

Structure of 2-oxo-3-deoxygalactonate kinase from  
*Klebsiella pneumoniae*

Karolina Michalska,<sup>a</sup> Marianne E. Cuff,<sup>a,b</sup> Christine Tesar,<sup>a</sup> Brian Feldmann<sup>a</sup> and Andrzej Joachimiak<sup>a,b,c\*</sup>

<sup>a</sup>Midwest Center for Structural Genomics, Biosciences Division, Argonne National Laboratory, USA, <sup>b</sup>Structural Biology Center, Biosciences Division, Argonne National Laboratory, USA, and <sup>c</sup>Department of Biochemistry and Molecular Biology, University of Chicago, USA

Correspondence e-mail: andrzejj@anl.gov

In most organisms, efficient D-galactose utilization requires the highly conserved Leloir pathway that converts D-galactose to D-glucose 1-phosphate. However, in some bacterial and fungal species alternative routes of D-galactose assimilation have been identified. In the so-called De Ley–Doudoroff pathway, D-galactose is metabolized into pyruvate and D-glyceraldehyde 3-phosphate in five consecutive reactions carried out by specific enzymes. The penultimate step in this pathway involves the phosphorylation of 2-oxo-3-deoxygalactonate to 2-oxo-3-deoxygalactonate 6-phosphate catalyzed by 2-oxo-3-deoxygalactonate kinase, with ATP serving as a phosphoryl-group donor. Here, a crystal structure of 2-oxo-3-deoxygalactonate kinase from *Klebsiella pneumoniae* determined at 2.1 Å resolution is reported, the first structure of an enzyme from the De Ley–Doudoroff pathway. Structural comparison indicates that the enzyme belongs to the ASKHA (acetate and sugar kinases/hsc70/actin) family of phosphotransferases. The protein is composed of two  $\alpha/\beta$  domains, each of which contains a core common to all family members. Additional elements introduced between conserved structural motifs define the unique features of 2-oxo-3-deoxygalactonate kinase and possibly determine the biological function of the protein.

### 1. Introduction

D-Galactose is a monosaccharide most commonly occurring in milk in the form of the disaccharide lactose, in which it is linked to D-glucose. Additionally, it is found in many plant tissues, including fruits and vegetables, which contain both monomeric galactose and  $\beta$ -1,4-linked galactosyl residues that form galactolipids in chloroplast membranes and galactan in cell walls (Acosta & Gross, 1995; Gross & Acosta, 1991). Galactose, together with xylose, arabinose and mannose, is also a constituent of hemicellulose (Schädel *et al.*, 2010). In most organisms, galactose utilization is governed by the well conserved Leloir pathway (Fig. 1), which starts with the phosphorylation of  $\alpha$ -D-galactose to D-galactose 1-phosphate (Gal-1-P) in a reaction controlled by galactokinase (GALK; EC 2.7.1.6; Trucco & Caputto, 1948). Galactose-1-phosphate uridylyltransferase (GALT; EC 2.7.7.12) then catalyses sugar exchange between Gal-1-P and UDP-glucose (UDP-G), leading to the formation of UDP-galactose (UDP-Gal) and free glucose 1-phosphate (G-1-P) (Caputto & Leloir, 1949; Caputto *et al.*, 1950). In a subsequent event, the UDP-Gal product is converted to UDP-glucose by UDP-galactose-4-epimerase (GALE; EC 5.1.3.2; Leloir, 1951). Cycling through GALT and GALE reactions results in a complete conversion of Gal-1-P to G-1-P. Phosphoglucomutase then converts G-1-P to D-glucose 6-phosphate, which enters the glycolytic pathway as an energy source.

Received 8 April 2011

Accepted 6 June 2011

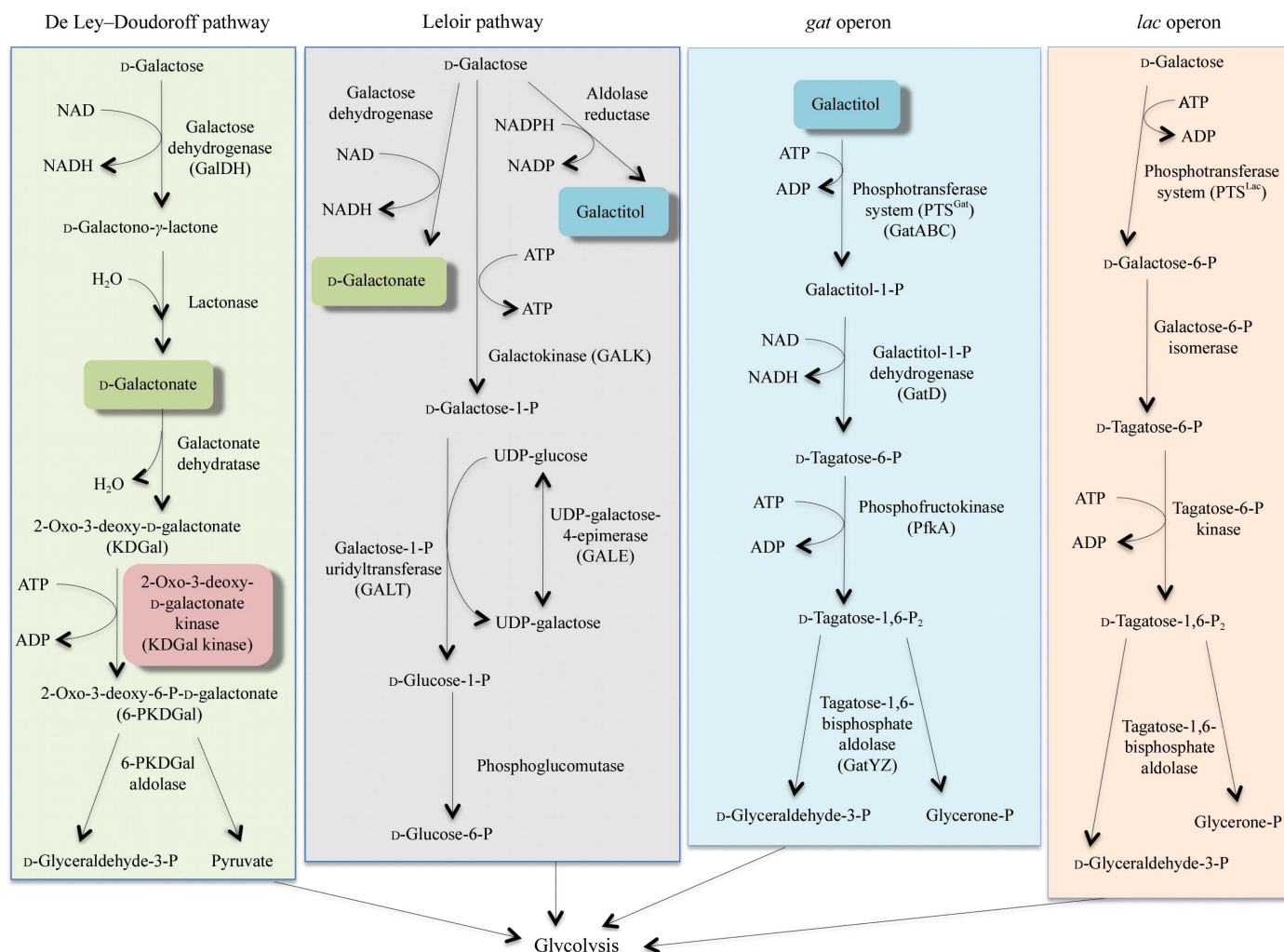
**PDB Reference:** 2-oxo-3-deoxygalactonate kinase, 3r1x.

Perturbations in the Leloir pathway lead to galactose-induced toxicity, which has been described in humans and yeast (de Jongh *et al.*, 2008; Mehta *et al.*, 1999). In humans, deficiency of GALK, GALT or GALE activity, which is usually caused by mutations in the corresponding genes, is diagnosed as one of three types of a disease known as galactosemia (Lai *et al.*, 2009; Leslie, 2003; Fridovich-Keil, 2006). The clinical syndromes of the disorder result from accumulation of Leloir-pathway intermediates, especially Gal-1-P, or of compounds synthesized in secondary routes of galactose metabolism, which include galactitol and D-galactonate. The pathways producing the two latter substances involve aldose reductase catalyzing galactose-to-galactitol conversion (Yabe-Nishimura, 1998) and galactose dehydrogenase participating in D-galactonate synthesis (Cuatrecasas & Segal, 1966a,b). Since galactitol and D-galactonate are not further metabolized they are predominantly excreted in the urine (Jakobs *et al.*, 1995; Wehrli *et al.*, 1997), but some is deposited in tissues.

