PROTEINS: Structure, Function, and Bioinformatics 58:764–768 (2005)

STRUCTURE NOTE

Crystal Structure of the Copper Homeostasis Protein (CutCm) From *Shigella flexneri* at 1.7 Å Resolution: The First Structure of a New Sequence Family of TIM Barrels

De-Yu Zhu,[1,3†] Yong-Qun Zhu,[1,3†] Ren-Huai Huang,[1] Ye Xiang,[1] Na Yang,[1,3] Hong-Xia Lu,[1] Gen-Pei Li,[1] Qi Jin,[2] and Da-Cheng Wang[1*]

[1]*Center for Structural and Molecular Biology, Institute of Biophysics, Chinese Academy of Sciences, Beijing, People's Republic of China*

[2]*State Key Laboratory for Molecular Virology and Genetic Engineering, Beijing, People's Republic of China*

[3]*Graduate School of the Chinese Academy of Sciences, Beijing, People's Republic of China*

Introduction. Copper is an essential heavy metal trace element to organisms, and organisms have obtained their own copper homeostasis mechanisms in their evolution.[1] Two types of gene families appear to be associated with the copper homeostasis in bacteria. One is the*cop* genes family, which is a well-understood system of active transport efflux pumps, and the other is the *cut* gene family, which has 6 *cut* gene members (*cut*A, *cut*B, *cut*C, *cut*D, *cut*E, *cut*F).[2,3] So far, the detailed functions of the proteins produced by the *cut* genes are not very clear.

The putative copper homeostasis protein (CutCm) produced by the *cut*Cm gene of *Shigella flexneri 2a str. 301* consists of 248 residues and belongs to the CutC family (Pfam-PF03932) (Fig. 1).[4,5] Some studies have implicated that CutCm may play a role in intracellular trafficking of Cu(I),[8,9] but the actual function of CutCm is still unknown.

Here we report the crystal structure of CutCm determined using multiple-wavelength anomalous dispersion (MAD) method, which is the first three-dimensional (3D) structure of CutC family. The tertiary structure of CutCm adopts a classical triosephosphate isomerase (TIM) barrel fold.[10] The structure and sequence comparisons show that the CutCm structure is the representative of a new sequence family of TIM barrels.

Materials and methods. *Cloning, expression, and purification:* Polymerase chain reaction (PCR) primers, including NdeI and XhoI restriction sites, were designed to amplify the *cut*Cm gene (GenBank: NP_707762) from the genomic DNA of *S. flexneri 2a strl 301*.[4] The gene was inserted into the pET22b(+) vector (Novagen), leading to a C-terminal His-tagged protein. The protein was overexpressed in *Escherichia coli* BL21-Codon Plus(DE3)-RIL (Stratagene). The cells were grown in Luria–Bertani (LB) medium at 37°C and induced with 1 m*M* isopropyl-β-D-thiogalactopyranoside (IPTG) when the culture reached an OD_{600} of 0.6–0.8. The harvested cells were resuspended in 30 mL lysis buffer (50 m*M* NaH_2PO_4, pH8.0; 300 m*M* NaCl; 10 m*M* imidazole) with 0.1 m*M* phenyl methyl sulfonyl fluoride (PMSF) and disrupted using FRENCH Pressure Cell. The lysate was centrifugated at 12,500 rpm at 4°C for 20 min. The supernatant was loaded directly into a Ni-nitriloacetic acid (Ni-NTA) column (Novagen) pre-equilibrated with lysis buffer. Then the column was washed with wash buffer (50 m*M* NaH_2PO_4, pH8.0; 300 m*M* NaCl; 20 m*M* imidazole) and the target protein was eluted with elution buffer (50 m*M* NaH_2PO_4, pH 8.0; 300 m*M* NaCl; 250 m*M* imidazole). The eluate was concentrated by ultrafiltration (Millipore) and then loaded into Superdex75 HR16/60 column (Amersham Pharmacia) pre-equilibrated with buffer of 50 m*M* NH_4HCO_3 at 20°C. Further purification was performed with a global Mono Q HR5/5 (Amersham Pharmacia). Finally, CutCm was desalted using the buffer of 50 m*M* Tris-HCl, pH 8.0, and concentrated for crystallization.

