Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Streptococcus pneumonia YlxR at 1.35 Å shows a putative new fold

Jerzy Osipiuk,[a] Piotr Górnicki,[b] Luke Maj,[a] Irina Dementieva,[a] Roman Laskowski[c] and Andrzej Joachimiak[a]*

[a]Argonne National Laboratory, Structural Biology Center, Biosciences Division, 9700 South Cass Avenue, Argonne, IL 60439, USA, [b]Department of Molecular Genetics and Cell Biology, University of Chicago, 920 East 58th Street, Chicago, IL 60637, USA, and [c]The Department of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX, England

Correspondence e-mail: andrzejj@anl.gov

Received 1 May 2001
Accepted 23 August 2001

PDB Reference: YlxR, 1g2r.

The structure of the YlxR protein of unknown function from *Streptococcus pneumonia* was determined to 1.35 Å. YlxR is expressed from the *nusA*/*infB* operon in bacteria and belongs to a small protein family (COG2740) that shares a conserved sequence motif GRGA(Y/W). The family shows no significant amino-acid sequence similarity with other proteins. Three-wavelength diffraction MAD data were collected to 1.7 Å from orthorhombic crystals using synchrotron radiation and the structure was determined using a semi-automated approach. The YlxR structure resembles a two-layer α/β sandwich with the overall shape of a cylinder and shows no structural homology to proteins of known structure. Structural analysis revealed that the YlxR structure represents a new protein fold that belongs to the α–β plait superfamily. The distribution of the electrostatic surface potential shows a large positively charged patch on one side of the protein, a feature often found in nucleic acid-binding proteins. Three sulfate ions bind to this positively charged surface. Analysis of potential binding sites uncovered several substantial clefts, with the largest spanning 3/4 of the protein. A similar distribution of binding sites and a large sharply bent cleft are observed in RNA-binding proteins that are unrelated in sequence and structure. It is proposed that YlxR is an RNA-binding protein.

1. Introduction

Structural genomics is flourishing owing to tremendous progress in genome sequencing as well as recent advances in computer and software technology and third-generation synchrotron beamlines for macromolecular crystallography. One of the foremost goals of structural genomics is to map the entire protein-folding space. This can be accomplished by solving the structures of a large number (15 000–20 000) of carefully selected proteins that show no significant sequence homology to each other and are therefore likely to include the majority of unique protein folds (Vitkup *et al.*, 2001). It is anticipated that this effort will expand knowledge of protein structure and will facilitate solving the structures of other proteins. For many proteins, function has not yet been established. It is expected that structural genomics will help to assign functions to proteins when assignment is not possible with amino-acid sequence comparisons alone (Shazand *et al.*, 1993; Christendat *et al.*, 2000). This is especially important as thousands of newly identified open reading frames (ORFs) representing putative protein genes became available from genome-sequencing programs. Structural information may provide important functional clues.

The selection of proteins for structure determination is key to the structural genomics approach (Linial & Yona, 2000). In this work we applied the following three criteria: (i) uniqueness of amino-acid sequence, to increase chances of identifying a new protein fold, (ii) unknown function, to aid function assignment to a new class of proteins, and (iii) origin from a pathogenic bacterium, to provide a basis for future investigation of the protein as a potential target for new drugs. These criteria were met by the YlxR homologue from *S. pneumonia*: a 97 amino-acid protein of unknown function whose gene is located 21 bp downstream of the *nusA* gene and 248 bp upstream of the *infB* gene. The YlxR name was given to the *S. pneumonia* ORF to reflect its similarity to the YlxR protein of *Bacillus subtilis* (Fig. 1) (Overbeek *et al.*, 2000; http://selkov.mcs.anl.gov/WIT2/CGI/prot.cgi?prot=RPN00578&user). YlxR belongs to a small 13-member protein family (COG2740; Tatusov *et al.*, 2001; http://www.ncbi.nlm.nih.gov/cgi-bin/COG/palox?COG2740) predicted to be nucleic acid-binding proteins and implicated in transcription termination. YlxR shows no significant sequence similarity to any other protein,

Table 1
Summary of crystal and MAD data.

