**, 8174–8176.
Petrova, T. & Podjarny, A. (2004). *Rep. Prog. Phys.* **67**, 1565.
Rudresh, Ramakumar, S., Ramagopal, U. A., Inai, Y., Goel, S., Sahal, D. & Chauhan, V. S. (2004). *Structure*, **12**, 389–396.
Schäfer, M., Schneider, T. R. & Sheldrick, G. M. (1996). *Structure*, **4**, 1509–1515.
Schäfer, M., Sheldrick, G. M., Bahner, I. & Lackner, H. (1998). *Angew. Chem. Int. Edit.* **37**, 2381–2384.
Sheldrick, G. M. (1990). *Acta Cryst.* A**46**, 467–473.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Sheldrick, G. M. & Gould, R. O. (1995). *Acta Cryst.* B**51**, 423–431.
Shioi, S., Ose, T., Maenaka, K., Shiroishi, M., Abe, Y., Kohda, D., Katayama, T. & Ueda, T. (2005). *Biochem. Biophys. Res. Commun.* **326**, 766–776.
Thaimattam, R., Tykarska, E., Bierzynski, A., Sheldrick, G. M. & Jaskolski, M. (2002). *Acta Cryst.* D**58**, 1448–1461.
Usón, I., Sheldrick, G. M., de La Fortelle, E., Bricogne, G., Di Marco, S., Priestle, J. P., Grütter, M. G. & Mittl, P. R. (1999). *Structure*, **7**, 55–63.
Xiang, Y., Huang, R.-H., Liu, X.-Z., Zhang, Y. & Wang, D.-C. (2004). *J. Struct. Biol.* **148**, 86–97.
Yeh, J. I., Pohl, E., Truan, D., He, W., Sheldrick, G. M., Du, S. & Achim, C. (2010). *Chem. Eur. J.* **16**, 11867–11875.