Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystal structure analysis of icosahedral lumazine synthase from *Salmonella typhimurium*, an antibacterial drug target

Pankaj Kumar, Mirage Singh and Subramanian Karthikeyan*

Institute of Microbial Technology, Council of Scientific and Industrial Research (CSIR), Sector 39-A, Chandigarh 160 036, India

Correspondence e-mail: skarthik@imtech.res.in, skarthik14@gmail.com

Received 12 August 2010
Accepted 20 December 2010

PDB Reference: lumazine synthase, 3mk3.

Riboflavin biosynthesis is an essential pathway in bacteria, in contrast to animals, which obtain riboflavin from their diet. Therefore, the enzymes involved in the riboflavin-biosynthesis pathway are potential targets for the development of antibacterial drugs. Lumazine synthase, an enzyme that is involved in the penultimate step of riboflavin biosynthesis, catalyzes the formation of 6,7-dimethyl-8-ribityllumazine from 3,4-dihydroxy-2-butanone 4-phosphate and 5-amino-6-ribitylamino-2,4-(1*H*,3*H*)-pyrimidinedione. Lumazine synthase from *Salmonella typhimurium* (sLS) has been cloned, overexpressed and purified and was crystallized in three forms, each with different crystal packing. The crystal structure of sLS in the monoclinic space group $P2_1$ has been determined with 60 subunits per asymmetric unit, packed as an icosahedron, at 3.57 Å resolution. Interestingly, sLS contains an N-terminal proline residue (Pro11) which had previously been suggested to disrupt the formation of the icosohedral assembly. In addition, comparison of the structure of sLS with known orthologous lumazine synthase structures allowed identification of the amino-acid residues involved in substrate binding and catalysis. The sLS structure reported here could serve as a starting point for the development of species-specific antibacterial drugs.

1. Introduction

Riboflavin (vitamin B_2) is biosynthesized in bacteria, fungi and plants, but animals are dependent on nutritional resources for this vitamin (Young, 1986; Bacher, 1991). Flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD), which are derivatives of riboflavin, are essential cofactors in living cells as they play important roles in many redox reactions, including amino-acid metabolism, DNA repair, light sensing and bioluminescence *etc.* (Meighen, 1991, 1993; O'Kane & Prasher, 1992; Briggs & Huala, 1999; Salomon *et al.*, 2001; Thompson & Sancar, 2002). Gram-negative bacteria such as *Escherichia coli* and *Salmonella* sp. are absolutely dependent on the endogenous synthesis of riboflavin (Bacher *et al.*, 1996). Therefore, the enzymes involved in the riboflavin-biosynthesis pathway can be considered to be potential antibacterial drug targets.

Lumazine synthase (LS) is involved in the penultimate step of the riboflavin-biosynthesis pathway and catalyses the condensation of 3,4-dihydroxy-2-butanone-4-phosphate (DHBP) and 5-amino-6-ribitylamino-2,4-(1*H*,3*H*)-pyrimidinedione (ARAPD) to 6,7-dimethyl-8-ribityllumazine (DMRL; Fig. 1; Volk & Bacher, 1991). The dismutation reaction of two

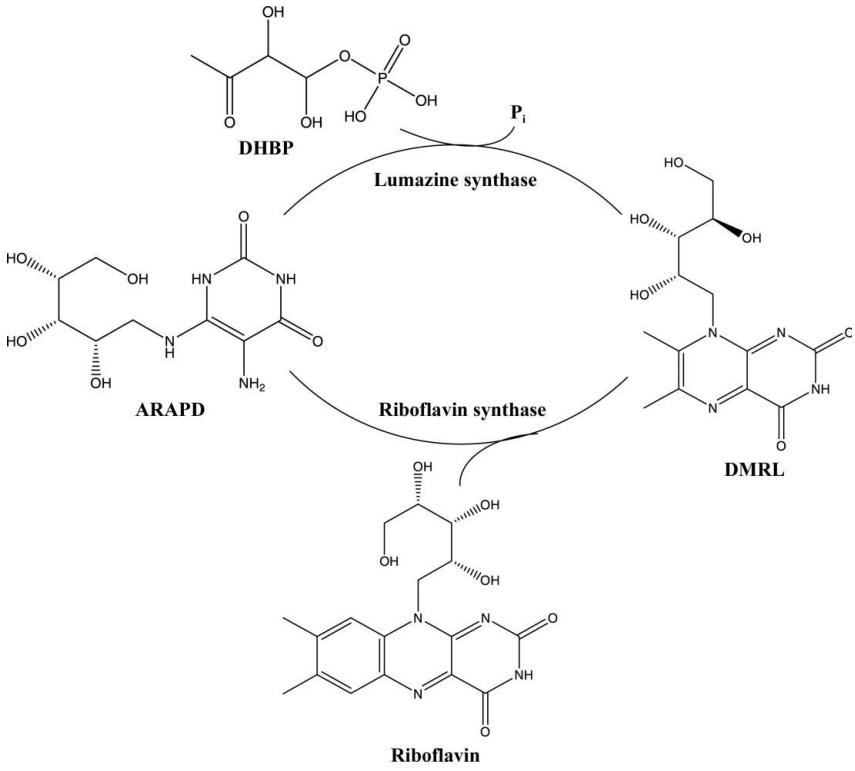

Figure 1
Chemical reaction catalyzed by lumazine synthase. sLS catalyzes the formation of 6,7-dimethyl-8-ribityllumazine (DMRL) using 3,4-dihydroxy-2-butanone 4-phosphate (DHBP) and 5-amino-6-ribitylamino-2,4-(1*H*,3*H*)-pyrimidinedione (ARAPD) as substrates.

molecules of DMRL results in one molecule of riboflavin and one molecule of ARAPD and is catalyzed by riboflavin synthase (RS; Plaut *et al.*, 1970; Plaut & Harvey, 1971). The ARAPD formed in the second half of the reaction is reutilized by lumazine synthase as a substrate. A proposed mechanism based on the experimentally observed regiochemistry of catalysis suggests that the reaction starts with substrate binding, followed by formation of the Schiff-base intermediate and elimination of phosphate with subsequent ring closure that results in the formation of DMRL (Kis *et al.*, 1995). Mutational and solution studies suggest that the catalytic function of the enzyme is tightly correlated with its quaternary structure (Zhang *et al.*, 2006).