Crystal data.

Unit-cell parameters (Å, °)	a = 28.053, b = 48.747, c = 73.695, $\alpha = \beta = \alpha$ = 90.00
Space group	$P2_12_12_1$
MW (100 residues)	11521
Molecules per asymmetric unit	1
SeMet per asymmetric unit	2

MAD data collection.

	Edge	Peak	Remote	High resolution
Wavelength (Å)	0.9779	0.97770	0.9701	1.0332
Resolution range (Å)	1.70	1.70	1.70	30–1.35 (1.37–1.35)
No. of unique reflections	11456	11456	11467	22942
Completeness (%)	99.4	99.4	99.4	93.5 (61.3)
$R_{\rm merge}$ (%)	6.2	5.6	5.8	6.3 (26.9)

including proteins of known structure, as determined by analysis of the ProtoMap Database (Portugaly & Linial, 2000; http://www.protomap.cs.huji.ac.il/).

YlxR is located in the putative *nusA*/*infB* operon of *S. pneumonia* which consists of seven genes. Three of these genes, *rbfA*, *nus*A and *infB*, are present in other bacteria. RbfA is a conserved protein that binds to the 30S ribosomal subunit and is believed to be involved in ribosomal maturation (it is essential for 16SrRNA processing) and/or translation initiation (Dammel & Noller, 1995). NusA, which is unique to prokaryota, is essential for bacterial viability and is involved in the modulation of RNA polymerase processivity leading to transcription antitermination (Vogel & Jensen, 1997; Mah *et al.*, 1999). Homologues of NusA are also found in archaebacteria. IF2 (product of *infB*) is a translation-initiation factor associated with ribosomes and plays a crucial role in translation initiation (Grill *et al.*, 2000). NusA, IF2 and RbfA are cold-shock proteins (Bae *et al.*, 2000). No function has been assigned to the remaining four proteins encoded by the operon, including YlxR (Overbeek *et al.*, 2000).

2. Material and methods

The *ylxR* gene was amplified by PCR with the *Nde*I and *Bam*HI sites engineered at the translation start codon and immediately downstream of the translation stop codon, respectively, and cloned between the *Nde*I and *Bam*HI sites of the pET15b vector (Novagen) in frame with the His tag and the thrombin cleavage site. Expression of the His-tagged fusion protein in *Escherichia coli* strain BL21[DE3] carrying the pMAGIC vector was induced with isopropyl-β-D-thiogalactoside. Cells were harvested after 4 h culture at 310 K, suspended in 50 m*M* sodium phosphate buffer pH 8.0, 300 m*M* NaCl, 10 m*M* imidazole, 10 m*M* β-mercaptoethanol, 10% glycerol and lysed by sonication. The fusion protein was purified by affinity chromatography using Ni-NTA Superflow resin (Qiagen). The His tag was removed by digestion with thrombin and the resulting protein was purified following the manufacturer's protocol (Novagen). In this design, three amino-acid residues were added at the N-terminus of YlxR. The protein was further purified on an SP Sepharose Fast Flow column (Pharmacia) using 0.5 and 1.0 *M* NaCl two-step elution and concentrated with simultaneous buffer exchange using Centriplus-3 (3 kDa cut-off; Amicon). A 2 m*M* protein stock solution in 10 m*M* Tris–HCl pH 7.4, 20 m*M* NaCl and 1 m*M* DTT was used for crystallization. Selenomethionine (SeMet) labeled YlxR protein was prepared by a standard procedure using methionine-biosynthesis inhibition (Walsh *et al.*, 1999).

