Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of the putative mutarotase YeaD from *Salmonella typhimurium*: structural comparison with galactose mutarotases

Sagar Chittori,[a] Dhirendra K. Simanshu,[a] H. S. Savithri[b] and M. R. N. Murthy[a]*

[a]Molecular Biophysics Unit, Indian Institute of Science, Bangalore, India, and [b]Department of Biochemistry, Indian Institute of Science, Bangalore, India

Correspondence e-mail: mrn@mbu.iisc.ernet.in

Received 26 July 2006
Accepted 2 November 2006

PDB References: YeaD, orthorhombic, 2hta, r2htasf; monoclinic, 2htb, r2htbsf.

Salmonella typhimurium YeaD (*st*YeaD), annotated as a putative aldose 1-epimerase, has a very low sequence identity to other well characterized mutarotases. Sequence analysis suggested that the catalytic residues and a few of the substrate-binding residues of galactose mutarotases (GalMs) are conserved in *st*YeaD. Determination of the crystal structure of *st*YeaD in an orthorhombic form at 1.9 Å resolution and in a monoclinic form at 2.5 Å resolution revealed this protein to adopt the β-sandwich fold similar to GalMs. Structural comparison of *st*YeaD with GalMs has permitted the identification of residues involved in catalysis and substrate binding. In spite of the similar fold and conservation of catalytic residues, minor but significant differences were observed in the substrate-binding pocket. These analyses pointed out the possible role of Arg74 and Arg99, found only in YeaD-like proteins, in ligand anchoring and suggested that the specificity of *st*YeaD may be distinct from those of GalMs.

1. Introduction

The data on protein sequences represented by the open reading frames (ORF) of a large number of organisms is increasing exponentially as a consequence of several completed and ongoing genome-sequencing projects. Many of the gene products encoded by these sequences have been identified in various organisms. The functions of several of these gene products have been deduced from comparative sequence analysis. However, it has not been possible to assign specific functions to a large fraction of the predicted gene products because of the absence of reliable sequence similarity to proteins of known function. Such functionally unassigned proteins are classified as hypothetical proteins. Some hypothetical proteins are well conserved among many organisms and presumably perform a common biological role that is yet to be defined. Other hypothetical proteins are specific to only one or few organisms, where they presumably fulfil a specialized role. A more powerful means of functional characterization is determination of the three-dimensional structure, as protein function is intimately coupled to structure.

Mutarotases catalyze the interconversion of the α- and β-anomers of sugars (Hucho & Wallenfels, 1971; Bailey *et al.*, 1975; Wurster & Hess, 1972; Shirokane & Suzuki, 1995; Beebe & Frey, 1998; Assairi *et al.*, 2004). Galactose mutarotases (GalMs) are known to possess a characteristic motif (NS)-*x*-T-N-**H**-*x*-Y-(FW)-N-(LI) centred on a conserved histidine [shown in bold; Prosite PS00545; where (*XY*) means either *X*

or *Y* and *x* refers to any amino acid; Hulo *et al.*, 2006; de Castro *et al.*, 2006]. A search for this motif in the complete genome of *Escherichia coli* and *Salmonella typhimurium* revealed that these genomes possess an ORF in addition to *galM*, called *yeaD* (alternatively *yzzQ*), annotated as an aldose 1-epimerase. The stretch of sequence T-S-A-L-**H**-S-Y-F-N-V (amino acids 160–169), resembling the characteristic motif of mutarotases, is present in *S. typhimurium* YeaD (*st*YeaD), suggesting that YeaD may also belong to the galactose mutarotase-like superfamily (SCOP unique identifier code 74650; Murzin *et al.*, 1995). The presence of YeaD in *S. typhimurium* in addition to GalM is enigmatic. *st*YeaD is a protein of 294 residues with a theoretical molecular weight of 32.5 kDa. This report describes the crystal structure of *st*YeaD and presents initial efforts towards structural analysis of YeaD-like proteins in order to annotate their specific function. The structure and active-site architecture of *st*YeaD were found to be similar to those of GalMs, suggesting that this protein might also be a mutarotase.

2. Experimental procedures

2.1. Sequence analysis

The sequence homologues of *st*YeaD were identified from the SWISS-PROT database of protein sequences using the *BLASTP* program (Altschul *et al.*, 1997). The sequences of *Homo sapiens* GalM (*hs*GalM), *Saccharomyces cerevisiae* GalM domain (*sc*GalM), *Caenorhabditis elegans* GalM (*ce*GalM) and *Lactococcus lactis* GalM (*lac*GalM) were extracted from the PDB (Berman *et al.*, 2002). Multiple sequence alignment including YeaD and GalM homologues was achieved using the *T-Coffee* server (Notredame *et al.*, 2000). The alignment was further analyzed by the Evolutionary Trace Server (*TraceSuite* II; Innis *et al.*, 2000; Lichtarge *et al.*, 1996) and clustered into two groups. The *ESPript* server was used to produce a representation of the alignment (Gouet *et al.*, 2003).

2.2. Cloning and purification of the protein

The 782 bp ORF encoding the putative aldose 1-epimerase (*st*YeaD) was PCR-amplified from *S. enterica* serovar Typhimurium strain IFO12529 genomic DNA as the template using high-fidelity KOD HiFi DNA polymerase (Novagen), a sense primer (5′-**GCT AGC** CAT ATG ATT AAT AAA ATT TTT GCC CTC C-3′) containing an *Nhe*I site (shown in bold) and an antisense primer (5′-**GGA TCC** TTA CTC GAG GCG TTT AGC GAC GCA GAT AG-3′) containing a *Bam*HI site (shown in bold). The PCR-amplified fragment was digested with *Nhe*I and *Bam*HI and cloned into pRSET-C vector (Invitrogen) previously digested with the same restriction enzymes. The sequence of the recombinant *styeaD* clone obtained was determined by nucleotide sequencing and confirmed by comparing it with the sequence of the corresponding gene in *S. typhimurium* LT2. The final plasmid construct encodes *st*YeaD polypeptide with an additional 15 amino acids from the vector. These additional residues contained a hexahistidine tag at the N-terminus to facilitate protein purification using Ni–NTA affinity column chromatography.