LS is very diverse in terms of its structural assembly both in crystal structures and in solution; it shows a pentameric form in *Magnaporthe grisea* (Persson *et al.*, 1999), *Saccharomyces cerevisiae* (Meining *et al.*, 2000), *Schizosaccharomyces pombe* (Gerhardt *et al.*, 2002), *Mycobacterium tuberculosis* (Morgunova *et al.*, 2005) and *Candida albicans* (Morgunova *et al.*, 2007), dimers of pentamers in *Brucella abortus* (Zylberman *et al.*, 2004) and icosahedral capsids consisting of 60 subunits (12 pentamers) in *Bacillus subtilis* (Ladenstein *et al.*, 1994), *E. coli* (Mörtl *et al.*, 1996), *Spinacia oleracea* (Persson *et al.*, 1999) and *Aquifex aeolicus* (Zhang *et al.*, 2001). However, comparison of the three-dimensional structures of LS from different species reveals a common flavodoxin-like fold regardless of the quaternary structure of the enzyme. The folding pattern of LS comprises a central four-stranded β-sheet flanked by two α-helices on one side and three on the other.

A comprehensive study has been carried out in order to understand the subunit assembly of lumazine synthases from different organisms (Persson *et al.*, 1999). Structural analysis of LS from a fungal (*Ma. grisea*) source forming a pentameric assembly and from a plant (*Sp. oleracea*) source forming an icosahedral assembly identified two potential structural determinants that may contribute to the formation of an icosahedral assembly (Persson *et al.*, 1999). Firstly, the presence of a proline residue in the N-terminal region would cause a distorted conformation that may hinder formation of the icosahedral assembly. Secondly, a five-residue loop connecting the last two C-terminal α-helices (α4 and α5) may play a role in formation of the icosahedral assembly. In addition, a systematic sequence analysis of lumazine synthases that form pentamers and icosahedral assemblies identified eight sequence sites that appear to be determinants of icosahedral assembly formation (Fornasari *et al.*, 2004). In spite of these studies, the driving force and structural elements that are responsible for the formation of pentamers and icosahedra still remain unclear (Morgunova *et al.*, 2007). In the current study, we have cloned, expressed, purified and crystallized the lumazine synthase from *Salmonella typhimurium* in order to understand its structure–function relationship, which could be helpful in rational drug design against this pathogen.

2. Materials and methods

2.1. Cloning, expression and purification of lumazine synthase

The *ribH* gene encoding lumazine synthase was amplified by polymerase chain reaction (PCR) from the genomic DNA of *S. typhimurium* using forward 5′-TGA TAT ACA CAT ATG AAC ATT ATT AAA GCT-3′ and reverse 5′-TTA TAA TCA CTC GAG TCA GGC CTT AAT TGC-3′ primers (IDT, USA). The amplified PCR product was digested with *Nde*I and *Xho*I restriction enzymes (New England Labs, USA) and ligated into pET28c vector (Novagen, USA). The integrity of the *ribH* gene in the vector was confirmed by DNA sequencing. The resulting plasmid (sLS-pET28c) expresses lumazine synthase with an N-terminal 6×His tag to enable protein purification by affinity chromatography. The clone sLS-

Table 1
Data-collection and refinement statistics for *S. typhimurium* lumazine synthase.

Values in parentheses are for the highest resolution shell.

	Crystal form *A*	Crystal form *B*	Crystal form *C*
Data collection			
Wavelength (Å)	1.542	1.542	1.542
Resolution (Å)	50.0–3.57 (3.64–3.57)	87.1–3.50 (3.70–3.50)	77.83–4.11 (4.33–4.11)
Space group	$P2_1$	$I2$	$I222$
Molecules per asymmetric unit	60	30	15
Unit-cell parameters (Å, °)	a = 154.23, b = 151.50, c = 235.03, $\alpha = \gamma = 90$, β = 97.08	a = 174.43, b = 157.49, c = 202.79, $\alpha = \gamma = 90$, β = 91.58	a = 153.0, b = 155.66, c = 213.82, $\alpha = \beta = \gamma = 90$
Unique reflections	105485	45638	18225
Multiplicity	1.7 (1.6)	2.2 (2.0)	4.1 (3.1)
Completeness (%)	83.2 (80.8)	66.5 (70.2)	90.2 (70.5)
$R_{\rm merge}$† (%)	9.1 (55.7)	25.8 (39.2)	11.9 (48.6)
$\langle I/\sigma(I)\rangle$	9.4 (2.0)	3.2 (2.2)	7.1 (2.0)
$V_{\rm M}$ (Å^3 Da^{-1})	2.6	2.9	2.8
Solvent content (%)	52.5	57.6	56.5
Refinement			
Resolution range (Å)	32.03–3.57		
$R_{\rm cryst}$‡ (%)	22.6		
$R_{\rm free}$§ (%)	26.4		
R.m.s.d. from ideality			
Bonds (Å)	0.009		
Angles (°)	0.983		
Average B factor, protein (Å^2)	87.6		
Ramachandran plot (%)			
Most favoured	89.3		
Additionally allowed	9.9		
Generously allowed	0.8		

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$. ‡ $R_{\rm cryst} = \sum_{hkl} ||F_{\rm obs}| - |F_{\rm calc}|| / \sum_{hkl} |F_{\rm obs}|$. § $R_{\rm free}$ is the cross-validation R factor computed for the test set, which consisted of 1% of the reflections that were not used in refinement.

pET28c was transformed into *E. coli* BL21 (DE3) strain for expression of lumazine synthase. An overnight culture (10 ml) of single transformant was inoculated into 1 l fresh Luria–Bertani (LB) medium containing kanamycin (30 µg ml^{-1}) and was allowed to grow further at 303 K until the absorbance at 600 nm reached a value of about 0.8–1.0. At this stage, *ribH* gene expression was induced by adding isopropyl β-D-1-thio-galactopyranoside (IPTG) to a final concentration of 0.5 m*M* and the cells were allowed to grow at 303 K for a further 1 h. The cells were harvested by centrifuging the culture at 5000 rev min^{-1} for 15 min at 277 K. The supernatant was discarded and the cell pellet was stored at 193 K until further processing. The cell pellet was resuspended in 25 ml buffer *A* (50 m*M* Tris pH 8.0, 150 m*M* NaCl, 1 m*M* imidazole) followed by the addition of a cocktail of protease inhibitors (Roche, USA). The resuspended cells were lysed by sonication for 20 min with 30 s pulses at 277 K (Sonics, USA). The cell debris was removed by centrifugation at 14 000*g* for 60 min at 277 K and the supernatant was passed through a nickel–nitrilotriacetic acid (Ni–NTA) column (Qiagen, Germany) pre-equilibrated with buffer *B* (50 m*M* Tris pH 8.0, 150 m*M* NaCl). The unbound proteins were washed with 25 column volumes of buffer *C* (50 m*M* Tris pH 8.0, 150 m*M* NaCl, 20 m*M* imidazole) and the bound protein was eluted using buffer *D* (50 m*M* Tris pH 8.0, 150 m*M* NaCl, 300 m*M* imidazole). The eluted protein was subsequently dialyzed against buffer *B* and concentrated to 8 mg ml^{-1} as measured by the Bradford method (Bradford, 1976) using an Amicon concentrator (10 kDa cutoff, Millipore, USA). The purity of the enzyme was checked by 15% SDS–PAGE (Laemmli, 1970).

