Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure and *in silico* substrate-binding mode of ADP-L-*glycero*-D-*manno*-heptose 6-epimerase from *Burkholderia thailandensis*

Mi-Sun Kim, Areum Lim, Seung Won Yang, Jimin Park, Daeun Lee and Dong Hae Shin*

The Center for Cell Signaling and Drug Discovery Research, College of Pharmacy, Division of Life and Pharmaceutical Sciences, Ewha Womans University, Seoul 120-750, Republic of Korea

Correspondence e-mail: dhshin55@ewha.ac.kr

Received 6 September 2012
Accepted 11 January 2013

PDB Reference:
ADP-L-*glycero*-D-*manno*-heptose 6-epimerase, 4ej0

ADP-L-*glycero*-D-*manno*-heptose 6-epimerase (AGME), the product of the *rfaD* gene, is the last enzyme in the heptose-biosynthesis pathway; it converts ADP-D-*glycero*-D-*manno*-heptose (ADP-D,D-Hep) to ADP-L-*glycero*-D-*manno*-heptose (ADP-L,D-Hep). AGME contains a catalytic triad involved in catalyzing hydride transfer with the aid of $NADP^+$. Defective lipopolysaccharide is found in bacterial mutants lacking this gene. Therefore, it is an interesting target enzyme for a novel epimerase inhibitor for use as a co-therapy with antibiotics. The crystal structure of AGME from *Burkholderia thailandensis* (*Bt*AGME), a surrogate organism for studying the pathogenicity of melioidosis caused by *B. pseudomallei*, has been determined. The crystal structure determined with co-purified $NADP^+$ revealed common as well as unique structural properties of the AGME family when compared with UDP-galactose 4-epimerase homologues. They form a similar architecture with conserved catalytic residues. Nevertheless, there are differences in the substrate- and cofactor-binding cavities and the oligomerization domains. Structural comparison of *Bt*AGME with AGME from *Escherichia coli* indicates that they may recognize their substrate in a 'lock-and-key' fashion. Unique structural features of *Bt*AGME are found in two regions. The first region is the loop between β8 and β9, affecting the binding affinity of *Bt*AGME for the ADP moiety of ADP-D,D-Hep. The second region is helix α8, which induces decamerization at low pH that is not found in other AGMEs. With the E210G mutant, it was observed that the resistance of the wild type to acid-induced denaturation is related to the decameric state. An *in silico* study was performed using the *Surflex-Dock GeomX* module of the *SYBYL-X* 1.3 software to predict the catalytic mechanism of *Bt*AGME with its substrate, ADP-D,D-Hep. In the *in silico* study, the C7″ hydroxymethyl group of ADP-D,D-Hep is predicted to form hydrogen bonds to Ser116 and Gln293. With the aid of these interactions, the hydroxyl of Tyr139 forms a hydrogen bond to O6″ of ADP-D,D-Hep and the proton at C6″ orients closely to C4 of $NADP^+$. Therefore, the *in silico* study supports a one-base mechanism as a major catalytic pathway, in which Tyr139 solely functions as a catalytic acid/base residue. These results provide a new insight into the development of an epimerase inhibitor as an antibiotic adjuvant against melioidosis.

1. Introduction

The lipopolysaccharide (LPS) of the outer membrane of Gram-negative bacteria blocks the entry of toxic hydrophobic compounds by maintaining the structural integrity of the bacterial outer membrane (Nikaido & Vaara, 1985). It is composed of a tripartite structure consisting of lipid A,

a conserved core oligosaccharide region and a repeating saccharide O antigen (Raetz & Whitfield, 2002). It plays key molecular roles in the survival of many infectious diseases as well as in the virulence of many human pathogens such as *Haemophilus influenzae*, *Salmonella typhimurium*, *Borrelia burgdorferi etc.* (Heumann *et al.*, 1998). The carbohydrate L-*glycero*-D-*manno*-heptose is an important component of LPS and is synthesized in a complex series of enzymatic steps. ADP-L-*glycero*-D-*manno*-heptose 6-epimerase (AGME or HldD, formerly RfaD), the last enzyme in the heptose-biosynthesis pathway, is involved in the epimerization at the C6″ position that converts ADP-D-*glycero*-D-*manno*-heptose (ADP-D,D-Hep) to ADP-L-*glycero*-D-*manno*-heptose (ADP-L,D-Hep). Mutants lacking the *rfaD* gene accumulate ADP-D,D-Hep and synthesize defective LPS (Schnaitman & Klena, 1993). AGME belongs to the extended short-chain dehydrogenases/reductases (eSDRs), which form a large functionally heterogeneous protein family with a similar structural architecture despite low sequence identity (Persson *et al.*, 2003). The key catalytic triad of serine, tyrosine and lysine (Deacon *et al.*, 2000), which is highly conserved in the eSDR family, functions together with the tightly bound $NADP^+$ cofactor of AGME to oxidize directly at the C6″ position of ADP-D,D-Hep (Morrison *et al.*, 2005). Structural study of the Y140F mutant of AGME from *Escherichia coli* (*Ec*AGME) in complex with ADP-β-D-mannose led to the proposal of a