Equal volumes of YlxR protein stock solution and buffers were mixed in hanging drops and equilibrated against 1 ml of solutions from Hampton Research sparse-matrix crystallization screening kits. YlxR was crystallized from 0.2 *M* potassium sodium tartrate, 100 m*M* sodium acetate pH 5.6 and 2 *M* ammonium sulfate at 283 K. Crystals (0.2 × 0.2 × 0.1 mm) were briefly rinsed in cryoprotectant solution consisting of 25% glycerol in the crystallization solution and flash-frozen in liquid nitrogen. Diffraction data were collected at 100 K at the 19BM beamline of the Structural Biology Center at the Advanced Photon Source, Argonne National Laboratory. Crystals of native YlxR protein and its SeMet derivative diffracted to 1.35 and 1.7 Å, respectively. MAD data were collected to 1.7 Å resolution from a single crystal containing SeMet-labeled protein at three different X-ray wavelengths near the Se edge. The inverse-beam strategy was used. The absorption edge was determined by a fluorescent scan of the crystal as described in Walsh *et al.* (1999). The data were processed using the *HKL*2000 suite (Otwinowski & Minor, 1997). Crystal characteristics and data-collection statistics are presented in Table 1.

The structure of YlxR was determined using the MAD approach. A single SeMet site was selected from the asymmetric unit and MAD phases were calculated using the *CNS* suite (Brunger *et al.*, 1998) (Table 2). These MAD phases were improved using density modification as implemented in *CNS*. Electron-density maps were high quality and allowed autotracing of the amino-acid chain using the *wARP* program (Perrakis *et al.*, 1999). The procedure provided an initial model containing 91 out of 100 amino-acid residues. The model was refined using 1.35 Å data with the *REFMAC* program from the *CCP*4 suite (Murshudov *et al.*, 1999). Manual adjustment and model building using the program *O* (Jones *et al.*, 1991) allowed the addition of three more amino-acid residues. Application of anisotropic refinement of *B* factors as implemented by *REFMAC* (Murshudov *et al.*, 1999) improved the *R* factor and $R_{\rm free}$ to 15.7

```
                          1               16              31              46            60
S. pneumoniae, RPN00578       MKTRKIPLRKS VVSNEVIDKRDLLRI VKNKEGQ----VFID PTGKANGRGAYIKLD
S. pyrogenes, RST00356      MSKVKKIPLRKS LVSGEIIAKRDLLRI VKTKDGQ----VFID PTGKQNGRGAYIKLD
L. lactis, gi|12723683|       MKQKKTPMRKS LVSNEQFPKKDLLRI AYNKEGE----ISID PSGKAHGRGAYIAIL
E. faecalis, REF02732|        MKKRKIPMRKS VVSGEMKPKKELVRI TRSKEGE----VALD PTGKLPGRGAYVDLD
B. subtilis, gi|418461|      MNKHKKIPLRKC VVTGEMKPKKELIRV VRSKEGE----ISVD PTGKKNGRGAYLTLD
B. halodurans, gi|10175035|   MKQRKIPLRKC VVTNEMKPQELIRV VRSPEGN----VFID PTGKQNGRGAYISNN
M. tuberculosis, gi|7477398|        MRTC VGCRKRGLAVELLRV VAVSTGNGNYAVIVD TATSLPGRGAWLHPL
M. leprae, gi|4455677|              MRTC VGCRKRELAVELLRV VAPSTGKGSYAVIVD TASSLSGRGAWLHPD
S. coelicolor, gi|7801277|  MSGRTRTRACPERTC VGCRERAVKSELLRT VAV-EGH----CTPD PRGTLPGRGAYVHPA
T. aquaticus, gi|1072955|       MKHVPIRMC VACRRRRPKGELLRI LVTEEGF-----VID PSGKRPGRGAYVCPD
                                     *              * *       *        *       ****

                          61              76              91              106
S. pneumoniae, RPN00578   NAEALEAKKKKVFNRSFS MEVE-E-SFYDE LIAYVDHKVKRRELG LE               97
S. pyrogenes, RST00356    NQEALMAKKKQVFNRSFS MDIP-E-SFYDD LIAYVDHKIKRRELG LD               98
L. lactis, gi|12723683|   NKEAREA-KKRVFDRAFQ TKIA-D-EFYDE LITYVEHLVARRELE SATYSADLAPDYD  107
E. faecalis, REF02732     PAEVQKAWDKKILDRVLE TKLS-D-EFYQE LLDYVTHQKARKELF GDGK             99
B. subtilis, gi|418461|   KECILAAKKKNTLQNQFQ SQID-D-QIFDE LLELAEKVKK                       91
B. halodurans, gi|10175035| KECFELAKKKDILSKHLN VKVS-D-DVYDQ LEEARQRGSSK                    91
M. tuberculosis, gi|7477398| RQCAQQAIRRRAFARALR IAGSPDTSAVVE YLESLGELEPPGNRT GSNRT         99
M. leprae, gi|4455677|    MQCVQQAIRRRAFTGALR IAGSPDTSAVVE HIEFLSELDRPGNRT GSKEHEHTVKSR  106
S. coelicolor, gi|7801277| PDCVDQAVRRKAFPRALR VPGPLDVKALRH YVEQAEGCSKVAESS M              101
T. aquaticus, gi|1072955|  SPECRSEKRLRRFAGA-K ARALAE--ALAA YLGGEDGQGKDLPAG               91
```