In order to produce the selenomethionine (SeMet) substituted derivate, the method of inhibition to methionine metabolism pathway was used.[11] The transformed BL21-Codon Plus(DE3)-RIL cells were grown in minimal medium at 37°C. When OD_{600} of the culture reached 0.6–0.8, solid amino acid supplements (Lys, Phe, Thr, Ile, Leu, Val, SeMet) were added to the culture. After 15 min, the expression was induced as usual. The purification of SeMet protein was performed using the same procedure as the described above for the native protein. Additionally, the reducing reagent β-mercaptoethanol (10 m*M*) was

†Both authors contributed equally to this work.

Grant sponsor: MOST of China; Grant numbers: 863-2002BA711A13 and 973-G1999075064. Grant sponsor: Chinese Academy of Science; Grant numbers: KSCX2-SW-322 and KSCX1-SW-17. Grant sponsor: KEK, Japan; Grant number: 00G209 (for X-ray data collection).

*Correspondence to: Da-Cheng Wang, Center for Structural and Molecular Biology, Institute of Biophysics, Chinese Academy of Sciences, 15 Datun Road, Beijing 100101, People's Republic of China. E-mail: dcwang@sun5.ibp.ac.cn

Received 30 August 2004; Accepted 5 September 2004

Published online 28 December 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20362

