Acta Crystallographica Section D Biological Crystallography

ISSN 0907-4449

#### Dong Hae Shin,<sup>a</sup>‡ In-Geol Choi,<sup>b</sup>‡ Didier Busso,<sup>b</sup> Jaru Jancarik,<sup>b</sup> Hisao Yokota,<sup>a</sup> Rosalind Kim<sup>a</sup> and Sung-Hou Kim<sup>a,b</sup>\*

<sup>a</sup>Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA, and <sup>b</sup>Department of Chemistry, University of California, Berkeley, California 94720-5230, USA

**‡** These authors contributed equally.

Correspondence e-mail: shkim@cchem.berkeley.edu

Structure of OsmC from *Escherichia coli*: a salt-shock-induced protein

The crystal structure of an osmotically inducible protein (OsmC) from Escherichia coli has been determined at 2.4 Å resolution. OsmC is a representative protein of the OsmC sequence family, which is composed of three sequence subfamilies. The structure of OsmC provides a view of a salt-shock-induced protein. Two identical monomers form a cylindrically shaped dimer in which six helices are located on the inside and two six-stranded  $\beta$ -sheets wrap around these helices. Structural comparison suggests that the OsmC sequence family has a peroxiredoxin function and has a unique structure compared with other peroxiredoxin families. A detailed analysis of structures and sequence comparisons in the OsmC sequence family revealed that each subfamily has unique motifs. In addition, the molecular function of the OsmC sequence family is discussed based on structural comparisons among the subfamily members.

Received 12 December 2003 Accepted 3 March 2004

**PDB Reference:** OsmC, 1nye, r1nyesf.

#### 1. Introduction

Bacterial cells often encounter unfavorable growth conditions in natural environments. Escherichia coli, a non-sporulating enterobacteriae, undergoes a global programmed modification of its gene-expression pattern and this results in the acquisition of resistance to chemical and physical stresses, such as heat, oxidative agents or hyperosmotic shock (Kolter et al., 1993). These properties provide a better chance of survival of the cells under adverse conditions. One key regulator of this genetic programme is the product of the rpoS gene, an RNA polymerase sigma factor called RpoS or  $\sigma^{s}$  (Lange & Hengge-Aronis, 1991). Many genes controlled by  $\sigma^{s}$  are inducible by elevated osmolarity (Hengge-Aronis, 1996). These include genes such as otsA and otsB that encode enzymes for trehalose biosynthesis (Kaasen et al., 1992) and several members of yet unknown function such as OsmB and OsmE that encode outer membrane lipoproteins (Jung et al., 1990), OsmY that encodes a periplasmic protein (Yim & Villarejo, 1992) and OsmC that encodes a putative inner membrane protein (Völker et al., 1998). These four genes are induced upon the entry of the cell into stationary phase in a  $\sigma^{s}$ -dependent manner and also by elevated osmolarity. Recently, genes homologous to OsmC have been identified in a variety of bacterial species (Fig. 1).

The OsmC sequence family, which is found only in prokaryotes, is composed of three subfamilies in the Pfam database (Bateman *et al.*, 2000): subfamily I, also called the OsmC subfamily (seven members), subfamily II, called the Ohr subfamily (organic hydrogen peroxide resistance protein, 20 members), and subfamily III of unknown function (39 members). Of these, crystal structures of members of subfamilies II and III have been reported. The crystal structure

Printed in Denmark - all rights reserved

© 2004 International Union of Crystallography

#### Table 1

Statistics of the peak-wavelength SAD data set.

Values in parentheses refer to the highest resolution shell, which is 2.44–2.40 Å.

Wavelength (Å)	0.9793
Resolution (Å)	46.6–2.4
Redundancy	3.8 (3.4)
Unique reflections	74939 (3736)
Completeness (%)	98.2 (97.9)
$I/\sigma(\hat{I})$	25.6 (2.2)
$R_{ m sym}$ † (%)	6.0 (57.7)

†  $R_{\text{sym}} = \sum_{hkl} \sum_{i} |I_{hkl,i} - \langle I \rangle_{hkl}| / \sum |\langle I \rangle_{hkl}|.$ 

and biochemical assay of Ohr revealed that it has peroxiredoxin (Prx) activity and utilizes highly reactive cysteine thiol groups to elicit hydroperoxide reduction (Lesniak *et al.*, 2002). Two protein crystal structures from subfamily III, MPN625 (gi|1673883) from *Mycoplasma pneumoniae* (Choi *et al.*, 2003) and yhfA from *E. coli* (gi|15803870; PDB code 1ml8), have been determined. However, the three-dimensional structures and molecular (biochemical and biophysical) functions of members of salt-shock-induced subfamily I have not been elucidated until now.

In order to obtain the structure and possible inference of the molecular function of a member of subfamily I, we have determined the crystal structure of OsmC from *E. coli* (EcOsmC) and compared it with those of subfamilies II and III.

While this paper was in preparation, another manuscript appeared presenting the same structure (Lesniak *et al.*, 2003). A comparison of the two structures is presented.