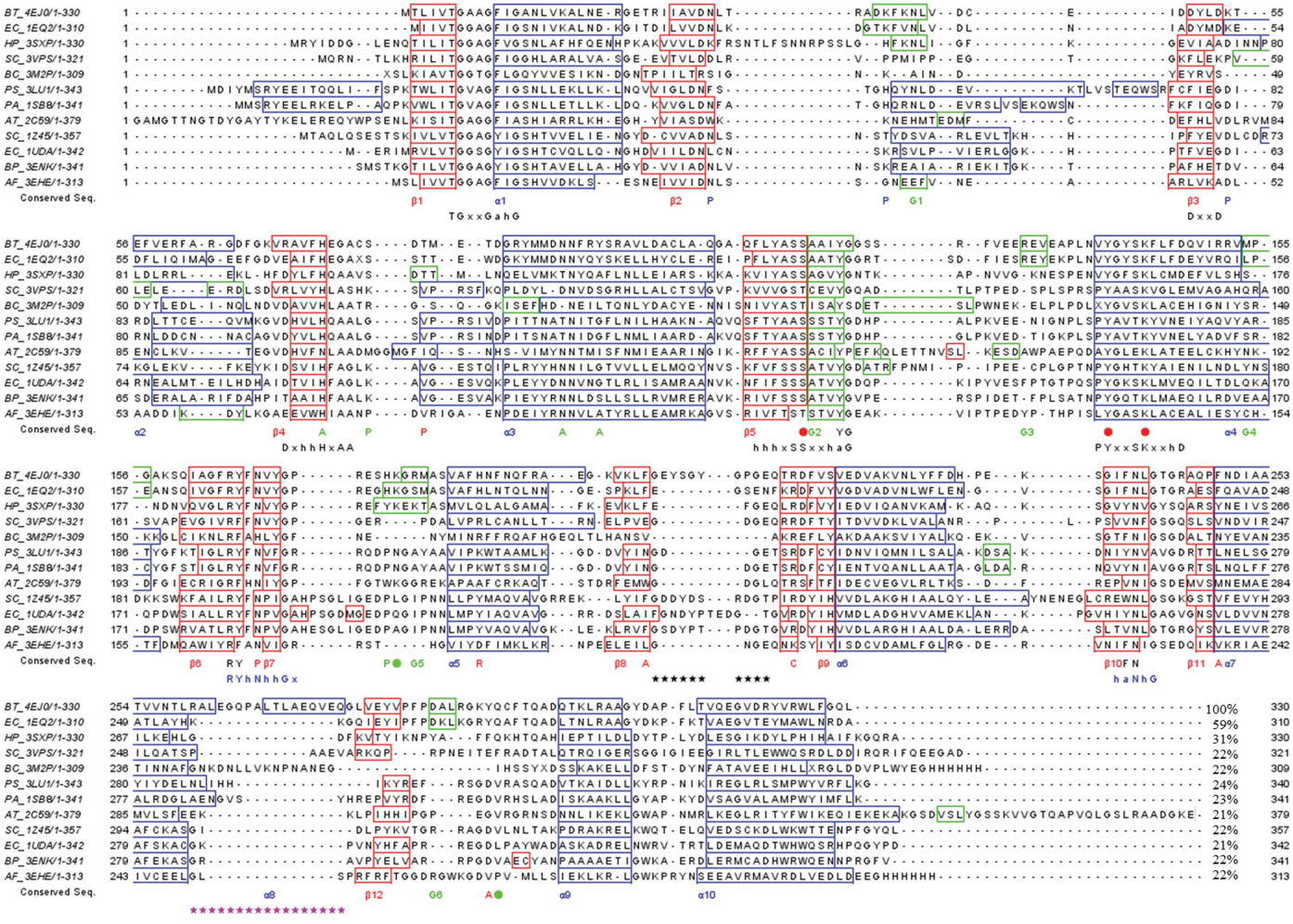


Figure 1
Sequence comparison between *Bt*AGME and GalE homologues. Abbreviations are as follows: AGME from *B. thailandensis*, BT_4EJ0; AGME from *E. coli*, EC_1EQ2; AGME from *H. pylori* G27, HP_3SXP; TunA from *Streptomyces chartreusis* NRRL 3882, SC_3VPS; WbgU from *Bacillus cereus* ATCC 14579, BC_3M2P; WbgU from *Plesiomonas shigelloides*, PS_3LU1; GalE from *Saccharomyces cerevisiae*, SC_1Z45; GalE from *E. coli*, EC_1UDA; GalE from *B. pseudomallei* 1710b, BP_3ENK; GalE from *A. fulgidus*, AF_3EHE. Secondary-structure elements of AGMEs and GalE homologues are boxed and coloured blue for α-helix, red for β-strand and green for 3_{10}-helix. Green and red characters represent residues comprising the $NADP^+$- and substrate-binding cavities, respectively, with the notation P for the pyrophosphate moiety, A for the adenine moiety, R for the ribose moiety and C for the carbohydrate moiety of the substrate or $NADP^+$. A blue P indicates residues interacting with the phosphate attached to the ribose moiety of $NADP^+$. The black and pink stars represent the two unique regions UR1 and UR2 of *Burkholderia* species, respectively. The catalytic triad (Ser116, Tyr139 and Lys143) are represented by red filled circles, and green circles represent the residues Lys177 and Gln293 that form hydrogen bonds to the substrate ADP-D,D-Hep in the docking model. Percentage sequence identities are listed compared with the sequence of *B. thailandensis*. The signature motifs of the eSDR family are represented using the following key: a denotes an aromatic residue, c a charged residue, h a hydrophobic residue, p a polar residue and x any residue. This figure was drawn with the *Jalview* program (Waterhouse *et al.*, 2009). The consensus amino acid is coloured according to the *ClustalX* colour scheme.

one-base mechanism requiring only one catalytic acid/base, Tyr140, for epimerization (Kowatz *et al.*, 2010). In this case, rotation of the 6″-keto intermediate of ADP-D,D-Hep around the C5″—C6″ bond is suggested for epimerization. However, a two-base mechanism based on dismutase activity towards ADP-α-D-*manno*-hexodialdose requires two catalytic acid/base residues, Tyr140 and Lys178, located in two catalytic pockets in the epimerase active site (Morrison & Tanner, 2007). In both cases Ser116 and Lys144 interact with the substrate and the cofactor, respectively (Deacon *et al.*, 2000).

Burkholderia pseudomallei, a soil-dwelling Gram-negative bacillus, is notorious for its pathogenicity and causes melioidosis (Wiersinga *et al.*, 2006), an endemic disease found throughout southeast Asia and northern Australia. Since mortality from melioidosis septic shock remains high despite antimicrobial therapy, the development of more effective antibiotics or antibiotic adjuvants is required. *B. thailandensis*, a relatively avirulent microbe, shares high genomic similarity and the majority of its virulence factors with *B. pseudomallei*. *B. thailandensis* is therefore a useful surrogate organism for studying the pathogenicity of melioidosis.

An open reading frame of *B. thailandensis* E264 codes for AGME (*Bt*AGME; ∼37 kDa); more than 10 000 sequence homologues with 23–99% amino-acid sequence identity were obtained from a *PSI-BLAST* (http://www.ncbi.nlm.nih.gov) search. Among these, more than 6000 are UDP-galactose 4-epimerase (GalE) homologues with a sequence identity of around 25%. A structure-based alignment using the *jCE* server at the Research Collaboratory for Structural Bioinformatics (http://www.rcsb.org) showed a highly conserved catalytic triad and a TG*XX*G*XX*G motif for NAD(P)$^+$ binding (Persson *et al.*, 2003). These are the signature sequences of the eSDR family (Fig. 1). *Bt*AGME also contains the conserved catalytic triad Ser116, Tyr139 and Lys143. Sequence alignment of various AGMEs reveals two unique regions that are only found in *Burkholderia* species (Supplementary Fig. S1[1]). Therefore, the elucidation of the unique structural features of *Bt*AGME is a critical step in the development of an epimerase inhibitor as an antibiotic adjuvant against melioidosis. In order to achieve this goal, we have determined the three-dimensional crystal structure of *Bt*AGME. We have also performed an *in silico* study of *Bt*AGME and *Ec*AGME complexed with ADP-D,D-Hep in order to predict the catalytic mechanism of the AGME family.

2. Materials and methods

2.1. Cloning of *Bt*AGME

The cloning primers (Genotech, Daejeon, Republic of Korea) prepared for ligation-independent cloning (LIC) were 5′-GGCGGTGGTGGCGGCATGACCCTCATCGTTACCG-GCG-3′ for the forward strand and 5′-GTTCTTCTCCTT-TGCGCCCCTACAGCTGGCCGAACAGCCAAC-3′ for the reverse strand. The *Bt*AGME gene was amplified by polymerase chain reaction (PCR) using 109 ng *B. thailandensis* E264 genomic DNA template and 25 μ*M* of primers. PrimeSTAR HS DNA polymerase with GC buffer (Takara Bio Inc., Shiga, Japan) designed for high-GC-content genomic DNA was used. The insert was ligated into the amplified LIC expression vector pB2 (Kim *et al.*, 2005), a construct that contains a noncleavable N-terminal His_6 tag. After transformation into DH5α, a plasmid containing the gene insert was isolated, confirmed and transformed into *E. coli* BL21 (DE3) for protein expression.

2.2. Protein expression

E. coli BL21 (DE3) cells transformed with the cloned vector harbouring the *Bt*AGME gene were grown on Luria–Bertani (LB) agar plates containing 150 μg ml^{-1} ampicillin. The prepared frozen cell stock was grown in 5 ml LB medium and diluted into 1000 ml fresh LB medium. The culture was incubated at 310 K with shaking until an OD_{600} of 0.6–0.8 was reached. At this point, expression of *Bt*AGME was induced by isopropyl β-D-1-thiogalactopyranoside at a final concentration of 1 m*M*. The culture was grown for a further 16 h at 298 K in a shaking incubator. Cells were harvested by centrifugation at 7650*g* (6500 rev min^{-1}) for 10 min in a high-speed refrigerated centrifuge at 277 K.