The recombinant plasmid was transformed into *E. coli* BL21(DE3)pLysS cells and transformants were selected on LB agar plates containing 100 µg ml^{-1} ampicillin. A single colony was picked and 20 ml primary culture was grown overnight at 310 K followed by inoculation of 1 l Terrific Broth (Hi-media) containing 2 ml glycerol until the OD at 600 nm reached 0.6. Protein expression was then induced by the addition of 0.3 m*M* isopropyl β-D-thiogalactopyranoside (IPTG). The culture was further incubated at 303 K for 4 h. Cells were harvested by centrifugation, resuspended in lysis buffer containing 50 m*M* Tris pH 8.0 and 200 m*M* NaCl and lysed by sonication on ice. All the following purification steps were performed at 277 K. The supernatant obtained after centrifugation of the cell lysate was applied onto an Ni–NTA affinity column (His-bind resin, Novagen) equilibrated with lysis buffer. Nonspecifically bound proteins were washed from the column using 50 ml lysis buffer containing 5 m*M* imidazole. The recombinant protein was eluted with 10 ml lysis buffer containing 100 m*M* imidazole. To remove imidazole and excess salt, the protein was dialysed for 24 h against 25 m*M* Tris pH 8.0 containing 100 m*M* NaCl. The protein was concentrated using a 10 kDa molecular-weight cutoff Centricon (Amicon). The typical yield of the protein was 10 mg per litre of *E. coli* culture as determined by measurement of the absorbance at 280 nm using the theoretically calculated molar extinction coefficient (E_{280}) of 1.420 (http://www.expasy.org/tools/protparam.html).

2.3. Initial characterization of the protein

The purity and molecular weight of the eluted protein were checked by 12% SDS–PAGE analysis. For gel-permeation chromatography experiments, 200 µl of 2 mg ml^{-1} protein solution was loaded onto a Superdex S-200 column (exclusion limit 600 kDa) with a bed volume of 24 ml and a void volume of 8 ml. 25 m*M* Tris pH 8.0 containing 100 m*M* NaCl was used as the equilibration buffer at a flow rate of 0.2 ml min^{-1}. 1 ml fractions were collected from the column and the absorbance was monitored at 280 nm. The molecular weight of *st*YeaD was determined using β-amylase (200 kDa), alcohol dehydrogenase (150 kDa), bovine serum albumin (66 kDa), carbonic anhydrase (29 kDa) and cytochrome *c* (12.4 kDa) as standards. Dynamic light-scattering (DLS) measurements on *st*YeaD were performed on a DynaPro DLS instrument using 20 µl of 2 mg ml^{-1} *st*YeaD with a data-acquisition time of 10 s. The protein sample was filtered through a 0.2 µm filter before measurements. *DYNAMICS* v.6 software was used to calculate the hydrodynamic radius of *st*YeaD. Circular-dichroism (CD) measurements for *st*YeaD were recorded at a protein concentration of 0.1 mg ml^{-1} in a buffer containing 25 m*M* Tris pH 8.0 and 100 m*M* NaCl using a Jasco J715 spectropolarimeter.