(1) Sul	bfamily I (	OsmC Subfam:	ily)							
ScOamC	1		MATTRSAHT	VWEGNLLEGN	GVVTFDSSGI	GEQPVSWP	SRAEQANGKT	SPEELIAAAH	SSCFSMA	LSH
DrOsmC	14 SRRAAPR	SEHTGRMADI	ARKASAHWEG	DLKSGN	GTITTESGVL	SQAQYSF K	TRFENGKG-T	NPEELLASAH	AGCETMO	LSA
PaOsmC	1	MSIHSSG	<b>GVDMKKTASA</b>	VWQGGLKDGK	GTLSTESGAL	KDNPYGF N	TRFEGAPG-T	NPEELIGAAH	AGCESMA	LSM
RIOsmC	1	MT	IREASAKWOG	TLKEGS	GRMKLGSG-V	FEGAYSPP	SRFENGPG-T	NPEELIAAAH	AGCESMA	LSA
EcOsmC	1	MTI	HKKGOAHWEG	DIKRGK	GTVSTESGVL	NOOPYGE N	TRFEGEKG-T	NPEELIGAAH	AACESMA	LSL
(2) Sul	family II	(Ohr Subfam	ilv)							
XcOhr	1		MASPEKVLY	TAHATATGER	RERAVSSOKA	LDAKLSTPRE	LGGAGGDG-T	NPROLFAAGY	AACFIGA	MKA
PaOhr	ī		MOTT-KALY	TATATATGOR	DERAVSSDOV	LOVELSTPRE	LGGOGGAA-T	NPROLEAAGY	SACETGA	LEF
Drohr	1		MANUY	TAFATATOR	AGTTRSSDDR	LNLDLSVPAR	MCGDCGPC-T	NPEOLEAAGY	AACFOGA	LOV
CcOhr	ĩ		MTT-LY	TTPATWYCER	EGHARTEDGI.	LOVOLSMPKS	LOCK-ETG-T	NPROLEAAGY	AACEOSAI	MGH
Matha	1 M	T.PNTPTTT.C	TLINMALTY	TVAOTET. OR	POCUNTI DC-	POTVLOPPVD	DL-SUOTE-N	NEPOLEACAV	ACCEGON	UTU
MoOhr	1 1	DENTEIKIDO	MAUTVE	TTAUACA-OP	ROUVOTUDG-	PUCIAPDYD	-CATHODY-N	NEPOLENCAV	ACCREON	VDV
(2) Cul	familie TTT		MAVIIN	TTARASA- OK	POAAA TAPO-	FIVOLAFFER	-GAINQUA-N	HE BYDE ADAT	NOCEDEN	VAV
(3) SU	bramily III			MOREVOTION	THER COMPANY	CDOPOTET DA	DEPUTATO P	ODI NALL COL	A A CIDY A C	
MPN025	1			MDKKIDIIAV	LNEDSSMIAI	SDUFUTTEDA	RPARIANG-P	GPLAALLSGL	AACELAL	ANL
MgenHi	1			MDKKYDITAV	LNDDSSINAV	SUNFQITLDA	RPKEKSKG-1	NPLSAFLAGL	AACELAT	ANA
MgalHY	1		MYIM	ARKEYAFTAK	LNPDRTVTGK	TPKHELLMDS	SATKG-P	SPLELMMNGL	MGCELSV	ISY
EcYhfA	1			MQARVKWV	EGLTFLGESA	SGHQILMD	GNSGDKAP	SPMEMVLMAA	GGCSAID	vvs
VcYhfA	1			MQAQVKWV	EDFRFIGLSN	SGHSIVMD	GNGGASAP	SPMEMVLMAA	GGCSSVD	VVD
BbYhfA	1		MO	MECTIDWGGP	AGMLFTASTG	SGHVAVMDGA	VDGGGHDLAP	RPMEMLLAGT	GGCTAYD	VVL
TvYhfA	1			MOVSFVYKDG	EGFDSDDGKE	TVKITYSO	GGDPNRH	SPTELLLAI	GGCTSDD	VLS
				M	otif I	Mo	tif I'	Motif II N	Aotif III	
ScOsmC	GLAGAGTPPT	-K-LTTSADV	TFQPGEG	IKGIHLTVEG	TVPGLDNDAF	VAAAE-DAKK	NCPVSQALTG	-TTITLSAKL	λ	141
ScOsmC DrOsmC	GLAGAGTPPT LLAEHGHEIK	-K-LTTSADV -A-LDTDATC	TFQPGEG EMVKDGPGFK	IKGIHLTVEG INHMHLRVRA	TVPGLDNDAF	VAAAE-DAKK EAHVK-DAAE	NCPVSQALTG KCPLSRIMQG	-TTITLSAKL -NVEVTHEAI	A LEG	141
ScOsmC DrOsmC PaOsmC	GLAGAGTPPT LLAEHGHEIK MLGEAGLTAE	-K-LTTSADV -A-LDTDATC -R-IETRAEV	TFQPGEG EMVKDGPGFK TLDKQSDGFA	IKGIHLTVEG INHMHLRVRA ITAVHLVLRA	TVPGLDNDAF QLTGSDQADF RVPGADAQTF	VAAAE-DAKK EAHVK-DAAE EQIAN-KAKA	NCPVSQALTG KCPLSRIMQG GCPVSKVLNA	-TTITLSAKL -NVEVTHEAI -KISLDASLD	A LEG G	141 172 151
ScOsmC DrOsmC PaOsmC RlOsmC	GLAGAGTPPT LLAEHGHEIK MLGEAGLTAE ILGTAHHIPA	-K-LTTSADV -A-LDTDATC -R-IETRAEV -S-ISTVAKV	TFQPGEG EMVKDGPGFK TLDKQSDGFA DLGATVAGPT	IKGIHLTVEG INHMHLRVRA ITAVHLVLRA ITRIELETRA	TVPGLDNDAF QLTGSDQADF RVPGADAQTF EIPGLAPDEF	VAAAE-DAKK EAHVK-DAAE EQIAN-KAKA QNLAE-RAKT	NCPVSQALTG KCPLSRIMQG GCPVSKVLNA TCLVSRALAG	-TTITLSAKL -NVEVTHEAI -KISLDASLD -VASITLKAE	A LEG G LIATTAQ	141 172 151 147
ScOsmC DrOsmC PaOsmC RlOsmC EcOsmC	GLAGAGTPPT LLAEHGHEIK MLGEAGLTAE ILGTAHHIPA MLGEAGFTPT	-K-LTTSADV -A-LDTDATC -R-IETRAEV -S-ISTVAKV -S-IDTTADV	TFQPGEG EMVKDGPGFK TLDKQSDGFA DLGATVAGPT SLDKVDAGFA	IKGIHLTVEG INHMHLRVRA ITAVHLVLRA ITRIELETRA ITKIALKSEV	TVPGLDNDAF QLTGSDQADF RVPGADAQTF EIPGLAPDEF AVPGIDASTF	VAAAE-D <mark>AKK</mark> EAHVK-DAAE EQIAN-KAKA QNLAE-RAKT DGIIQ-KAKA	NCPVSQALTG KCPLSRIMQG GCPVSKVLNA TCLVSRALAG GCPVSQVLKA	-TTITLSAKL -NVEVTHEAI -KISLDASLD -VASITLKAE -EITLDYQLK	A LEG G LIATTAQ S	141 172 151 147 143
ScOsmC DrOsmC PaOsmC RlOsmC EcOsmC	GLAGAGTPPT LLAEHGHEIK MLGEAGLTAE ILGTAHHIPA MLGEAGFTPT VAAODKLKLP	-K-LTTSADV -A-LDTDATC -R-IETRAEV -S-ISTVAKV -S-IDTTADV GE-VSIDSSV	TFQPGEG EMVKDGPGFK TLDKQSDGFA DLGATVAGPT SLDKVDAGFA GIGOIPGGF-	IKGIHLTVEG INHMHLRVRA ITAVHLVLRA ITRIELETRA ITKIALKSEV	TVPGLDNDAF QLTGSDQADF RVPGADAQTF EIPGLAPDEF AVPGIDASTF	VAAAE-DAKK EAHVK-DAAE EQIAN-KAKA QNLAE-RAKT DGIIQ-KAKA OTLVD-KAHO	NCPVSQALTG KCPLSRIMQG GCPVSKVLNA TCLVSRALAG GCPVSQVLKA	-TTITLSAKL -NVEVTHEAI -KISLDASLD -VASITLKAE -EITLDYQLK -NIDVTLTLA	A G LIATTAQ S	141 172 151 147 143
ScOsmC DrOsmC PaOsmC RlOsmC EcOsmC XcOhr PaOhr	GLAGAGTPPT LLAEHGHEIK MLGEAGLTAE ILGTAHHIPA MLGEAGFTPT VAAQDKLKLP VAGOKKOTLP	-K-LTTSADV -A-LDTDATC -R-IETRAEV -S-ISTVAKV -S-IDTTADV GE-VSIDSSV AD-ASITGKV	TFQPGEG EMVKDGPGFK TLDKQSDGFA DLGATVAGPT SLDKVDAGFA GIGQIPGGF- GIGOIPGGF-	IKGIHLTVEG INHMHLRVRA ITAVHLVLRA ITRIELETRA ITKIALKSEV GIVVELRI GLEVELHI	TVPGLDNDAF QLTGSDQADF RVPGADAQTF EIPGLAPDEF AVPGIDASTF AVPGMDKAEL NLPGLEFEAA	VAAAE-DAKK EAHVK-DAAE EQIAN-KAKA QNLAE-RAKT DGIIQ-KAKA QTLVD-KAHQ EALVA-AAHQ	NCPVSQALTG KCPLSRIMQG GCPVSKVLNA TCLVSRALAG GCPVSQVLKA VCPYSNATRG	-TTITLSAKL -NVEVTHEAI -KISLDASLD -VASITLKAE -EITLDYQLK -NIDVTLTLA -NIDVTLTLA	A G LIATTAQ S	141 172 151 143 143
ScOsmC DrOsmC PaOsmC RlOsmC EcOsmC XcOhr PaOhr DrOhr	GLAGAGTPPT LLAEHGHEIK MLGEAGLTAE ILGTAHHIPA MLGEAGFTPT VAAQDKLKLP VAGQRKQTLP VAGROKLDVP	-K-LTTSADV -A-LDTDATC -R-IETRAEV -S-ISTVAKV -S-IDTTADV GE-VSIDSSV AD-ASITGKV AD-ASITGKV	TFQPGEG EMVKDGPGFK TLDKQSDGFA DLGATVAGPT SLDKVDAGFA GIGQIPGGF- GIGQIPGGF- GLOKAGLAF-	IKGIHLTVEG INHMHLRVRA ITAVHLVLRA ITRIELETRA ITKIALKSEV GIVVELRI GLEVELHI ALDVELEG	TVPGLDNDAF QLTGSDQADF RVPGADAQTF EIPGLAPDEF AVPGIDASTF AVPGMDKAEL NLPGLEREAA	VAAAE-DAKK EAHVK-DAAE EQIAN-KAKA QNLAE-RAKT DGIIQ-KAKA QTLVD-KAHQ EALVA-AAHQ EGLMH-AAHQ	NCPVSQALTG KCPLSRIMQG GCPVSKVLNA TCLVSRALAG GCPVSQVLKA VCPVSNATRG VCPVSNATRG	-TTITLSAKL -NVEVTHEAI -KISLDASLD -VASITLKAE -EITLDYQLK -NIDVTLTLA -NIDVRLNVS -NVDVRLKVB	A G LIATTAQ S V	141 172 151 147 143 142 142