2.3. Protein purification

The cultured cell paste (5.7 g) was resuspended in 25 ml buffer consisting of 50 m*M* Tris–HCl pH 8.0, 100 m*M* NaCl, 10 m*M* imidazole, 1 m*M* PMSF, 10 μg ml^{-1} DNase I and Roche protease-inhibitor cocktail (Roche Applied Science, Indianapolis, Indiana, USA). The cell suspension was disrupted using a Digital Sonifier 450 (Branson Ultrasonics Co., Danbury, Connecticut, USA). Cell debris was pelleted by centrifugation at 24 900*g* (15 000 rev min^{-1}) for 30 min in a high-speed refrigerated centrifuge at 277 K. The supernatant was affinity purified using a HisTrap column on an ÄKTA-explorer system (GE Healthcare, Piscataway, New Jersey, USA). The column was equilibrated with a buffer consisting of 50 m*M* Tris–HCl pH 8.0, 300 m*M* NaCl, 10 m*M* imidazole. The target protein was eluted with a buffer consisting of 50 m*M* Tris–HCl pH 8.0, 100 m*M* NaCl with a gradient from 10 to 500 m*M* imidazole. *Bt*AGME was further purified by ion-exchange chromatography using a 5 ml Hi-Trap Q column (GE Healthcare, Piscataway, New Jersey, USA) equilibrated with buffer containing 20 m*M* Tris–HCl pH 8.0. The protein was eluted at 0.43 *M* NaCl in a linear NaCl gradient. SDS–PAGE showed one band at around 38 kDa corresponding to the molecular weight of His_6-tagged *Bt*AGME. The purified protein contained a noncleavable N-terminal His_6 tag followed by five glycine residues. For crystallization, the protein was concentrated to 8.8 mg ml^{-1}, as determined using a UV–Vis spectrophotometer with an extinction coefficient of 0.77 M^{-1} cm^{-1} at 280 nm, in a buffer consisting of 0.43 *M* NaCl, 20 m*M* Tris–HCl pH 8.0.

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: CB5020). Services for accessing this material are described at the back of the journal.

2.4. Crystallization

Screening for crystallization conditions was performed at room temperature using the sparse-matrix method (Jancarik & Kim, 1991) with several screens from Hampton Research (Laguna Niguel, California, USA). A Hydra-Plus-One crystallization robot (Matrix Technologies, Hudson, New Hampshire, USA) was used to set up the screens using the sitting-drop vapour-diffusion method in a 96-well Intelli-Plate (Art Robbins Instrument, Salt Lake City, Utah, USA). Sitting drops were produced by mixing 0.2 µl protein solution (8.8 mg ml^{-1}) and 0.2 µl reservoir solution and were equilibrated against 50 µl reservoir solution. A VDX48 plate (Hampton Research) was used to optimize the crystallization conditions using hanging drops produced by mixing 0.8 µl protein solution and 0.8 µl reservoir solution and equilibrated against 200 µl reservoir solution. The final crystallization solution consisted of 35% PEG 200, 0.1 *M* bis-tris–HCl pH 5.5.

2.5. Data collection and reduction

Before flash-cooling in liquid nitrogen, crystals were soaked in LV CryoOil (MiTeGen, Ithaca, New York, USA). X-ray diffraction data were collected at a single wavelength on the MX2 beamline at the Australian Synchrotron using a Quantum 315r CCD detector (Area Detector System Co., Poway, California, USA) placed 350 mm from the sample. The oscillation range per image was 0.5° with 1 s exposures, and 720 oscillation images were collected with no overlap between two contiguous images. X-ray diffraction data were processed and scaled using *iMosflm* (Battye *et al.*, 2011) and *SCALA* from the *CCP*4 program suite (Winn *et al.*, 2011). Details of the data-collection statistics are presented in Table 1.

2.6. Structure determination

Initial phases were determined by molecular replacement with *EPMR* (Kissinger *et al.*, 2001) using the crystal structure of *Ec*AGME (PDB entry 1eq2; Deacon *et al.*, 2000) as a search model. A decamer composed of two pentamers was found in the asymmetric unit. The initial electron-density map was interpretable and the protein sequence of *Bt*AGME was fitted. The preliminary model was built and refined using *Coot* (Emsley *et al.*, 2010) and *REFMAC*5 (Murshudov *et al.*, 2011) with stepwise mutation of the residues. 5% of the data were assigned for calculation of the free *R* factor. Noncrystallographic symmetry (NCS) matrices were obtained for ten protomer molecules in the asymmetric unit. The NCS restraint was applied during refinement and was released in the final round of refinement. Reflection data between 20.0 and 2.6 Å resolution were included throughout the refinement calculations. The isotropic *B* factors for individual atoms were initially fixed at 20 Å^2 and were refined in the last stages. The final refinement and the addition of NADP$^+$ and water molecules were performed using *PHENIX* (Adams *et al.*, 2010; Afonine *et al.*, 2012). The refinement statistics are shown in Table 1. All residues lie in the allowed regions of the Ramachandran plot produced with *PROCHECK* (Laskowski *et al.*, 1993). The coordinates and structure factors have been deposited in the Protein Data Bank (PDB entry 4ej0).

Table 1
Data-collection statistics and refinement parameters.

Values in parentheses are for the highest resolution shell.

PDB code	4ej0
Data-collection statistics	
Wavelength (Å)	0.95369
Resolution (Å)	20.0–2.6
Multiplicity	14.4 (13.9)
Unique reflections	131360 (18791)
Completeness (%)	99.6 (98.6)
$\langle I/\sigma(I)\rangle$	26.4 (5.1)
$R_{\rm sym}$† (%)	9.3 (53.8)
Crystal parameters and refinement statistics	
Space group	$P2_12_12_1$
Unit-cell parameters (Å)	a = 158.3, b = 160.9, c = 169.6
Solvent content (%)	57.76
$V_{\rm M}$ (Å^3 Da^{-1})	2.91
Total No. of residues	3300
Total non-H atoms	27040
No. of NADP$^+$ molecules	10
No. of water molecules	456
Average temperature factors (Å^2)	
Protein	50.52
Solvent	41.49
NADP$^+$	36.44
Resolution range of reflections used (Å)	20.0–2.6 (2.64–2.60)
Amplitude cutoff	0.0σ
R factor (%)	18.5 (25.1)
Free *R* factor (%)	24.1 (32.2)
Stereochemical ideality	
Bonds (Å)	0.008
Angles (°)	1.135
Chirality (°)	0.072
Planarity (°)	0.004
Dihedrals (°)	16.16
Ramachandran plot, residues in (%)	
Most favoured regions	90.2
Additional allowed regions	9.5
Generously allowed regions	0.3
Disallowed regions	0.0

† The data were obtained from a single crystal.