On the other hand, in prokaryotes or fungi these metabolites are not useless as they represent intermediate products of fully functional energy-yielding routes alternative to the

Leloir pathway. For example, there are reports of galactitol metabolism through L-tagatose (Schneider *et al.*, 1995) or L-sorbose (Fekete *et al.*, 2004). Moreover, in enteric bacteria galactitol can be transported and phosphorylated *via* the galactitol-specific phosphoenolpyruvate-dependent transferase system (PTS<sup>Gat</sup>) to galactitol 1-phosphate, which is then converted to D-tagatose 6-phosphate in a reaction catalyzed by galactitol-1-phosphate dehydrogenase (Nobelmann & Lengeler, 1996; Shakeri-Garakani *et al.*, 2004). Further metabolism resembles the tagatose 6-phosphate pathway governed by the *lac* operon encoding the lactose-phosphotransferase system (PTS<sup>Lac</sup>; Fig. 1; van Rooijen *et al.*, 1991). PTS<sup>Lac</sup> represents yet another way to metabolize lactose-derived galactose and is utilized mostly by lactic acid bacteria (de Vos *et al.*, 1990; Zeng *et al.*, 2010).

Galactonate, a second byproduct of defects in the Leloir pathway enzymes, is an intermediate in the De Ley–Doudoroff (DD) pathway of galactose utilization (De Ley & Doudoroff, 1957), which resembles the Entner–Doudoroff (ED) pathway of glucose metabolism (Entner & Doudoroff, 1952) but involves reactions and enzymes entirely different



**Figure 1**  
D-Galactose metabolic pathways.

from those engaged in the ED and Leloir pathways. In an initial step of the DD pathway D-galactose is converted by galactose dehydrogenase (GalDH; EC 1.1.1.48) to D-galactono- $\gamma$ -lactone, which is subsequently hydrolyzed by lactonase (EC 3.1.1.25) to D-galactonate (Fig. 1). D-Galactonate then undergoes dehydration carried out by galactonate dehydratase (GalDH; EC 4.2.1.6), leading to the formation of 2-oxo-3-deoxygalactonate (KDGal). In the next step, this compound is phosphorylated by 2-oxo-3-deoxygalactonate kinase (KDGal kinase; EC 2.7.1.58) to 2-oxo-3-deoxygalactonate 6-phosphate (6-P-KDGal), which decomposes into pyruvate and D-glyceraldehyde 3-phosphate. Both products are utilized by the glycolytic pathway. To date, the De Ley–Doudoroff pathway has only been described for a few bacterial species, including *Pseudomonas saccharophila* (De Ley & Doudoroff, 1957), *Gluconobacter liquefaciens* (Stouthamer, 1961), *Azotobacter vinelandii* (Wong & Yao, 1994), *Rhizobium meliloti* (Arias & Cerveñansky, 1986) and *Stenotrophomonas maltophilia* (Brechtel *et al.*, 2002). In addition, a nonphosphorylative variation of the DD pathway has been reported for the filamentous fungus *Aspergillus niger* (Elshafei & Abdel-Fatah, 2001). A shorter version of the pathway, which lacks the first two enzymes, has been discovered in *Escherichia coli* (Deacon & Cooper, 1977) and *Mycobacterium butyrium* (Szumilo, 1981). Detailed analysis of the *E. coli* route has shown that the genes encoding the key enzymes are organized into the *dgo* operon (Cooper, 1978; Babbitt *et al.*, 1995). Transcription from the *dgo* operon is induced exclusively by D-galactonate. Genome-sequencing projects have identified other organisms that possibly utilize the DD pathway, including several human pathogens such as *Klebsiella pneumoniae*.

As the most common metabolic pathway of galactose assimilation with human health implications, the Leloir pathway has been extensively studied. A significant fraction of research has been focused on the characterization of the enzymes catalyzing individual reactions, including the determination of their three-dimensional structures (Holden *et al.*, 2003). Less attention has been attracted by alternative routes such as the De Ley–Doudoroff pathway and others. In particular, none of the proteins belonging to this metabolic network have been characterized structurally. In this study, we present the first crystal structure of an enzyme belonging to the DD pathway, namely 2-oxo-3-deoxygalactonate kinase from the human pathogen *K. pneumoniae*, at 2.1 Å resolution. KDGal kinase belongs to a large protein family with over 1100 members mainly found in bacteria and also in viruses.

## 2. Materials and methods

### 2.1. Protein cloning, expression and purification

The 2-oxo-3-deoxygalactonate kinase gene from *K. pneumoniae* subsp. *pneumoniae* MGH 78578 was amplified by PCR using genomic DNA as a template and the following primers: 5'-TACTTCCAATCCAATGCCATGACGGCTCGCTACATCGCAATA-3' and 5'-TTATCCACTTCCAATGTTAGTT-

TGCCACTGCATAAGCGATGCT-3'. The PCR product was cloned into the pMCSG19 vector (Donnelly *et al.*, 2006) according to the ligation-independent cloning procedure (Aslanidis & de Jong, 1990; Eschenfeldt *et al.*, 2009). The expression vector pAPC38818 was transformed into *E. coli* strain BL21 (DE3) harboring the plasmid pRK1037 (Nallamsetty *et al.*, 2004). This construct allows the expression of a fusion protein containing an N-terminal maltose-binding protein (MBP) followed by a tobacco vein mottling virus (TVMV) protease cleavage site, a His<sub>6</sub> tag, a TEV protease cleavage site and a target protein. MBP cleavage takes place *in vivo* owing to constitutive expression of TVMV protease from the pRK1037 plasmid, generating the His<sub>6</sub>-tagged target protein.

To produce the protein, the bacterial culture was grown at 310 K and shaken at 200 rev min<sup>-1</sup> in enriched M9 medium (Donnelly *et al.*, 2006) until it reached an OD<sub>600</sub> of 1. After cooling the culture to 277 K, selenomethionine (SeMet) and a mixture of amino acids inhibiting the metabolic pathway of methionine synthesis were added (Van Duyne *et al.*, 1993; Walsh *et al.*, 1999). After an additional 60 min, kinase expression was induced with 0.5 mM isopropyl  $\beta$ -D-1-thiogalactoside (IPTG). The cells were incubated overnight at 291 K, harvested and resuspended in lysis buffer [500 mM NaCl, 5%(v/v) glycerol, 50 mM HEPES pH 8.0, 20 mM imidazole, 10 mM  $\beta$ -mercaptoethanol]. The SeMet-labeled protein was purified as described previously (Kim *et al.*, 2004). Specifically, the protocol included Ni<sup>2+</sup>-affinity chromatography on an ÄKTExpress system (GE Healthcare Life Sciences) followed by His<sub>6</sub>-tag cleavage using recombinant His-tagged TEV protease and an additional step of Ni<sup>2+</sup>-affinity chromatography performed to remove the protease, the uncut protein and the affinity tag. The pure protein was concentrated using Centricon filtration (Millipore, Bedford, Massachusetts, USA) in 20 mM HEPES pH 8.0 buffer, 250 mM NaCl and 2 mM dithiothreitol (DTT). The final protein yield was ~11 mg per litre of culture.

### 2.2. Size-exclusion chromatography

Size-exclusion chromatography was performed on a GF Superdex-200 (10/300) analytical column using FPLC (GE Healthcare Life Sciences). The column was equilibrated with buffer consisting of 20 mM HEPES pH 8.0, 250 mM NaCl and 2 mM DTT. A 200  $\mu$ l protein sample at 24 mg ml<sup>-1</sup> was injected onto the column. After the run, the column was equilibrated with the same buffer supplemented with 15 mM ATP. A 200  $\mu$ l protein sample at a concentration of 23 mg ml<sup>-1</sup> and supplemented with 10 mM ATP was injected onto the column. The chromatography was carried out at room temperature at a flow rate of 0.3 ml min<sup>-1</sup>.

### 2.3. Protein crystallization

The SeMet-labeled 2-oxo-3-deoxygalactonate kinase was screened for crystallization conditions with the help of a Mosquito liquid dispenser (TTP LabTech, Cambridge, Massachusetts, USA) using the sitting-drop vapor-diffusion

**Table 1**

Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

Data collection	
Space group	$P2_12_12_1$
Unit-cell parameters (Å)	$a = 85.5, b = 134.0, c = 139.9$
Temperature (K)	100
Radiation source	APS, BM-19
Wavelength (Å)	0.97921
Resolution (Å)	50.0–2.10 (2.14–2.10)
Reflections collected	520395
Unique reflections	94834
$R_{\text{merge}}^\dagger$	0.091 (0.684)
$\langle I/\sigma(I) \rangle$	14.5 (2.2)
Completeness (%)	99.7 (99.7)
Multiplicity	5.5 (5.3)
Refinement	
Resolution (Å)	35.0–2.1
Reflections	
Work set	92571
Test set	1902
$R_{\text{work}}/R_{\text{free}}^\ddagger$	0.174/0.199
No. of atoms	
Protein	9135
Ligands	114
Water	540
Average $B$ factor (Å <sup>2</sup> )	
Protein	51.3
Ligands	72.7
Water	54.5
R.m.s. deviations from ideal	
Bond lengths (Å)	0.016
Bond angles (°)	1.476
Ramachandran statistics of $\varphi/\psi$ angles (%)	
Most favored	97.8
Outliers	0
PDB code	3r1x

$^\dagger R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$ , where  $I_i(hkl)$  is the intensity of observation  $i$  of reflection  $hkl$ .  $^\ddagger R = \sum_{hkl} ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum_{hkl} |F_{\text{obs}}|$ , where  $F_{\text{obs}}$  and  $F_{\text{calc}}$  are observed and calculated structure factors, respectively.  $R_{\text{free}}$  is calculated analogously for the test reflections, which were randomly selected and excluded from the refinement.

technique in 96-well CrystalQuick plates (Greiner Bio-One, Monroe, North Carolina, USA). For each condition, 0.4  $\mu\text{l}$  protein solution (30 mg ml<sup>-1</sup>) and 0.4  $\mu\text{l}$  crystallization formulation were mixed and the mixture was equilibrated against a 135  $\mu\text{l}$  reservoir. Three crystallization screens were used: Index (Hampton Research, Aliso Viejo, California, USA), ANL-1 (Qiagen, Valencia, California, USA) and ANL-2 (Qiagen) at both 297 and 277 K. Diffraction-quality crystals appeared at 277 K in 0.1 M magnesium formate, 15% (w/v) PEG 3350, corresponding to condition H8 from the Index screen.