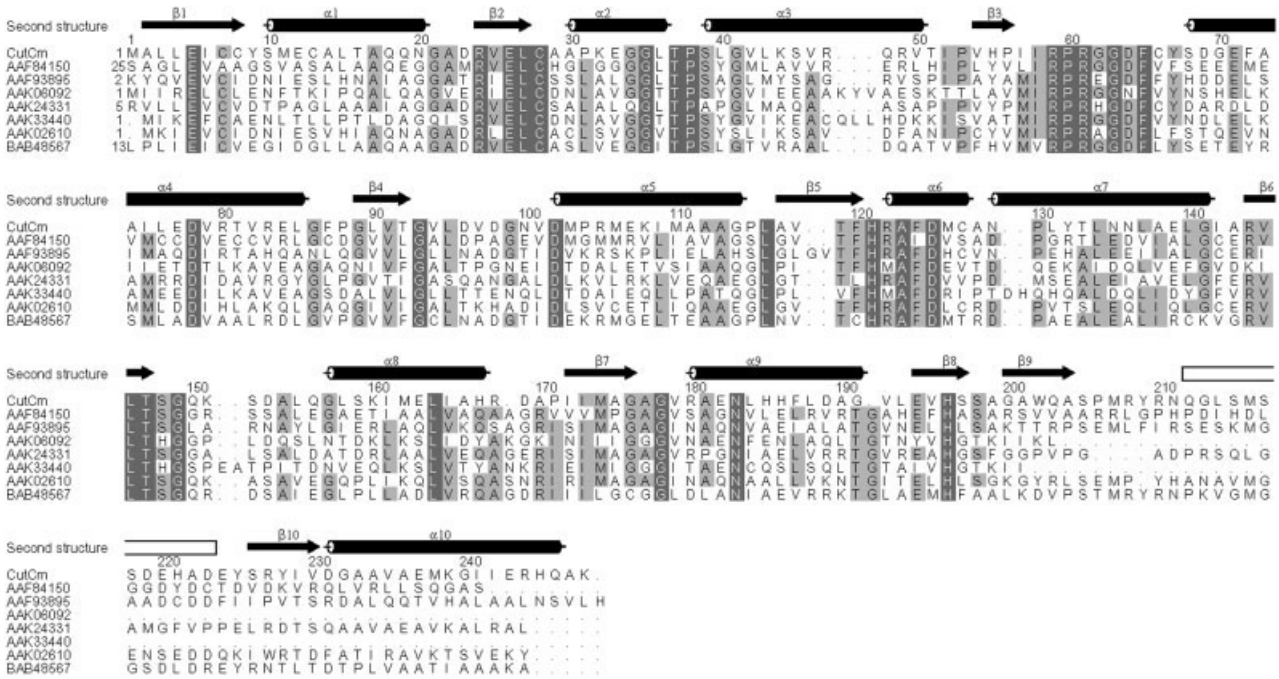

Fig. 1. Secondary structure of CutCm of *S. flexneri* and sequence alignment with its homologous proteins of CutC family. The numbering above the alignment corresponds to *S. flexneri* CutCm. The secondary-structure elements (β-sheets, α-helices) found in *S. flexneri* CutCm are shown above the sequences. The box represents the missing residues in the crystal structure of CutCm owing to lack of electron density. Conserved residues are shaded light gray and the highly conserved residues are dark gray. Seven aligned homologous proteins come from Pfam[5] (the names of proteins are from the GenBank accession ID): AAF84150 (*Xylella fastidiosa 9a5c*), AAF93895 (*Vibrio cholerae O1 biovar eltor str. N16961*), AAK06092 (*Lactococcus lactis* subsp. *lactis Il1403*), AAK24331 (*Caulobacter crescentus CB15*), AAK33440 (*Streptococcus pyogenes M1 GAS*), AAK02610 (*Pasteurella multocida* subsp. *multocida str. Pm70*), and BAB48567 (*Mesorhizobium loti*). Sequence alignment was performed with CLUSTALW[6] and the figure was prepared with ALSCRIPT.[7]

added to all purification buffers except the final buffer (50 m*M* Tris-HCl, pH 8.0) for desalting.

Crystallization: The crystallization experiments were conducted using the hanging-drop vapor-diffusion method at 20°C, with 2 μL drops containing a 50:50 (v/v) mixture of reservoir solution and protein solution (10 mg/mL) equilibrated against 0.5 mL reservoir solution. The reservoir solution for native crystal growth was 20% polyethylene glycol (PEG8000), 0.1 *M* sodium cacodylate, pH 6.0, and 0.15 *M* calcium acetate. The qualified native crystals were obtained by successive microseeding and macroseeding. The qualified SeMet derivative crystals were obtained by a combined procedure of cross-seeding (the seeds come from the native crystal), microseeding, and macroseeding,[12] and the reservoir solution was 20% PEG8000, 0.1 *M* *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid (HEPES)-Na, pH 7.0, and 0.1 *M* calcium acetate.

Data collecting and processing: All diffraction data were collected at the beamline 6A (BL6A) at the Photon Factory (Tsukuba, Japan) using an ADSC Quantum-4 charge-coupled device (CCD) detector. All crystals were briefly soaked in the paraffin oil (Hampton Research) after being mounted in nylon cryoloops (Hampton Research) and then flash-cooled in a nitrogen-gas stream at 95 K. The native data were collected at a wavelength of 0.9780 Å with a 1.7 Å resolution. The 3-wavelength MAD data diffracted up to 2.1 Å (peak: 0.97850 Å, inflection: 0.97938 Å, high-energy remote: 0.9700 Å) were collected from a SeMet derivative crystal. The peak and inflection point wavelengths were determined by recording an X-ray fluorescence spectrum. The native and SeMet derivative data were integrated using MOSFLM[13] and scaled using SCALA.[13] Data statistics are summarized in Table I.

Structure determination and refinement: The structure of CutCm was determined using the MAD method. Twenty-two possible selenium sites were found by SHELXD.[16] The best experimental phases for the 20 reasonable selenium sites (2 unreasonable selenium sites were removed) were calculated by SHARP.[17] After phasing, density modification was achieved by density modificiation (DM).[18] The initial model consisting of 85% of the structure of CutCm was built automatically by using ARP/wARP.[13,19] Regions not constructed were manually built using O.[20] The model was refined initially using the Crystallography & NMR System (CNS) version 1.1[21] against the data recorded at the peak wavelength and restrained by the experimental phases. With the R_{free} dropping below 30%, the model was refined against the 1.7 Å native data without phase restraints. The final model was manually achieved by using O.[20] The final R factor was 19.1% with R_{free} of 22.2%. The model quality was checked using PROCHECK[15] (Table I).