2.2. Crystallization

Purified recombinant sLS (8 mg ml^{-1} concentration) in 50 m*M* Tris pH 8.0, 150 m*M* NaCl was used for crystallization by the sitting-drop vapour-diffusion method in a 96-well plate (MRC plates, Molecular Dimensions, UK). Initially, a screening kit from Jena Bioscience (Germany) was used to screen for crystallization conditions by mixing 1 µl protein solution with 1 µl reservoir buffer, equilibrating against 60 µl precipitant solution and incubating at 293 K. Plate-like crystals appeared after 2 d in 1.6 *M* ammonium sulfate, 0.1 *M* Tris pH 8.0. To improve the crystal quality, the initial condition was expanded by the hanging-drop method using a 24-well plate with a 4 µl drop consisting of 2 µl protein solution and 2 µl reservoir buffer, which was equilibrated against 500 µl precipitant solution and incubated at 293 K. Three types of crystals were obtained when the reservoir buffer consisted of 1.6 *M* ammonium sulfate, 0.1 *M* Tris buffer with slightly different pH values. At pH 7.75 plate-like crystals (crystal form *A*) appeared after 3 d. At pH 8.0 tetragonal crystals (crystal form *B*) appeared after two months. At pH 8.5 pyramidal crystals (crystal form *C*) appeared after two months.

2.3. Data collection and processing

X-ray diffraction data sets for all three crystal forms (*A*, *B* and *C*) were collected on a MAR345dtb image-plate detector mounted on a Rigaku MicroMax-007 HF microfocus rotating-anode X-ray generator operated at 40 kV and 30 mA. All data sets were collected at 100 K using an Oxford cryostream. Prior to diffraction, crystals were soaked in a cryoprotectant solution consisting of 30% glycerol with the respective components of the precipitating buffer. For crystal form *A*, X-ray diffraction data were collected to 3.50 Å resolution as a total of 67 frames each with 1° oscillation. For crystal form *B*, a complete data set extending to 3.57 Å resolution was collected as a total of 71 frames each with 1° oscillation. For crystal form *C*, a complete data set extending to 4.11 Å resolution was collected as a total of 139 frames each with 1° oscillation. The diffraction images for all the data sets were integrated and scaled using the *HKL*-2000 suite of programs (Otwinowski & Minor, 1997). As the data for crystal form *B* were more complete and of better quality than those for the *A* and *C*

forms, structural analysis was only carried out using the data obtained from crystal form *B*. The data-collection statistics for all of the data sets are given in Table 1.

2.4. Structure determination and refinement

The structure of sLS (crystal form *B*) was solved by the molecular-replacement method using *Phaser* (McCoy *et al.*, 2005) with lumazine synthase from *B. subtilis* (Ritsert *et al.*, 1995) with 30 subunits as a search model (PDB entry 1rvv; 52% sequence identity with sLS). The final solution from *Phaser* yielded two ensembles with 30 subunits each, corresponding to 60 subunits in the asymmetric unit. The initial model was refined by rigid-body refinement using *REFMAC*5 (Murshudov *et al.*, 1997) as implemented in the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994). The model was further refined using *PHENIX* (Adams *et al.*, 2010) by applying 60-fold strict noncrystallographic symmetry (NCS). The sLS model was built using the program *Coot* (Emsley & Cowtan, 2004) and refined iteratively until the model was completely built. The final model was validated using the program *PROCHECK* (Laskowski *et al.*, 1993) from the *CCP*4 suite.

3. Results and discussion

3.1. Cloning, expression and purification of lumazine synthase from *S. typhimurium*

The *ribH* gene encoding lumazine synthase from *S. typhimurium* was amplified by polymerase chain reaction, cloned into the pET28c vector and expressed in *E. coli*. The recombinant protein was purified to homogeneity using 6×His-tag and Ni–NTA affinity chromatography. The *ribH* gene encodes a 156-amino-acid protein with a calculated mass of 16 008 Da. The molecular mass of purified sLS, including 20 extra amino acids contributed from the cloning vector, was estimated as 17–18 kDa by SDS–PAGE and 18 100 Da by MALDI analysis.

3.2. Crystal structure determination, refinement and quality of the model

Lumazine synthase from *S. typhimurium* was crystallized in the monoclinic space group $P2_1$, with unit-cell parameters