ScOsmC DrOsmC PaOsmC RlOsmC EcOsmC XcOhr PaOhr DrOhr CcOhr	GLAGAGTPPT LLAEHGHEIK MLGEAGLTAE ILGTAHHIPA MLGEAGFTPT VAAQDKLKLP VAGQRKQTLP VAGQRKQTLP VAGQRKIDVP	-K-LTTSADV -A-LDTDATC -R-IETRAEV -S-ISTVAKV -S-IDTTADV GE-VSIDSSV AD-ASITGKV AD-STITARV -G-STVTGOV	TFQPGEG EMVKDGPGFK TLDKQSDGFA DLGATVAGPT SLDKVDAGFA GIGQIPGGF- GIQQTPGGF- GIQQTAGLAF- GLATAFVGF-	IKGIHLTVEG INHMHLRVRA ITAVHLVLRA ITRIELETRA ITKIALKSEV GIVVELRI GLEVELHI ALDVELEG DLEVALEY	TVPGLDNDAF QLTGSDQADF RVPGADAQTF EIFGLAPDEF AVPGIDASTF AVPGMDKAEL NLPGLEREAA HFPGLSREQA	VAAAE-DAKK EAHVK-DAAE EQIAN-KAKA QNLAE-RAKT DGIIQ-KAKA QTLVD-KAHQ EALVA-AAHQ EGLMH-AAHE EALVA-AAHQ	NCPVSQALTG KCPLSRIMQG GCPVSKVLNA TCLVSRALAG GCPVSQVLKA VCPYSNATRG VCPYSAATRG VCPYSAATRG	-TTITLSAKL -NVEVTHEAI -KISLDASLD -VASITLKAE -EITLDYQLK -NIDVTLTLA -NIDVRLNVS -NUDVRLKVR	A G LIATTAQ S V E	141 172 151 142 142 142 142
ScOsmC DrOsmC PaOsmC RlOsmC EcOsmC XcOhr PaOhr DrOhr CcOhr	GLAGAGTPPT LLAEHGHEIK MLGEAGLTAE ILGTAHHIPA MLGEAGFTPT VAAQDKLKLP VAGQRKQTLP VARQCKLDP VARTQKIALA VMOQHOSES	-K-LTTSADV -A-LDTDATC -R-IETRAEV -S-ISTVAKV -S-IDTTADV GE-VSIDSSV AD-ASITGKV AD-STITARV -G-STVTGQV	TFQPGEG EMVKDGPGFK TLDKQSDGFA DLGATVAGPT SLDKVDAGFA GIGQIPGGF- GIGQIPGGF- GLQKAGLAF- SLATAEVGF- ELLQCMGLF-	IKGIHLTVEG INHMHLRVRA ITAVHLVLRA ITRIELETRA ITRIALKSEV GIVVELRI GLEVELHI ALDVELEG RLEVALEV	TVPGLDNDAF QLTGSDQADF RVPGADAQTF EIFGLAPDEF AVPGIDASTF AVPGMDKAEL NLPGLEREAA HFPGLSREQA ETQGLSQAD TTNENDOEVG	VAAAE-DAKK EAHVK-DAAE EQIAN-KAKA QNLAE-RAKT DGIIQ-KAKA QTLVD-KAHQ EALVA-AAHE EALVA-AAHE EALVA-AAHE	NCPVSQALTG KCPLSRIMQG GCPVSKVLNA TCLVSRALAG GCPVSQVLKA VCPYSNATRG VCPYSNATRG VCPYSNATRG	-TTITLSAKL -NVEVTHEAI -KISLDASLD -VASITLKAE -BITLDYQLK -NIDVTLTLA -NIDVRLNVS -NVDVRLKVR -NIDVAITTK -ENVLGTIN	λ G LLATTAQ S V E λλ	141 172 151 147 143 142 142 139 138
ScOsmC DrOsmC PaOsmC RlOsmC EcOsmC XcOhr PaOhr DrOhr CcOhr MgOhr MgOhr	GLAGAGTPPT LLAEHGHEIK MLGEAGLTAE ILGTAHHIPA MLGEAGFTPT VAAQDKLKLP VAGQRKQTLP VSRRQKIDVP VSRRQKIDVP VARTQKIALA VMQQHQFSFS	-K-LTTSADV -A-LDTDATC -R-IETRAEV -S-ISTVAKV -S-IDTTADV GE-VSIDSSV AD-ASITGKV AD-ASITGKV -G-STVTGQV -KKPVVSVKV -TADIVGVSV	TFQPGEG EMVKDGPGFK TLDKQSDGFA DLGATVAGPT SLDKVDAGFA GIGQIPGGF- GIGQIPGGF- GLQKAGLAF- SLATAEVGF- ELHQENGLF-	IKGIHLTVEG INHMHLRVRA ITAVHLVLRA ITRIELETRA ITKIALKSEV GIVVELRI GLEVELHI ALDVELEG RLEVALEV HIKAGVEL	TVPGLDNDAF QLTGSDQADP RVPGADAQTF EIPGLAPDEP AVPGIDASTF AVPGMDKAEL NLPGLEREAA HFPGLSREQA ETQGLSQADA TTNSNDQEVG	VAAAE-DAKK EAHVK-DAAE EQIAN-KAKA QNLAE-RAKT DGIIQ-KAKA QTLVD-KAHQ EALVA-AAHQ EGLMH-AAHB EALVA-AAHA KKLIQ-KAHQ OTUIT-AAHA	NCPUSQALTG KCPLSRIMQG GCPUSKVLNA TCLUSRALAG GCPUSQVLKA VCPUSNATRG VCPUSNATRG VCPUSNATRG MCPFSRLIRN	-TTITLSAKL -NVEVTHEAI -KISLDASLD -VASITLKAE -EITLDYQLK -NIDVTLTLA -NIDVRLNVS -NUDVRLKVR -NIDVAITTK -ENFLGLTIN -ENFLGLTIN	A G LIATTAQ S E AA GIKL	141 172 151 142 143 142 142 133 133 133
ScOsmC DrOsmC PaOsmC RlOsmC EcOsmC XcOhr PaOhr DrOhr CcOhr MgOhr MgOhr	GLAGAGTPPT LLAEHGHEIK MLGEAGLTAE ILGTAHHIPA MLGEAGFTPT VAAQDKLKLP VSRRQKIDVP VSRRQKIDVP VARTQKIALA VMQQHQFSFS VLQQHQLQLA	-K-LTTSADV -A-LDTDATC -R-IETRAEV -S-ISTVAKV -S-IDTTADV GE-VSIDSSV AD-ASITGKV AD-STITARV -G-STVTGQV -KKPVVSVKV -TQPIVGVSV	TFQPGEG EMVKDGPGFK TLDRQSDGFA DLGATVAGPT SLDKVDAGFA GIGQIPGGF- GIQQIPGGF- GLQRAGLAF- SLATAEVGF- ELHQENGLF- ELHQDGLF-	IKGIHLTVEG INHMHLRVRA ITAVHLVLRA ITRIELETRA ITKIALKSEV - GIVVELRI - GLEVELHI - ALDVELEG - RLEVALEV - HIKAGVEL - HIKAGVEL	TVPGLDNDAF QLTGSDQADF RVPGADAGTF EIPGLAPDEF AVPGIDASTF AVPGIDASTF HPGLSRQA ETQGLSQADA TTNSNDQEVG AITGVDQTTA	VAAAE-DAKK EAHVK-DAAE EQIAN-KAKA QNIAE-RAKT DGIIQ-KAKA QTLVD-KAHQ EALVA-AAHQ EGIAM-AAHB EALVA-AAHA KKLIQ-KAHB QTVIT-AAHA	NCPVSQALTG KCPLSRIMQG GCPVSKVLNA TCLVSRALAG GCPVSQVLKA VCPVSNATRG VCPVSNATRG VCPVSNATRG MCPPSRLIK-	- TTITLSAKL - NVEVTHEAI - KISLDASLD - VASITLKAE - EITLDYQLK - NIDVTLTLA - NIDVTLTLA - NVDVRLKVR - NVDVRLKVR - NIDVAITTK - ENFLGLTLN PENFLGLTLN	λ G LIATTAQ S Ε Aλ GIKL GAKL	141 172 151 142 142 142 142 132 132 132 140
ScOsmC DrOsmC PaOsmC RIOsmC EcOsmC XcOhr PaOhr DrOhr CcOhr MgOhr MpOhr MPN625	GLAGAGTPPT LLAEHGHEIK MLGEAGLTAE ILGTAHHPA MLGEAGFTPT VAAQDKLKLP VAAQQKKQTLP VSRQKIDVP VARTQKIALA VMQQHQFSFS VLQQHQLQLA MAPAKMITIN	-K-LTTSADV -A-LDTDATC -R-IETRAEV -S-ISTVAKV -S-IDTTADV GE-VSIDSSV AD-ASITGKV AD-STITARV -G-STVTGQV -KKPVVSVKV -TQPIVGVSV KLLMNVTGSR	TFQPGEG EMVKDGPGFK TLDKQSDGFA DLGATVAGPT SLDKVDAGFA GIGQIPGGF- GIQXAGLAF- GLQKAGLAF- ELHQENGLF- ELHQQDGLF- STNFTDGYFG	IKGIHLTVEG INHMHLRVRA ITAVHLVLRA ITRIELETRA ITKIALKSEV GIVVELRI GLEVELHI ALDVELEG RLEVALEV HIKAGVEL LREINLHWEI	TVPGLDNDAF QLTGSDQADF RVPGADAQTF EIPGLAPDEP AVPGIDASTF AVPGMDKAEL NLPGLEREAA HFPGLSREQA ATTNSNDQEVG AITGVDQTTA HSPN-SETEI	VAAAE-DAKK EAHVK-DAAE EQIAN-KAKA QNLAE-RAKT DGIIQ-KAKA QTLVD-KAHQ EALVA-AAHQ EGLMH-AAHE EALVA-AAHA KKLIQ-KAHE QTVIT-AAHA KEFID-FV <u>S</u> K	NCPVŠQALTG KCPLŠRIMQG GCPVŠVLNA TCLVŠRALAG GCPVŠQVLKA VCPYŠNATRG VCPVŠNATRG VCPVŠATRN MCPPŠRLIRN MCPPŠRLIK- RCPAHNTLQG	-TTITLSAKL -NVEVTHEAI -KISLDASLD -VASITLKAE -EITLDYQLK -NIDVTLTLA -NIDVRLNVS -NUDVRLXVR -NIDVAITTK -ENFLGLTLN PENFLGLTLN VSQLKINVNV	A LEG LIATTAQ S V E AA GIKL GAKL TLVH	141 172 151 142 142 142 132 132 132 140