#### 2.4. Data collection

An X-ray diffraction data set extending to 2.1 Å resolution was collected on the Structural Biology Center beamline BM-19 at the Advanced Photon Source, Argonne National Laboratory, USA. Prior to data collection, the SeMet-labeled protein crystal was cryoprotected in mother liquor supplemented with 20% glycerol and flash-cooled in liquid nitrogen. A single-wavelength anomalous diffraction (SAD) data set was collected near the Se  $K$  absorption edge (0.97921 Å) at 100 K. The diffraction data were integrated and scaled with

the *HKL-3000* suite (Minor *et al.*, 2006). The processing statistics are given in Table 1.

#### 2.5. Structure solution and refinement

The structure was determined by the SAD method using selenium peak data and the *HKL-3000* software suite (Minor *et al.*, 2006), utilizing *SHELXD* for the heavy-atom search and *SHELXE* for initial phasing (Sheldrick, 2008). Next, heavy-atom positions were refined and phases improved by iterations of *MLPHARE* (Otwinowski, 1991) and *DM* (Cowtan, 1994) with NCS averaging included in calculations. The improved experimental maps were used to build an initial model with *RESOLVE* (Terwilliger, 2003). Further model building was performed manually in *Coot* (Emsley & Cowtan, 2004), while crystallographic maximum-likelihood and TLS refinement (six groups per protein monomer; Winn *et al.*, 2001) were carried out using *PHENIX* (Adams *et al.*, 2010) and *REFMAC5* (Murshudov *et al.*, 2011; Winn *et al.*, 2011). The refinement statistics are shown in Table 1.

#### 2.6. PDB accession code

The atomic coordinates and structure factors were originally deposited in the Protein Data Bank (PDB) with accession code 3nuw, which was later substituted by 3r1x.

### 3. Results and discussion

#### 3.1. Expression and purification

2-Oxo-3-deoxygalactonate kinase is composed of 292 amino-acid residues with a molecular weight of 31.1 kDa. SeMet-labeled protein was expressed and purified to homogeneity using the Midwest Center for Structural Genomics pipeline. For efficient production, KDGal kinase was expressed as a fusion protein containing additional elements at the N-terminus, namely MBP, which was introduced to enhance solubility, and a His<sub>6</sub> tag that facilitates purification by Ni<sup>2+</sup>-affinity chromatography. The MBP was removed *in situ* during protein expression in *E. coli* by a TVMV protease-catalyzed reaction; the His<sub>6</sub> tag was removed after metal-affinity chromatography using TEV protease. The final protein used for crystallization bears a non-native N-terminal SNA sequence.

#### 3.2. Structure solution and refinement

The structure of KDGal kinase was determined by the SAD method using a diffraction data set collected from a crystal of SeMet-labeled protein containing five selenium sites per monomer. The crystal belonged to the orthorhombic space group  $P2_12_12_1$  and contained four monomers (*A*, *B*, *C* and *D*) in the asymmetric unit. The final model fitted into electron-density maps comprises of residues Ser(–2)–Asn292 for protein molecule *A* and residues Met1–Asn292 for molecules *B*, *C* and *D*. The N-terminal Ser(–2)–Asn(–1)–Ala0 residues are cloning artifacts. In chains *B*, *C* and *D* these fragments have no visible electron density and have not been modeled.

In addition to the polypeptide chains, the asymmetric unit contents include 540 water molecules, 17 glycerol molecules and four formate ions. The structure was refined to final  $R$  and  $R_{\text{free}}$  factors of 0.174 and 0.199, respectively. The final model has an 0.016 Å r.m.s. deviation from ideal bond lengths. The quality of the stereochemistry of the structure was validated using the *MolProbity* server (Chen *et al.*, 2010), which indicated the expected distribution of the main-chain torsion angles on the Ramachandran plot.

### 3.3. Overall fold

The structure of KDGal kinase is arranged into N- and C-terminal  $\alpha/\beta$  domains separated by a deep cleft (Fig. 2). The core of the N-terminal domain (residues 1–82, 85–131 and 275–292) is established by a mixed  $\beta$ -sheet (E1) containing six strands, S3, S2, S1, S4, S8 and S7, where the S1, S4 and S8 strands are parallel. The  $\beta$ -sheet is sandwiched by two layers of helices. The solvent-exposed layer is formed by  $\alpha$ -helix H1 and a short fragment adopting a  $3_{10}$ -helix conformation (G2), while the layer neighboring a second domain is built from  $\alpha$ -helices H3 and H4 and the C-terminal  $\alpha$ -helix H11. Additionally, the N-terminal domain possesses a second anti-parallel  $\beta$ -sheet (E2) created by strands S5, S9 and S10 that flanks one opening of the interdomain groove. The C-terminal domain (residues 83–84 and 132–274) is composed of a mixed  $\beta$ -sheet (E3) localized at the domain interface and a cluster of helices (H5,  $3_{10}$  G6, H7, H8, H9 and H10) forming an external part of the domain. The E3 sheet includes strands S6, S13, S12, S11, S14 and S15, the last three of which interact with each other in a parallel manner. Within the helical cluster a central role is played by the H9 helix, around which all other helices are regularly distributed. The secondary-structure elements are connected by 25 loops (L1–L25), with six of them (L1, L2, L10, L12, L15 and L16) representing  $\beta$ -hairpins between consecutive strands of the  $\beta$ -sheet. By analogy to similar proteins, a large groove located at the interface between the two domains encompasses a catalytic center composed of nucleotide-binding and substrate-binding sites (see below).

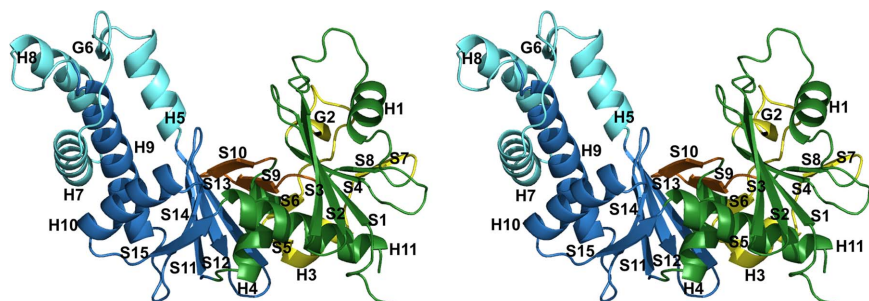
Comparison of the individual KDGal kinase molecules (*A*, *B*, *C* and *D*) does not show any significant differences between

the monomers. Not surprisingly, the fragments that exhibit some flexibility belong to coil regions, with the most pronounced conformational variations localized to the N-termini and the L3 and L4 loops. Pairwise superpositions of  $C^\alpha$  chains reveal the largest deviations to be between molecules *A* and *B* (r.m.s.d. of 0.56 Å), while the best agreement is observed for molecules *C* and *D* (r.m.s.d. of 0.36 Å).