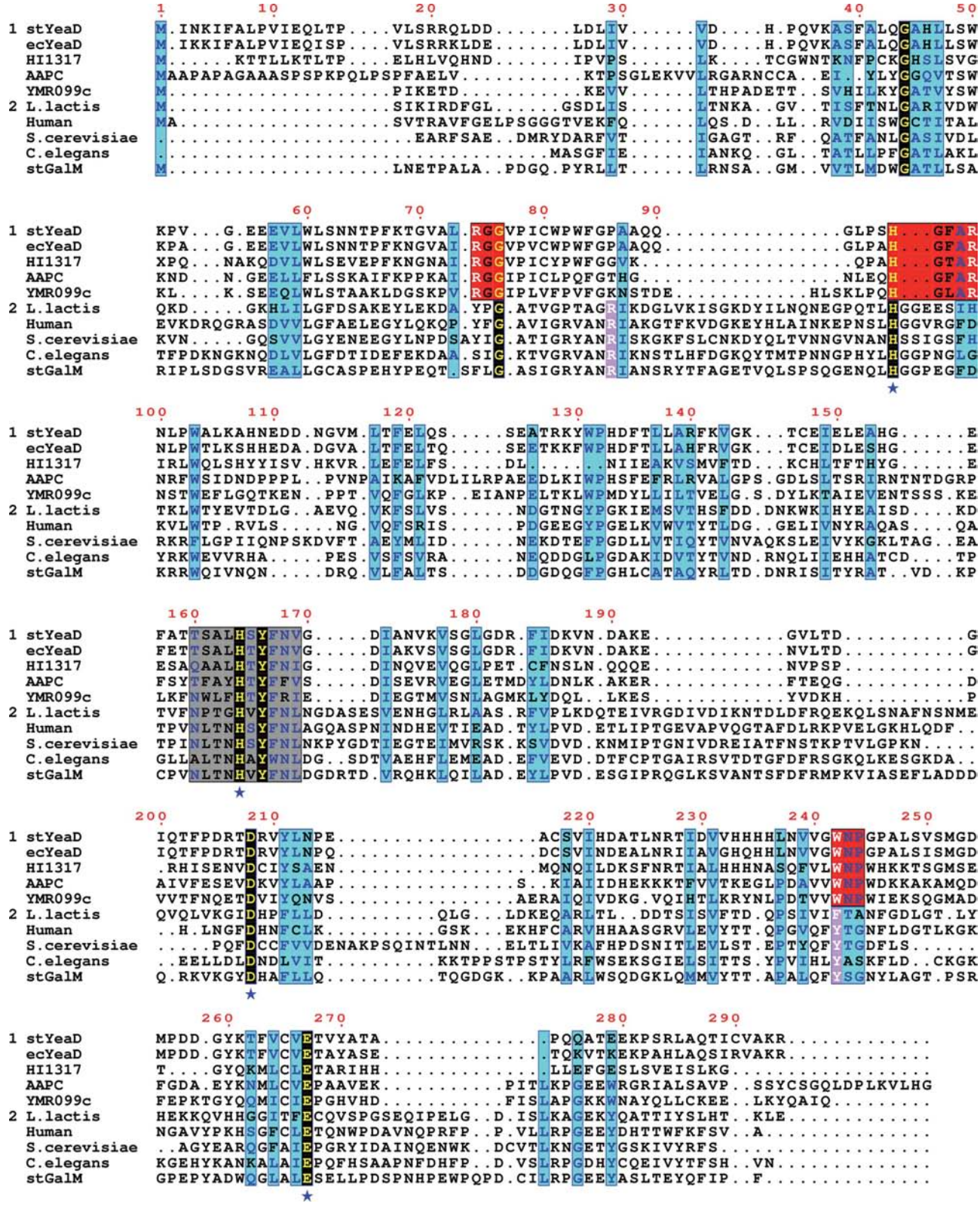

Figure 1
Representation of the multiple sequence alignment generated using the *ESPript* server (Gouet *et al.*, 2003). Sequences stYeaD (*S. typhimurium* putative mutarotase), ecYeaD (*E. coli* putative mutarotase), HI1317 (*H. influenzae* hypothetical UPF0010 protein), AAPC [*Pennisetum ciliare* (buffelgrass) putative apospory-associated protein C] and YMR099c (Yeast-hypothetical UPF0010 protein) constitute the group of YeaD-like proteins (group 1) while human (hsGalM), *S. cerevisiae* (scGalM), *C. elegans* (ceGalM), L. lactis (lacGalM) and stGalM (*S. typhimurium* GalM) constitute the GalM group of proteins (group 2). Residue colour code: yellow, identical in both groups; blue, similar residues in both groups; white, active-site residues in one group. Background colour code: red, motif representation; cyan, similar residues in both the groups; black, identical residues in both the groups. Catalytic residues are indicated by a star below the alignment.

2.4. Protein crystallization and X-ray diffraction data collection

Crystallization experiments on native *st*YeaD were carried out by the hanging-drop vapour-diffusion method. The crystallization droplet contained 3 µl of 10 mg ml^{-1} protein solution and 3 µl crystallization cocktail. Two morphologically distinct crystal forms, orthorhombic and monoclinic, were obtained using ammonium sulfate and PEG 8000, respectively, as precipitants (Table 1). In both experiments, crystals appeared within 5 d and grew to a size suitable for data collection in 15 d. Cocrystallization of *st*YeaD with various ligands (glucose, galactose, mannose, fructose, fucose, xylose, lyxose, lactose, maltose, sucrose and glucose 6-phosphate) was attempted. In these trials, protein was incubated prior to crystallization with an excess of ligand (1:10 molar ratio) for 12 h at 277 K.

X-ray diffraction data for the orthorhombic crystal were collected at beamline BL44XU at SPring8, Hyogo, Japan at a wavelength of 0.9 Å using a DIP2040b detector. The monoclinic data were collected at a wavelength of 1.5418 Å on a Rigaku RU-200 rotating-anode X-ray generator equipped with a 300 µm focal cup. The images were recorded using a MAR 345 image-plate detector. Both data sets were collected at 100 K from a single crystal mounted on a cryoloop. Data sets were processed using *DENZO* and scaled using *SCALEPACK* from the *HKL*-2000 suite (Otwinowski & Minor, 1997).

Table 1
Crystallization details and data-collection statistics.

Values in parentheses refer to the highest resolution shell.

	Monoclinic	Orthorhombic
Crystallization conditions	15–25% PEG 8K, 0.1 *M* Na citrate pH 5.5, 0.2–0.5 m*M* NaCl	1.5–2 *M* ammonium sulfate, 25 m*M* Mg formate, 0.1 *M* Na cacodylate pH 5.5
Resolution range (Å)	30.0–2.50 (2.59–2.50)	50.0–1.90 (1.97–1.90)
Space group	*C*2	$P2_12_12_1$
Unit-cell parameters		
a (Å)	47.59	49.75
b (Å)	138.08	88.77
c (Å)	219.55	179.29
β (°)	91.816	
Protomers in the ASU	4	2
Observed reflections	1117646	871675
Unique reflections	50310	61215
Data completeness	99.0 (91.9)	95.9 (93.8)
Data redundancy	22.22	14.2
$I/\sigma(I)$	16.4 (2.3)	22.4 (4.3)
$R_{\rm merge}$† (%)	8.5 (44.3)	7.5 (13.7)
Matthews coefficient (Å^3 Da^{-1})	2.73	2.77
Solvent content (%)	54.52	55.21

† $R_{\rm merge} = \sum |I_{hj} - \langle I_h \rangle| / \sum I_{hj}$, where I_{hj} is the jth measurement of the intensity of reflection h and $\langle I_h \rangle$ is its mean value.

2.5. Structure solution and refinement

The structure of orthorhombic *st*YeaD was solved by molecular replacement with *AMoRe* (Navaza, 1994) using the atomic coordinates of a hypothetical protein, HI1317 from *Haemophilus influenzae* (PDB code 1jov; N. Bonander, M. Tordova, A. J. Howard, E. Eisenstein & G. Gilliland, unpublished results), as the search model. A monomer of orthorhombic *st*YeaD was used as the search model for the structure solution of the monoclinic form. Both structures were refined using *REFMAC*5 (Murshudov *et al.*, 1997). The protein models were improved by visual inspection and manual model building using the graphics program *Coot* (Emsley & Cowtan, 2004). The progress of refinement was monitored by calculation of $R_{\rm free}$ using 5% of the total independent reflections. Stereochemical qualities of the models were verified using *PROCHECK* (Laskowski *et al.*, 1993).

2.6. Structural analysis of *st*YeaD

PROMOTIF was used to estimate the secondary-structure content of the protein (Hutchinson & Thornton, 1996). The Protein Interfaces, Surfaces and Assemblies (*PISA*) server (Krissinel & Henrick, 2005) was used for calculation of buried surface areas. The program *CONTACT* from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994) was used for the identification of hydrogen bonds and salt bridges. Polar atoms within a distance range of 2.5–3.5 Å with proper hydrogen-bond geometry were considered to form hydrogen bonds, while oppositely charged atoms within 4 Å distance were considered to be salt bridges. Structural alignment was carried out using the program *ALIGN* (Cohen, 1997).