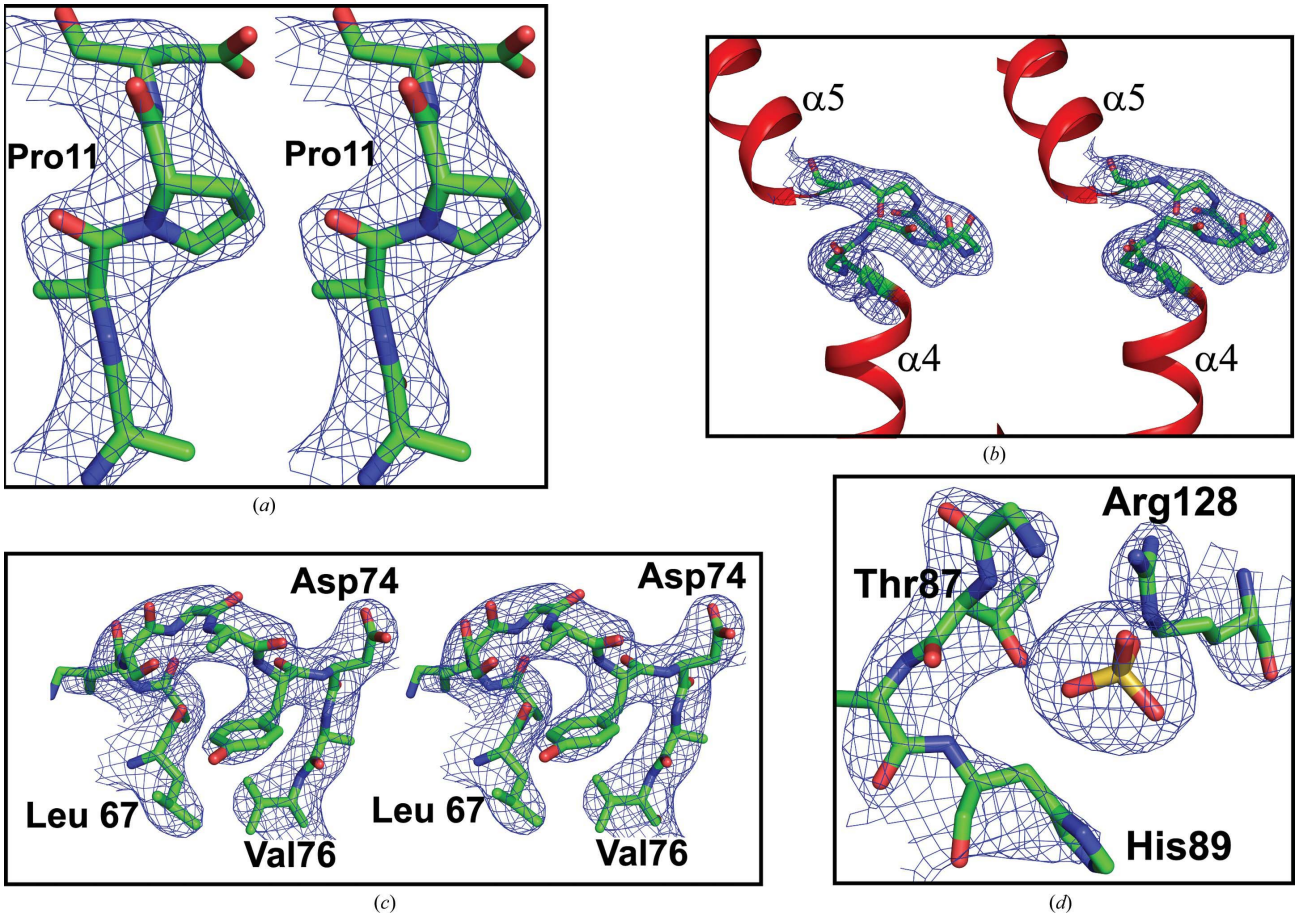

Figure 2
Stereoview showing the final $2F_o - F_c$ electron-density map contoured at the 1.0σ level (*a*) for residue Pro11, (*b*) for the loop connecting the two C-terminal helices $\alpha4$ and $\alpha5$ (the side chains are removed for clarity), (*c*) for the loop region covering residues 67–76 and (*d*) for the sulfate ion

a = 154.23, b = 151.50, c = 235.03 Å, β = 97.08°, and data were collected to 3.57 Å resolution. Assuming the molecular weight of sLS to be about 16 kDa and that there are 60 monomers per asymmetric unit, the calculated Matthews coefficient corresponds to 2.84 Å^3 Da^{-1}, with a solvent content of 56.7%, which is within the normal limits for protein crystals (Matthews, 1968). The structure of sLS was solved by the molecular-replacement method using *B. subtilis* LS as a search model, which shows 52% sequence identity to sLS. The 30 subunits of LS from *B. subtilis* were considered as one single ensemble and *Phaser* yielded a solution with two ensembles, giving a total of 60 subunits per asymmetric unit. The 60 subunits were initially refined by rigid-body refinement using *REFMAC*5 as implemented in *CCP*4. The structure together with the $2F_o - F_c$ map was displayed in *Coot* and the first subunit of the model was manually mutated according to the *S. typhimurium* sequence. The complete 60-subunit structure corresponding to the *S. typhimurium* sequence was generated using *LSQMAN* (Kleywegt & Jones, 1994) by superposing the first monomer onto the remaining 59 subunits. From this point

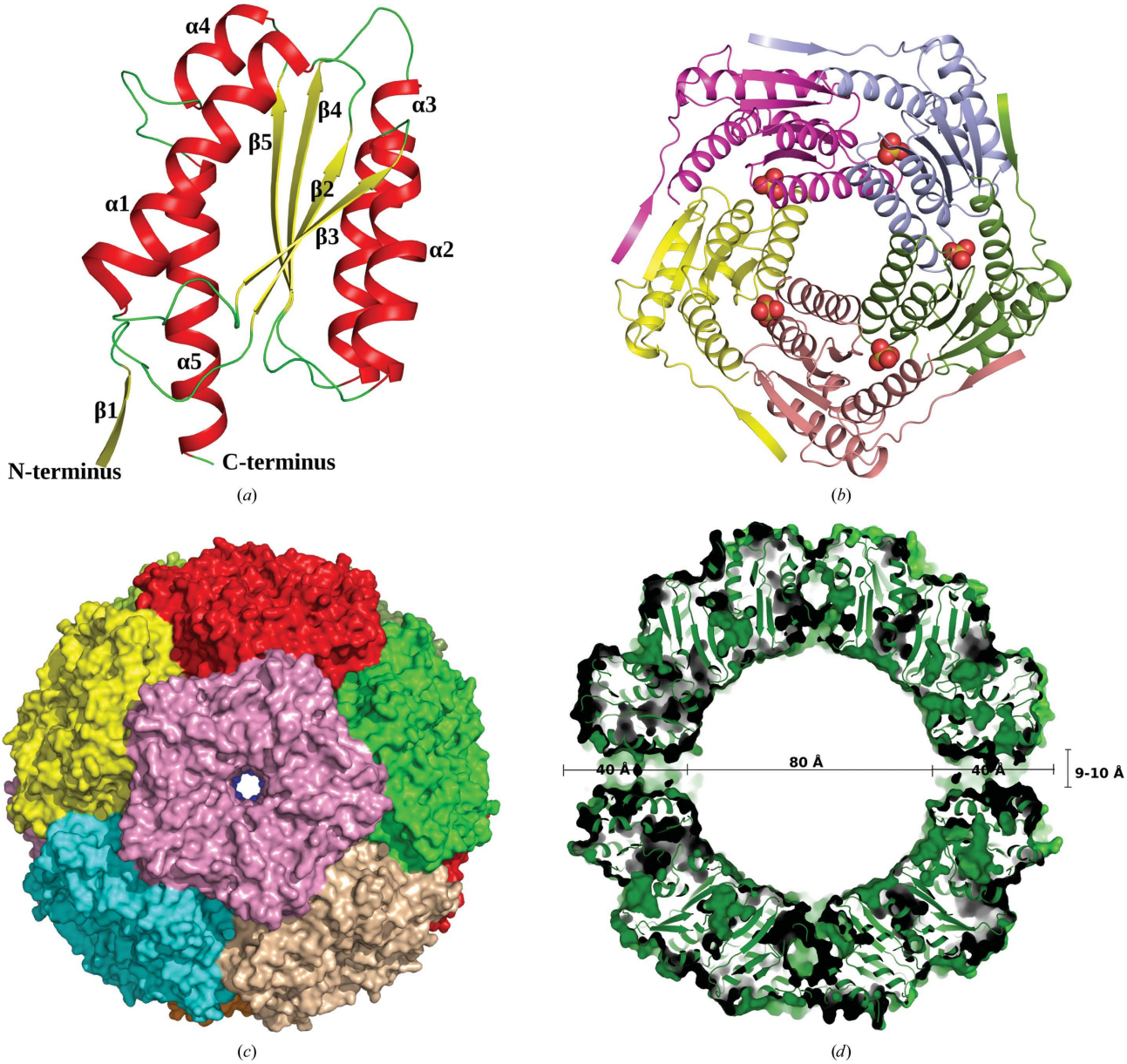

Figure 3
(*a*) Cartoon representation of the subunit fold of sLS with secondary-structure elements labelled. (*b*) Pentamer assembly of sLS; each subunit is represented in a different colour. The active site is indicated by the bound sulfate ion, which is shown as spheres. (*c*) Surface representation of the icosahedral assembly of sLS, showing each pentamer in a different colour. (*d*) Cross-section of the icosahedral assembly of sLS, showing the dimensions of each region.

onwards, the structure was refined using *PHENIX* by applying 60-fold strict NCS. The NCS-averaged map was calculated and displayed in *Coot* for model building.