Results and Discussion. The final native structure model of CutCm includes 2 protein molecules (residues 1–210 and 223–248 for chain A, residues 1–211 and 224–250 for

TABLE I. Data Collection, Phasing, and Structure Refinement Statistics of CutCm

	Native	Inflect · point		SeMet Peak	High-Remote	
A. Data collection						
Space group (molecules/asu)	$C222_1$ (2)			$C222_1$ (2)		
Unit cell dimensions						
a (Å)	75.327			74.579		
b (Å)	97.672			97.579		
c (Å)	132.691			131.846		
			$\alpha = \beta = \gamma = 90°$			
Station	BL6A, PF			BL6A, PF		
Wavelength (Å)	0.978	0.97938		0.97850	0.9700	
Resolution range (Å)	45.83–1.70	59.26–2.10		59.26–2.10	59.26–2.10	
Observation ($I/\sigma(I) \geq 0$)	229712	296347		287695	300890	
Unique reflections ($I/\sigma(I) > 0$)	52869	28435		28370	28481	
Last shell (Å)	1.79–1.70	2.21–2.10		2.21–2.10	2.21–2.10	
R_{sym} (%)[a]: overall (last shell)	5.7 (15.9)	7.6 (20.9)		7.1 (16.4)	7.3 (18.1)	
$\langle I/\sigma(I)\rangle$: overall (last shell)	8.2 (4.4)	8.0 (3.4)		8.3 (4.3)	8.1 (4.0)	
Completeness (%): overall (last shell)	97.9 (91.7)	99.6 (97.5)		99.6 (97.6)	99.8 (98.4)	
Redundancy: overall (last shell)	4.4 (3.6)	10.4 (7.0)		10.5 (7.1)	10.6 (7.4)	
B. Phasing						
Selenium atom sites				20		
Resolution range of data used (Å)				59.26–2.10		
Phasing power		Fried	Iso	Fried	Fried	Iso
		0.406	1.61	2.418	2.055	0.62
Figure of merit (FOM)				0.66		
Density modification, FOM (2.1 Å)				0.88		
C. Structure refinement						
Resolution range (last shell) (Å)	45.83–1.70 (1.81–1.70)					
Sigma cutoff	0.0					
R-factor (last shell) (%)	19.8 (21.2)					
R_{free}[b] (last shell) (%)	22.6 (25.4)					
RMSD[c]						
Bond length (1–2) (Å)	0.004					
Angle (°)	1.2					
Dihedral (°)	22.0					
Improper (°)	0.80					
Wilson B-factor (Å^2)	14.0					
Average B-factor (Å^2)						
All atoms	14.10					
Protein atoms (3581)	12.39					
Protein A-chain (1782)	12.15					
Protein B-chain (1799)	12.63					
Ca (1)	15.48					
Water (547)	25.0					
Ramachandran statistics (%)[d]						
Most favored regions	91.3					
Additionally allowed regions	7.7					
Generously allowed regions	0.5					
Disallowed regions	0.5					

[a] $R_{sym} = \Sigma\,(|I| - |\langle I\rangle|)/\Sigma\langle I\rangle$.
[b] R_{free} calculated using 5% of total reflections omitted from refinement.
[c] RMSDs from ideal bond lengths/angles with respect to Engh and Huber.[14]
[d] Ramachandran statistics calculated using PROCHECK.[15]

chain B), 547 water molecules, and 1 Ca^{2+} atom. In addition, 1 disulfide bond is formed between the Cys8 and the Cys13 in each molecule. Some residues were missed in the model due to lack of electron density, which included the hexa-histine tag and 12 residues for each molecule (Asn A211 to Ala A222, and Gln B212 to Ala B223 for chains A and B, respectively), in addition to 2 residues (Leu A249, Glu A250) for chain A. The 2 molecules in the asymmetric unit are related by a noncrystallographic 2-fold axis [Fig. 2(a)]. They form a dimer with a buried surface area of 3688 Å^2 per monomer and 60 hydrogen bonds in between [root-mean-square deviation (RMSD) 0.49Å for their Cα atoms].