3. Results and discussion

3.1. Sequence analysis of *st*YeaD

Four homologous sequences to *st*YeaD were identified in the SWISS-PROT database: *ec*YeaD, HI1317, AAPC and YMR099c (together referred to in the following as 'YeaD-like proteins') from *E. coli*, *H. influenzae*, *Pennisetum ciliare* (buffelgrass) and *Saccharomyces cerevisiae*, which share sequence identities of 82, 27, 26 and 22% with *st*YeaD, respectively. All these proteins have been annotated as enzymes related to aldose 1-epimerase. Of these four homologues, only HI1317 has had its crystal structure determined (PDB code 1jov; N. Bonander, M. Tordova, A. J. Howard, E. Eisenstein & G. Gilliland, unpublished results). Multiple sequence alignment including the YeaD-like proteins and GalMs (*Lactococcus lactis* GalM, *lac*GalM; *Homo sapiens* GalM, *hs*GalM; *S. cerevisiae* GalM domain, *sc*GalM; *C. elegans* GalM, *ce*GalM; *S. typhimurium* GalM, *st*GalM) showed the conservation of the aldose 1-epimerase (mutarotase) motif (NS)-*x*-T-N-**H**-*x*-Y-(FW)-N-(LI) in all the proteins with slight variations (Fig. 1, grey background). The histidine shown in bold has been shown to be involved in the catalytic mechanism of GalMs (Beebe & Frey, 1998; Thoden *et al.*, 2003). The alignment revealed three consensus motifs R-G-G, [H-G-(F,T,L)-A-R] and W-N-P, shown in Fig. 1 with red background, to be present in YeaD-like proteins. Although no long stretch of significant sequence identity could be found among these proteins, the alignment showed the conservation of several residues: Gly44, Gly76, His95, His164, Tyr166, Asp208 and

Glu267 (*st*YeaD numbering; Fig. 1). Of these residues, His164 and Glu267 correspond to the strictly conserved catalytic residues of GalMs (*i.e.* His170 and Glu304 of *lac*GalM; Fig. 1), while His95 and Asp208 of *st*YeaD were found to be equivalent to His96 and Asp243 of *lac*GalM, residues which are implicated as playing a critical role in ligand binding in *lac*GalM (Thoden & Holden, 2002; Thoden *et al.*, 2002). Therefore, the catalytic machinery in *st*YeaD may be considered to be equivalent to that of *lac*GalM and hence it is likely that YeaD-like proteins also catalyze the mutarotation reaction by a mechanism similar to that of GalM. This analysis showed that though there is no significant identity in the primary structure of YeaD and GalM (Table 3), they are likely to be functionally related proteins.

3.2. Initial characterization of the protein

The *yeaD* gene from *S. typhimurium* was cloned in an IPTG-inducible vector, overexpressed in *E. coli* and purified by a one-step protocol using Ni–NTA affinity chromatography as described in §2.2. SDS–PAGE analysis revealed a single band with a molecular weight of 34 kDa, which agreed with the theoretically calculated molecular weight from the sequence including the additional amino acids containing the hexahistidine tag (http://www.expasy.org/tools/protparam.html). Gel-permeation chromatography measurements on a Superdex S-200 column showed a peak corresponding to 34 kDa, indicating that the protein is predominantly a monomer in solution. DLS measurements showed monodisperse distribution (polydispersity value = 13.4%) and hence the protein was considered to be suitable for crystallization. The estimated hydrodynamic radius (R_h) of the protein was 2.8 nm, corresponding to a molecular weight of the *st*YeaD monomer consistent with gel-filtration experiments. A trough at 215 nm and a minor dip at 225 nm in the CD spectrum of *st*YeaD indicated a predominantly β-strand structure with a low α-helical content.

The enzyme activity was monitored by estimating the rate of conversion of α-D-glucose and α-D-glucose 6-phosphate to their β-anomers in a coupled reaction catalyzed by β-D-glucose dehydrogenase and β-D-glucose 6-phosphate dehydrogenase, respectively, by following the rate of formation of NADH spectrophotometrically at 340 nm. However, consistent values for the activity of the enzyme could not be obtained as the variations observed in the catalytic activities of protein purified in different batches were too large.

3.3. Crystal structure of *st*YeaD

The Matthews coefficients (Matthews, 1968) for the monoclinic and orthorhombic forms of *st*YeaD were 2.91 and 2.7 Å^3 Da^{-1}, respectively, assuming the presence of four and two monomers in the respective asymmetric units. Table 1 shows the crystallization details and data-collection statistics for both crystal forms. The best solution obtained by molecular replacement for the orthorhombic form had a correlation coefficient of 0.401 and an R factor of 48.1% with no short contacts between symmetry-related molecules. Building of side chains followed by the addition of water molecules resulted in a final R_{work} of 18.5% and R_{free} of 22.2% for the orthorhombic form and an R_{work} of 18.8% and an R_{free} of 23.9% for the monoclinic form. The bond lengths and angles in the final models of the orthorhombic and monoclinic forms agree well with the standard values of these parameters and the main-chain torsion angles correspond well to the expected values (Table 2). In both crystal forms, the electron density accounts for most of the protein. The electron density is less well defined for a few exposed loops. 87.4% and 84.8% of the residues in the orthorhombic and monoclinic forms, respectively, are in the most favoured regions of the Ramachandran map. In both structures, Asp24, which has a well defined electron density, is present in the disallowed region. The carbonyl O atom of Asp24 makes a strong hydrogen bond with the NE of Gln22, while the main-chain N atom of Asp24 is hydrogen bonded to the carboxylate of its own side chain. The side chain of Asp24 is also involved in a salt bridge with Lys68 and a hydrogen bond with a water molecule. These interactions appear to compensate for the strain that arises owing to the disallowed Ramachandran angles. Another interesting observation was the *cis*-peptide found between Trp81 and Pro82. This *cis*-peptide bond is conserved in the known structures of both YeaD-like proteins and GalMs, suggesting a plausible structural or functional role of this *cis*-peptide.