SCOSMC DrOSMC PAOSMC RIOSMC ECOSMC XCOhr PAOhr DrOhr CcOhr MgOhr MpOhr MPN625 MgenHY	GLAGAGTPPT LLAEHGHEIK MLGEAGLTAE ILGTAHHIPA MLGEAGETPT VAQORKLUP VARQKKIDVP VARTQKIALA VMQOHQFSFS VLQOHQLQLA MAPAKMITIN MAAAKMITIN	-K-LTTSADV -A-LDTDATC R-IETRAEV -S-ISTVAKV -S-IDTTADV GE-VSIDSSV AD-ASITGKV -G-STVTGQV -KEPVVSVKV -TQPIVGVSV KLLMNVTGSR KALINIKGYR	TFQPGEG EMVKDGPGFA TLDKQSDGFA DLGATVAGPT SLDKVDAGFA GIGQIPGGF- GIGQIPGGF- GLQKAGLAF- SLATAEVGF- ELHQQBGLF- ELHQQDGLF- STNPTDGYFG STNPTDGYFG	IKGIHLTVEG INHMRLRVRA ITAVHLVLRA ITAVHLVLRA ITKIELETRA ITKIALKSEV - GIVVELRI - GLEVELHI - ALDVELEG - RLEVALEV - HIKAGVEL LREINLHWEI LREINLHWEI	TVPGLDNDAF QLTGSDQADF RVPGADAQTP EIPGLAPDEF AVPGIDASTF AVPGMDKAEL NLPGLEREAA ETQGLSQADA TTNSNDQEVG AITGVDQTTA HSPN-SETEI HSPN-SETEI	VAAAE-DAKK EAHVK-DAAE EQIAN-KAKA QNLAE-RAKT DGIIQ-KAKA QTLVD-KAHQ EALVA-AAHQ EGLMH-AAHE EALVA-AAHA KKLIQ-KAHE QTVIT-AAHA KEFID-FVSK KEFID-FVSK	NCPVSQALTG KCPLSRIMGG GCPVSKVINA TCLVSRALAG GCPVSQVLKA VCPYSNATRG VCPVSNATRG VCPVSNATRG MCPPSRLIKN MCPPSRLIKN RCPAHNTLQG RCPAHNTLQG	-TTITLSAKL -NVEVTHERI -KISLDASLD -VASITLKAE -EITLDYQLK -NIDVTLTLA -NIDVLTLLA -NUDVRLKVR -NUDVAITTK -ENFLGLTLN PENFLGLTLN VSQLKINVNV TSNFKINISV	A G G S E GIKL GAKL TLVH TLVH	141 172 151 147 143 142 142 138 159 140 141
SCOSMC DrOSMC PAOSMC RIOSMC ECOSMC XCOhr PAOhr DrOhr CCOhr MgOhr MpOhr MPN625 MgenHY MgalHY	GLAGAGTPPT LLAEHGHEIK MLGEAGLTAE ILGTAHHIPA MLGEAGETPT VAAQDKLKLP VASRQKIDP VARTQKIALA VMQ0HQFSFS VLQ0HQLFGL MAPAKMITIN YGPKLFGLQL	-K-LTTSADV -A-LDTDATC R-IETRAEV -S-ISTVAKV -S-IDTTADV GE-VSIDSSV AD-ASITGKV -G-STVTGQV -KEPVVSVKV -TQPIVGVSV KLLMNVTGSR KALINIKGYR EDLEMKVQAW	TFQPGEG EMVKDGPGFK TLDKQSDGFA DLGATVAGFT SLDKVDAGFA GIGQIFGGF- GIGQIFGGF- GLQKAGLAF- SLATAEVGF- ELHQENGLF- ELHQENGLF- STNFTDGYFG STNFTDGYFG RDFEPQDEYY	IKGIHLTVEG INHMRLRVRA ITAVHLVLRA ITRIELETRA ITKIALKSEV GIVVELRI GLEVELHI ALDVELEG RLEVALEV HIKAGVEL LREINLHWEI LREINLHWEI GLRKIEIEWF	TVPGLDNDAF QLTGSDQADF RVPGADAQTP EIPGLAPDEF AVPGIDASTF AVPGMDKAEL NLPGLEREDA ETQGLSQADA TTNSNDQEVG AITGVDQTTA HSPN-SETEI HSPN-SETEI VKSNKSLEEV	VAAAE-DAKK EAHVK-DAAE EQIAN-KAKA QNLAE-RAKT DGIIQ-KAKA QTLVD-KAHQ EALVA-AAHQ KGLMI-AAHE EALVA-AAHA KKLIQ-KAHE QTVIT-AAHA KEFID-FVSK KEFID-FVSK KEFIK-FVAHK	NCPVSQALTG KCPLJRIMGG GCPVSKVINA TCLVSRALAG GCPVSQVLKA VCPYSNATRG VCPVSNATRG VCPVSNATRG MCPPSRLIKN MCPPSRLIKN RCPAHNTLQG RCPAHNTLQG VCPVFNSLSG	-TTITLSAKL -NVEVTHERI -KISLDASLD -VASITLKAE -EITLDYQLK -NIDVTLTLA -NIDVTLTLA -NIDVRLNVS -NVDVRLKVR -NIDVAITTK -ENFLGLTLN VSQLKINVNV TSNFKINISV RIVIDEKFTL	A G LIATTAQ S V GIKL GAKL TLVH K	141 172 153 147 143 142 133 133 133 155 140 141 141 140
SCOSMC DrOSMC PAOSMC RIOSMC ECOSMC XCOhr PAOhr DrOhr CcOhr MgOhr MpOhr MPN625 MgenHY MgalHY EcYhfA	GLAGAGTPPT LLAEHGHEIK MLGEAGLTAE ILGTAHHIPA MLGEAGFTPT VAAQDKLKLP VSRQKIDVP VARTQKIALA VMQQHQFSPS VLQCHQLQLA MAPARMITIN MAAAKMITIN MAAAKMITIN ILOKGRODVV	-K-LTTSADV -A-LDTDATC R-IETRAEV -S-IETVAKV -S-IETVAKV -S-IETVAKV -D-STITARV -G-STVTGQV -KEPVVSVKV -TQPIVGVSVK -TQPIVGVSVK KALINKGYR KALINKGYR EDLEMKVQAW -D-CEVKLTS	TFQPGEG EMVKDGPGFK TLDRQSDGFA DLGATVAGFT SLDKVDAGFA GIGQIPGGF- GLQYKAGLAF- SLATAEVGF- ELHQQDGLF- ELHQQDGLF- STNFTDGYFG LTNPSDGYFG LTNPSDGYFG EDFEPQDEYY ERREADTRL-	IKGIHLTVEG INNHMLRVRA ITAVHLVLRA ITRIELETRA ITKIALKSEV GLVVELRI GLVVELRI ALDVELEG RLEVALEV LRELNLHWEI LRELNLHWEI LRELNLHWEI FTHINLHFIV	TVPGLDNDAF QLTGSDQADF EIPGLAPDEF AVPGNDAGTF EIPGLAPDEF AVPGMDKAEL HPPGLSREQA ETQGLSQADA HSPN-SETEI HSPN-SETEI HSPN-SETEI HSPN-SETEI TGRDLKDAAV	VAAAE-DAKK EAHWK-DAAE EQIAN-KAKA QNLAE-BAKT DGIIQ-KAKA QTLVD-KAHQ EGLMH-AAHQ EGLMH-AAHQ EGLMH-AAHQ EGLMH-AAHQ KKLIQ-KAHE QTVIT-AAHA KEFID-PY9K KEFIL-FY9K KEFIL-YAHK ARAVDLSAEK	NCPVSQALTG KCPLERINGG GCPVSKVLNA TCLVSRALAG GCPVSQVLKA VCPVSQALKA VCPVSQATRG VCPVSQATRG VCPVSQATRG VCPVSQATRG RCPAINTLQG RCPAINTLQG RCPAINTLQG YCPVFNSLSG VCPVFNSLSG	-TTITLSAKL -NVEVTHEAI -KISLDASLD -VASITLKAE -EITLDYQLK -NIDVTLTLA -NIDVTLKNS -NVDVRLKVR -NIDVRLKVR -NIDVAITTK -ENFLGLTLN PENFLGLTLN VSQLKINVNV TSNFKINISV A-VNITHSYE	λ LEG G LIATTAQ S E Aλ GIKL TLVH K VVAλ	141 172 153 147 143 142 142 133 138 140 141 140 141 140
ScOsmC DrOsmC PaOsmC RlOsmC EcOsmC XcOhr PaOhr DrOhr DrOhr MgOhr MgOhr MpOhr MPN625 MgenHY MgalHY EcYhfA VcYhfA	CLACAGTPPT LLAEHGHEIK MLGRAGLTAE ILGTAHHIPA MLGRAGFTPT VAAQDKLKLP VAGQRKQTLP VSRQKIDV VARTQKIALA VMQQHQFSFS VLQQHQLGLA MAPAKMITIN YGPKLFGLQL ILQKRQDVV	-K-LTTSADV -A-LDTDATC R-IETRAEV -S-ISTVAKV -S-IDTTADV GE-VSIDSSV AD-STITARV -G-STVTGQV -KKPVVSVKV -TQPIVGVSVK -TQPIVGVSVK KLLMNVTGSR KALLMIKGYR EDLEMKVQAW -D-CEVKLTS	TFQPGEG EMVKDGPGFA TLDRQSDGFA DLGATVAGPT SLDKVDAGFA GIQQIPGGF- GLQKAGLAF- SLATAEVGF- ELHQQDGLF- ELHQDGLF- STNPTDGYFG RDFEPQDEYY ERREADTRL- ERRDTAFKL-	IKGIHLTVEG INHMRLRVRA ITAVHLVLRA ITRIELETRA ITRIELETRA ITRIALKSEV GLEVELHI ALDVELEG RLEVALEV HIKAGVEL LREINLHWEI GLRKIEIEWP FTHINLHFIV	TVPGLDNDAF QLTGSDQADF RVPGDDAOTF EIPGLAPDEF AVPGIDASTF AVPGIDASTF HPGLSRQA ETQGLSQAD ETQGLSQAD HSPN-SETEI HSPN-SETEI HSPN-EEEEI VKSNKSLEEV TGRDLKDAAV	VAAAE-DAKK EAHVK-DAAE EQIAN-KAAA QULAE-RAKT DGIIQ-KAKA QTLUD-KAHQ EALVA-AAHQ EGLMH-AAHE EALVA-AAHA KKLIQ-KAHE QTVIT-AAHA KEFID-FVSK KEFID-FVSK KEFID-FVSK KEFID-FVSK KEFID-FVSK KEFID-FVSK	NCPVSQALTG KCPLGRIMOG GCPVGKVLNA TCLVSRALAG GCPVGQVLKA VCPVSNATRG VCPVSNATRG VCPVSNATRG VCPVSNATRG MCPFGRLKA RCPAHNTLQG RCPAHNTLQG VCPVNLSG VCSVALML6K	-TTITLSAKL -NVEVTHERI -KISLDASLD -VASITLKAE -EITLDYQLK -NIDVTLTLA -NIDVRLNVS -NIDVRLNVS -NIDVAITTK -ENFLGLTLN PENFLGLTLN VSQLKINVNV TSNFKINISV RIVIDEKFTL A-VNITHSWE	A LEG G LIATTAG S V E GIKL GIKL GAKL TLVH K VVAA IRTE	141 172 153 147 143 142 142 133 133 140 142 144 144 144 144 144 144