### 3.4. Structural relatives

The overall architecture of KDGal kinase classifies it as a member of the ASKHA (Acetate and Sugar Kinases/Hsc70/Actin) superfamily (Buss *et al.*, 2001), which includes acetate (Buss *et al.*, 2001) and sugar kinases (Anderson *et al.*, 1978), heat-shock cognate 70 (hsc70; Flaherty *et al.*, 1990) and actin (Kabsch *et al.*, 1990). Members of this family either perform simple ATP hydrolysis (hsc70, actin) or catalyze the transfer of a phosphoryl group from ATP to their cognate substrates (kinases). Both activities depend on the presence of specific metal cations. Despite unambiguous fold similarity between KDGal kinase and other proteins belonging to this family, a *DALI*-based search for homologous structures (Holm & Rosenstrom, 2010) did not reveal any strong matches. The identified structural homologs show an amino-acid sequence identity of 12% or less. The closest relatives are butyrate kinase (PDB entry 1x9j;  $Z$  score 13.4, sequence identity 12%; Diao *et al.*, 2009), type III pantothenate kinase (PDB entry 3djc;  $Z$  score 13.5, sequence identity 11%; Y. Patskovsky, J. B. Bonanno, R. Romero, M. Dickey, C. Logan, M. Maletic, S. Wasserman, J. Koss, M. J. Sauder, S. K. Burley & S. C. Almo, unpublished work) and L-rhamnulose kinase (PDB entry 2cgg;  $Z$  score 13.0, sequence identity 11%; Grueninger & Schulz, 2006). Additional homologs with  $Z$  scores above 12 include propionate, acetate, glycerol, gluconate and xylulose kinases. Such low sequence similarity suggests that KDGal kinase is a relatively distant cousin of other ASKHA representatives and bears some unique structural motifs. Indeed, detailed comparison reveals significant differences.

Typically, both domains of the ASKHA proteins contain a common core with the topology  $\beta_1\beta_2\beta_3\alpha_1\beta_4\alpha_2\beta_5\alpha_3$ , while a distinctive feature of each family member is a specific pattern of insertions localized between the conserved elements. In the KDGal kinase structure the N-terminal fragment slightly breaks this canon because its ASKHA core is not fully preserved (Fig. 2). Specifically, there is no equivalent of helix  $\alpha_2$ . Instead, between the  $\beta_4$  and  $\beta_5$  elements (S4–S8 in KDGal kinase description) is an insertion contributing helices G2 and H3, strands S7 and S6 that extend sheets E1 and E3, and strand S5 from the E2 sheet. A second insertion is localized between the  $\beta_5$  and  $\alpha_3$  elements (S8–H4) and provides major strands to the E2 sheet (S9 and S10). Therefore, the topology of the ASKHA core of the KDGal kinase N-terminal domain is

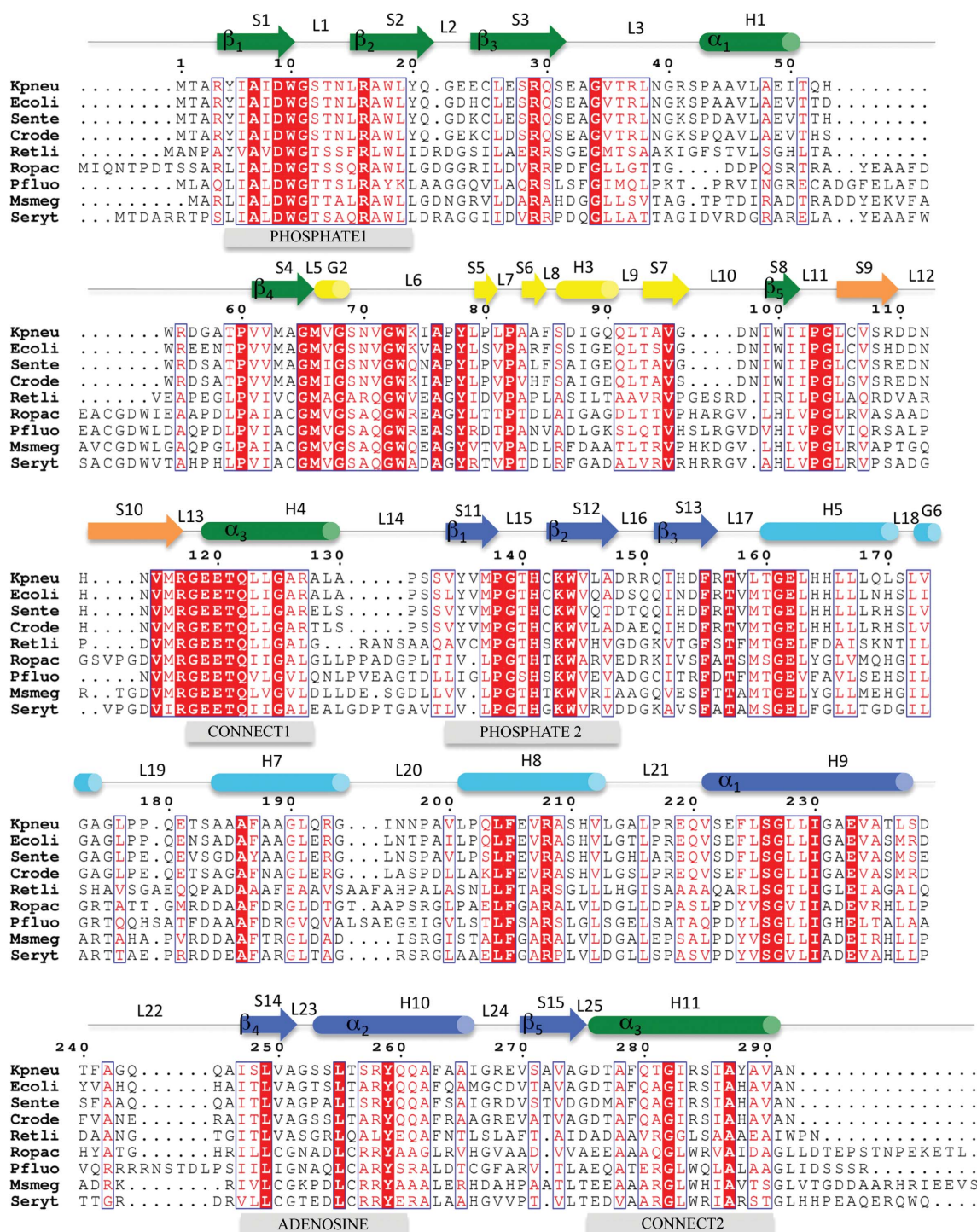


**Figure 2**

Overall structure of 2-oxo-3-deoxygalactonate kinase (stereoview). Green and blue colors indicate the conserved  $\beta\beta\beta\alpha\beta\beta\alpha$  ASKHA core within the N- and C-terminal domains, respectively. The  $\beta_4$ – $\beta_5$  and  $\beta_5$ – $\alpha_3$  insertions in the N-terminal domain are shown in yellow and orange. Cyan elements represent the  $\beta_3$ – $\alpha_1$  insertion in the C-terminal domain. Secondary-structure elements have been annotated according to *DSSP* (Kabsch & Sander, 1983).

$\beta_1\beta_2\beta_3\alpha_1\beta_4\beta_5\alpha_3$ . The C-terminal domain, on the other hand, is composed of the standard ASKHA motif with an insertion

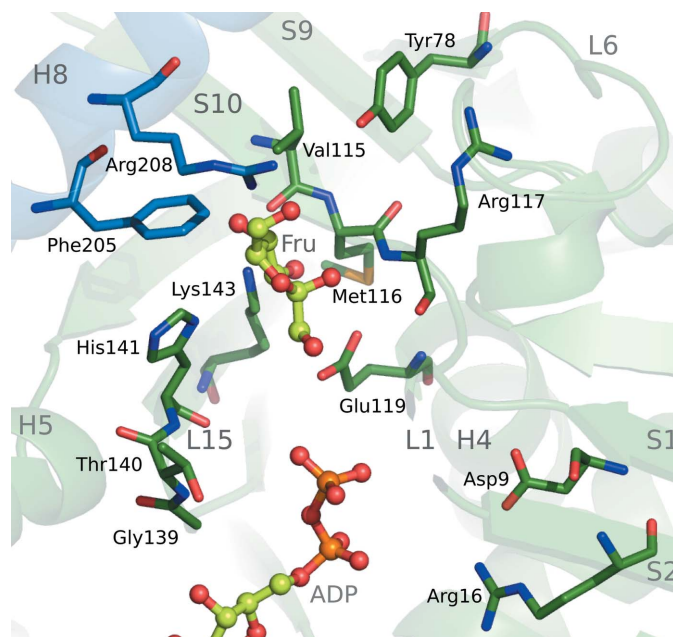
between the  $\beta_3$  and  $\alpha_1$  elements (S13–H9) consisting of four extra helices (H5, G6, H7 and H8).



**Figure 3**

Multiple sequence alignment of selected KDGal kinase homologs. Identical residues are highlighted in red and similar residues are shown as red letters. Secondary-structure elements derived from *K. pneumoniae* kinase are depicted by arrows (representing  $\beta$ -strands) and cylinders (representing helices). The color code is as in Fig. 2. All  $\beta$ -strands (S) and helices (H for  $\alpha$ -helix, G for  $3_{10}$ -helix) are numbered consecutively. Additionally, secondary-structure elements building the common ASKHA-family core are labeled with Greek letters. The positions of ASKHA signature motifs are marked. The sequences belong to Gram-positive bacteria (Kpneu, *K. pneumoniae*, current structure, UniProt accession code A6TFZ6; Ecoli, *Escherichia coli*, B5QUP1; Sente, *Salmonella enterica*, B5QUP1; Crode, *Citrobacter rodentium*, D2THW7; Retli, *Rhizobium etli*, Q2KBX1) and Gram-negative bacteria (Pfluo, *Pseudomonas fluorescens*, UniProt accession code Q3K8P2; Ropac, *Rhodococcus opacus*, C1AVX8; Msmeg, *Mycobacterium smegmatis*, A0R5F7; Seryt, *Saccharopolyspora erythraea*, A4FCZ7).