Table 2
Refinement statistics.

Parameter	Monoclinic	Orthorhombic
No. of atoms		
Total	9434	5094
Protein	9112	4550
Water	322	528
SO_4	—	10
Glycerol	—	6
R factors (%)		
Overall	19.1	18.7
R_{work}	18.8	18.5
R_{free}	23.9	22.2
Model quality		
Average B factors (Å^2)		
Overall (r.m.s.d.)	32.47 (1.80)	26.21 (1.35)
Protein	33.17	25.44
Water	32.47	35.42
R.m.s. deviations		
Bond lengths (Å)	0.028	0.019
Bond angles (°)	2.314	1.705
Chiral centre restraints	0.157	0.135
General planes (Å)	0.010	0.008
Residues in Ramachandran map		
Most favoured region	868 (84.8%)	449 (87.4%)
Allowed region	150 (14.7%)	59 (11.5%)
Generously allowed region	3 (0.3%)	1 (0.2%)
Disallowed region	2 (0.2%)	5 (1.0%)

PROMOTIF analysis (Hutchinson & Thornton, 1996) showed that *st*YeaD consists mainly of β-strands (45.6%; 136 residues), with a low α-helical (7%; 22 residues) and 3_{10}-helical content (3%; nine residues) (Fig. 2*a*), consistent with the CD spectral results. The polypeptide chain folds into three antiparallel β-sheets (consisting of a total of 22 β-strands of length 2–15 amino acids), six helices (of length 3–10 amino acids) and approximately 25 loop regions containing 28

β-turns, ten β-hairpins and six β-bulges (Figs. 2*a* and 2*b*). The three antiparallel β-sheets (Fig. 2*b*) constitute the core of *st*YeaD. The core consists of a β-sandwich fold with overall dimensions of approximately 55 × 42 × 52 Å (Fig. 2*a*). The root-mean-square deviation (r.m.s.d.) values of corresponding C^{α} atoms after structural superposition of the monomers in the same crystal form as well as monomers from the two different crystal forms were in the range 0.24–0.31 Å, indicating that there are no significant conformational differences between the two crystal forms. Therefore, for further structural analysis only subunit *A* of the orthorhombic form was considered.

3.4. Oligomeric state of the protein

*st*YeaD was found to be a monomer in solution by gel-permeation chromatography and DLS experiments. However, two monomers *A* and *B* related by a noncrystallographic twofold axis were found in the asymmetric unit of the orthorhombic form, while two pairs of such monomers (*AB* and *CD*) were observed in the monoclinic crystal form. Inspection of the *A*–*C* and *A*–*D* interfaces of the monoclinic form indicated that the interactions were not extensive and thus they were completely ruled out as dimer interfaces. Superposition of the *AB* or *CD* dimers from the two crystal forms showed no significant differences (r.m.s. deviations of 0.42–0.69 Å) in the corresponding C^{α} positions. This suggests that the *A*–*B* and *C*–*D* dimer interfaces are relatively stable and thus that the protein could function as a dimer. However, the buried surface area (~4.08% of the total solvent-accessible surface area) at the interface is small compared with those of other stable dimeric proteins. The *A*–*B* and *C*–*D* interfaces contain seven or eight weak hydrogen bonds (formed by Ile11, Asp27, His108, Gln13 and Leu105 of one monomer with the same region of the other monomer), five salt bridges of higher range and a few water-mediated hydrogen bonds.

For further examination of the oligomeric status of *st*YeaD, gel-permeation chromatography experiments were repeated with 10 mg ml^{-1} protein solution pre-incubated for 24 h with the crystallization cocktail. Under these conditions, a dimeric