ScOsmC DrOsmC PaOsmC RlosmC EcosmC XcOhr DrOhr DrOhr CcOhr MgOhr MpOhr MpN625 MgalHY EcYhfA BoyhfA	GLAGAGTPPT LLAEHGHEIK MLGEAGLTAE ILGTAHHIPA MLGEAGFTPT VAAQRKQTLP VSRQKLTLP VSRQKIDLP VARTQKIALA VMQQHQFSFS VLQUHQLQLA MAPAKMITIN YGPKLFGLQL ILQKGRQDVV GMKKAGQKIH	-K-LTTSADV -A-LDTDATC R-IETRAEV -S-IETVAKV -S-IETVAKV -S-IETVAKV -S-IETVAKV AD-ASITGKV AD-STITARV -G-STVTGQV -KKPVVSVKV -TQPIVGVSV KLIMNVTGSR EDLEMKVQAW -D-CEVKLTS -G-CTAQLSA -G-CTAQLSA	TFQPGEG EMVKDGPGFK TLDRQSDGFA DLGATVAGFT SLDKVDAGFA GIGQIPGGF- GLQKAGLAF- ELHQENGLF- ELHQQDGLF- STNFTDGYFG RDPEPQDEYY ERREADTRL- ERREADTRL- ERREADTRL-	IKGIHLTVEG INHMHLRVRA ITAVHLVLRA ITTIELETRA ITKIALKSEV GLVVELRI ALDVELEG RLEVALEV HIKAGVEL LRELNLHWEI GLRKIEIEWP FTHINLHFIV FTQVNIHFVV FTCHIPAFTV	TVPGLDNDAF QLTGSDQADF RVPGADAQTF EIPGLAPDEF AVPGIDASTF AVPGMDKAEL HPPGLSREQA ETQGLSREQA AITGVDQTTA HSPN-SETEI HSPN-SETEI HSPN-SETEI UKSNKSLEEV TGRDLKDAAV SGEDLDQSIV	VAAAE-DAKK EAHVK-DAAE EQIAN-KAKA QNLAE-BAKT DGIIQ-KAKA QTLUD-KAHQ EGLMK-AAHQ EGLMK-AAHQ EGLMK-AAHQ QTVIT-AAHA QTVIT-AAHA QTVIT-AAHA KEFID-FYGK KEFID-FYGK KEFID-FYGK ARVTADSLEA ARVTADSLEA	NCPVSQALTG KCPLERINGG GCPVSKVLNA TCLVSRALAG GCPVSQVLKA VCPVSQALKA VCPVSQATRG VCPVSQATRG VCPVSQATRG MCPPSRLIRN MCPPSRLIRN MCPPSRLIRN MCPPSRLIRN VCPVSNSLSG VCPVFNSLSG VCSVSALMLEK YCSVSALMLEK YCSVSALMLEK	-TTITLSAKL -NVEVTHEAI -KISLDASLD -VASITLKAE -EITLDYQLK -NIDVRLNVS -NDDVRLNVS -NDVRLNVS -NDVRLNVS -NDVRLNV -ENFLGLTLN PENFLGLTLN YSQLKINVNV RIVLDEKFTL A-VNITHSYE G-VENTHSWE -AELSFSVD	A LEG G LIATTAQ S P P AA GAKL TLVH TLVH K VVAA IRTE IVDTOAA	141 172 151 147 143 142 142 143 142 143 142 144 144 144 134 145
SCOSMC DrOSMC PaOSMC RIOSMC ECOSMC XCOhr PaOhr DrOhr CCOhr MgOhr MgOhr MgOhr MgOhr MgalHY EcYhfA BbYhfA TvYhfA	GLAGAGTPPT LLABHGHEIK MLGEAGLTAE ILGTAHHIPA MLGEAGFTPT VAAQQKLKLPP VSRRQKIDVP VARQKQYLP VAQQRQFSFS VLQQHQLQLA MAPAKMITIN MAAAKMITLN YGPKLFGLQL ILQKRQDVV GMKKAGQKIH ILKKRRAVT	-K-LTTSADV -A-LDTDATC R-IETRAEV -S-IETVAKV -S-IETVAKV -S-IDTTADV dD-STITARV -G-STVTQQV -KKPVVSVKV -TQPIVGVSVK -TQPIVGVSVK KLLMNVTGSR KALLNIKCYR EDLEMKVQAW -D-CEVKLTS -G-CTAQLSA -G-CSVKLQA -S-YECVVEG	TFQPGEG EMVKDGPGFA TLDKQSDGFA DLGATVAGPT SLDKVDAGFA GIGQIPGGF- GLQKAGLAF- SLATAEVGF- ELHQQGLF- ELHQQDGLF- STNPTDGYFG RDPEPQDEYY ERREADTRL- ERRDTAPKL- ERRDTAPKL- ERRDADPKV-	IKGIHLTVEG INHMHLRVRA ITAVHLVLRA ITRIELETRA ITRIELETRA ITKIALKSEV GIEVVELKI GIEVVELKI ALDVELEG RIEVALEV HIKAGVEL LREINLHWEI LREINLHWEI GLRKIEIEWF FTHINLHFIV FTQVNIHFVV FTRIHFAFTV KKFANVSYII	TVPGLDNDAF QLTGSDQADF RVPGADAQTF EIPGLAPDEF AVPGIDASTF AVPGIDASTF NLPGLEREAA HEPGLSRQA ETQGLSQADA HSPN-SETEI HSPN-SETEI HSPN-SETEI VKSNKSLEEV VKSNKSLEEV SGEDLDQBIV TGSRLKDAAV SGEDLDQBIV	VAAAE-DAKK EAHWK-DAAE EQIAN-KAKA QNLAE-RAKT DGIIQ-KAKA GTLUD-KAHQ EALVA-AAHQ EGLMH-AAHE EGLMH-AAHE QTVIT-AAHA QTVIT-AAHA KEFID-FVSK KEFID-FVSK KEFID-FVSK KEFID-FVSK KEFIK-YAHK ARAVDLSAEK ARVTADSLEK ERAVQLSHEK	NCPVSQALTG KCPLERINGG GCPV3EVUNA TCLV3RALAG GCPV3QVLKA VCPV3RATRG VCPV3RATRG VCPV3RATRG VCPV3RATRG VCPV3RATRG MCPP3RLIN- MCPP3RLIN- MCPP3RLIN- RCPAHNTLQG RCPAHNTLQG YCVVFNSLSG YCCVLMLGK YCSVALMLEK YCSVALMLEK YCSVALMLEK YCSVALMLEK	-TTITLSAKL -NVEVTHEAI -KISLDASLD -VASITLKAE -EITLDYQLK -NIDVTLTLA -NIDVRLNVS -NUDVRLKVR -NIDVAITK -ENFLGLTLN PENFLGLTLN VSQLKINVNV TSNFKINISV RIVIDEKFTL A-VNITHSYE G-VENTHSWE T-AELSFSVD	A LEG G LIATTAG S F GIKL GIKL GAKL TLVH TLVH K VVAA IRTE IVDTQAA LNGRPID	141 172 151 147 143 142 142 139 138 155 140 141 141 140 134 138
SCOSMC DrOSMC PAOSMC RIOSMC ECOSMC XCOhr PAOhr DrOhr MgOhr MgOhr MPN625 MgenHY MgalHY ECYhfA BbYhfA TvYhfA	GLAGAGTPT LLAEHGHEIK MLGEAGLTAE ILGTAHHIPA MLGEAGFTPT VAAQDKLKLP VSRQKILVP VSRQKILVP VSRQKILALA VMQQHQFSFS VLQQHQLQLA MAPAKMITIN MAAAKMITIN YGPKLFGLQL ILQKGRQDVV GMKKAGQXIH ILKRGRHAVT ILKRKMQDVK	-K-LTTSADV -A-LDTDATC -R-IETRAEV -S-ISTVAKV -S-IDTTADV dD-ASITCKV AD-STITARV -G-STVTQV -KEPVVSVKV -TQPIVGVSV KLLMNVTGSR EDLEMKVQAW -D-CEVKLTS -G-CTXQLSA -S-YRCVVEG	TFQPGEG ENVKDGPGFK TLDRQSDGFA DLGATVAGFT SLDKVDAGFA GIGQIPGGF- GLQKAGLAF- ELHQENGLF- ELHQENGLF- ELHQENGLF- STNPTDGYFG RDFEPQDEYY ERREADTRL- ERREADTRL- ERREADTRL- ERREADTRL- ERREADTRL-	IKGIHLTVEG INHMHLRVRA ITAVHLVLRA ITTIELETRA ITKIALKSEV GEVVELRI ALDVELEG RLEVALEV HIKAGVEL LREINLHWEI GLRKIEIEWF FTHINLHFIV FTQINIHVU LKFANVSYII	TVPGLDNDAF QLTSSDQADF RVPGADAQTP EIPGLAPDEP AVPGIDASTP AVPGMDKAEL NLPGLERRAA HFPGLSRRQA AITGVDQTTA HSPN-SETEI VKSNKSLEEV TGRDLKDAAV SGEDLDQEIV TGSLLPRAAV NG-DVDPDKA	VAAAE-DAKK EAHVK-DAAE EQIAN-KAKA QNLAE-RAKT QNLAE-RAKT QTLVD-KAHQ EGLWA-AAHQ EGLWA-AAHQ EGLWA-AAHA KKLIQ-KAHE QTVIT-AAHA KKEFID-FVSK KEFID-FVSK KEFID-FVSK KEFID-FVSK KEFID-FVSK KEFID-FVSK KEFID-FVSK KEFID-FVSK KEFID-FVSK KEFID-FVSK KEFID-FVSK	NCPUSQALTG KCPLERIMGG GCPUSKVLNA TCLUSRALAG GCPUSQVLKA VCPUSNATRG VCPUSATRG VCPUSATRG MCPFSRLIK- RCPAENTLGG VCPUFNSLSG VCPUFNSLSG VCSVSLMLEK YCSVSIMAER	-TTITLSAKL -NVEVTHEAI -KISLDASLD -VASITLKAE -BITLDYQLK -NIDVILTLA -NIDVILVS -NUDVILVS -NUDVILVS -NUDVILVS -NUDVITK -ENFLGLTLN PENFLGLTLN VSQLKINVNV RIVIDEKFTL A-VNITHSYE G-VENTHSWE G-VENTHSWE G-VENTHSWE	λ LEG G LIATTAQ S V E GIKL GIKL GIKL TLVH K VVAλ IRTE IVDTQAA LNGRPID	141 172 151 147 143 142 142 139 138 155 140 141 141 140 134 138