Figure 1
Sequence alignment of YlxR proteins. Completely conserved residues are marked with an asterix and the GRGA(Y/W) motif is underlined.

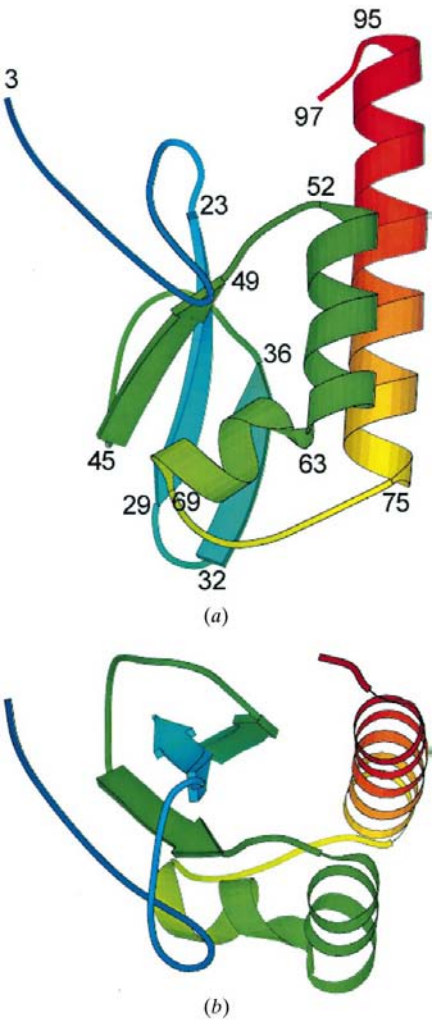

Figure 2
Structure of the YlxR protein. The ribbon diagrams of the C^{α} backbone were prepared with the *Molscript* (Kraulis, 1991) and Adobe *Photoshop* programs. The view in (*b*) was obtained by a 90° rotation of the view in (*a*) around the *x* axis. Numbers indicate amino-acid residues.