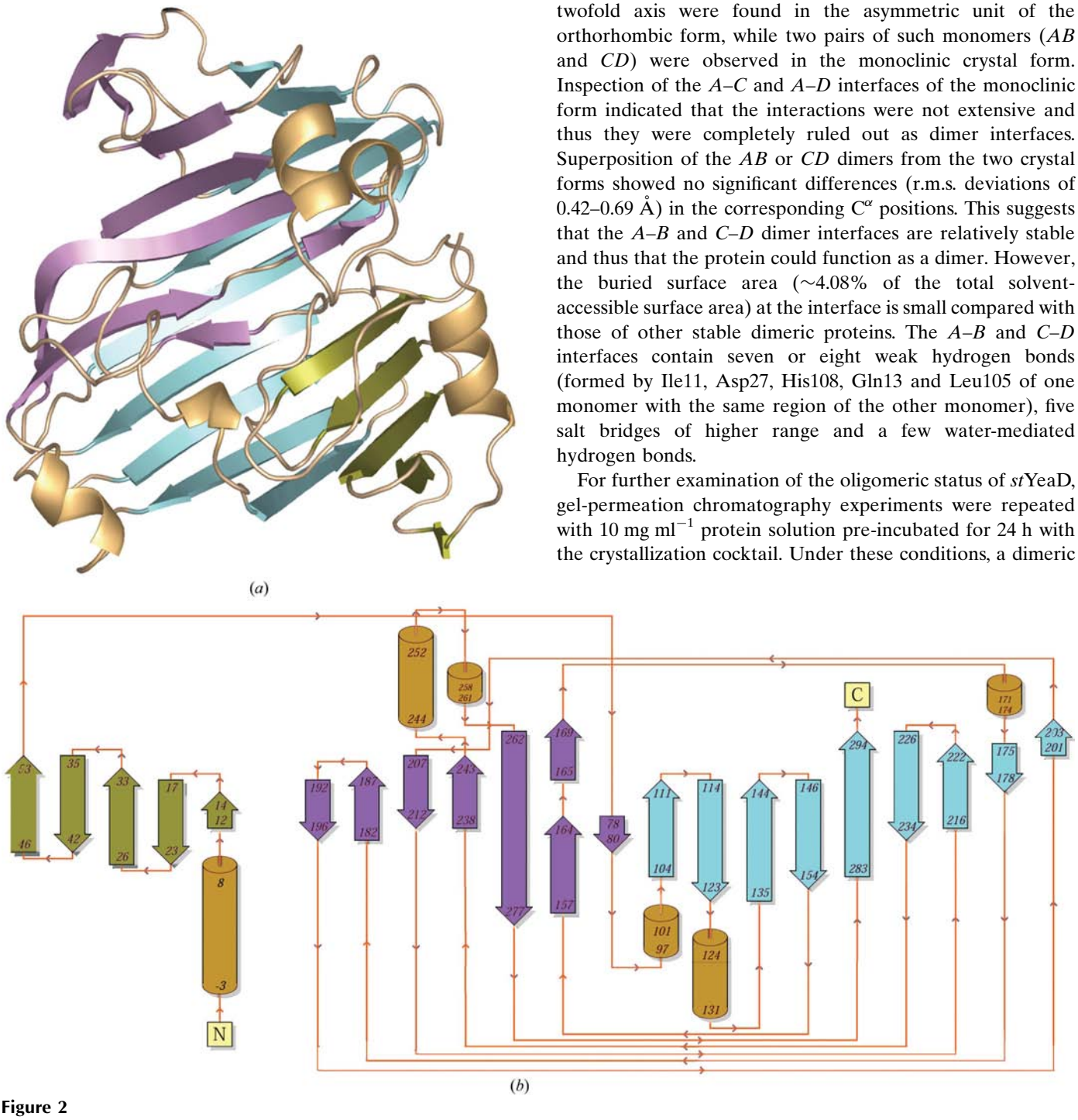

Figure 2
(*a*) Ribbon diagram illustrating the tertiary structure of *st*YeaD. This figure was generated using *PyMOL* (DeLano, 2002). (*b*) A topology diagram of *st*YeaD generated using the *PDBsum* server (Laskowski *et al.*, 2005). Colour code: lime green, sheet *A*; violet, sheet *B*; cyan, sheet *C*; light orange, helix; wheat, loop.

Table 3
R.m.s.d. value/sequence identity (%) obtained after structural and sequence alignments.

Values in parentheses are the number of superposed C^{α} atoms in the alignment.

Protein†	*st*YeaD	HI1317	*lac*GalM	*hs*GalM	*sc*GalM	*ce*GalM
*st*YeaD	0.00/100 (294)	0.88/33 (227)	2.48/10 (249)	2.62/8 (248)	2.51/12 (247)	2.42/6 (238)
HI1317	0.88/33 (227)	0.00/100 (269)	2.49/7 (226)	2.54/9 (222)	2.58/4 (231)	2.36/7 (213)
*lac*GalM	2.48/10 (249)	2.49/7 (226)	0.00/100 (339)	1.19/28 (307)	1.39/22 (275)	1.26/27 (290)
*hs*GalM	2.62/8 (248)	2.54/7 (222)	1.19/28 (307)	0.00/100 (342)	1.23/28 (284)	1.04/30 (301)
*sc*GalM	2.51/12 (247)	2.58/4 (231)	1.39/22 (275)	1.23/28 (284)	0.00/100 (339)	1.39/25 (279)
*ce*GalM	2.42/6 (238)	2.36/7 (213)	1.26/27 (290)	1.04/30 (301)	1.39/25 (279)	0.00/100 (329)

† *st*YeaD, *S. typhimurium* putative mutarotase; HI1317, *H. influenzae* hypothetical UPF0010 protein; *lac*GalM, *L. lactis* GalM; *hs*GalM, *H. sapiens* GalM; *sc*GalM, *S. cerevisiae* GalM domain; *ce*GalM, *C. elegans* GalM.

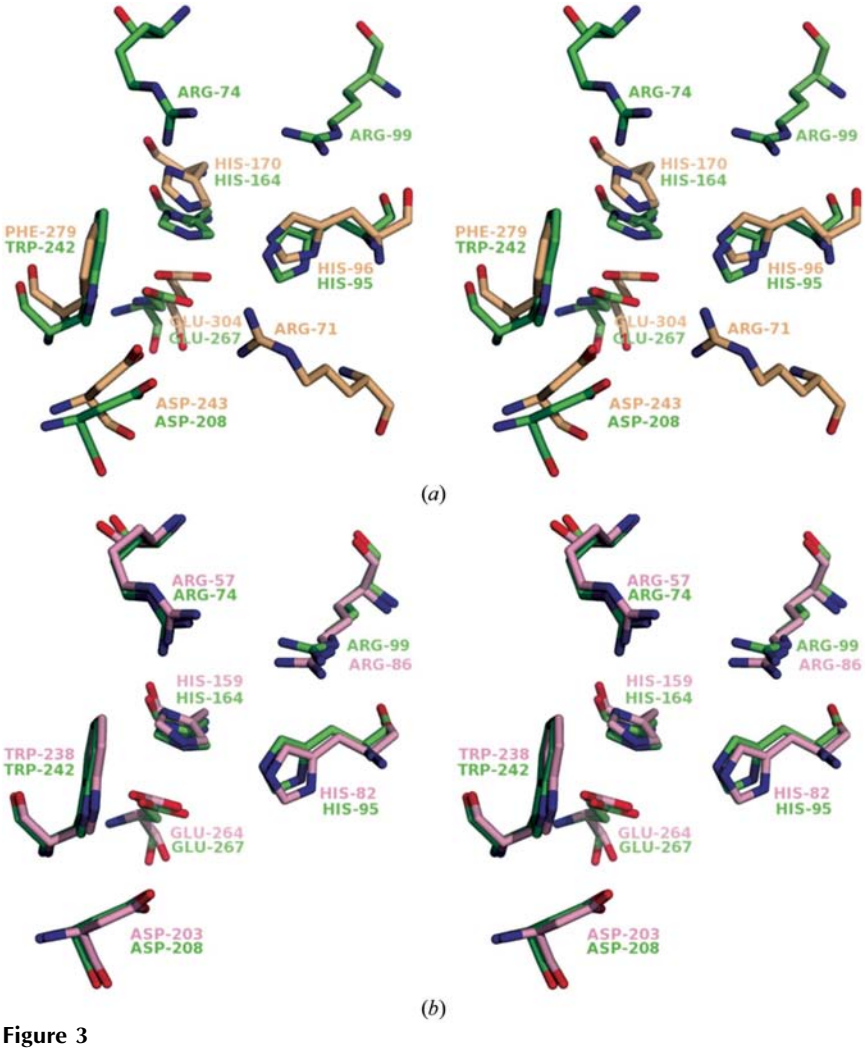

Figure 3
Superposition of the active-site residues of (*a*) *st*YeaD and *lac*GalM and (*b*) *st*YeaD and YMR099cp. Colour code: green, *lac*GalM; wheat, *st*YeaD; pink; YMR099cp.