Another common feature of the ASKHA proteins is that the phosphoryl-group transfer is coupled to conformational rearrangement leading to domain closure (Hurley, 1996; Peters & Neet, 1978; Hoggett & Kellett, 1976; Miki & Kouyama, 1994). The transformation from the open to the closed state occurs through rotation around the hinge region established by the crossover  $\alpha_3$  helices (Diao *et al.*, 2009; Grueninger & Schulz, 2006; Nishimasu *et al.*, 2007; Anderson *et al.*, 1978; Bennett & Steitz, 1980). Crystal structures of different kinases complexed with substrates, nucleotides, metal ions or a combination of all ligands have demonstrated that it is substrate binding that triggers protein motion (Nishimasu *et al.*, 2007; Buss *et al.*, 2001; Diao *et al.*, 2009). Simultaneous binding of nucleotide, although not essential for domain closure, may stimulate it further. On the other hand, trapping ATP alone does not induce any conformational change in these enzymes (Nishimasu *et al.*, 2007; Simanshu *et al.*, 2005; Diao & Hasson, 2009). We note, however, that for those ASKHA representatives which bind only ATP and  $Mg^{2+}$  ions to perform their functions, such as actin and hsc70, nucleotide alone is able to induce closure of the interdomain cleft (Chik *et al.*, 1996; Wilbanks *et al.*, 1995). Since in the current KDGal kinase structure neither nucleotide nor substrate is bound, it is very likely that it represents an open conformation. This hypothesis is further supported by comparisons with the closest structural relatives. For example, superposition of KDGal kinase onto the structure of ligand-free L-rhamnulose kinase (PDB code 2cgg), illustrating an open state of the protein (Grueninger & Schulz, 2006), shows



**Figure 4**  
The putative substrate- and phosphate-binding site of KDGal kinase. Green color represents molecule *A* and blue color corresponds to molecule *B*. To indicate general positions for nucleotide and substrate binding, a  $\beta$ -L-fructose molecule (Fru) and a fragment of an ADP molecule from the L-rhamnulose kinase structure (PDB code 2cgg, molecule *A*) are shown in ball-and-stick representation. Selected secondary-structure elements are marked.

better agreement than superposition with the fructose-bound closed form of L-rhamnulose kinase (PDB code 2cgg), especially in the regions involved in ligand binding (data not shown).

### 3.5. ASKHA signature motifs in KDGal kinase

Analysis of several structures and sequences of ASKHA proteins identified five sets of conserved residues that establish signature motifs of the superfamily (Bork *et al.*, 1992). They are denoted as PHOSPHATE1 (P1), PHOSPHATE2 (P2), ADENOSINE (A), CONNECT1 and CONNECT2. The first three motifs are involved in nucleotide binding, whereas the two latter segments define an interdomain hinge. The P1 and P2 regions correspond to the  $\beta_1$ - $\beta_2$  hairpins from the N- and C-terminal domains, respectively. The most conservative fragment of the P1 motif bears either a DxG or ExG sequence (where *x* is any residue). The P2 fragment is more variable. For instance, actin, hsc70 and pantothenate kinase have DxG sequences and sugar kinases have a conserved GT block, whereas butyrate kinases have a triplet of glycine residues. Sequence alignment of KDGal kinases reveals that their P1 motif includes a DWG sequence (residues 9–11, loop L1), while P2 possesses a sugar kinase-like GT signature (residues 139–140, loop L15) (Figs. 3 and 4). By analogy to other ASKHA members, Asp9 from the P1 fingerprint (S1 strand) is likely to coordinate a metal cation bound to ATP (see below), while an invariant Arg16 from the same motif (S2 strand) may interact with a phosphoryl group in a manner similar to that observed in glycerol kinase (Hurley *et al.*, 1993) and L-rhamnulose kinase (Grueninger & Schulz, 2006).

The ADENOSINE motif is localized in the  $\beta_4$ - $\alpha_2$  region of the C-terminal domain and the most important fragment of this segment contains the sequence *-hhhGfxfh-* [where *h* is purely hydrophobic (VLIFWY), *f* is partly hydrophobic (VLIFWYMCGATKHR) and *x* is any residue; Bork *et al.*, 1992], with the *-Gf-* block usually consisting of two consecutive glycine residues. In several ASKHA-family representatives the loop linking elements  $\beta_4$  and  $\alpha_2$  is  $\sim 7$  residues long and part of it folds into an  $\alpha$ -helix (Buss *et al.*, 2001; Diao & Hasson, 2009; Simanshu *et al.*, 2005) or  $3_{10}$ -helix (Nishimasu *et al.*, 2007; Grueninger & Schulz, 2006). The equivalent region of KDGal kinase, namely the L23 loop (Gly252–Ser253), is not only much shorter (two residues) but also does not have the consensus sequence (Fig. 3).

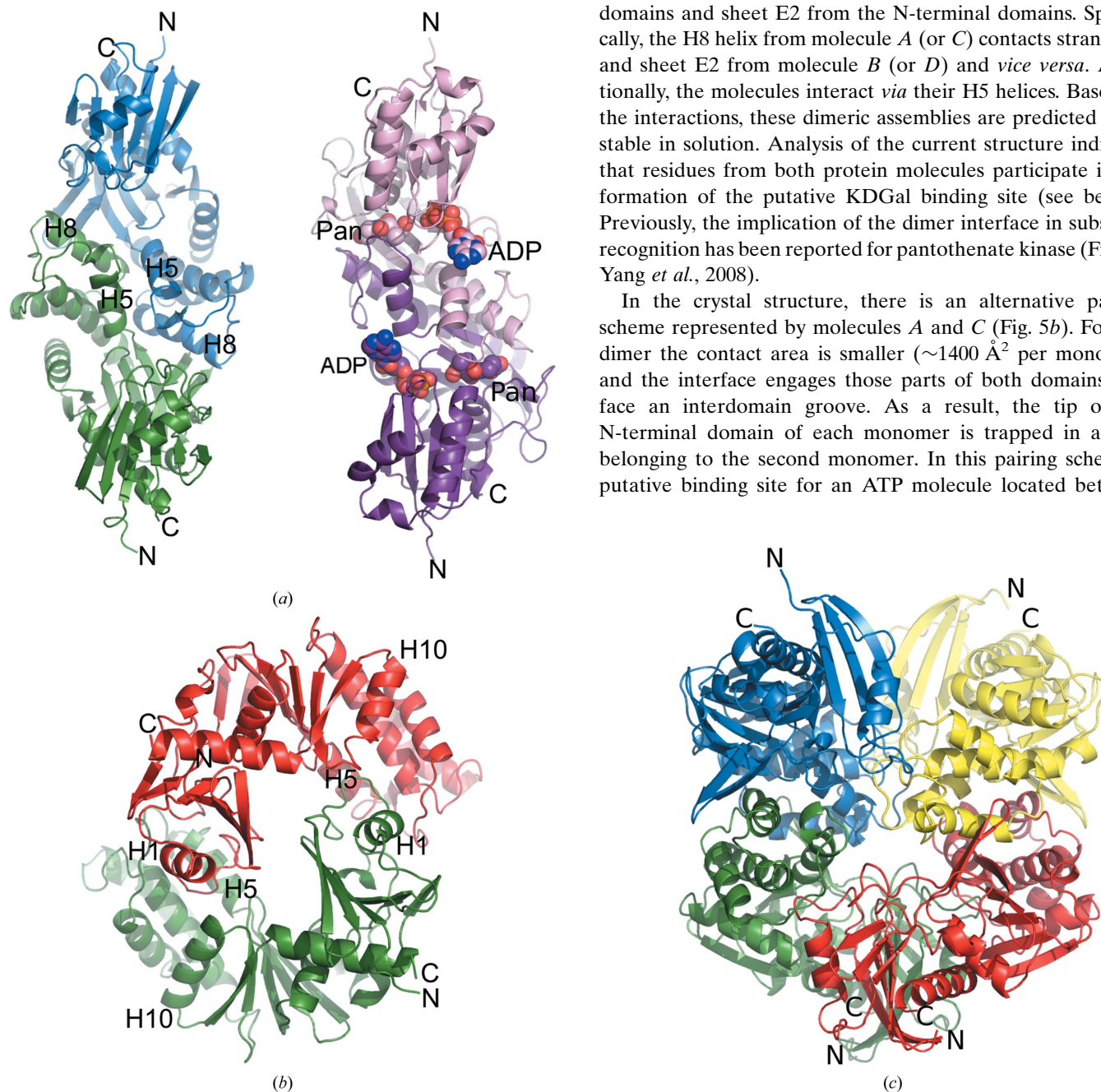
The CONNECT1 and CONNECT2 motifs correspond to the  $\alpha_3$  helices from the N- and C-terminal domains, respectively, and fragments directly preceding these structural elements. The CONNECT1 fragment crosses from the N-terminal domain to the C-terminal domain, while the CONNECT2 motif crosses in the opposite direction, resulting in a close helix–helix contact. Upon substrate binding, these helices move with respect to each other and to the core  $\beta$ -sheets, leading to significant conformational changes that are required for appropriate positioning of reactants within the catalytic apparatus. Typically, the CONNECT motifs contain conserved alanine and glycine residues, which are

localized at the helix–helix interface. These residues also have equivalents in KDGal kinase, where CONNECT1 bears an invariant Gly125 residue while CONNECT2 contains Ala278 and Gly282 (Fig. 3). In addition, it has been shown that the N-terminus of helix  $\alpha_3$  from the CONNECT1 fragment contributes a residue (Asp, Glu or Gln) that functions as a general base during catalysis or is involved in metal binding. The corresponding residue in KDGal kinase is Glu119 (Figs. 3 and 4).