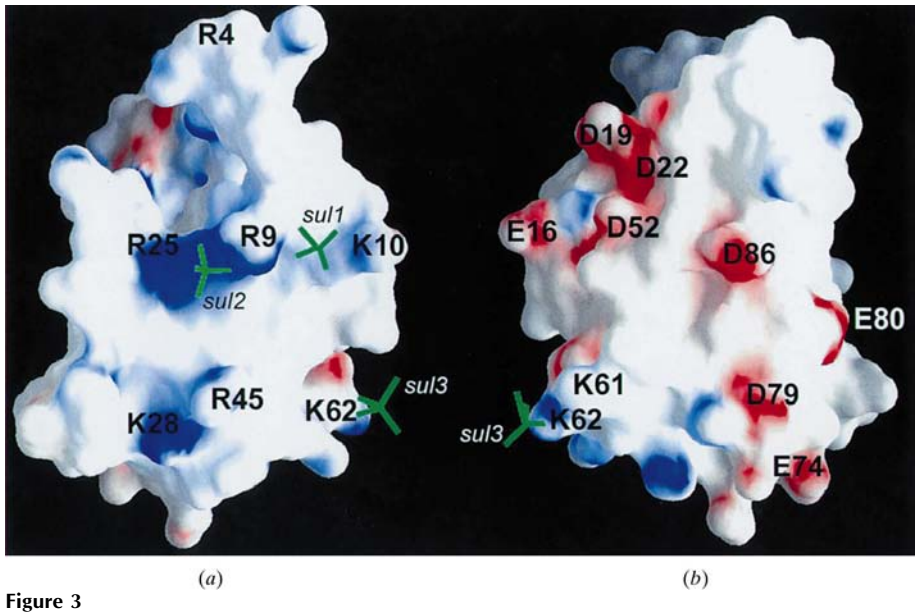

Figure 3
Electrostatic surface potential of the YlxR protein calculated with the *GRASP* program (Nicholls *et al.*, 1993). Both positive (*a*) and negative (*b*) sides of the protein are shown. Sulfate ions are shown in green. Numbers indicate amino-acid residues.

and 18.5%, respectively. The final structure included 94 amino-acid residues, three sulfate ions and 131 water molecules. Six amino-acid residues present at the N-terminus of the protein were not visible in the electron density: the three residues remaining from the affinity-tag fusion and the first three residues of YlxR. Residues 4 and 5 of YlxR were modeled using a partial occupancy of 62% because of the disorder of the protein N-terminus. 13 amino-acid side chains (residues 6, 12, 21, 30, 31, 39, 41, 45, 51, 64, 74, 75 and 88) showed double occupancy. Partial occupancies of alternate conformers were calculated with the program *SHELXL* (Sheldrick & Schneider, 1997).

3. Results and discussion

The protein structure resembles a two-layer α/β sandwich with an overall cylindrical shape (Fig. 2). N-terminal residues 1–19 do not form any regular secondary structure. This segment is followed by a very short 3_{10}-helix (residues 20–22). The central part of the protein consists of three antiparallel β-strands (β1, β2 and β3). The C-terminal part of the protein forms two relatively long α-helices (α1 and α2). α1 is bent at Lys63 by approximately 60°, giving rise to two sub-helices (α1a and α1b). Helix α1b interacts with β3; helix α1a is parallel to and interacts with α2. Helix α2 also interacts with β2 and closes the cylindrical structure (Fig. 2*b*). A well defined hydrophobic core is formed by the residues of helices α1a and α1b, the loop between α1 and α2, the N-terminal part of α2, all three β-strands and Val12 and Val13.

The Protein Data Bank was searched (*DALI* server; Holm & Sander, 1993) to identify proteins with structural similarity to YlxR. The best match, with a *Z* score of 3.1 [with a positional root-mean-square deviation (r.m.s.d.) of superimposed C^{α} atoms of 2.9 Å for 64 equivalenced residues], indicating rather low structural similarity, was to domain A of guanosine pentaphosphate synthetase (PDB code 1e3h). This domain has a structural motif consisting of three β-strands and two α-helices; however, the orientation of these elements is different to the orientation of α-helices and β-strands in YlxR. Second on the list was ColE1 ROP protein (PDB code 1nkd), which shows a *Z* score of 2.9 (with an r.m.s.d. of superimposed C^{α} atoms of 4.3 Å