state of the protein was observed. DLS experiments also indicated an increase in the hydrodynamic radius from the usual 2.8 nm (equivalent to the YeaD monomer) for 10 mg ml^{-1} protein in 25 m*M* Tris pH 8.0 containing 100 m*M* NaCl to 3.7 nm for the protein incubated with the crystallization cocktail in a 1:1 ratio. The larger hydrodynamic radius corresponds to a dimeric form of the enzyme. Thus, the observed dimeric structure in both crystal forms is most probably a consequence of oligomerization induced during crystallization.

3.5. Structural comparisons of *st*YeaD

Superposition of *st*YeaD against a database of structures using the *DALI* server (Holm & Sander, 1993) shows similarities to proteins containing a β-sandwich fold. The structural domains identified by the server include HI1317, *lac*GalM, a hypothetical 37.9 kDa protein bio3-hxt17, domain 5 of β-galactosidase, copper amine oxidase, chondroitinase, hyaluronate lyase and maltose phosphorylase. Further inspection of the *DALI* results shows that the structures close to *st*YeaD (other than HI1317) belong to the GalMs. The conservation of the mutarotase motif and the closer structural similarity to GalMs than to any other proteins containing β-sandwich folds suggest that *st*YeaD is a probable mutarotase. For further characterization, the C^{α} traces of all structurally known GalMs and HI1317 were superposed on *st*YeaD using the program *ALIGN* (Cohen, 1997). These superpositions showed structural conservation for the core region, although significant deviations were observed in the helical and loop regions. For further evaluation, pairwise structural superpositions were performed for *st*YeaD, HI1317, *hs*GalM, *sc*GalM, *ce*GalM and *lac*GalM. The range of pairwise r.m.s.d. values found within the YeaD-like proteins or GalMs was less than 1.40 Å, while the r.m.s.d. values across YeaD-like proteins and GalMs were greater than 2.35 Å (Table 3), suggesting that YeaD and GalM belong to different groups of mutarotases. Major deletions in YeaD-like proteins compared with GalMs include the regions corresponding to residues 73–90 (containing two small β-strands and a helix), 214–239 (containing two small helices) and 309–318 (containing a small

helix) of *lac*GalM. There is also an insertion observed at the N-terminus of YeaD-like proteins compared with GalMs. Therefore, it is likely that the higher r.m.s.d. values between the two families reflect insertions/deletions as well as flexibility of the polypeptides which are not part of the core region. These rearrangements might be an adaptation required to perform the catalytic activity on specific substrates. Significant sequence similarity in the core structure coupled with the excellent agreement in the core region of YeaD-like proteins and GalMs suggest that these proteins are products of function-driven divergent evolution, despite their low overall sequence identity.

3.6. Active-site pocket and ligand building

As the catalytic and substrate-positioning residues of GalM (His96, His170, Asp243 and Glu304 of *lac*GalM, corresponding to His95, His164, Asp208 and Glu267 of *st*YeaD) appear to be conserved in YeaD-like proteins (Fig. 1), the putative active site of *st*YeaD was identified by structural superposition. Although the catalytic (His170 and Glu304) and substrate-positioning (His96 and Asp243) residues of *lac*GalM do not exactly coincide with the corresponding catalytic (His164 and Glu267) and substrate-positioning (His95 and Asp208) residues of *st*YeaD, their relative geometry remains invariant (Fig. 3*a*). These residues showed excellent superposition when the alignment was repeated by considering only their coordinates (data not shown). An unexplained but similar electron density was observed at the active site of native *st*YeaD in both crystal forms. None of the compounds that could originate from purification (Tris, NaCl or imidazole), the crystallization cocktail (ammonium sulfate, formate or cacodylate) or the cryoprotectant (glycerol) were suitable to account for the observed density. The size and shape of the density was strongly suggestive of a bound sugar. However, significant density was not present for the hydroxyl groups of sugar. This density was connected to a blob of density in the direction of Arg74 (part of the R-G-G motif) and Arg99 [part of the **H**-G-(F,T,L)-A-R motif] conserved only in YeaD-like proteins (Fig. 1). This density was suitable for an anion such as sulfate or phosphate. However, a similar blob of density was also observed in the monoclinic form, where sulfate ions were not part of the crystallization cocktail. Therefore, a reasonable explanation for the density is that it represents a phosphate group attached to the bound sugar. Therefore, the identification of the additional density as a sugar phosphate should be treated as tentative.

In mutarotases a conserved arginine residue (*e.g.* Arg71 of *lac*GalM) is found in the active site, while in the present structure there are two conserved arginines (Arg74 and Arg99). Sequentially, Arg74 of *st*YeaD is close to Arg71 of *lac*GalM. However, the spatial position of Arg74 differs from that of Arg71 (Fig. 3*a*), implying that they play different roles. Arg99 is present at a distance of approximately 6.6 Å from Arg74. Arg74 and Arg99 of *st*YeaD are at appropriate distances to make significant interactions with the putative phosphate attached to the sugar ring and hence may have important roles in ligand binding. To confirm these observations, cocrystallization of *st*YeaD was attempted with several ligands. Crystals of *st*YeaD could be obtained in the presence of a number of sugars, while an attempt to crystallize *st*YeaD in the presence of glucose-6-phosphate did not lead to diffraction-quality crystals. However, the density at the active site of crystals obtained in the presence of hexose sugars was similar to that obtained in their absence. No extra density that could be attributed to the added sugar could be located. Therefore, the added sugar presumably did not bind or displace the *in vivo* bound ligand. This observation may also account for the inconsistent results obtained in activity assays. The subtle but significant variations found at the active sites of *st*YeaD and *lac*GalM suggest that the substrate specificities of these enzymes might be different.