2.7. Docking study

Docking studies to predict the catalytic mechanism of *Bt*AGME and *Ec*AGME were carried out with their substrate, ADP-D,D-Hep, using the *Surflex-Dock* program of *SYBYL-X* 1.3. The coordinates of ADP-D,D-Hep were retrieved from the PubChem database (unique chemical structure identifier CID 46173177; http://pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?cid=46173177) and were energy minimized using the Tripos force field with Gasteiger–Hückel charges. Docking was guided by the protomol, an idealized representation of a ligand that makes every potential interaction with the binding site (Kellenberger *et al.*, 2004). In this study, prior to docking, the H atoms were added to predicted models (chain *A* of PDB entry 4ej0 for *Bt*AGME and chain *D* of PDB entry 1eq2 for *Ec*AGME) using the Biopolymer modulators of *SYBYL-X* 1.3. The MMFF94 atom charges were assigned to protein atoms. Protomols for *Surflex-Dock* were generated according to the software protocol. Two important factors, 'Bloat(A)', which determines how far the protomol extends into the concavity of the target site, and 'Threshold', which impacts on

how far the protomol extends outside the concavity, were set to 0.5 and 0, respectively. Other parameters were employed with default settings in all runs. Protomols were visualized with *SYBYL-X* 1.3 to ensure proper coverage of the desired target area. The *Surflex-Dock* scoring function, which contains hydrophobic, polar, repulsive, entropic and solvation terms, was trained to estimate the dissociation constant (K_d) expressed in $-\log(K_d)$ units (Jain, 2007). Since the *syn* conformation of the cofactor observed in the crystal structures of *Bt*AGME and *Ec*AGME is known to be an oxidized form (Thoden *et al.*, 1996), $NADP^+$ was used during the whole docking study. The initial docking models were obtained after running *Surflex-Dock* and the scores of the docked conformers were ranked in a molecular spreadsheet (Supplementary

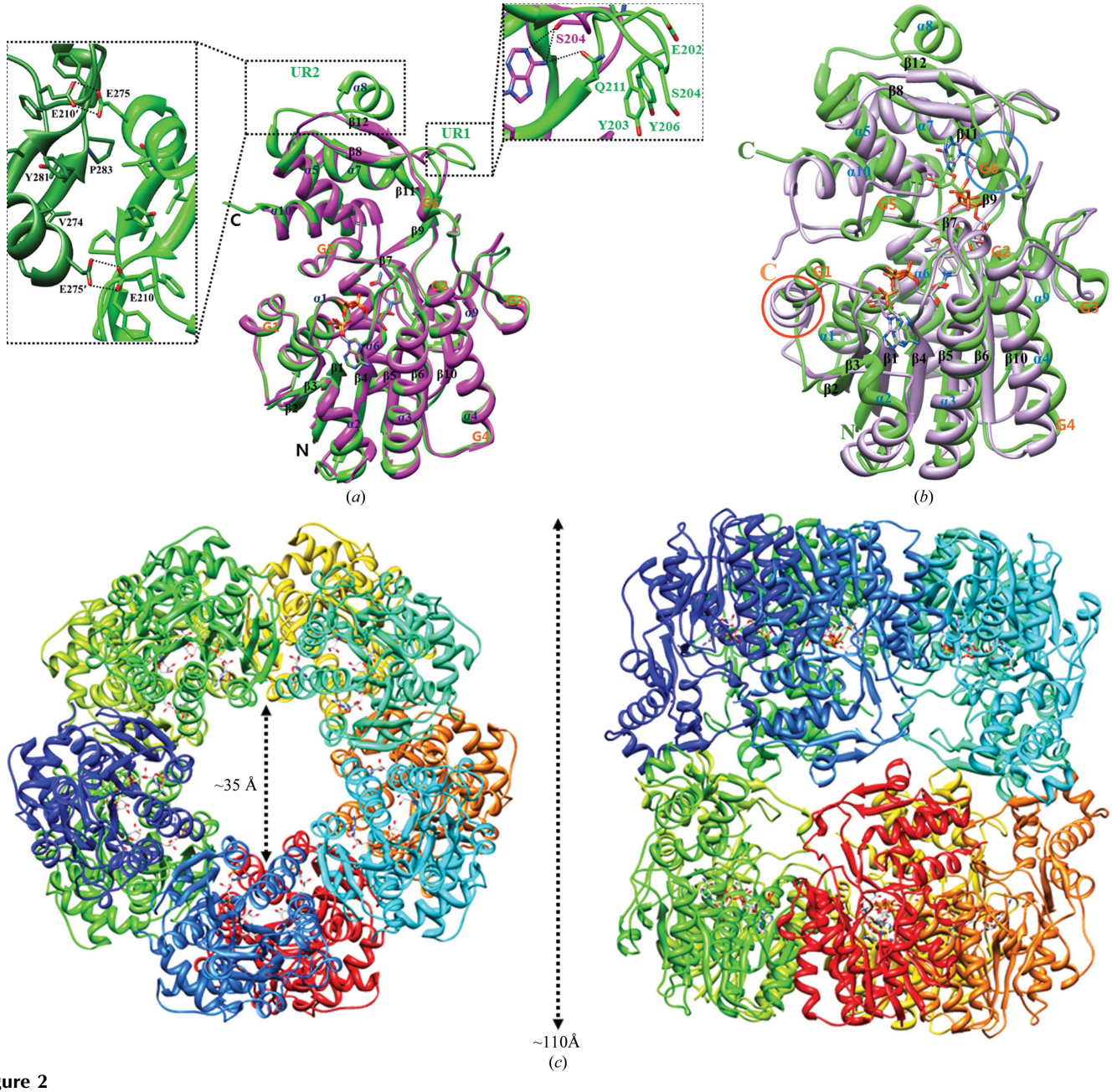


Figure 2
Crystal structure of *Bt*AGME. (*a*) The crystal structure of *Bt*AGME superimposed on that of *Ec*AGME. Green represents *Bt*AGME and pink represents *Ec*AGME. The $NADP^+$ molecules are represented as stick models (blue represents N atoms, red O atoms, green C atoms and scarlet P atoms). Secondary-structure elements are labelled. The box on the right represents the first unique region (UR1), in which Ser116 of *Ec*AGME forms two hydrogen bonds to ADP-α-D-glucose whereas Gln211 of *Bt*AGME forms one hydrogen bond to the modelled ADP-α-D-glucose. The left box represents the decamerization interface (UR2) of *Bt*AGME, in which two glutamates from each subunit (green and blue) form hydrogen bonds with average distances of 3.12 and 3.86 Å obtained from ten protomers, respectively. The figure was drawn with *UCSF Chimera* (Pettersen *et al.*, 2004). (*b*) *Bt*AGME complexed with *in silico* modelled ADP-D,D-Hep superimposed on GalE from *E. coli* (PDB entry 1a9z). The cofactors and substrates are shown as stick models. The red circle indicates the helix-blocking pentamerization in GalE from *E. coli*. The blue circle represents the region around 3_{10}-helix G6 which decreases the size of the substrate-binding site of *Bt*AGME. (*c*) Diagrams of the decameric form of *Bt*AGME. Left, top view; right, side view. Each subunit is coloured differently. The $NADP^+$ molecules are represented by stick models. The dimensions of the decamer are represented.

Tables S1*a* and S1*c*). The five best-scoring conformers were selected and applied to *Surflex-Dock GeomX*, which uses multiple starting conformations and produces up to 20 docked poses per ligand with a more exhaustive docking-accuracy parameter set. Since the positions of active-site residues and $NADP^+$ of *Bt*AGME were almost unchanged compared with those of substrate-bound crystal structures of *Ec*AGME, only a ligand-flexibility option (Ring Flexibility) in *Surflex-Dock GeomX* was adopted for extensive searches. The best scores from *Surflex-Dock GeomX* were for the fifth conformer for *Bt*AGME and the first conformer for *Ec*AGME as obtained from their *Surflex-Dock* runs (Supplementary Tables S1*b* and S1*d*). The coordinates of the top-ranked models from the *Surflex-Dock GeomX* runs have been deposited in the Protein Model Data Base (PMDB; PM0078686 for *Bt*AGME complexed with ADP-D,D-Hep and PM0078687 for *Ec*AGME complexed with ADP-D,D-Hep). These coordinates were used in the discussion.