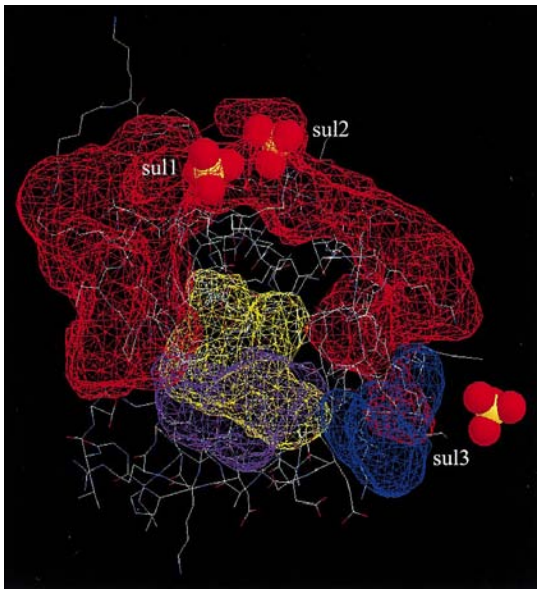

Figure 4
Clefts and cavities in the structure of YlxR protein were calculated using the program *SURFNET* (Laskowski, 1995) and displayed with *RasMol* (Sayle & Milner-White, 1995). Only the four largest cavities are shown (red, 2885 $Å^3$; magenta, 497 $Å^3$; yellow, 455 $Å^3$; blue, 348 $Å^3$). The major cavity (red) includes two sulfate ions (sul1 and sul2).

Table 2
Crystallographic statistics.

Phasing.

	Centric		Acentric		All		
Resolution range (Å)	FOM	Phasing power	FOM	Phasing power	No.	FOM	Phasing power
20.0–1.7	0.62	1.9229	0.48	1.6433	20584	0.49	1.6788

Refinement.

Resolution range (Å)	30–1.35
No. of reflections	21402
σ cutoff	None
R value (%)	15.7
Free R value (%)	18.5 (1573 reflections)
R.m.s. deviations from ideal geometry (Å)	
Bond length (1–2)	0.013
Angle distance (1–3)	0.031
Planar distance (1–4)	0.034
No. of atoms	
Protein	835
Sulfates	15
Water	131
Mean B factor (Å^2)	
All atoms	20.4
Protein atoms	18.0
Protein main chain	15.7
Protein side chain	20.2
Sulfate ions	24.7
Water	34.3
Ramachandran plot statistics (%)	
Residues in most favored regions	92.9
Residues in additional allowed regions	7.1
Residues in disallowed region	0.0

for 48 equivalenced residues). Other matches found by the *DALI* program showed even lower similarity (structures with a Z score of 2 or less are dissimilar).

CATH analysis (Pearl *et al.*, 2000) showed that domain 1 (residues 1–47) of inositol polyphosphate 1-phosphatase (PDB code 1inp) shows a very distant structural homology to YlxR. This domain has two β-strands and two α-helices and it belongs to the α–β plait folds (Cort *et al.*, 1999). The fold of this domain is described as 'irregular', having little secondary structure. Approximately half of the YlxR fold matches other similar motifs in the plait folds. The plaits appear to share motifs in common with other folds. Therefore, YlxR seems to be an addition to this group.

Comparison of YlxR topology with the TOPS domain database and PDB domains (http://tops.ebi.ac.uk/tops/) shows only one domain (A2 of iron superoxide dismutase; rank = 5; PDB code 1mng) with a rank less than 10. This domain has a structural motif consisting of three β-strands and one α-helix, but its structure (the length of β-strands and α-helix, and their relative orientation) is very different to that of YlxR. Based on structural comparisons, we postulate that the YlxR protein structure represents a new protein fold, *e.g.* it has a unique arrangement and a connectivity of secondary-structure elements not found in protein folds deposited in the PDB.