During the review of this manuscript, a paper describing the structure of *S. cerevisiae* protein YMR099cp related to aldose 1-epimerase was published (Graille *et al.*, 2006). This protein shares 22% sequence identity with *st*YeaD. Superposition of 244 structurally equivalent C^{α} atoms of the two polypeptides shows an r.m.s.d. value of 1.32 Å, suggesting that this protein is closer to YeaD-like proteins than to GalMs (2.2–2.6 Å for ∼240 corresponding C^{α} atoms of GalMs). Comparison of the active-site pockets of YMR099cp and *st*YeaD showed the conservation of catalytic residues. Also, two arginine residues similar to those found in *st*YeaD (Fig. 3*b*) are found in YMR099cp. The authors have demonstrated that glucose-6-phosphate and tagatose-6-phosphate (galactose-6-phosphate) bind at the active site of YMR099cp by studies of the respective complexes. These studies give additional and strong support to the interpretation of the density found in the active site of stYeaD as a hexose-6-phosphate. The discussion presented here on *st*YeaD is also applicable to the YeaD-like proteins discussed here.

4. Conclusions

To gain insight into the function of *st*YeaD, systematic sequence and structural analyses were carried out. Here, we have reported the crystal structure of *st*YeaD determined in orthorhombic and monoclinic crystal forms at 1.9 and 2.5 Å resolution, respectively. *st*YeaD possesses a fold similar to those of GalMs. Structural superposition revealed strict conservation of the catalytic residues and hence suggested that the catalytic mechanism of *st*YeaD may be similar to those of GalMs. Comparative analysis of the ligand-anchoring sites of *st*YeaD and *lac*GalM yielded information on the residues involved in conserved interactions. The additional density observed at the active site could correspond to a bound sugar phosphate. Therefore, the substrate specificity of YeaD-like proteins seems to be distinct from those of GalMs.

The intensity data were collected at the X-ray Facility for Structural Biology at the Molecular Biophysics Unit, Indian Institute of Science (IISc), supported by the Department of Science and Technology (DST) and the Department of

Biotechnology (DBT). MRNM and HSS thank the DST and the DBT of the Government of India for financial support. We thank the staff of the X-ray laboratory and Supercomputer Education and Research Centre of IISc for their cooperation during the course of these investigations. We thank Gayathri, Lokanath and the beamline staff at SPring-8 for synchrotron data collection and Subash for gel-permeation chromatography. This work was performed under the International Collaborative Research Program of the Institute of Protein Research, Osaka University.

References

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.
Assairi, L., Bertrand, T., Ferdinand, J., Slavova-Azmanova, N., Christensen, M., Briozzo, P., Schaeffer, F., Craescu, C. T., Neuhard, J., Barzu, O. & Gilles, A. M. (2004). *Protein Sci.* **13**, 1295–1303.
Bailey, J. F., Fishman, P. H., Kusiak, J. W., Mulhern, S. & Pentchev, P. G. (1975). *Methods Enzymol.* **41**, 471–484.
Beebe, J. A. & Frey, P. A. (1998). *Biochemistry*, **37**, 14989–14997.
Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K., Feng, Z., Gilliland, G. L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J. D. & Zardecki, C. (2002). *Acta Cryst.* D**58**, 899–907.
Castro, E. de, Sigrist, C. J., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P. S., Gasteiger, E., Bairoch, A. & Hulo, N. (2006). *Nucleic Acids Res.* **34**, W362–W365.
Cohen, G. E. (1997). *J. Appl. Cryst.* **30**, 1160–1161.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* http://www.pymol.org.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Gouet, P., Robert, X. & Courcelle, E. (2003). *Nucleic Acids Res.* **31**, 3320–3323.
Graille, M., Baltaze, J. P., Leulliot, N., Liger, D., Quevillon-Cheruel, S. & van Tilbeurgh, H. (2006). *J. Biol Chem.* **281**, 31075–30185.
Holm, L. & Sander, C. (1993). *J. Mol. Biol.* **233**, 123–138.
Hucho, F. & Wallenfels, K. (1971). *Eur. J. Biochem.* **23**, 489–496.
Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., De Castro, E., Langendijk-Genevaux, P. S., Pagni, M. & Sigrist, C. J. (2006). *Nucleic Acids Res.* **34**, D227–D230.
Hutchinson, E. G. & Thornton, J. M. (1996). *Protein Sci.* **5**, 212–220.
Innis, C. A., Shi, J. & Blundell, T. L. (2000). *Protein Eng.* **13**, 839–847.
Krissinel, E. V. & Henrick, K. (2005). *CompLife 2005*, edited by M. R. Berthold, R. Glen, K. Diederichs, O. Kohlbacher & I. Fischer, pp. 67–78. Berlin Heidelberg: Springer-Verlag.
Laskowski, R. A., Chistyakov, V. V. & Thornton, J. M. (2005). *Nucleic Acids Res.* **33**, D266–D268.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lichtarge, O., Bourne, H. R. & Cohen, F. E. (1996). *J. Mol. Biol.* **257**, 342–358.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995). *J. Mol. Biol.* **247**, 536–540.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Notredame, C., Higgins, D. G. & Heringa, J. (2000). *J. Mol. Biol.* **302**, 205–217.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Shirokane, Y. & Suzuki, M. (1995). *FEBS Lett.* **367**, 177–179.
Thoden, J. B. & Holden, H. M. (2002). *J. Biol. Chem.* **277**, 20854–20861.
Thoden, J. B., Kim, J., Raushel, F. M. & Holden, H. M. (2002). *J. Biol. Chem.* **277**, 45458–45465.
Thoden, J. B., Kim, J., Raushel, F. M. & Holden, H. M. (2003). *Protein Sci.* **12**, 1051–1059.
Wurster, B. & Hess, B. (1972). *FEBS Lett.* **23**, 341–344.