2.8. Site-directed mutagenesis

The expression plasmid pB2-*Bt*AGME was constructed and used as a template for mutagenesis. The expression plasmid including the E210G mutant was constructed using the QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent Technologies, La Jolla, California, USA). The sequences of the mutagenic primers were 5′-GGCCCGGG-CGGGGCAGACGCGC-3′ for the forward strand and 5′-GCGCGTCTGCCCGCCCGGGCC-3′ for the reverse strand. Briefly, the PCR mixture (50 µl) was composed of 1 µl of template, 0.4 µl of each primer (∼50 p*M* each), 1 µl dNTP mixture, 5 µl 10× QuikChange Lightning Buffer, 1.5 µl QuikSolution reagent, 1 µl QuikChange Lightning Enzyme and 39.7 µl dH_2O. The reaction was initiated for 2 min at 368 K to pre-denature the template, followed by 18 cycles of 20 s at 368 K, 10 s at 333 K and 204 s at 341 K. After the PCR reaction, 1.6 µl *Dpn*I was added to the 40 µl reaction product and the mixture was incubated at 301 K for 5 min to degrade the original unmodified plasmid templates. After *Dpn*I digestion, 10 µl of the mixture was used to transform 45 µl of XL10-Gold competent cells treated with 2 µl β-mercaptoethanol. The transformed *E. coli* XL10-Gold competent cells were grown in 500 µl NZY broth and were incubated at 310 K for 1 h. A 100 µl aliquot of the NZY broth was spread onto LB plates containing 50 mg ml^{-1} ampicillin. After isolating the plasmids, the DNA sequences were confirmed by gene sequencing and the confirmed recombinant plasmids were introduced into *E. coli* BL21 (DE3). The mutant protein was purified using the same protocol as was used for the wild type.

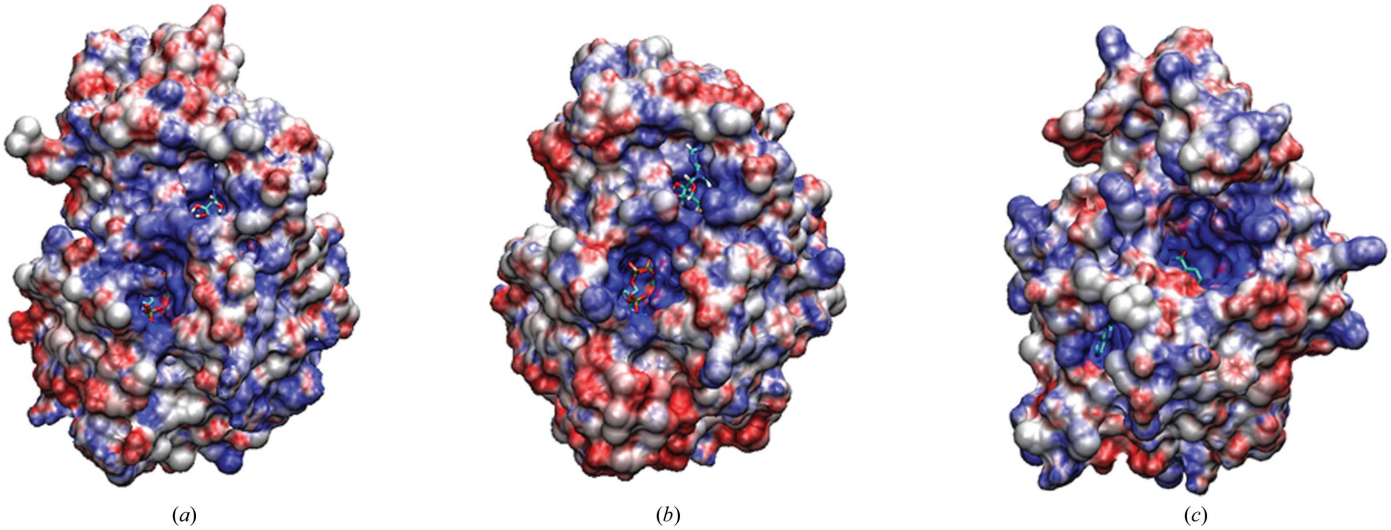
2.9. Dynamic light-scattering experiment

Dynamic light scattering (DLS) was used to determine the hydrodynamic radius of native *Bt*AGME and its E210G mutant under various pH conditions. It was performed using a DynaPro Titan instrument (Wyatt Technology Co., Santa Barbara, California, USA). Protein concentrations of 0.2–0.5 mg ml^{-1} were used. A micro cuvette with a volume of 12 µl and a path length of 0.15 cm was used for the measurements.

3. Results and discussion

3.1. Overall structure and structural comparison with *Ec*AGME

The crystal structure of *Bt*AGME (Figs. 2*a* and 2*b*) was solved by molecular replacement using the crystal structure of *Ec*AGME (Deacon *et al.*, 2000) as a model. In the asymmetric unit, ten *Bt*AGME protomers were observed in a dimeric form with dimensions of ∼110 × 110 × 110 Å comprised of two pentamers (Fig. 2*c*). Each protomer contains the cofactor $NADP^+$, which co-purified during protein purification. The



(*a*) (*b*) (*c*)

Figure 3
Electrostatic surface potentials of AGMEs. (*a*) *Bt*AGME complexed with *in silico* modelled ADP-D,D-Hep. The dimensions of the entrances of its cofactor-binding and substrate-binding pockets are approximately 16 × 8 and 13 × 8 Å, respectively. (*b*) *Ec*AGME complexed with ADP-α-D-glucose. (*c*) *Hp*AGME. The flexible loop Lβ13α9 is not shown in the surface model. Molecular surfaces were created by the *VMD* molecular-graphics software package (Humphrey *et al.*, 1996) after electrostatic calculations using *APBS* (Baker *et al.*, 2001; red, negative; blue, positive; white, uncharged). The bound ligands located in the upper cavities and the cofactors, $NADP^+$ or NAD^+, located in the lower cavities are represented by stick models.

final model exhibited good stereochemical geometry and was refined to R and R_{free} values of 18.9% and 24.5%, respectively (Table 1). The root-mean-square deviations (r.m.s.d.s) among the ten protomers are within the range 0.31–0.44 Å, indicating that there is no apparent conformational difference among them. Architecturally, *Bt*AGME is composed of two domains: a predominant N-terminal domain with a modified Rossmann fold and a smaller C-terminal globular domain. The latter domain is involved in decamerization in the asymmetric unit (Fig. 2*c*). The previously solved structures of *Ec*AGME have been crystallized in the presence of substrate homologues such as ADP-α-D-glucose (Deacon *et al.*, 2000) or ADP-β-D-mannose (Kowatz *et al.*, 2010). However, *Bt*AGME was crystallized in a substrate-free form. Structural analyses of ligand-free *Bt*AGME and ligand-bound *Ec*AGME revealed that their overall architecture is almost conserved, with an r.m.s.d. of 0.59 Å for 281 C^{α} atoms (henceforth, structural comparison is between chain *A* of *Bt*AGME and chain *D* of *Ec*AGME unless mentioned otherwise). When the $NADP^+$-binding pockets of the two structures are compared along the bound $NADP^+$ molecule, the residues of *Bt*AGME that interact with the adenine-moiety region (Asp31, Asn32, Lys38, Lys53, Tyr92 and Tyr96) and the nicotinamide-moiety region (Ile11, Ser79, Ala114, Lys144, Tyr167 and His177) are almost conserved (Supplementary Fig. S1). In addition, when the substrate-binding pockets were compared using the bound ADP-α-D-glucose as a reference molecule, the residues that interact with the carbohydrate-moiety region (Thr81, Ser116, Tyr139, Asn168 and Lys177) and the adenine-moiety region (Val183, His186, Leu199, Phe200, Arg213, Phe248 and Tyr292) are also almost conserved. Interestingly, the atomic positions of the above residues of *Bt*AGME are almost unchanged compared with those of *Ec*AGME. These features indicate that both AGMEs recognize their substrate in a 'lock-and-key' fashion instead of an 'induced-fit' fashion. The electrostatic surface potentials of *Bt*AGME and *Ec*AGME support the lock-and-key recognition in that the entrances for their cofactor and substrate are wide open, with approximate