### 3.6. KDGal kinase oligomerization

In the crystal lattice, the KDGal kinase molecules assemble into two types of noncrystallographic dimer (Figs. 5*a* and 5*b*). In the more compact dimer, molecule *A* interacts with molecule *B*. The same module is formed by molecule *C* and molecule *D*. These associations bury about 15% of solvent-accessible surface per monomer, which is 1719 Å<sup>2</sup> for the *A*–*B* pair and 1759 Å<sup>2</sup> for the *C*–*D* pair, as calculated by the PISA server (Krissinel & Henrick, 2007). The intermolecular interface is established by external fragments of the C-terminal domains and sheet E2 from the N-terminal domains. Specifically, the H8 helix from molecule *A* (or *C*) contacts strand S13 and sheet E2 from molecule *B* (or *D*) and *vice versa*. Additionally, the molecules interact *via* their H5 helices. Based on the interactions, these dimeric assemblies are predicted to be stable in solution. Analysis of the current structure indicates that residues from both protein molecules participate in the formation of the putative KDGal binding site (see below). Previously, the implication of the dimer interface in substrate recognition has been reported for pantothenate kinase (Fig. 5*a*; Yang *et al.*, 2008).

In the crystal structure, there is an alternative pairing scheme represented by molecules *A* and *C* (Fig. 5*b*). For this dimer the contact area is smaller (~1400 Å<sup>2</sup> per monomer) and the interface engages those parts of both domains that face an interdomain groove. As a result, the tip of the N-terminal domain of each monomer is trapped in a cleft belonging to the second monomer. In this pairing scheme a putative binding site for an ATP molecule located between



**Figure 5** Oligomerization of 2-oxo-3-deoxygalactonate kinase. (a) *A*–*B* dimer (left). Molecule *A* is shown in green and molecule *B* is shown in blue. For comparison, the functional dimer of pantothenate kinase in complex with ADP and pantothenate (Pan) is presented (right; PDB entry 3bf1; chain *A*, purple; chain *B*, pink; Yang *et al.*, 2008). Ligand molecules are depicted in sphere representation. Selected secondary-structure elements are marked. (b) *A*–*C* dimer. Molecule *A* is shown in green and molecule *C* is shown in red. (c) KDGal kinase tetramer (molecule *A*, green; molecule *B*, blue; molecule *C*, red; molecule *D*, yellow).



the H5 and H10 helices is blocked (see below), locking the enzyme in an inactive state. Although the contact area as well as the network of interactions for the *A–C* dimer are comparable to those for *A–B* and *C–D* dimers, the limited number of hydrophobic interactions at the *A–C* interface suggests that this pairing may represent a crystal-packing artifact.

The combination of the two types of dimers generates a tetrameric assembly which can be described as a dimer of dimers, *AB + CD* (Fig. 5*c*). Size-exclusion chromatography has demonstrated that the homologous protein from *Mycobacterium* sp. indeed exists as a tetramer in solution (Szumilo, 1983). However, size-exclusion chromatography of the *Klebsiella* KDGal kinase performed in the presence and absence of ATP shows only dimers. This suggests that the minimal biological unit is a dimer with an *A–B*-type interface.

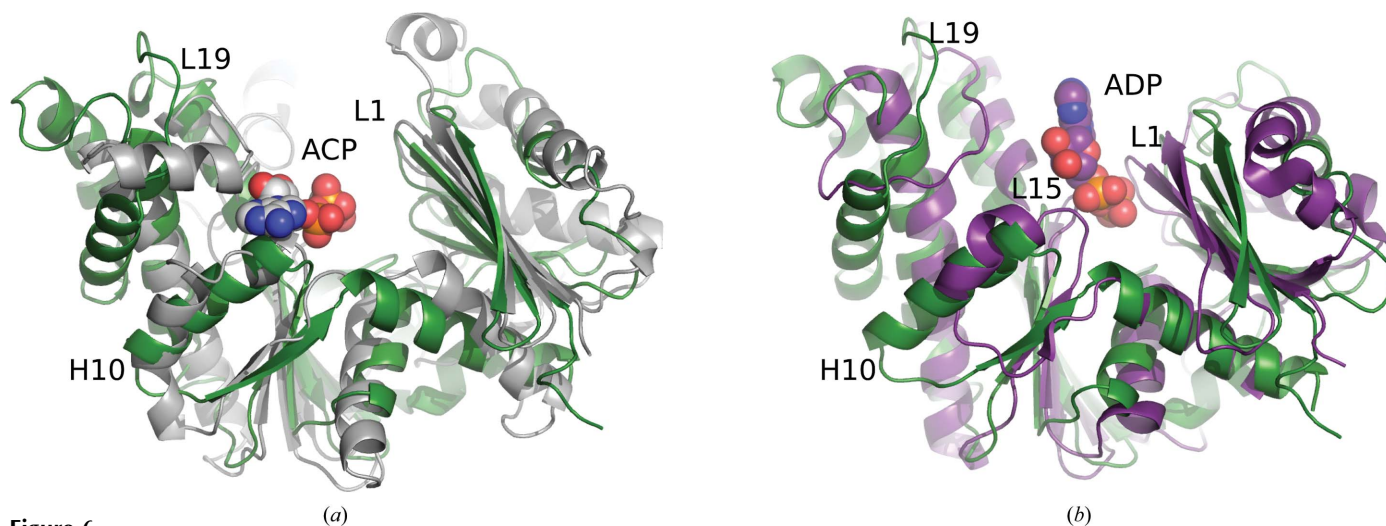
### 3.7. Nucleotide-binding site

In ASKHA phosphotransferases the nucleotide molecule binds in a subcavity of the C-terminal domain, with the adenine ring sandwiched in a hydrophobic pocket created by the ADENOSINE motif and one of the helices from the  $\beta_3\text{--}\alpha_1$  insertion (Fig. 6*a*). The ring is primarily anchored by hydrophobic interactions and only in rare cases does it form water-mediated hydrogen bonds with a protein molecule, as is observed, for example, in the structure of L-rhamnulose kinase (Grueninger & Schulz, 2006). The adenine amino group is typically exposed to solvent but occasionally forms hydrogen bonds with a protein molecule, as illustrated by the structure of hexokinase (Nishimasu *et al.*, 2007). The ribose moiety usually interacts with the insertion elements from the C-terminal domain. These fragments of ASKHA proteins differ between family members; as a consequence, there is no universal pattern of sugar binding. The most conserved contacts between nucleotide and protein molecules involve phosphoryl groups, which typically bind to main-chain amide

groups from the P2 loop (Buss *et al.*, 2001; Diao & Hasson, 2009; Simanshu *et al.*, 2005). Additional interactions with side chains have also been observed (Nishimasu *et al.*, 2007). After domain closure, the P1 motif also stabilizes a nucleotide molecule within the active site. Interactions between the P1 element and phosphoryl groups occur either through direct hydrogen bonds, for example in L-rhamnulose (Grueninger & Schulz, 2006) and glycerol kinases (Hurley *et al.*, 1993), or are mediated by a metal cation which is coordinated by a conserved carboxylate residue from the  $\beta_1$  strand, as has been observed in the structures of hexokinase (Nishimasu *et al.*, 2007) and hsc70 (Flaherty *et al.*, 1994).

Although most of the above elements are conserved in the KDGal kinase structure (see above), one significant difference is noticeable. Specifically, the protein has no well defined cavity for adenine-ring binding. First of all, as mentioned above, the region corresponding to the ADENOSINE motif does not have the usual length and amino-acid composition. In consequence, it does not superpose well with the equivalent fragments from related structures (Fig. 6). In addition to this, KDGal kinase does not contain an insertion helix that builds a second face of the groove. This part of the enzyme adopts a loop conformation (L19) and is localized far from the ADENOSINE motif. As a result, the putative adenine-binding site has no clear pocket definition and remains wide open. This feature is utilized by the *A–C* dimer, in which the putative nucleotide-binding site of one monomer is occupied by a fragment of the complementary molecule.