#### Table 2

Crystal parameters and refinement statistics.

P2 <sub>1</sub>		
a = 49.57, b = 90.39,		
$c = 112.71, \beta = 93.93$		
46.4		
2.39		
855		
56		
6911		
6332		
492		
87		
61.6		
87.0		
59.2		
20.0-2.4		
$0.0\sigma$		
22.7		
28.7		
0.010		
1.64		
0.88		
25.01		
98.9		
1.1		

#### 2. Materials and methods

#### 2.1. Cloning

Cloning primers (Operon, Alameda, CA, USA) for *OsmC* gene PCR amplification from genomic DNA contained an *NdeI* restriction site in the forward primer (5'- CATATGA-

#### Figure 1

Sequence comparison among OsmC sequence family. OsmC sequence subfamilies are aligned with several homologs. Abbreviations are as follows. (1) OsmC from Streptomyces coelicolor (ScOsmC), Deinococcus radiodurans (DrOsmC), Pseudomonas aeruginosa (PaOsmC), Rhizobium loti (RlOsmC) and E. coli (EcOsmC). (2) Ohr from Xanthomonas campestris (XcOhr), Pseudomonas aeruginosa (PaOhr), D. radiodurans (DrOhr), Caulobacter crescentus (CcOhr), Mycoplasma genitalium (MgOhr) and M. pneumoniae (MpOhr). (3) MPN625 homologs from M. genitalium (MgenHY) and M. gallisepticum (MgalHY). (4) YhfA from E. coli (EcYhfA) and YhfA homologs from Vibrio cholerae (VcYhfA), Bordetella bronchisepticac (BbYhfA) and Thermoplasma volcanium (TvYhfA). The red characters represent residues highly conserved in OsmC and Ohr subfamilies and the residues in blue represent sequences highly conserved in each subfamily. The yellow and green boxes represent for the conserved motifs in each subfamily primarily based on the OsmC and Ohr subfamilies (Atichartpongkul et al., 2001). Underlined characters represent highly conserved histidine and serine residues neighboring the two conserved cysteines. The amino-acid sequence from residues 139 to 153 is not shown in the case of TvYhfA.

CAATCCATAAGAAAGGTCAGGC) and a *Bam*HI site in the reverse primer (5'- GGATCCTTACGATTTCAACTGG-

TAATCCAGC). PCR was performed using Deep Vent DNA Polymerase (New England Biolabs, Inc., Beverly, MA, USA)



#### Figure 2

(a) Steroview of the putative active site of EcOsmC. The  $2F_o - F_c$  map from the final refined phase was calculated using all reflection data between 20 and 2.4 Å. The figure was generated using the program *RIBBONS* (Carson, 1991). The red net represents the electron-density map. The residues are represented by a ball-and-stick model. PHE represents the phenylalanine residue from the His<sub>6</sub>-tag. Residues Arg239, Phe240 and Glu249 of one monomer correpsond to Arg39, Phe30 and Glu49 of the other monomer, respectively. Blue represents N atoms, red O atoms and green C atoms. (b) A stereo drawing of a C<sup> $\alpha$ </sup> trace of EcOsmC. Two subunits are represented by a thick line with different colors. Conserved cysteines (Cys59 and Cys125, light green) and Arg39 (blue) are represented by a ball-and-stick model. Every 20th residue is numbered and represented by a dot. The N-terminus (residue Met1) and C-terminus (residue Ser143) are labeled. The figure was generated using *MOLSCRIPT* (Kraulis, 1991).

and *E. coli* genomic DNA. The PCR product was cloned into the pCR-BluntII-TOPO vector (Invitrogen) and the *OsmC* gene insert was confirmed by DNA sequencing. The amplified TOPO vector was cut with *NdeI* and *Bam*HI and the gene insert was purified by agarose gel electrophoresis extraction. This insert was cloned in a pSKB3 expression system (a gift from Steve Burley, Rockefeller University, NY, USA) and transformed into B834(DE3)/ pSJS1244 (Kim *et al.*, 1998).

## 2.2. Protein expression, purification and crystallization

A selenomethionine derivative of the protein was expressed in a methionine auxotroph, E. coli strain B834(DE3)/ pSJS1244, grown in M9 medium supplied with selenomethionine (Kim et al., 1998). The cloned OsmC gene containing an N-terminal His<sub>6</sub>-tag was affinity purified from the soluble fraction using Talon metalaffinity resin (Clonetech, Palo Alto, CA, USA). The target protein was bound onto a 5 ml HiTrap Q column (Pharmacia Biotech, Uppsala, Sweden) and eluted with 150 mM NaCl. Since the His<sub>6</sub>-tag was not cleaved, the 25-residue tag MGSSHHHHHHDY-DIPTTENLYFQGH remained fused to the N-terminus of the protein. Using the sparsematrix crystal-screening method (Jancarik & Kim, 1991), rod-shaped crystals appeared at room temperature. The optimized crystallization condition was 1 µl protein solution (9 mg ml<sup>-1</sup>) in 50 mM Tris–HCl pH 7.5, 100 mM NaCl and 5 mM DTT, mixed with  $1 \,\mu$ l reservoir solution containing  $0.2 \,M$ magnesium formate and 20% PEG 3350.

#### 2.3. Data collection and reduction

X-ray diffraction data sets were collected at one wavelength corresponding to the selenium absorbance peak ( $\lambda = 0.9793$  Å) at the Macromolecular Crystallography Facility beamline 5.0.2 at the Advanced Light Source at Lawrence Berkeley National Laboratory using an Area Detector System Co. (Poway, CA, USA) Quantum 4 CCD detector placed 200 mm from the sample. X-ray diffraction data were processed and scaled using *DENZO* and *SCALEPACK* from the *HKL* program

#### Table 3

Structural comparison of OsmC sequence family structures.

All the values are from the results of structural comparison performed using the Combinatorial Extension (CE) Method (http://cl.sdsc.edu/ce.html; Shindyalov & Bourne, 1998).

	R.m.s. deviation (Å)	Aligned amino acids	Z score	Sequence identity (%)
OsmC/Ohr	2.2	128	6.1	21.1
OsmC/MPN625	2.4	129	5.0	17.1
OsmC/yhfA	2.3	118	5.6	15.3
Ohr/MPN625	2.6	133	6.0	15.8
Ohr/yhfA	3.3	125	5.3	8.8
MPN625/yhfA	1.7	128	6.0	17.2

suite (Otwinowski & Minor, 1997). Data statistics are summarized in Table 1.

#### 2.4. Structure determination and refinement

The program SOLVE (Terwilliger & Berendzen, 1999) was used to locate the selenium sites in the crystal and to calculate initial phases at 2.4 Å resolution. The partial models were built based on an initial map using the program O (Jones *et al.*, 1991). The presence of six molecules in the asymmetric unit was found based on partial models and this was used to find the non-crystallographic symmetry (NCS) matrices. Sixfold NCS density averaging was carried out using the *DM* program in the *CCP*4 package (Dodson *et al.*, 1997). The improved density map clearly revealed the presence of three homodimers in the asymmetric unit. The model was then refined using the program *CNS* (Brünger *et al.*, 1998), with 10% of the data randomly chosen for free *R*-factor cross-validation. The NCS restraints were used for the entire refinement steps and

were released for the final cycle of refinement. In one monomer model we could not observe residues 41–43, but we could see all 143 residues in the other five monomer models. Four out of the six monomer models display a His<sub>6</sub>-tag and the lengths of the His<sub>6</sub>-tags observed are eight, ten, 19 and 19 residues, respectively. The refinement statistics are shown in Table 2.

#### 3. Results and discussion

#### 3.1. Quality of the model

Five of the six final models in the asymmetric unit contain all 143 aminoacid residues. One monomer model is missing residues 41–43. Most of the residues are well defined by electron density in the refined models of EcOsmC, including some of the His<sub>6</sub>tag residues (Fig. 2*a*). The final model has been refined to 2.4 Å resolution to a crystallographic *R* factor of 22.7%. The root-mean-square (r.m.s.) deviations

**906** Shin et al. • OsmC

from ideal stereochemistry are 0.010 Å for bond lengths, 1.64° for bond angles and 0.88° for improper angles. The averaged *B* factors for main-chain atoms and side-chain atoms are 57.5 and 64.7 Å<sup>2</sup>, respectively. The averaged *B* factor of all modeled His<sub>6</sub>-tag residues is 87.0 Å<sup>2</sup>. Table 2 summarizes the refinement statistics as well as the model-quality parameters. The mean positional error in atomic coordinates for the refined model is estimated to be ~0.3 Å by a Luzzati plot (Luzzati, 1952). All residues lie in the allowed region of the Ramachandran plot produced with *PROCHECK* (Laskowski *et al.*, 1993). The C<sup> $\alpha$ </sup> r.m.s. deviations among the three dimers in the asymmetric unit are in the range 0.4–0.8 Å.