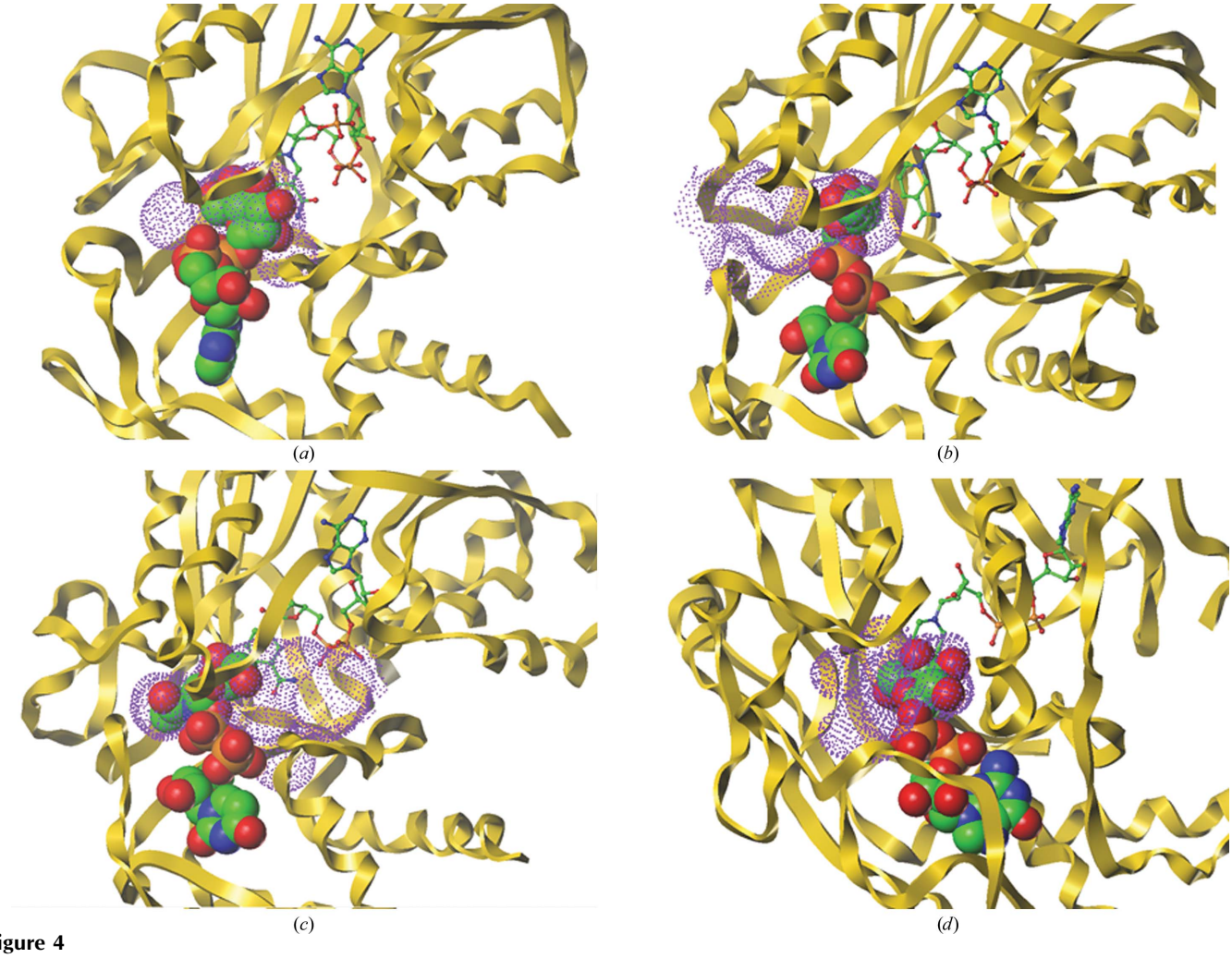


Figure 4
The epimerase active sites of AGMEs and GalE homologues. (*a*) *Bt*AGME complexed with modelled ADP-D,D-Hep. (*b*) *E. coli* GalE complexed with UDP-α-D-glucose. (*c*) WbgU from *Pseudomonas aeruginosa* complexd with UDP-*N*-acetylglucosamine. (*d*) *A. thaliana* GME complexed with GDP-β-L-galactose. All of the cavities and their volumes were detected and calculated using the *MOLCAD* module in *SYBYL-X* 1.3 with their substrates after removing the carbohydrate moieties. The $NADP^+$ molecule and substrates are represented by ball-and-stick and sphere models, respectively. The cavities are represented by violet dots.

Table 2
Analysis of the substrate-binding cavities of AGMEs and GalE homologues.

Species (PDB code)	Protein	Substrate	*Z*-score†	R.m.s.d.‡	SeqID§ (%)	SeqSim¶ (%)	Cov†† (%)	Cavity volume (Å^3)	Molecular volume (Å^3)	Cavity/ molecule	Related PDB entries
B. thailandensis (4ej0)	AGME	ADP-D,D-Hep	8.03	0	100	100	100	416.0	213.5	1.9	
E. coli (2x6t)		ADP-β-D-mannose	7.64	1.08	59	71	92	375.6	181.5	2.1	
E. coli (1eq2)		ADP-α-D-glucose	7.64	1.13	59	71	92	396.6	181.5	2.2	
H. pylori (3sxp)		N/A	7.13	1.76	31	51	87	N/A	N/A	N/A	
S. chartreusis (3vps)	TunA	UDP-*N*-acetylglucosamine	7.02	2.29	22	39	88	326.8	237.4	1.4	
B. cereus (3m2p)	WbgU	N/A	6.47	2.37	22	41	76	N/A	N/A	N/A	
P. shigelloides (3lu1)		UDP-*N*-acetylglucosamine	6.92	2.46	24	41	92	663.4	237.4	2.8	3ru9, 3ru7, 3ruc
P. aeruginosa (1sb8)		UDP-*N*-acetylglucosamine	7.24	2.58	23	38	93	742.5	237.4	3.1	
A. thaliana (2c59)	GME	GDP-β-L-galactose or GDP-α-D-mannose	7.02	2.46	21	40	88	463.5	181.5	2.6	2c5e, 2c54, 2c5a
S. cerevisiae (1z45)	GalE	UDP-α-D-glucose	6.81	2.52	22	42	87	503.7	181.5	2.8	
E. coli (1a9z)		UDP-α-D-glucose	7.13	2.65	20	34	93	615.5	181.5	3.4	1a9y
E. coli (1uda)		UDP-4-deoxy-4-fluoro-α-D-galactose	7.13	2.68	21	34	93	673.9	176.0	3.8	1udb
B. pseudomallei (3enk)		UDP-α-D-glucose	7.13	2.72	22	37	93	830.8	181.5	4.6	
A. fulgidus (3ehe)		N/A	6.81	2.92	22	41	85	N/A	N/A	N/A	
E. coli (1lrk)		UDP-*N*-acetylglucosamine	7.02	2.97	21	34	94	756.5	237.4	3.2	1lrj, 1kvs, 1kvq, 1kvr, 1kvu, 1kvt

† Measure of the statistical significance of the result relative to an alignment of random structures. ‡ R.m.s.d. value of the alignment. § Percentage sequence identity in the alignment. ¶ Percentage sequence similarity in the alignment †† The coverage of aligned residues in chain 1.

dimensions of 12 × 8 Å and 13 × 8 Å, respectively, and are positively charged to attract the cofactor and substrate easily without conformational changes (Fig. 3).