The CutCm monomer structure adopts a common TIM barrel $(\beta/\alpha)_8$.[10] It is composed of 10 β-strands (β1–β10) and 10 helices (α1–α10), where the 8 β-strands (β1–β8) are correspondingly surrounded by 8 helices (α1, α3–α5, and α7–α10) , while α2 and α6 are 3_{10}-helices [Figs. 1 and 2(b)]. Inside the barrel, there are 10 charged residues (Glu A5, Arg A24, Glu A26, His A55, Arg A59, His A121, Arg A122, Arg A144, Glu A194, and His A196). Additionally, 6 hydrogen bonds are formed between 6 of the 10 residues (A5 OE2…NE A24, 2.72 Å; A5 OE2…NH1 A24, 3.11 Å; A26 OE2…NE2 A196, 3.31 Å; A194 OE1…NH1 A24, 2.84 Å; A194 OE1…NE A144, 2.75 Å; A194 OE2…NH1 A144, 2.98 Å) [Fig. 2(c)]. At the N-terminal end of the barrel, all of the turns located between the α-helices and the subsequent β-strands are composed of only 3 or 4 residues except the loop (residues 167–172) between α8 and β7, whose length is uncertain owing to disorder. At the

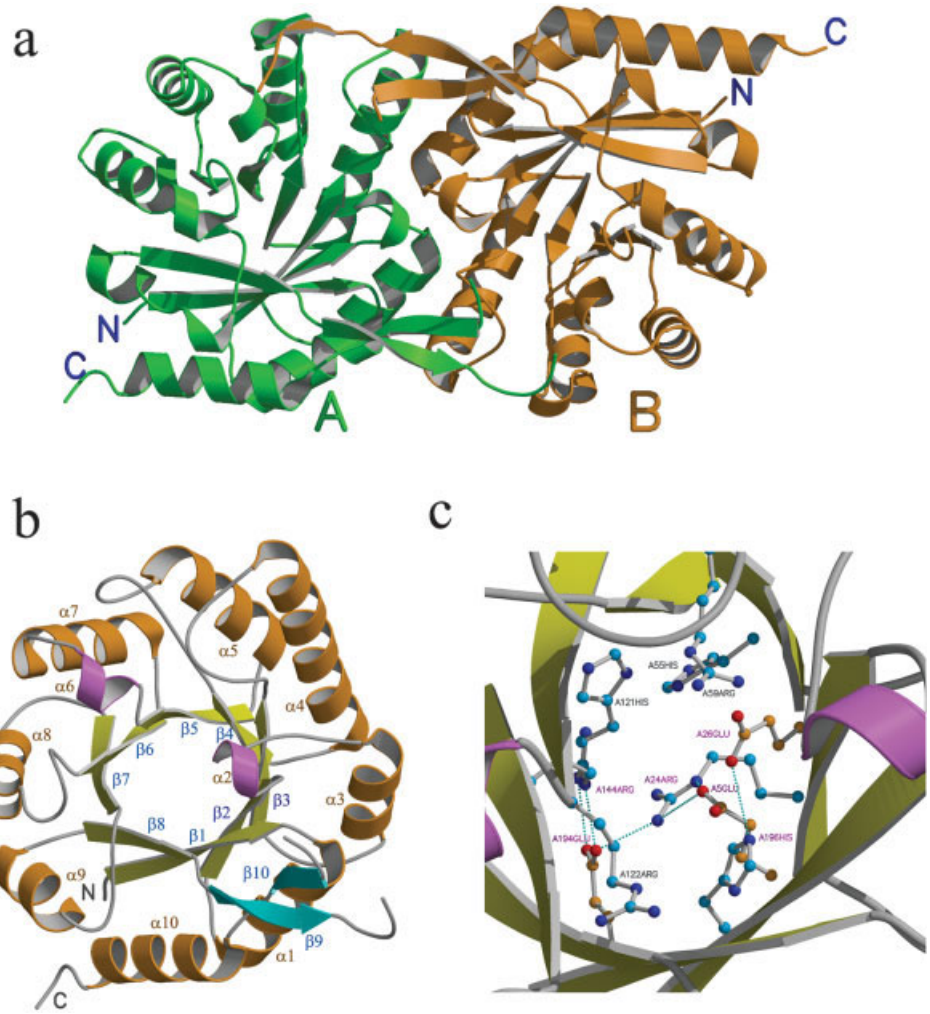

Fig. 2. (**a**) Ribbon diagram of the CutCm dimer. Ribbons are green and orange for molecules A and B, respectively. (**b**) The ribbon drawing of the CutCm molecule A structure. α-helices and 3_{10}-helices are shown in orange and magenta, respectively; β1–β8 and β9–β10 outside the barrel are shown in yellow and cyan, respectively. (**c**) The 10 charged residues inside the TIM barrel are presumed to be functionally important. The oxygen and nitrogen atoms are red and blue, respectively, and the hydrogen bonds are cyan. The figures were prepared with MOLSCRIPT.[22]