The sLS structure was refined at 3.57 Å resolution to an R value of 22.6% and a free R value of 26.4%. Each subunit consisted of 154 amino-acid residues covering the *ribH* sequence of *S. typhimurium*. The quality of the final 60-fold averaged electron-density map is generally good and the complete polypeptide chain could be traced in the electron-density map (Fig. 2). A few residues located on the surface of the protein had poor electron density, which is expected at this resolution. In addition, owing to a lack of electron density the residues from the cloning vector at the N-terminal region and the last two residues at the C-terminus of sLS were not included in the model. The difference Fourier map at 3.0σ clearly showed extra electron density which could be modelled as a sulfate ion at the interface of two subunits. The final model consisted of 9240 residues and 79 sulfate ions (including ten sulfate ions bound nonspecifically to some chains). The Ramachandran plot (Ramachandran & Sasisekharan, 1968) analysis for a monomer in the final model shows that 89.3% of the residues are in the most favoured region, 9.9% are in the additionally allowed region and 0.8% are in the generously allowed region. The refinement statistics are shown in Table 1.

3.3. Overall subunit structure of lumazine synthase

The crystal structure of the sLS monomer consists of 154 residues forming a single domain belonging to the flavodoxin-like fold, similar to those observed in the icosahedral LS of *B. subtilis*, spinach and *A. aeolicus* (Ritsert *et al.*, 1995; Persson *et al.*, 1999; Zhang *et al.*, 2001). The core of the sLS has $\alpha/\beta/\alpha$ topology with four parallel β-strands arranged in the order $\beta3$-$\beta2$-$\beta4$-$\beta5$ forming a central β-sheet surrounded by five α-helices (Fig. 3*a*). Two helices, $\alpha1$ and $\alpha4$, are on one side of the β-sheet and are parallel to each other; helices $\alpha2$ and $\alpha3$ are on the other side of the β-sheet and are almost parallel to each other. The $\alpha5$ helix extends outside the protein structure and terminates at the C-terminal end. Both the N-terminal and C-terminal regions are closer on one side of the monomeric structure of sLS. In addition, the N-terminal residues 1–4 form a β-strand ($\beta1$). The overall secondary-structure elements are arranged in the order $\beta1$-$\beta2$-$\alpha1$-$\beta3$-$\alpha2$-$\beta4$-$\alpha3$-$\beta5$-$\alpha4$-$\alpha5$, as shown in Fig. 3(*a*). All of the β-strands and α-helices are interconnected by either loops or turns. The smallest turn is that between $\beta3$ and $\alpha2$ (54–57) and the longest is that between $\alpha4$ and $\alpha5$ (128–136). All of these loops and turns are conserved structurally, including the turn connecting the $\alpha4$ and $\alpha5$ helices (helix–turn–helix motif) in all species.

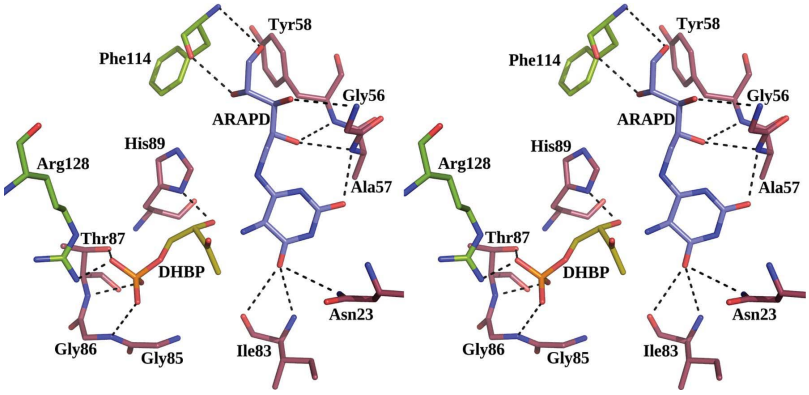

Figure 4
Stereoview showing the proposed model for binding of the substrates at the active site of sLS.

3.4. Pentameric substructure of sLS

The pentamer of sLS forming a central channel with a diameter of about 10 Å is shown in Fig. 3(*b*). The inner side of the channel is mainly occupied by hydrophilic residues (Glu91, Asn99 *etc.*) and is surrounded by $\alpha3$ helices, forming a left-handed twist. In addition, the five subunits interact with each other through hydrogen-bonding and hydrophobic interactions to form the assembly. Specifically, the Ile4 residues from strand $\beta1$ form hydrogen bonds with Val51 and Trp53 which belong to the $\beta3$ strand of the other subunit. Similarly, residues Glu91 ($\alpha3$) from one subunit interact with Tyr92 ($\alpha3$) from other subunit, Ser98 ($\alpha3$) interacts with Tyr58 ($\alpha2$) of the other subunit, Glu106 ($\alpha3$) interacts with Ser103 ($\alpha3$) and Glu65 ($\alpha2$) of the other subunit and Glu146 ($\alpha5$) interacts with Arg21 ($\beta2$) of the other subunit. There are some hydrophobic interactions between Leu151 ($\alpha5$) and Leu62 ($\alpha2$) of the other subunit: Thr143 ($\alpha5$) interacts with Pro55 (which belongs to a loop between $\beta3$ and $\alpha2$) of the other subunit and Phe114 ($\beta5$) interacts with Tyr58 ($\alpha2$) of the other subunit. Thus, interactions between the subunits are between $\beta1$–$\beta3$, $\alpha3$–$\alpha2$, $\alpha3$–$\alpha3$, $\alpha5$–$\beta2$ and $\beta5$–$\alpha2$. The total accessible surface area (Lee & Richards, 1971) calculated for each monomer is about 8095 Å^2; however, this area is reduced to 5450 Å^2 upon pentamer formation, corresponding to 32.6% buried surface area and suggesting that the pentamer is a stable complex.

In general, active lumazine synthase structures are either observed as pentamers, dimers of pentamers or dodecamers of pentamers, suggesting that the basic substructure is a pentamer for all LS. However, the sequence similarity among LS homologues with known structures is between 18% and 93% identity, suggesting that complementarity of the interface surfaces rather than conservation of the hydrogen-bonding pattern plays a role in pentamer formation (Persson *et al.*, 1999). In addition, burial of hydrophobic residues has also been suggested to play a role in the formation and stabilization of pentamers in *S. cerevisiae* LS (Meining *et al.*, 2000). Based on sequence similarity, we predict that the burial of hydrophobic residues will also play a role in the formation of pentamers in sLS.