#### 3.2. Overall structure

The *Ec*OsmC monomer has approximate dimensions of 50  $\times$  35  $\times$  30 Å with two domains (Fig. 2*b*). The N-terminal half is composed of three  $\beta$ -strands forming a  $\beta$ -sheet and the C-terminal half is a mixed  $\alpha$ - $\beta$  structure composed of a two-layer  $\alpha\beta$ -sandwich in the CATH classification (Orengo *et al.*, 1997). Two identical monomers form a cylindrically shaped dimer in which six helices are located on the inside and two six-stranded  $\beta$ -sheets formed by domain swapping wrap around these helices (Fig. 2*b*).

#### 3.3. Overall structural comparison among subfamilies

The sequence alignment based on tertiary structures indicates that the secondary-structures are well aligned with each other except for one helix (H1) and several loops (Fig. 3). Helix H1 is unique to the *EcOsmC* structure. Loops L2, L4 and L6 form the walls of the active-site pocket and vary in size and sequence, as shown in the sequence alignment and Fig. 4.



#### Figure 3

Sequence alignment based on tertiary structures. Sequence alignment of *Ec*OsmC, Ohr, MPN625 and yhfA based on three-dimensional structure comparisons using the Combinatorial Extension (CE) Method (http://cl.sdsc.edu/ce.html; Shindyalov & Bourne, 1998). The secondary structure derived from *Ec*OsmC is shown above the sequence. Blue characters represent amino-acid residues belonging to  $\alpha$ -helix, green 3<sub>10</sub>-helix and red  $\beta$ -strand. The yellow-shaded regions refer to the conserved motifs as in Fig. 1. GRXG (motif I) and FXXR (motif I') are motifs unique to the Ohr and *Ec*OsmC subfamilies, respectively, based on structural comparison and sequence alignment (Fig. 1). Underlined characters represent highly conserved histidine and serine residues neighboring the two conserved cysteines. The sequence numbers refer to that of *Ec*OsmC. The averaged r.m.s. deviations of  $C^{\alpha}$  atoms among the subfamilies are listed in Table 3. The structural comparison indicates that the overall structures are quite similar to each other, although the pairwise sequence identity only ranges between 8.8 and 21.1%.

A sequence alignment of the OsmC and Ohr subfamilies suggests the presence of four conserved regions (Völker *et al.*, 1998; Atichartpongkul *et al.*, 2001; Fig. 1): (i) a glycine residue near the amino-terminus, (ii) an NPEQ/EXL motif, (iii) a CF



#### Figure 4

Structural comparison. The stereo drawing was generated using the program *RIBBONS* (Carson, 1991). *EcOsmC* is colored green, Ohr (PDB code 1n2f) red, MPN625 (11ql) yellow and yhfA (1ml8) cyan. The two conserved cysteine residues in the active sites are represented by a ball-and-stick model with corresponding color for each structure. The three loops (L2, L4 and L4) that confine the active-site pocket are labeled. The N- and C-termini of one subunit of *EcOsmC* are labeled. (*a*) Monomer, (*b*) dimer. The figure was drawn with the operational matrix obtained from (*a*).

motif and (iv) an AXXXCPXS motif. The OsmC family structures reveal that the latter three motifs comprise the active site in all three subfamilies. Interestingly, the active site is located at the interface of two subunits (Fig. 5): two conserved cysteine residues (Cys59 and Cys125) from motif III and IV from one subunit are located at the bottom of the active sites and the NPEQ/EXL motif is contributed by the other subunit.

All the structures from the OsmC sequence family revealed

the presence of homodimers. The extensive and tight interaction between monomers is found in all four structures representing the three subfamilies of the OsmC sequence family. The approximate molecular surface area buried by the dimers of *Ec*OsmC, Ohr, MNP625 and yhfA are 2070, 2210, 2010 and 1360 Å<sup>2</sup> per monomer, respectively. The relatively small buried molecular surface area of yhfA is a consequence of the short N-terminal loop and the absence of three residues (Ala77, Asp78 and Thr79) that contact with the other subunit in other subfamily structures (Fig. 4).

# 3.4. Structural comparison of the active sites of *EcOsmC*, subfamily II and subfamily III

When the active-site residues of the four structures were compared, those of EcOsmC and Ohr were surprisingly similar despite their belonging to different subfamilies. Specifically, the arginine (Arg18) residue neighboring the active cysteine (Cys60) in Ohr (Lesniak et al., 2002) is also structurally conserved (Arg39) in EcOsmC (Fig. 6). Biochemical studies of Ohr revealed that Arg18 is important for removing reactive oxygen species (ROS; Lesniak et al., 2002). The role of a positively charged residue in stabilizing the reactive thiol is known to be important for Prx activity (Choi et al., 1998). Therefore, structural similarity of the active site of *Ec*OsmC with Ohr strongly suggests that one of the biochemical functions of EcOsmC may also be to reduce peroxide generated during stress. Although the side chain containing the functional group in the active site is positioned similarly in both structures, it comes from residues located at different positions in the two structures. For example, the functional Arg18 of Ohr comes from the glycine motif (motif I) of Ohr, but the equivalently located Arg39 of EcOsmC comes from the FXXR motif (motif I') of EcOsmC (Figs. 1 and 3).

The highly conserved histidine and serine side chains in motif IV are clustered near the two conserved cysteines in three of the structures, the exception being yhfA, although the functional reason for this motif is not clear (Fig. 6b). In the OsmC subfamily, this histidine residue comes from motif II rather than motif IV as found in other subfamilies (Figs. 1 and 3). YhfA only has a conserved serine residue (Ser109) in this region.

The environment of the cysteine located at the active site of MPN625 and yhfA is quite different from that in EcOsmC and Ohr. Firstly, the neighboring residues of the active cysteine of MPN625 and yhfA (Cys52 and Cys47 based on the Ohr structure, respectively) have no arginine next to them. Secondly, two conserved cysteines of these two proteins form a disulfide bond, unlike the thiol form in EcOsmC and Ohr (Figs. 1 and 4b). Functional studies of Ohr showed that DTT could regenerate the thiol form during the catalytic cycle (Lesniak et al., 2002). The crystallization conditions of EcOsmC, Ohr and MPN625 contained 5, 3 and 1 mM DTT, respectively (the crystallization conditions for yhfA are not available). Therefore, one of the reasons for the oxidized form of MPN625 in the crystal structure could be the result of the small amount of DTT in the crystallization conditions or because of a requirement for a different reducing agent. We



#### Figure 5

Molecular surface of EcOsmC. One subunit is colored yellow and the other is colored scarlet. The figure shows that the dimer interface confines the active-site pocket. Yellow color represents S atoms, blue nitrogen and green carbon. The molecule was rotated to show a better view of an active site. The figure was drawn with the program *GRASP* (Nicholls *et al.*, 1991).

have previously proposed that the function of MPN625 could be that of a structural or regulatory protein in the signal transduction pathway of the stress response, based on the absence of prominent characteristics at the active site (Choi et al., 2003). However, the different environment of the active site of MPN625 does not exclude the possibility of a Prx function. The reactive thiol is, if formed, still stable in a hydrophobic environment as shown in the reduced thioredoxin structure (Dyson et al., 1989; Jeng et al., 1994). The oxidized forms of MPN625 and yhfA indicate that the structural differences between the reduced and oxidized form among the OsmC sequence family members are very small (Fig. 4b). The Prx family generally shows a large conformational change from a dimer (reduced) to a decamer (oxidized) during a catalysis cycle (Wood et al., 2002). However, OsmC sequence family structures did not show this type of quaternary conformational change.

#### 3.5. Comparison of active-site pocket

A detailed look at the molecular surface and charge distribution of each structure revealed differences around the entrance to the active-site pocket, although the overall structure of the three subfamilies are similar (Fig. 7). The differences are thought to be related to the substrate specificity, assuming a Prx function. Since the entrance to the active site of Ohr is mostly surrounded by hydrophobic side chains, Ohr prefers the metabolism of hydrophobic rather than inorganic hydroperoxides (Lesniak et al., 2002). The EcOsmC structure shows that the entrance to the active site has more aromatic residues than in the other two subfamily structures. Interestingly, an aromatic residue (phenylalanine) from the uncleaved His<sub>6</sub>-tag is bound to the entrance of the active-site pocket in the *Ec*OsmC structure (Fig. 2*a*). This is reminiscent of the crystal structure of peroxiredoxin 5 (Prx 5), in which a benzoate molecule was bound around the active site in the crystal structure (Declercq et al., 2001). Prx 5 has an aromatic residue at the entrance of the active site that conferred a preference for aromatic hydroperoxides as substrates (Declercq *et al.*, 2001). Similarly, *EcOsmC* may have a possible preference for aromatic hydroperoxides as substrates as suggested by the crystal structure. The active sites of EcOsmC and Ohr have a diameter of ~4 Å owing to the location of aromatic amino acids around the active-site pockets: Phe36, Phe40 and Phe93 in the case of EcOsmC, and Phe67 and Phe95 (each from a different subunit) in the case of Ohr. The active-site pockets of MPN625 and yhfA are apparently wider and more open. The diameters of the putative active-site pockets of MPN625 and yhfA are  $\sim 10$  Å (Fig. 7). This data, together with the different charged distributions next to the active sites (Fig. 7), would indicate that if MPN625 and yhfA have Prx activity, their substrate could be different from those that bind to *Ec*OsmC and Ohr.