3.2. Structural comparison with UDP-galactose 4-epimerase homologues

A *PSI-BLAST* search displayed more than 10 000 sequence homologues of *Bt*AGME. Among these, 29 crystal structures from 11 unique proteins complexed with six different types of substrate have been deposited in the Protein Data Bank (Table 2). Two of these are *Ec*AGME and AGME from *Helicobacter pylori* (*Hp*AGME). The remainder are UDP-galactose 4-epimerase (GalE) homologues such as GalEs, NAD-dependent NDP-hexosamine 5,6-dehydratase, UDP-*N*-acetylglucosamine 4-epimerases (WbgUs/WbpPs) and GDP-mannose-3′,5′-epimerases (GMEs). Since AGMEs and GalE homologues belong to the eSDR family, their overall structures are quite similar to each other (Fig. 2*b*), with several highly conserved eSDR signature motifs (Persson *et al.*, 2003). In addition, there is one modified motif ^{165}RY*h*N*hh*G*x*^{172} instead of the known motif *h*(KR)*xx*NGP that is common to the eSDR family and one new motif ^{237}*ha*N*h*GA^{241} in the sequence alignment of AGMEs and GalE homologues (Fig. 1), where *a* denotes an aromatic residue, *h* a hydrophobic residue and *x* any residue. Regardless of their similarity, AGMEs prefer NADP$^+$ and GalE homologues prefer NAD$^+$ as a cofactor. Therefore, the residues Asn33, Lys39 and Lys54 that attract the phosphate moiety of NADP$^+$ are only conserved in the AGME family (Fig. 1 and Supplementary Fig. S1). The penatameric form of the AGME family is not observed in the GalE homologues. This is owing to the deviation of the helix or loop of GalE homologues from the corresponding position of the 3_{10}-helix G1 of the AGME family (Figs. 1 and 2*b*).

Analysis of the epimerase active sites of AGMEs and GalE homologues revealed that there is a noticeable variation in volume around the carbohydrate moiety of their substrates. The volume of each cavity around the carbohydrate moiety was calculated using the *MOLCAD* module implemented in *SYBYL-X* 1.3 (Fig. 4). Interestingly, the volume depends on the enzymatic mechanism of each epimerase. As listed in Table 2, that of the GalE homologues is almost one and a half times larger than that of the AGMEs (Figs. 4*a* and 4*b*). The spacious room required for the GalE epimerization mechanism is inferable from the 180° rotation step of the carbohydrate moiety of UDP-α-D-galactose along the phosphate backbone of UDP (Thoden *et al.*, 1996). WbgUs/WbpPs, which have a similar epimerization mechanism to GalE (Ishiyama *et al.*, 2004), also have a large cavity of ∼700 Å^3, which is three times larger than the volume of the carbohydrate moieties of their substrates (Fig. 4*c*). In contrast, that of GME from *Arabidopsis thaliana* (*At*GME), which adopts a ring-flip mechanism that requires a smaller volume compared with that of a ring-rotation mechanism (Major *et al.*, 2005) occupies ∼460 Å^3 (Fig. 4*d*). Since the volumes of the two AGMEs are smaller than that of *At*GME, a minor conformational shift of ADP-D,D-Hep is expected during epimerization. Therefore, the proposed mechanism (Morrison & Tanner, 2007) requiring a smaller space for rotation of the 6″-keto intermediate around the C5″—C6″ bond of ADP-D,D-Hep instead of rotation of the whole heptose along the phosphate backbone of ADP matches our analysis quite well. The relatively small size of the substrate-binding pocket of AGMEs compared with that of GalEs also enables Gln293 to hold ADP-D,D-Hep as suggested in the docking study below. This architecture, which is important for the ADP-D,D-Hep affinity, is constructed by the 3_{10}-helix G6 which is unique to AGMEs (Figs. 1 and 2*b*).

3.3. Structural comparison with *Hp*AGME

The crystal structure of AGME from *H. pylori* (*Hp*AGME) has recently been determined (Shaik *et al.*, 2011). A *PSI-BLAST* search showed that *Hp*AGME comprises a species-specific family (Supplementary Fig. S2). Therefore, it shows quite a low sequence similarity to other AGMEs. Despite its low sequence identity of 25% and 32% to *Bt*AGME and *Ec*AGME, respectively, its structural similarity is quite high, with an r.m.s.d. of 0.93 Å for 213 C^{α} atoms with *Bt*AGME and of 0.92 Å for 208 C^{α} atoms with *Ec*AGME. However, there are several local conformations of *Hp*AGME that deviate severely from those of the other two AGMEs. Prominent differences are observed in the three loops LN, Lβ2G1 and Lβ13α9 of *Hp*AGME (Fig. 5*a* and Supplementary Fig. S1). These loops have a flexible nature, with sequences that are nonconserved in other AGMEs but that are highly conserved in *H. pylori* substrains. Lβ2G1 of *Hp*AGME does not have the conserved asparagine and lysine (Asn33 and Lys54 in *Bt*AGME), the role of which is to bind to the phosphate moiety of $NADP^+$ in the cofactor-binding site. Instead, the side chain of Ser58 of *Hp*AGME occupies the position of the phosphate moiety of $NADP^+$ (Fig. 5*b*). Consequently, NAD^+ (or NAD) was detected in the crystal structure of *Hp*AGME. Lβ13α9, which covers the substrate-binding pocket of *Hp*AGME, is also very flexible owing to the absence of the ^{283}PFP^{285} motif stabilizing this loop region in *Bt*AGME (Fig. 4*a* and Supplementary Fig. S1). As a result, the space for the adenosine portion of ADP-α-D-glucose is totally exposed in the crystal structure of *Hp*AGME (Fig. 3*c*). These species-specific sequence and structural properties indicate that the substrate-binding affinity and epimerization mechanism of *Hp*AGME may differ from those of *Bt*AGME and *Ec*AGME.

3.4. Structural properties of two unique regions

The crystal structure of *Bt*AGME provides information on the structural properties of two unique regions that are only found in *Burkholderia* species. The first region is the loop between β8 and β9 (residues 200–210) located next to the adenine moiety of ADP-α-D-glucose docked in *Bt*AGME. Gln211 is the closest residue, forming a hydrogen bond to the N6 atom of the adenine moiety with a distance of 3.5 Å. In contrast, Ser204 forms two hydrogen bonds to the N1 and N6 atoms of the adenine moiety with distances of 3.6 and 3.2 Å, respectively, in *Ec*AGME. Consequently, the binding affinity of ADP-D,D-Hep for *Bt*AGME is expected to be slightly decreased compared with that for *Ec*AGME. In the cases of GalE homologues, a similar deviation is detected in this region in some species (Fig. 1). Since this loop next to the substrate-binding site is unique in *Burkholderia* species, an inhibitor targeting this region is strongly recommended for development as an anti-melioidosis adjuvant.