The N-terminal half of YlxR shows significant sequence similarity to other members of the COG2740 family. It includes a conserved sequence motif GRGA(Y/W). Some of the highly conserved residues are located in the hydrophobic core (Val12, Leu23, Leu24, Ile26, Ile36, Ile49, Phe65, Phe69, Leu81); however, some are charged and located on the protein surface (Arg4, Arg9, Asp19, Arg22, Asp37, Arg45, Lys61, Lys62). We propose that these conserved charged residues on the surface are likely to play a role in the protein's function.

The distribution of the electrostatic surface potential shows one side of the protein charged positively and the other charged negatively (Fig. 3). Several conserved residues contribute to this charge distribution. Such a large positively charged patch is a typical feature of nucleic acid-binding proteins such as trp repressor (Lawson *et al.*, 1988) and YrdC (Teplova *et al.*, 2000). Moreover, three sulfate ions were found in the YlxR structure which bind to the positively charged surface (Fig. 3). Two sulfate ions interact with the conserved Arg9 and Arg25 and with Tyr48, which is a part of the GRGA(Y/W) motif. The third sulfate ion coordinates to Lys61 and Lys62, which are also conserved. The distances between sulfate ions (7.04, 25.0 and 28.3 Å) may correspond to distances between phosphate groups in the RNA duplex. Binding of sulfate ions could reflect interaction between YlxR and nucleic acid phosphate groups. A potentially relevant example is provided by the crystal structure of the E2 DNA-binding domain from the human papillomavirus (PDB entry 1a7g). In this structure, there are two sulfate ions bound to the E2 protein. One sulfate ion contacts Arg309 and Thr325. These residues are in equivalent positions to Arg342 and Thr359 in the bovine papillomavirus E2 DNA-binding domain. In the complex of bovine E2 with DNA target, these two residues contact two consecutive phosphate groups of DNA duplex (Hegde *et al.*, 1992; PDB entry 2bop). Therefore, binding of sulfate ions may indicate potential interaction of proteins such as YlxR with DNA/RNA phosphate groups.

We have searched for potential binding sites on the surface of YlxR using the program *SURFNET* (Laskowski, 1995). This analysis revealed several clefts (gap regions) on the protein surface (Fig. 4). The largest cleft (labeled in red in Fig. 4), with a volume of 2885 Å^3, runs around 3/4 of protein surface, encompasses two sulfate ions and is near the third sulfate ion. The cleft is sharply bent ($\sim$100°) near conserved Arg9, Arg25 and Tyr48 and could accommodate an L-shaped RNA (tRNA-like) molecule. Sulfate ions 1 and 2 are located at the position where the cleft is bent. A strikingly similar distribution of binding sites and a large sharply bent cleft is observed in RNA-binding protein U1A of *Hepatitis delta* virus ribozyme (PDB code 1cx0; Ferre-D'Amare *et al.*, 1998).

Furthermore, because the majority of genes that flank *ylxR* gene in the 13 known *nusA*/*infB* operons code for proteins that bind RNA and/or participate in processes involving RNA (Vogel & Jensen, 1997; Mah *et al.*, 1999; Dammel & Noller, 1995; Bylund *et al.*, 1998; Grill *et al.*, 2000), we propose that YlxR, consistent with the prediction made for COG2740 (Tatusov *et al.*, 2001; http://www.ncbi.nlm.nih.gov/cgi-bin/COG/palox?COG2740), is very likely to be an RNA-binding protein.

We thank Natalia Maltsev for help with target selection, Young-Chang Kim and Martin Walsh for stimulating discussions, and Joanna Jelenska for constructing the YlxR expression system. This work was supported by National Institutes of Health Grant GM62414-01 and by the US Department of Energy, Office of Health and Environmental Research under contract W-31-109-Eng-38.

References

Bae, W., Xia, B., Inouye, M. & Severinov, K. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 7784–7789.

Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Bylund, G. O., Wipemo, L. C., Lundberg, L. A. & Wikstrom, P. M. (1998). *J. Bacteriol.* **180**, 73–82.