## 3.6. The comparison between OsmC sequence family and peroxiredoxin

The structural comparison within the OsmC sequence family confirmed that members of this family may have Prx

activity and contain a unique fold and molecular characteristics that are distinct from the structure of the first reported Prx family (Pfam00578; Choi *et al.*, 1998), which is highly conserved in eukaryotes and prokaryotes. The crystal structure of hORF6 revealed an extended shape with a tenstranded  $\beta$ -sheet (in dimeric state) in the center that is distinct



#### Figure 6

(a) Active-site residues of EcOsmC. The residues around the active sites are labeled and represented by a ball-and-stick model. The view is from opposite the entrance to the active site. (b) Comparison of active-site residues. Green color represents EcOsmC, pink Ohr and yellow MPN625. The average distances between the S<sup> $\delta$ </sup> atoms of the two active-site cysteines of EcOsmC, Ohr, MPN625 and yhfA (not shown in this figure) are 3.78, 3.67, 2.07 and 2.05 Å, respectively.

from the barrel shape found in the OsmC sequence family structures. Unlike the OsmC sequence family, the two cysteines of the 2-Cys Prx family come from different subunits, which results in the large quaternary structural changes that occur during a catalysis cycle. The molecular surface area of  $\sim$ 1700 Å<sup>2</sup> buried by the dimer in hORF6 (Choi *et al.*, 1998) is

smaller than that of the averaged value of the OsmC sequence family ( $\sim$ 1910 Å<sup>2</sup>). In both Prx families, the interface of the dimercontact region confines the active-site pocket (Fig. 5).

#### 3.7. Possible molecular function of OsmC

Since OsmC is induced by salt stress, its molecular function as a Prx is not clear at first glance. However, OsmC-induced cells shows high viability during oxidative stress: an OsmC mutant exhibited a higher sensitivity to t-butanol in the exponential growth phase and to  $H_2O_2$  and *t*-butanol in the stationary phase (Conter et al., 2001). The relationship between osmotic pressure and the generation of reactive oxygen species (ROS) is not known in E. coli. However, in plants, during water stress brought about by salt stress, reduction of chloroplast stromal volume and generation of ROS result in the inhibition of photosynthesis (Price & Hendry, 1991). Therefore, the generation of ROS by osmotic pressure is a hazardous situation in plants (Price & Hendry, 1991). A similar phenomenon was reported during osmotic shock of E. coli cells in NaCl or sucrose medium, resulting in a large decrease in the cytoplasmic volume and in the inhibition of growth, of the electrontransfer chain and of four different types of sugar-transport systems (Houssin et al., 1991). Alhough direct evidence has not yet been reported for the production of ROS in these situations, the malfunction of cellular machinery during osmotic pressure might be the source of production of ROS in E. coli as well as in plants. Therefore, one of the molecular functions of OsmC may be Prx activity working as a scavenger for specific ROS.

#### 4. Conclusions

The structure of *Ec*OsmC provides a structural view of an OsmC subfamily I member. Our study has confirmed that the OsmC sequence family may have a Prx function. The MPN625 structure in an oxidized state clearly showed that the conformational

### structural genomics papers

change during a catalysis cycle might be very small in the OsmC sequence family compared with the drastic quaternary structural change shown in the first Prx family (Wood *et al.*, 2002; Choi *et al.*, 2003). Even though the substrates of the proteins are not known, the different entrance surface features in the subfamilies suggest different substrate specificities. The genome sequences of *Mycobacterium genitalium* and *M. pneumoniae* revealed a lack of known genes involved in peroxide metabolism (Mongkolsuk *et al.*, 1998) except for the



#### Figure 7

Electrostatic surface potential of EcOsmC, Ohr, MPN625 and yhfA. Electrostatic surface potentials of EcOsmC (*a*), Ohr (*b*), MPN625 (*c*) and yhfA (*d*) were drawn with the program *GRASP* (red, negative; blue, positive; white, uncharged; Nicholls *et al.*, 1991). The conserved cysteine residues are labeled and represented by a ball-and-stick model. Yellow color represents S atoms and green represents C atoms. Each molecule was rotated to show a better view of the active site.

OsmC sequence family. Since overcoming oxidative stress induced by the host is essential for survival against the host defense system, the OsmC sequence family may be a good target for a new class of antibacterial drugs.

During the preparation of this paper, another manuscript appeared presenting the same structure with experimentally proven Prx function (PDB code 1qwi; Lesniak *et al.*, 2003). The r.m.s. deviations of 278 C<sup> $\alpha$ </sup> atoms between the structural models (dimers) of the two *Ec*OsmCs are in the range 0.4–

0.8 Å and reveal no significant difference, although our structural models contain an uncleaved His<sub>6</sub>-tag model.

We thank Dr Keith Henderson (Advanced Light Source, Lawrence Berkeley National Laboratory) for assistance during data collection. We are also grateful to Dr Igor Grigoriev for target selection, Barbara Gold for cloning and Bruno A. Martinez and Marlene Henriquez for protein expression and cell growth. The work described here was supported by the National Institutes of Health GM 62412.

#### References

- Atichartpongkul, S., Loprasert, S., Vattanaviboon, P., Whangsuk, W., Helmann, J. D. & Mongkolsuk, S. (2001). *Microbiology*, 147, 1775–1782.
- Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L. & Sonnhammer, E. L. (2000). *Nucleic Acids Res.* **30**, 276–280.
- 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. D54, 905–921.
- Carson, M. (1991). J. Appl. Cryst. 24, 958-961.
- Choi, H. J., Kang, S. W., Yang, C. H., Rhee, S. G. & Ryu, S. E. (1998). *Nature Struct. Biol.* 5, 400–406.
- Choi, I. G., Shin, D. H., Brandsen, J., Jancarik, J., Busso, D., Yokota, H., Kim, R. & Kim, S.-H. (2003). J. Struct. Funct. Genomics, 4, 31–34.
- Conter, A., Gangneux, C., Suzanne, M. & Gutierrez, C. (2001). *Res. Microbiol.* 152, 17–26.
- Declercq, J. P., Evrard, C., Clippe, A., Stricht, D. V., Bernard, A. & Knoops, B. (2001). J. Mol. Biol. 311, 751–759.
- Dodson, E. J., Winn, M. D. & Ralph, A. C. (1997). *Methods Enzymol.* 277, 620–633.
- Dyson, H. J., Holmgren, A. & Wright, P. E. (1989). *Biochemistry*, **28**, 7074–7087.
- Hengge-Aronis, R. (1996). *Mol. Microbiol.* **21**, 887–893.

- Houssin, C., Eynard, N., Shechter, E. & Ghazi, A. (1991). *Biochim. Biophys. Acta*, **1056**, 76–84.
- Jancarik, J. & Kim, S.-H. (1991). J. Appl. Cryst. 24, 409-411.
- Jeng, M. F., Campbell, A. P., Begley, T., Holmgren, A., Case, D. A., Wright, P. E. & Dyson, H. J. (1994). *Structure*, **2**, 853–868.
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Acta Cryst. A47, 110–119.
- Jung, J. U., Gutierrez, C., Martin, F., Ardourel, M. & Villarejo, M. (1990). J. Biol. Chem. 265, 10574–10581.
- Kaasen, I., Falkenberg, P., Styrvold, O. B. & Strom, A. R. (1992). J. Bacteriol. 174, 889–898.
- Kim, R., Sandler, S. J., Goldman, S., Yokota, H., Clark, A. J. & Kim, S.-H. (1998). *Biotechnol. Lett.* 20, 207–210.
- Kolter, R., Siegele, D. A. & Tormo, A. (1993). Annu. Rev. Microbiol. 28, 855–874.
- Kraulis, P. J. (1991). J. Appl. Cryst. 24, 946-950.
- Lange, R. & Hengge-Aronis, R. (1991). Mol. Microbiol. 28, 49–59.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). J. Appl. Cryst. 26, 283–291.

- Lesniak, J., Barton, W. A. & Nikolov, D. B. (2002). *EMBO J.* **21**, 6649–6659.
- Lesniak, J., Barton, W. A. & Nikolov, D. B. (2003). Protein Sci. 12, 2838–2843.
- Luzzati, V. (1952). Acta Cryst. 5, 802-810.
- Mongkolsuk, S., Praituan, W., Loprasert, S., Fuangthong, M. & Chamnongpol, S. (1998). J. Bacteriol. 180, 2636–2643.
- Nicholls, A., Sharp, K. A. & Honig, B. (1991). *Proteins*, **11**, 281–296.
   Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B. & Thornton, J. M. (1997). *Structure*, **5**, 1093–1108.
- Otwinowski, Z. & Minor, W. (1997). Methods Enzymol. 276, 307–326.
- Price, A. H. & Hendry, G. A. F. (1991). Plant Cell Environ. 14, 477– 484.
- Shindyalov, I. N. & Bourne, P. E. (1998). Protein Eng. 11, 739-747.
- Terwilliger, T. C. & Berendzen, J. (1999). Acta Cryst. D55, 849–861.
- Völker, U., Andersen, K. K., Antelmann, H., Devine, K. M. & Hecker, M. (1998). J. Bacteriol. 180, 4212–4218.
- Wood, Z. A., Poole, L. B., Hantgan, R. R. & Karplus, P. A. (2002). Biochemistry, 41, 5493–5504.
- Yim, H. H. & Villarejo, M. (1992). J. Bacteriol. 174, 3637-3644.