3.5. Icosahedral assembly of sLS

In sLS, the 60 monomeric subunits are assembled to form an icosahedral capsid as shown in Fig. 3(*c*) which is similar to other reported icosahedral LS structures (Ritsert

et al., 1995; Persson *et al.*, 1999; Zhang *et al.*, 2001). Five neighbouring subunits interact with each other to form a pentamer unit and each pentagonal unit makes edge-to-edge contacts with five pentamer neighbours, corresponding to the arrangement of the faces of the pentagonal dodecahedron, in which three corners are joined at each threefold axis and the angle between each neighbouring pentamer is around 60°. Such an arrangement of pentameric capsomeres indicates a $T = 1$ icosahedral assembly according to icosahedral assembly nomenclature (Caspar & Klug, 1962; Johnson & Speir, 1997). The total diameter of the compact icosahedron is around 160 Å, with the diameter of the inner core being around 80 Å and the width of the pentameric capsomere along its central axis being around 40 Å (Fig. 3*d*). The volume of the central core of the icosahedral capsid corresponds to 2.90×10^5 Å^3. Each monomer in the capsid is arranged in such a way that the active sites are located towards the interior of the capsid.

To date, crystal structures of LS from *B. subtilis*, *S. oleracea* and *A. aeolicus* that form icosahedral assemblies have been reported (Ladenstein *et al.*, 1994; Persson *et al.*, 1999; Zhang *et al.*, 2001). In *B. subtilis*, two enzymes, namely lumazine synthase and riboflavin synthase (RS), form a 1 MDa complex composed of three subunits of RS and 60 subunits of LS. It has been proposed that the tight packing of LS and RS improves the catalytic efficiency by substrate channelling at low substrate concentrations (Kis & Bacher, 1995). In addition, it was shown that the LS catalytic activities from the native enzyme complex of *B. subtilis* and the reconstituted hollow icosahedral capsid are identical (Kis *et al.*, 1995). It was also reported that lumazine synthase from *E. coli* does not physically associate with any other enzyme of the riboflavin pathway and that the core of the icosahedral capsid is empty (Mörtl *et al.*, 1996). Moreover, it was shown that LS activities are similar in both the icosahedral and pentameric forms, suggesting that capsid

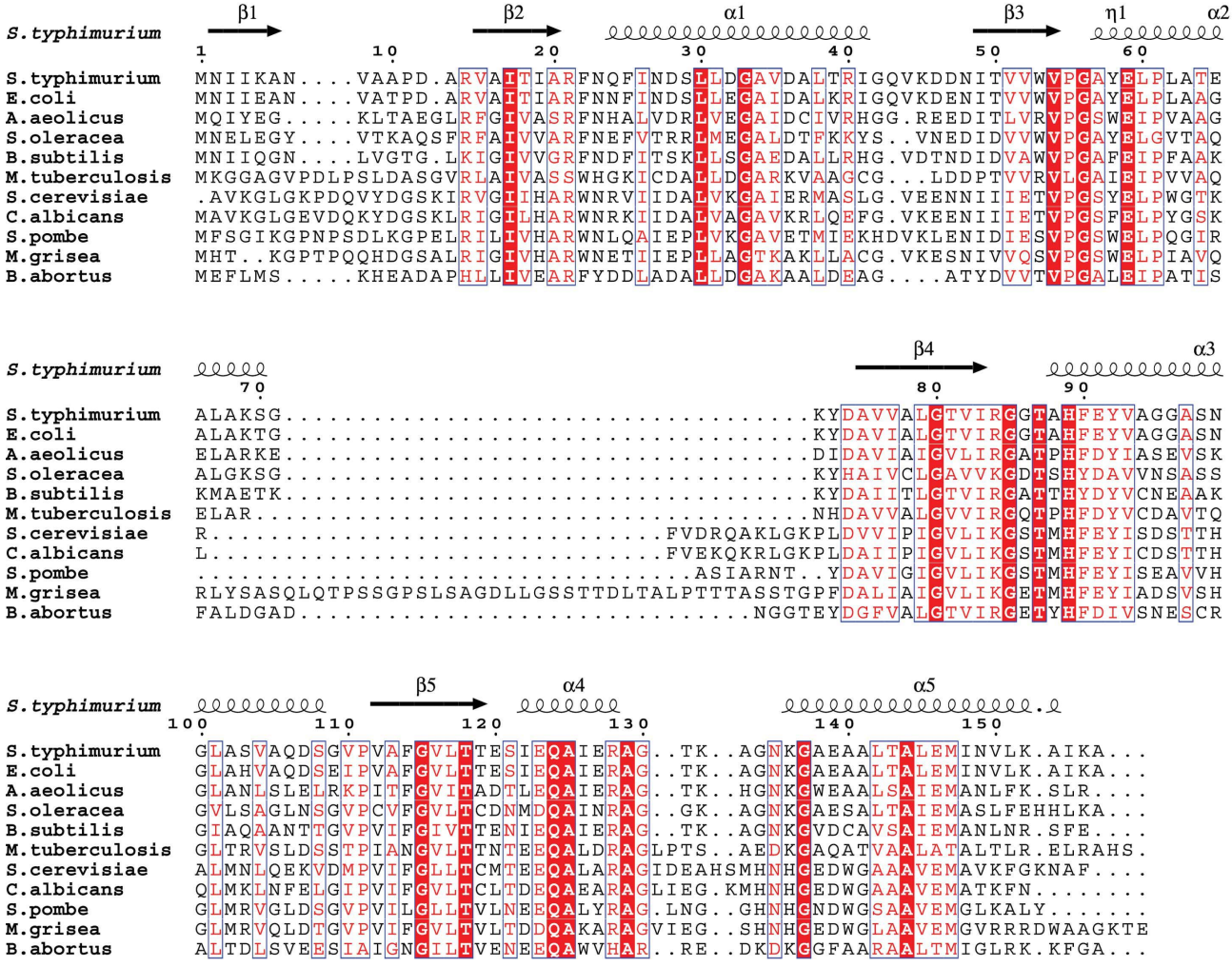

Figure 5
Multiple sequence alignment of lumazine synthases from different species. The secondary structures are indicated on the top row for *S. typhimurium*. Residues that are strictly conserved are shown with a red background, residues that are well conserved within a group are indicated by red letters and residues that are conserved between groups are boxed. This figure was generated using *ESPript* (Gouet *et al.*, 1999).

formation is not involved in the catalytic activity of the enzyme (Mörtl *et al.*, 1996; Persson *et al.*, 1999). However, studies to dissociate the capsid formation using mild denaturants resulted in larger capsid formation without any LS activity, suggesting that the quaternary structure is tightly correlated with catalytic function (Zhang *et al.*, 2006). Thus, the requirement for icosahedral assembly formation for certain species and the driving forces required for the assembly formation of lumazine synthase have remained unclear to date.