C-terminal end of the barrel, the long regions between the β-strands and the subsequent α-helices correspondingly include two 3_{10}-helices (α2, α6) [Fig. 2(b)]. Furthermore, the longer region between β8 and α10 was composed of 35 residues (A197–A231), which included 2 β-strands (β9, β10) and 12 missing residues. β9 and β10 are nearly parallel to the TIM barrel axis and are shown as the short handle of the TIM barrel.

So far, over 1800 TIM barrel domains have been found, which are classified into 29 homologous superfamilies and 139 sequence families (CATH version 2.5.1).[23,24] As expected, the combinatorial extension (CE) search[25] revealed that the relatively highly similar structures of CutCm all belong to the TIM barrel family. The PSI-BLAST[26] search indicated that there is no obvious sequence similarity between CutCm and all other structures of the TIM barrel family in the Protein Data Bank (PDB). For detecting more distant homologues and analogues of CutCm, the structure comparisons between the CutCm and the representatives of 29 homologous superfamilies of TIM barrel family were performed using the SSAP program.[27] The results showed that the closest structural homologue was phosphoribosyl anthranilate isomerase from *Thermotoga maritima*[28] (tPRAI; PDB ID: 1NSJ; SSAP score, 81.89%; overlap, 82%; RMSD, 3.05 Å for equivalent 194 Cα atoms; sequence identity, 14% for 205 aligned residues). The main differences between the 2 structures are located at the C-terminal end of the barrel, where the CutCm is more expansible and especially stretches out 2 additional β-strands (β9, β10).

According to the classification described by Nagano et al.[29] and the CATH levels,[24] the CutCm and tPRAI belong to the same homologous superfamily of TIM barrel, triose phosphate isomerase, flavin mononucleotide (FMN)-dependent oxidoreductases, phosphate-binding enzymes, and tryptophan biosynthesis enzymes homologous superfamily

(CATH Code: 3.20.20.90).[24] And within the homologous superfamily, there are 24 sequence families (http://www.biochem.ucl.ac.uk/bsm/cath_new/class3/20/20/90/index.html). Compared with their sequences using SSAP again, CutCm has less than 20% sequence identity to each representative member of the 24 sequence families for at least 60% residues of the smaller aligned protein. According to CATH criteria of S-level classification, CutCm should be assigned to a new sequence family of the homologous superfamily of TIM barrels.

The CutCm structure represents the first structure of the CutC family. So far, the specific function of CutCm is still not clear; the tertiary structure of CutCm reported in this article provides a sound basis for the in-depth study of its structure–function relationship. According to the functional character of common TIM barrel proteins, the loops at the C-terminal end of barrel are very important for their functions.[10] Mapping the conserved sequence motif region of CutC family (Fig. 1) onto the crystal structure of CutCm, one can find that the loop between β3 and α4 at the C-terminal end of the barrel, and 8 of 10 charge residues (Glu A5, Arg A24, Glu A26, His A121, Arg A144, Glu A194, His A196, and Arg A59) inside the barrel are conserved. It is plausible to speculate that the loop between β3 and α4 may be a part of the active site, and the polar property at the interior of the barrel is necessary for the function of CutCm.

Protein Data Bank accession code: Coordinates and structure factors for the structure of CutCm have been deposited at the PDB, Research Collaboratory for Structural Bioinformatics (RCSB, http://www.rcsb.org), with accession code 1X7I.

Acknowledgments. We thank Professor N. Sakabe for his help during data collection in Photon Factory in Japan, and Professor Jiang F. for providing the in-house X-ray facility for preliminary X-ray analysis.