Christendat, D., Yee, A., Dharamsi, A., Kluger, Y., Savchenko, A., Cort, J. R., Booth, V., Mackereth, C. D., Saridakis, V., Ekiel, I., Kozlov, G., Maxwell, K. L., Wu, N., McIntosh, L. P., Gehring, K., Kennedy, M. A., Davidson, A. R., Pai, E. F., Gerstein, M., Edwards, A. M. & Arrowsmith, C. H. (2000). *Nature Struct. Biol.* **7**, 903–908.

Cort, J. R., Koonin, E. V., Bash, P. A. & Kennedy, M. A. (1999). *Nucleic. Acids. Res.* **15**, 4018–4027.
Dammel, C. S. & Noller, H. F. (1995). *Genes Dev.* **9**, 626–637.
Ferre-D'Amare, A. R., Zhou, K., & Doudna, J. A. (1998). *Nature (London)*, **395**, 567–574.
Grill, S., Gualerzi, O., Londei, P. & Blasi, U. (2000). *EMBO J.* **19**, 4101–4110.
Hegde, R. S., Grossman, S. R., Laimins, L. A. & Sigler, P. B. (1992). *Nature (London)*, **359**, 505–512.
Holm, L. & Sander, C. (1993). *J. Mol. Biol.* **233**, 123–138.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Laskowski, R. A. (1995). *J. Mol. Graph.* **13**, 323–330.
Lawson, C. L., Zhang, R.-G., Otwinowski, Z., Marmorstein, R. Q., Schevitz, R. W., Luisi, B., Joachimiak, A. & Sigler, P. B. (1988). *Proc. 39th Mosbacher Colloquium*, **369**, 202.
Linial, M. & Yona, G. (2000). *Prog. Biophys. Mol. Biol.* **73**, 297–320.
Mah, T.-F., Li, J., Davidson, A. R. & Greenblatt, J. (1999). *Mol. Microbiol.* **34**, 523–537.
Murshudov, G. N., Lebedev, A., Vagin, A. A., Wilson, K. S. & Dodson, E. J. (1999). *Acta Cryst.* D**55**, 247–255.
Nicholls, A., Bharadwaj, R. & Honig, B. (1993). *Biophys. J.* **64**, 166–170.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **277**, 269–305.
Overbeek, R., Larsen, N., Pusch, G. D., D'Souza, M., Selkov, E., Kyrpides, N., Fonstein, M., Maltsev, N. & Selkov, E. (2000). *Nucleic Acids Res.* **28**, 123–125.
Pearl, F. M. G., Lee, D., Bray, J. E., Sillitoe, I., Todd, A. E., Harrison, A. P., Thornton, J. M. & Orengo, C. A. (2000). *Nucleic. Acids Res.* **28**, 277–282.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Portugaly, E. & Linial, M. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 5161–5166.
Sayle, R. A. & Milner-White, E. J. (1995). *Trends Biochem. Sci.* **20**, 374–376.
Shazand, K., Tucker, J., Grunberg-Manago, M., Rabinowitz, J. C. & Leighton, T. (1993). *J. Bacteriol.* **175**, 2880–2887.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Tatusov, R. L., Natale, D. A., Garkavtsev, I. V., Tatusova, T. A., Shankavaram, U. T., Rao, B. S., Kiryutin, B., Galperin, M. Y., Fedorova, N. D. & Koonin, E. V. (2001). *Nucleic Acids Res.* **29**, 22–28.
Teplova, M., Tereshko, V., Sanishvili, R., Joachimiak, A., Bushueva, T., Anderson, W. F. & Egli, M. (2000). *Protein Sci.* **12**, 2557–2566.
Vitkup, D., Melamud, E., Moult, J. & Sander, C. (2001). *Nature Struct. Biol.* **8**, 559–66.
Vogel, U. & Jensen, K. F. (1997). *J. Biol. Chem.* **272**, 12265–12271.
Walsh, M. A., Dementieva, I., Evans, G., Sanishvili, R. & Joachimiak, A. (1999). *Acta Cryst.* D**55**, 1168–1173.