3.6. Substrate-binding site

In sLS, the catalytic site was located by the bound sulfate ion which mimics the phosphate moiety of the DHBP substrate (Fig. 4). The active site of sLS is formed at the interior surface of the icosahedral interface between two subunits of each pentamer. The active site is composed mainly of residues from the loops connecting the β2 strand and α1 helix, the β3 strand and α2 helix and the β4 strand and α3 helix from one subunit and the β5 strand, α4 helix and α5 helix from the other subunit of the same pentamer.

We have predicted the residues of sLS involved in substrate binding and catalysis by modelling the substrates ARAPD and DHBP based on structures of LS of *M. tuberculosis* complexed with inhibitors (PDB entries 2c97 and 2vi5; Morgunova *et al.*, 2006; Zhang *et al.*, 2008). In sLS, the residues Phe22, Asn23, Gly56, Ala57 and Tyr58 from one subunit and Phe114 from the other subunit are residues that could potentially interact with the substrate ARAPD. The sulfate which occupies the same position as the phosphate of the substrate DHBP revealed that residues Gly86, Thr87 and His89 from one subunit and Arg128 from the other subunit may be involved in binding the DHBP substrate (Fig. 4). All of the residues involved in the active site are mostly conserved among species (Fig. 5); moreover, no difference in the catalytic activity between the pentamer and the icosahedral capsid could be established, suggesting a similar catalytic mechanism in all species.

3.7. Comparsion of sLS with other lumazine synthase structures

The sLS monomeric structure could be superimposed onto *B. subtilis* LS (PDB code 1rvv; Ritsert *et al.*, 1995) with a root-mean-square deviation (r.m.s.d.) of 0.93 Å for 153 C^{α} atoms, onto *A. aeolicus* LS (1nqu; Zhang *et al.*, 2003) with an r.m.s.d. of 1.02 Å for 152 C^{α} atoms, onto *S. oleracea* LS (1c2y; Persson *et al.*, 1999) with an r.m.s.d. of 1.0 Å for 150 C^{α} atoms, onto *S. pombe* LS (1kyv; Gerhardt *et al.*, 2002) with an r.m.s.d. of 1.20 for 120 C^{α} atoms, onto *M. tuberculosis* LS (2c92; Morgunova *et al.*, 2006) with an r.m.s.d. of 1.05 Å for 138 C^{α} atoms, onto *B. abortus* LS (2f59; Klinke *et al.*, 2007) with an r.m.s.d. of 1.10 Å for 130 C^{α} atoms and onto *C. albicans* LS (2jfb; Morgunova *et al.*, 2007) with an r.m.s.d. of 1.13 Å for 141 C^{α} atoms, reflecting the high similarity among LS structures.

Studies of lumazine synthase structures to identify the elements that are responsible for the formation of the icosahedral assembly suggested two regions that may be involved in icosahedral capsid formation (Mörtl *et al.*, 1996; Persson *et al.*, 1999). Firstly, the formation of a β-strand (β1) in the N-terminal region which interacts with the core β-sheet of the adjacent subunit of LS may potentially help in formation of the icosahedral assembly (Fig. 6). In the case of pentamer-forming LS, these N-terminal residues are either disordered or are observed in a conformation which is unlikely to form a β-strand with the neighbouring subunit owing to the presence of a proline residue at the N-terminus as observed in the *M. grisea* (Persson *et al.*, 1999), *S. cerevisiae* (Meining *et al.*, 2000), *M. tuberculosis* (Morgunova *et al.*, 2005) and *S. pombe* (Gerhardt *et al.*, 2002) structures. However, in the case of sLS a proline (Pro11) residue was found in the N-terminal region but it could still form an icosahedral assembly, suggesting that the proline residue may not play a role in the formation of the icosahedral assembly. Secondly, the size and orientation of the loop between helices α4 and α5 play a role in -assembly formation. In the icosohedral capsid, the loop is shorter and makes a turn which could be compatible with the formation of the

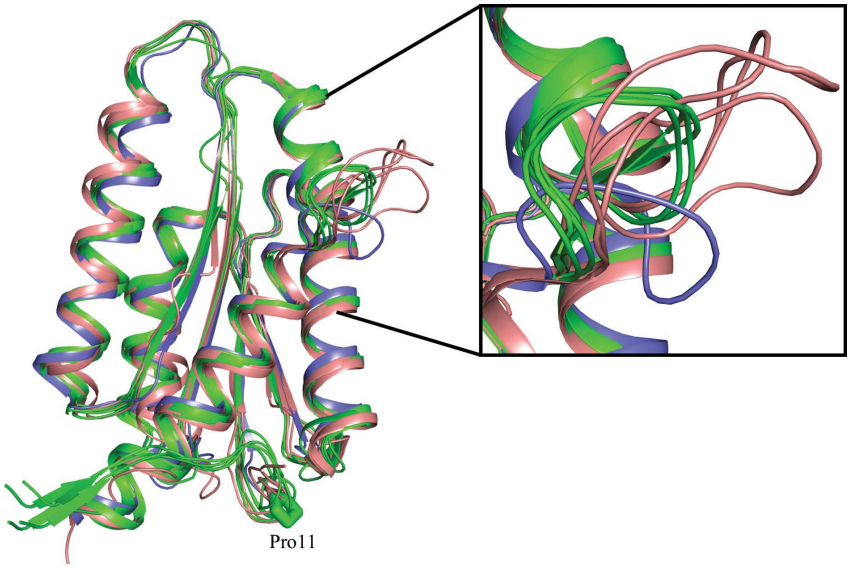

Figure 6
Superposition of LS structures from the organisms forming icosahedral assemblies are shown in green, those from the organisms forming pentameric assemblies are shown in pink and that from the organism forming a decamer (dimer of pentamers) is shown in blue. The proline residue of sLS is shown as a stick model (green). The inset shows a close-up view of the different conformations of the loop between the α4 and α5 helices.

icosohedral capsid. However, in the pentameric form this loop is longer and has a different orientation which prevents the formation of the icosahedral capsid (Fig. 6). In sLS the loop is short, similar to other icosahedral assembly-forming LS, and thus is compatible with the formation of the icosahedral capsid, suggesting that this loop region may play a role in the formation of the icosahedral assembly. However, further studies are required to identify the driving forces for the formation of the icosahedral assembly.