REFERENCES

1. Gupta SD, Lee BT, Camakaris J, Wu HC. Identification of cutC and cutF (nlpE) genes involved in copper tolerance in *Escherichia coli*. J Bacteriol 1995;177:4207–4215.
2. Tanaka Y, Tsumoto K, Nakanishi T, Yasutake Y, Sakai N, Yao M, Tanaka I, Kumagai I. Structural implications for heavy metal-induced reversible assembly and aggregation of a protein: the case of *Pyrococcus horikoshii* CutA. FEBS Lett 2004;556:167–174.
3. Silver S, Phung LT. Bacterial heavy metal resistance: new surprises. Annu Rev Microbiol 1996;50:753–789.
4. Jin Q, Yuan Z, Xu J, Wang Y, Shen Y, Lu W, Wang J, Liu H, Yang J, Yang F, Zhang X, Zhang J, Yang G, Wu H, Qu D, Dong J, Sun L, Xue Y, Zhao A, Gao Y, Zhu J, Kan B, Ding K, Chen S, Cheng H, Yao Z, He B, Chen R, Ma D, Qiang B, Wen Y, Hou Y, Yu J. Genome sequence of *Shigella flexneri 2a*: insights into pathogenicity through comparison with genomes of *Escherichia coli K12* and *O157*. Nucleic Acids Res 2002;30:4432–4441.
5. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C, Eddy SR. The Pfam protein families database. Nucleic Acids Res 2004;32:D138–D141.
6. Thompson JD, Higgins DG, Gibson TJ. CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22:4673–4680.
7. Barton GJ. ALSCRIPT: a tool to format multiple sequence alignments. Protein Eng 1993;6:37–40.
8. Rensing C, Grass G. *Escherichia coli* mechanisms of copper homeostasis in a changing environment. FEMS Microbiol Rev 2003;27:197–213.
9. Kimura T, Nishioka H. Intracellular generation of superoxide by copper sulphate in *Escherichia coli*. Mutat Res 1997;389:237–242.
10. Wierenga RK. The TIM-barrel fold: a versatile framework for efficient enzymes. FEBS Lett 2001;492:193–198.
11. Doublie S. Preparation of selenomethionyl proteins for phase determination. Methods Enzymol 1997;276:523–530.
12. Stura, EA. Seeding techniques. In: Ducruix A, Giege R, editors. Crystallization of nucleic acids and proteins. Oxford, UK: Oxford University Press; 1999. p 177–207.
13. Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 1994;50:760–763.
14. Engh RA, Huber R. Accurate bond and angle parameters for X-ray protein structure refinement. Acta Crystallogr A 1991;47:392–400.
15. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–291.
16. Schneider TR, Sheldrick GM. Substructure solution with SHELXD. Acta Crystallogr D Biol Crystallogr 2002;58:1772–1779.
17. de la Fortelle E, Bricogne G. Maximum-likelihood heavy-atom parameter refinement in the MIR and MAD methods. Methods Enzymol 1997;276:472–494.
18. Cowtan KD, Zhang KY. Density modification for macromolecular phase improvement. Prog Biophys Mol Biol 1999;72:245–270.
19. Morris RJ, Perrakis A, Lamzin VS. ARP/wARP and automatic interpretation of protein electron density maps. Methods Enzymol 2003;374:229–244.
20. Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron density maps and thelocation of errors in these models. Acta Crystallogr A 1991;47:110–119.
21. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 1998;54:905–921.
22. Kraulis PJ. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J Appl Crystallogr 1991;24:946–950.
23. Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, Thornton JM. CATH—a hierarchic classification of protein domain structures. Structure 1997;5:1093–1108.
24. Pearl FM, Bennett CF, Bray JE, Harrison AP, Martin N, Shepherd A, Sillitoe I, Thornton J, Orengo CA. The CATH database: an extended protein family resource for structural and functional genomics. Nucleic Acids Res 2003;31:452–455.
25. Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng 1998;11:739–747.
26. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–3402.
27. Orengo CA,Taylor WR. SSAP: sequential structure alignment program for protein structure comparison. Methods Enzymol 1996;266:617–635.
28. Hennig M, Sterner R, Kirschner K, Jansonius JN. Crystal structure at 2.0 Å resolution of phosphoribosyl anthranilate isomerase from the hyperthermophile *Thermotoga maritima*: possible determinants of protein stability. Biochemistry 1997;36:6009–6016.
29. Nagano N, Orengo CA, Thornton JM. One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. J Mol Biol 2002;321:741–765.