The lack of capping helix resembles the structure of type III pantothenate kinase (Fig. 6*a*; Yang *et al.*, 2008). However, in this case the insertion loop remains in close proximity to the ADENOSINE fragment, leaving the cavity much shallower than that observed in KDGal kinase. In consequence, the base moiety is not buried in a dedicated subcavity but is positioned above the P2 hairpin and oriented parallel to a helix that directly succeeds strand  $\beta_1$ . The only interactions between the



**Figure 6** Comparison of 2-oxo-3-deoxygalactonate kinase with similar structures. (a) Superposition of KDGal kinase (chain A, green) and butyrate kinase 2 in complex with 5'-adenylyl-( $\beta,\gamma$ -methylene)diphosphonate (ACP) (PDB entry 1saz; gray; Diao & Hasson, 2009). The nucleotide ligand is shown in sphere representation. (b) Superposition of KDGal kinase (chain A, green) and pantothenate kinase in complex with ADP and pantothenate (PDB code 3bf1; chain A, purple; Yang *et al.*, 2008). For clarity, the pantothenate moiety is not shown. The nucleotide ligand is shown in sphere representation.

adenine ring and the protein molecule involve stacking with a phenylalanine side chain from the helix. It is likely that in KDGal kinase the nucleotide also binds in an unusual way, but the discrepancy in the organization of putative ATP-binding sites between the current enzyme and related proteins does not permit reliable prediction of protein–nucleotide interactions.

In the absence of natural ligand, the nucleotide-binding sites of molecules *B* and *D* of the KDGal kinase structure are filled with glycerol molecules soaked in during cryoprotection and solvent molecules (Fig. 7). Molecule *B* binds four glycerol moieties denoted as GOL1B, GOL2B, GOL3B and GOL4B, whereas in molecule *D* five ligand molecules could be identified: GOL1D, GOL2D, GOL3D, GOL4D and GOL5D. In both protein monomers GOL1 sits in a small niche near the H5 and H9 helices and is hydrogen bonded to the hydroxyl group of the conserved Ser226 residue. The GOL2 molecules are loosely packed above GOL1, near the L19 loops. The L15 loop (PHOSPHATE1) interacts with the GOL3 moieties. Specifically, its two main-chain amide groups form hydrogen bonds to glycerol –OH groups. The GOL4 molecules are anchored by a direct hydrogen bond to Asp276 and additional water-mediated interactions. GOL5D is bound to the side chain of Asp9.

### 3.8. Substrate-binding site

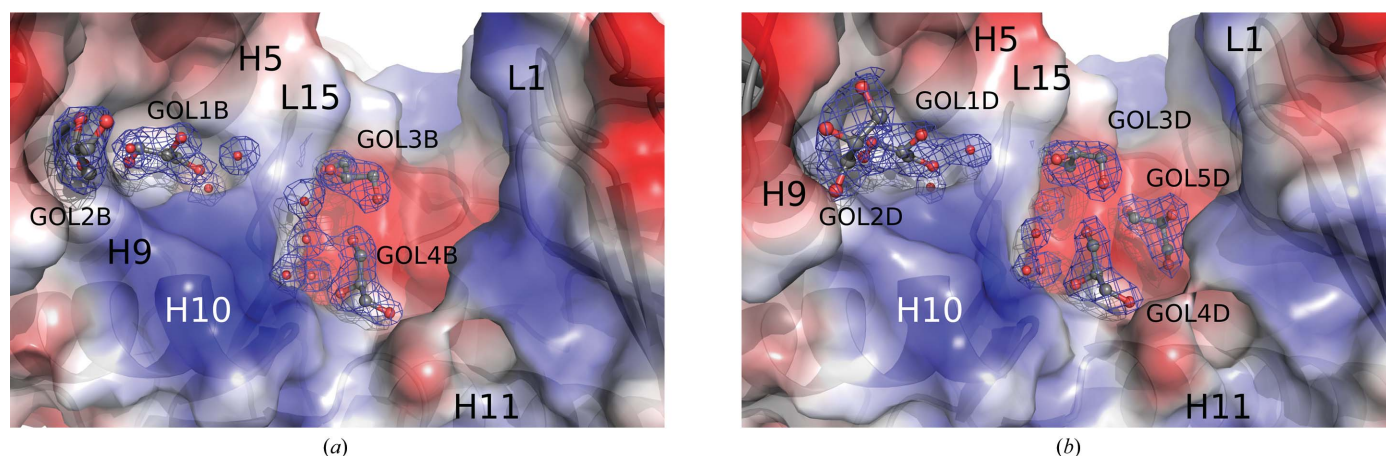
The comparison of KDGal kinase with similar structures suggests the location of the substrate-binding pocket. The putative 2-oxo-3-deoxygalactonate-binding site is localized near the dimer interface (*A–B* or *C–D* pairing) in the vicinity of the E2 sheet, which represents one of the insertions of the N-terminal domain (Fig. 4). The main residues surrounding the pocket belong to the S2 strand, the H4 helix and the loop between them (L13). The side chains of Val115 and Met116 contributed by S2 as well as Arg117 from L13 point outside the cavity; thus, only their main-chain atoms line the pocket

and possibly interact with a substrate molecule. Additional residues that are likely to participate in KDGal binding include the conserved Tyr78 from loop L6, His141 and Lys143 from L15 (PHOSPHATE1) as well as Glu119 from the H4 helix (CONNECT1). The latter amino acid may play the role of general base during catalysis and could be involved in Mg<sup>2+</sup> chelation. The second protein monomer provides further residues, such as the invariant Phe205 and Arg208, that potentially anchor a KDGal molecule.

### 4. Conclusions

In the present study, we have determined a 2.1 Å resolution crystal structure of 2-oxo-3-deoxygalactonate kinase, which represents one of the enzymes utilized in the De Ley–Doudoroff pathway of galactose metabolism. Our analysis demonstrates that the protein belongs to the functionally diverse ASKHA superfamily of phosphotransferases. As in the other members of the family, the kinase is composed of two domains separated by a deep groove in which the catalytic reaction takes place. In the crystal lattice, protein molecules associate into a dimer of dimers, but a size-exclusion chromatography experiment indicated the presence of only dimeric species.

The putative substrate-binding site is created mostly by the N-terminal domain, especially by those fragments that are unique to KDGal kinase. On the other hand, the pocket in which the nucleotide docks is formed by the C-terminal domain, in which a loop that is found in all family members participates in extensive interactions with phosphoryl groups. Despite a clear similarity between KDGal kinase and other ASKHA proteins, in particular pantothenate kinase, it is obvious that the ATP-binding site from the current enzyme does not resemble any known architecture. Thus, it is difficult to predict how the nucleotide moiety would bind. To answer this question, a structure of a kinase–ATP complex will need to be determined.



**Figure 7**

Putative nucleotide-binding site of KDGal kinase. The protein molecule is represented by electrostatic surface potential, in which the blue color corresponds to positively charged areas, red to negatively charged areas and white to neutral regions. In the absence of natural ligand, the pockets of molecules *B* (*a*) and *D* (*b*) are occupied by glycerol molecules soaked into the crystal lattice during cryoprotection. The glycerol moieties are shown in ball-and-stick representation in a  $2F_o - F_c$  electron-density map contoured at the  $1\sigma$  level.

We thank the members of the Structural Biology Center and the Midwest Center for Structural Genomics at Argonne National Laboratory for their help in conducting these experiments. This work was supported by National Institutes of Health grants GM074942 and GM094585, by the US Department of Energy, Office of Biological and Environmental Research under contract DE-AC02-06CH11357 (AJ) and by a scholarship from the Foundation for Polish Science (KM). The submitted manuscript has been created by UChicago Argonne LLC, Operator of Argonne National Laboratory ('Argonne'). Argonne, a US Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357.