We thank Ruchi Gautam for her assistance in protein purification. We also thank Sharanjit Kaur and Purnananda Guptasarma for the MALDI analysis and Dr Srikrishna Subramanian for carefully reading the manuscript. PK and MS thank the Council of Scientific and Industrial Research (CSIR), Government of India for doctoral fellowships. This work was supported by CSIR through the network project NWP05.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Bacher, A. (1991). *Chemistry and Biochemistry of Flavoenzymes*, Vol. 1, edited by F. Müller, pp. 215–259. Boca Raton: CRC Press.
Bacher, A., Eberhardt, S. & Richter, G. (1996). In *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2nd ed., edited by F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter & H. E. Umbarger. Washington: American Society for Microbiology.
Not cited in text – Bacher, A. & Ladenstein, R. (1990). *Chemistry and Biochemistry of Flavoenzymes*, Vol. 1, edited by F. Müller, pp. 293–316. Boca Raton: CRC Press.
Not cited in text – Bellido, D., Craig, P. O., Mozgovoj, M. V., Gonzalez, D. D., Wigdorovitz, A., Goldbaum, F. A. & Dus Santos, M. J. (2009). *Vaccine*, 27, 136–145.
Not cited in text – Braden, B. C., Velikovsky, C. A., Cauerhff, A. A., Polikarpov, I. & Goldbaum, F. A. (2000). *J. Mol. Biol.* 297, 1031–1036.
Bradford, M. M. (1976). *Anal. Biochem.* **72**, 248–254.
Briggs, W. R. & Huala, E. (1999). *Annu. Rev. Cell Dev. Biol.* **15**, 33–62.
Caspar, D. L. & Klug, A. (1962). *Cold Spring Harb. Symp. Quant. Biol.* **27**, 1–24
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Fornasari, M. S., Laplagne, D. A., Frankel, N., Cauerhff, A. A., Goldbaum, F. A. & Echave, J. (2004). *Mol. Biol. Evol.* **21**, 97–107.
Gerhardt, S., Haase, I., Steinbacher, S., Kaiser, J. T., Cushman, M. & Bacher, A. (2002). *J. Mol. Biol.* **318**, 1317–1329.
Gouet, P., Courcelle, E., Stuart, D. I. & Métoz, F. (1999). *Bioinformatics*, **15**, 305–308.
Johnson, J. E. & Speir, J. A. (1997). *J. Mol. Biol.* **269**, 665–675.
Not cited in text – Karshikoff, A. & Ladenstein, R. (1989). *Proteins*, 5, 248–257.
Kis, K. & Bacher, A. (1995). *J. Biol. Chem.* **270**, 16788–16795.
Kis, K., Volk, R. & Bacher, A. (1995). *Biochemistry*, **34**, 2883–2892.
Kleywegt, G. J. & Jones, T. A. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 9–14.
Klinke, S., Zylberman, V., Bonomi, H. R., Haase, I., Guimaraes, B. G., Braden, B. C., Bacher, A., Fischer, M. & Goldbaum, F. A. (2007). *J. Mol. Biol.* **373**, 664–680.
Ladenstein, R., Ritsert, K., Huber, R., Richter, G. & Bacher, A. (1994). *Eur. J. Biochem.* **223**, 1007–1017.
Laemmli, U. K. (1970). *Nature (London)*, **227**, 680–685.
Laskowski, R. A., Moss, D. S. & Thornton, J. M. (1993). *J. Mol. Biol.* **231**, 1049–1067.
Lee, B. & Richards, F. M. (1971). *J. Mol. Biol.* **55**, 379–400.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). *Acta Cryst.* D**61**, 458–464.
Meighen, E. A. (1991). *Microbiol. Rev.* **55**, 123–142.
Meighen, E. A. (1993). *FASEB J.* **7**, 1016–1022.
Meining, W., Mörtl, S., Fischer, M., Cushman, M., Bacher, A. & Ladenstein, R. (2000). *J. Mol. Biol.* **299**, 181–197.
Morgunova, E., Illarionov, B., Sambaiah, T., Haase, I., Bacher, A., Cushman, M., Fischer, M. & Ladenstein, R. (2006). *FEBS J.* **273**, 4790–4804.
Morgunova, E., Meining, W., Illarionov, B., Haase, I., Jin, G. & Bacher, A. (2005). *Biochemistry*, **44**, 2746–2758.
Morgunova, E., Saller, S., Haase, I., Cushman, M., Bacher, A., Fischer, M. & Ladenstein, R. (2007). *J. Biol. Chem.* **282**, 17231–17241.
Mörtl, S., Fischer, M., Richter, G., Tack, J., Weinkauf, S. & Bacher, A. (1996). *J. Biol. Chem.* **271**, 33201–33207.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
O'Kane, D. J. & Prasher, D. C. (1992). *Mol. Microbiol.* **6**, 443–449.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Persson, K., Schneider, G., Jordan, D. B., Viitanen, P. V. & Sandalova, T. (1999). *Protein Sci.* **8**, 2355–2365.
Plaut, G. W., Beach, R. L. & Aogaichi, T. (1970). *Biochemistry*, **9**, 771–785.
Plaut, G. W. & Harvey, R. A. (1971). *Methods Enzymol.* **18**B, 515–539.
Ramachandran, G. N. & Sasisekharan, V. (1968). *Adv. Protein Chem.* **23**, 283–438.
Ritsert, K., Huber, R., Turk, D., Ladenstein, R., Schmidt-Bäse, K. & Bacher, A. (1995). *J. Mol. Biol.* **253**, 151–167.
Salomon, M., Eisenreich, W., Durr, H., Schleicher, E., Knieb, E. & Massey, V. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 12357–12361.
Thompson, C. L. & Sancar, A. (2002). *Oncogene*, **21**, 9043–9056.
Volk, R. & Bacher, A. (1991). *J. Biol. Chem.* **266**, 20610–20618.
Not cited in text – Yang, Z., Bahar, I. & Widom, M. (2009). *Biophys. J.* 96, 4438–4448.
Young, D. W. (1986). *Nat. Prod. Rep.* **3**, 395–419.
Zhang, X., Konarev, P. V., Petoukhov, M. V., Svergun, D. I., Xing, L., Cheng, R. H., Haase, I., Fischer, M., Bacher, A., Ladenstein, R. & Meining, W. (2006). *J. Mol. Biol.* **362**, 753–770.
Zhang, X., Meining, W., Cushman, M., Haase, I., Fischer, M., Bacher, A. & Ladenstein, R. (2003). *J. Mol. Biol.* **328**, 167–182.
Zhang, X., Meining, W., Fischer, M., Bacher, A. & Ladenstein, R. (2001). *J. Mol. Biol.* **306**, 1099–1114.
Zhang, Y., Illarionov, B., Morgunova, E., Jin, G., Bacher, A., Fischer, M., Ladenstein, R. & Cushman, M. (2008). *J. Org. Chem.* **73**, 2715–2724.
Zylberman, V., Craig, P. O., Klinke, S., Braden, B. C., Cauerhff, A. & Goldbaum, F. A. (2004). *J. Biol. Chem.* **279**, 8093–8101.