## References

- Acosta, P. B. & Gross, K. C. (1995). *Eur. J. Pediatr.* **154**, S87–S92.
- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- Anderson, C. M., Stenkamp, R. E., McDonald, R. C. & Steitz, T. A. (1978). *J. Mol. Biol.* **123**, 207–219.
- Arias, A. & Cerveñansky, C. (1986). *J. Bacteriol.* **167**, 1092–1094.
- Aslanidis, C. & de Jong, P. J. (1990). *Nucleic Acids Res.* **18**, 6069–6074.
- Babbitt, P. C., Mrachko, G. T., Hasson, M. S., Huisman, G. W., Kolter, R., Ringe, D., Petsko, G. A., Kenyon, G. L. & Gerlt, J. A. (1995). *Science*, **267**, 1159–1161.
- Bennett, W. S. & Steitz, T. A. (1980). *J. Mol. Biol.* **140**, 183–209.
- Bork, P., Sander, C. & Valencia, A. (1992). *Proc. Natl Acad. Sci. USA*, **89**, 7290–7294.
- Brechtel, E., Huwig, A. & Giffhorn, F. (2002). *Appl. Environ. Microbiol.* **68**, 582–587.
- Buss, K. A., Cooper, D. R., Ingram-Smith, C., Ferry, J. G., Sanders, D. A. & Hasson, M. S. (2001). *J. Bacteriol.* **183**, 680–686.
- Caputto, R. & Leloir, L. R. (1949). *J. Biol. Chem.* **179**, 497.
- Caputto, R., Leloir, L. R., Cardini, C. E. & Paladini, A. C. (1950). *J. Biol. Chem.* **184**, 333–350.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* **D66**, 12–21.
- Chik, J. K., Lindberg, U. & Schutt, C. E. (1996). *J. Mol. Biol.* **263**, 607–623.
- Cooper, R. A. (1978). *Arch. Microbiol.* **118**, 199–206.
- Cowtan, K. (1994). *Jnt CCP4/ESF-EACBM Newsl. Protein Crystallogr.* **31**, 34–38.
- Cuatrecasas, P. & Segal, S. (1966a). *J. Biol. Chem.* **241**, 5904–5909.
- Cuatrecasas, P. & Segal, S. (1966b). *J. Biol. Chem.* **241**, 5910–5918.
- Deacon, J. & Cooper, R. A. (1977). *FEBS Lett.* **77**, 201–205.
- De Ley, J. & Doudoroff, M. (1957). *J. Biol. Chem.* **227**, 745–757.
- Diao, J. & Hasson, M. S. (2009). *J. Bacteriol.* **191**, 2521–2529.
- Diao, J., Ma, Y. D. & Hasson, M. S. (2009). *Proteins*, doi:10.1002/prot.22610.
- Donnelly, M. I., Zhou, M., Millard, C. S., Clancy, S., Stols, L., Eschenfeldt, W. H., Collart, F. R. & Joachimiak, A. (2006). *Protein Expr. Purif.* **47**, 446–454.
- Elshafei, A. M. & Abdel-Fatah, O. M. (2001). *Enzyme Microb. Technol.* **29**, 76–83.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Entner, N. & Doudoroff, M. (1952). *J. Biol. Chem.* **196**, 853–862.
- Eschenfeldt, W. H., Lucy, S., Millard, C. S., Joachimiak, A. & Mark, I. D. (2009). *Methods Mol. Biol.* **498**, 105–115.
- Fekete, E., Karaffa, L., Sándor, E., Bányai, I., Seiboth, B., Gyémánt, G., Sepsí, A., Szentirmai, A. & Kubicek, C. P. (2004). *Arch. Microbiol.* **181**, 35–44.
- Flaherty, K. M., DeLuca-Flaherty, C. & McKay, D. B. (1990). *Nature (London)*, **346**, 623–628.
- Flaherty, K. M., Wilbanks, S. M., DeLuca-Flaherty, C. & McKay, D. B. (1994). *J. Biol. Chem.* **269**, 12899–12907.
- Fridovich-Keil, J. L. (2006). *J. Cell. Physiol.* **209**, 701–705.
- Gross, K. C. & Acosta, P. B. (1991). *J. Inher. Metab. Dis.* **14**, 253–258.
- Grueninger, D. & Schulz, G. E. (2006). *J. Mol. Biol.* **359**, 787–797.
- Hoggett, J. G. & Kellett, G. L. (1976). *Eur. J. Biochem.* **66**, 65–77.
- Holden, H. M., Rayment, I. & Thoden, J. B. (2003). *J. Biol. Chem.* **278**, 43885–43888.
- Holm, L. & Rosenstrom, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
- Hurley, J. H. (1996). *Annu. Rev. Biophys. Biomol. Struct.* **25**, 137–162.
- Hurley, J. H., Faber, H. R., Worthylake, D., Meadow, N. D., Roseman, S., Pettigrew, D. W. & Remington, S. J. (1993). *Science*, **259**, 673–677.
- Jakobs, C., Schweitzer, S. & Dorland, B. (1995). *Eur. J. Pediatr.* **154**, S50–S52.
- Jongh, W. A. de, Bro, C., Ostergaard, S., Regenber, B., Olsson, L. & Nielsen, J. (2008). *Biotechnol. Bioeng.* **101**, 317–326.
- Kabsch, W., Mannherz, H. G., Suck, D., Pai, E. F. & Holmes, K. C. (1990). *Nature (London)*, **347**, 37–44.
- Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.
- Kim, Y., Dementieva, I., Zhou, M., Wu, R., Lezondra, L., Quartey, P., Joachimiak, G., Korolev, O., Li, H. & Joachimiak, A. (2004). *J. Struct. Funct. Genomics*, **5**, 111–118.
- Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
- Lai, K., Elsas, L. J. & Wierenga, K. J. (2009). *IUBMB Life*, **61**, 1063–1074.
- Leloir, L. F. (1951). *Arch. Biochem.* **33**, 186–190.
- Leslie, N. D. (2003). *Annu. Rev. Nutr.* **23**, 59–80.
- Mehta, D. V., Kabir, A. & Bhat, P. J. (1999). *Biochim. Biophys. Acta*, **1454**, 217–226.
- Miki, M. & Kouyama, T. (1994). *Biochemistry*, **33**, 10171–10177.
- Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. (2006). *Acta Cryst.* **D62**, 859–866.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.
- Nallamsetty, S., Kapust, R. B., Tözsér, J., Cherry, S., Tropea, J. E., Copeland, T. D. & Waugh, D. S. (2004). *Protein Expr. Purif.* **38**, 108–115.
- Nishimasu, H., Fushinobu, S., Shoun, H. & Wakagi, T. (2007). *J. Biol. Chem.* **282**, 9923–9931.
- Nobelmann, B. & Lengeler, J. W. (1996). *J. Bacteriol.* **178**, 6790–6795.
- Otwinowski, Z. (1991). *Proceedings of the CCP4 Study Weekend. Isomorphous Replacement and Anomalous Scattering*, edited by W. Wolf, P. R. Evans & A. G. W. Leslie, pp. 80–86. Warrington: Daresbury Laboratory.
- Peters, B. A. & Neet, K. E. (1978). *J. Biol. Chem.* **253**, 6826–6831.
- Rooijen, R. J. van, van Schalkwijk, S. & de Vos, W. M. (1991). *J. Biol. Chem.* **266**, 7176–7181.
- Schädel, C., Blöchl, A., Richter, A. & Hoch, G. (2010). *Plant Physiol. Biochem.* **48**, 1–8.
- Schneider, K. H., Jäkel, G., Hoffmann, R. & Giffhorn, F. (1995). *Microbiology*, **141**, 1865–1873.
- Shakeri-Garakani, A., Brinkötter, A., Schmid, K., Turgut, S. & Lengeler, J. W. (2004). *Mol. Genet. Genomics*, **271**, 717–728.
- Sheldrick, G. M. (2008). *Acta Cryst.* **A64**, 112–122.
- Simanshu, D. K., Savithri, H. S. & Murthy, M. R. (2005). *J. Mol. Biol.* **352**, 876–892.
- Stouthamer, A. H. (1961). *Biochim. Biophys. Acta*, **48**, 484–500.
- Szumilo, T. (1981). *J. Bacteriol.* **148**, 368–370.
- Szumilo, T. (1983). *Acta Microbiol. Pol.* **32**, 47–52.
- Terwilliger, T. C. (2003). *Acta Cryst.* **D59**, 38–44.
- Trucco, R. E. & Caputto, R. (1948). *Arch. Biochem.* **18**, 137–146.
- Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. (1993). *J. Mol. Biol.* **229**, 105–124.

- Vos, W. M. de, Boerrigter, I., van Rooyen, R. J., Reiche, B. & Hengstenberg, W. (1990). *J. Biol. Chem.* **265**, 22554–22560.
- Walsh, M. A., Dementieva, I., Evans, G., Sanishvili, R. & Joachimiak, A. (1999). *Acta Cryst.* **D55**, 1168–1173.
- Wehrli, S. L., Berry, G. T., Palmieri, M., Mazur, A., Elsas, L. III & Segal, S. (1997). *Pediatr. Res.* **42**, 855–861.
- Wilbanks, S. M., Chen, L., Tsuruta, H., Hodgson, K. O. & McKay, D. B. (1995). *Biochemistry*, **34**, 12095–12106.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). *Acta Cryst.* **D57**, 122–133.
- Wong, T. Y. & Yao, X.-T. (1994). *Appl. Environ. Microbiol.* **60**, 2065–2068.
- Yabe-Nishimura, C. (1998). *Pharmacol. Rev.* **50**, 21–33.
- Yang, K., Strauss, E., Huerta, C. & Zhang, H. (2008). *Biochemistry*, **47**, 1369–1380.
- Zeng, L., Das, S. & Burne, R. A. (2010). *J. Bacteriol.* **192**, 2434–2444.