The second unique region is around residues 260–280, where an additional 17 residues are inserted compared with *Ec*AGME and *Hp*AGME and form α-helix α8 with a long loop (Fig. 2*a*). Interestingly, this region seems to facilitate the decamerization of two pentamers, resulting in a decamer. This interaction is mediated by two hydrogen bonds between Glu210 of one subunit and Glu275 of another. It is noteworthy that crystals of *Bt*AGME began to appear below pH 5.5 and are well formed around pH 4.5, where glutamates are protonated to form hydrogen bonds easily. The decamerization is further stabilized by covering exposed hydrophobic residues from solvent as shown in Fig. 2(*a*). The accessible surface area occupied by the decamerization is 260 $Å^2$ per monomer. In order to prove that the decamerization is mediated by two glutamate residues, the E210G mutant of *Bt*AGME was constructed and its product was purified. At neutral pH, the dynamic light-scattering results confirmed that both the wild type and the mutant form a pentamer (Supplementary Fig. S3). However, at pH 5.0, unlike the

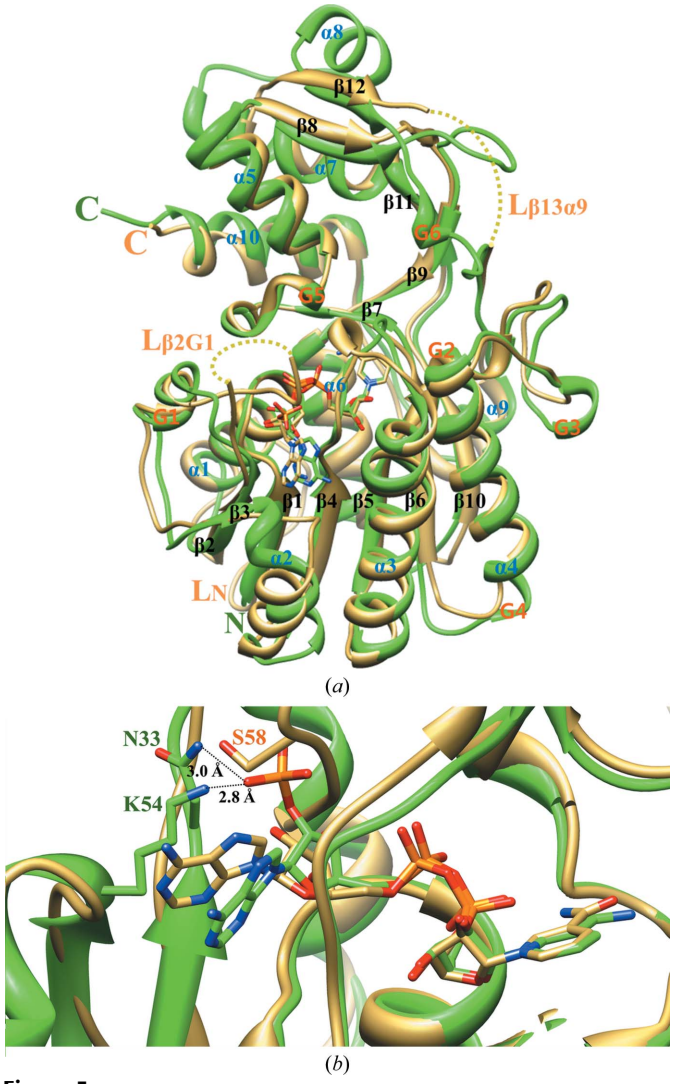


Figure 5
(*a*) The superimposed crystal structures of *Bt*AGME and *Hp*AGME. Green represents *Bt*AGME and gold represents *Hp*AGME. The $NADP^+$ and NAD^+ molecules are represented by stick models (blue represents N atoms, red O atoms, green C atoms and scarlet P atoms). Secondary-structure elements of *Bt*AGME are labelled. (*b*) Cofactor-binding pockets. $NADP^+$ and NAD^+ molecules are represented by stick models. The highly conserved residues that interact with the phosphate moiety of $NADP^+$ are labelled with their salt-bridge distances. The average distances were obtained from ten protomers.

E210G mutant which still forms a pentamer, the wild type forms a decamer as observed in the crystal structure. Interestingly, below pH 5.0 the mutant becomes unstable and starts to precipitate. This tendency is quite low in the wild type. Therefore, one of the molecular roles of decamerization is to protect *Bt*AGME from denaturation at low pH.

3.5. The catalytic mechanism inferred from *in silico* binding models

The catalytic mechanism of AGME has been studied by Tanner and coworkers (Kowatz *et al.*, 2010; Morrison & Tanner, 2007). A one-base mechanism was proposed to be the major catalytic pathway rather than a two-base mechanism. These hypotheses were deduced from crystal structures with the substrate mimics ADP-β-D-mannose (Kowatz *et al.*, 2010) and ADP-α-D-glucose (Deacon *et al.*, 2000), respectively. Therefore, we performed an *in silico* binding study with the substrate ADP-D,D-Hep using the *Surflex-Dock GeomX* module of the *SYBYL-X* 1.3 software. The active sites of *Bt*AGME and *Ec*AGME were used as templates for docking ADP-D,D-Hep. The total scores and consensus scores (CScore) were used to pick out the most probable configuration from a set of candidate configurations produced by *Surflex-Dock GeomX* docking runs. The total scores of the best poses of ADP-D,D-Hep were 12.3 for *Bt*AGME and 14.4 for *Ec*AGME (Supplementary Table S1). The better total score for *Ec*AGME may be the result of the better binding affinity for the ADP moiety, as discussed above. The poses of the heptose moiety in the active sites of both AGMEs are almost identical (Fig. 6*a*). The key point in the catalytic process of AGME is the inversion at the C6″ stereocentre of ADP-D,D-Hep mediated by the reversible redox reactions of NADP$^+$/NADPH. Therefore, the correct positioning of the proton at C6″ in ADP-D,D-Hep relative to the C4 atom of NADP$^+$ is important. The modelled structures show that Ser116, Tyr139, Lys177 and Gln293 act cooperatively to satisfy this geometrical condition of the substrate for the catalytic reaction. Lys177 forms hydrogen bonds to O2″ and O3″ of ADP-D,D-Hep with distances of 2.2 and 3.2 Å, respectively. The importance of the interaction of Lys177 with the heptose moiety is inferable from the 1000-fold reduction in activity on mutation of the corresponding residue in the K178M mutant of *Ec*AGME (Kowatz *et al.*, 2010). Ser116 makes hydrogen bonds to O6″ and O7″ of ADP-D,D-Hep with distances of 2.9 and 3.1 Å, respectively. The hydrogen bond from Ser116 to O7″ was not predicted in previous studies owing to the lack of the C7″ hydroxymethyl group in the substrate mimics ADP-β-D-mannose and ADP-α-D-glucose. The C7″ hydroxymethyl group also forms a hydrogen bond to the conserved Gln293 residue. The additional hydrogen bonds of Ser116 and Gln293 predicted in this study assist in placing the proton at C6″ of ADP-D,D-Hep geometrically closer to that of ADP-β-D-mannose rather than to that of ADP-α-D-glucose. This implies that the catalytic reaction may be carried out by the one-base mechanism as described by Deacon *et al.* (2000). The final pose of ADP-D,D-Hep with a short distance from O6″ to the hydroxyl group of Tyr139 and the orientation of the proton at C6″ with the C4 atom of NADP$^+$ at a distance of 2.7 Å

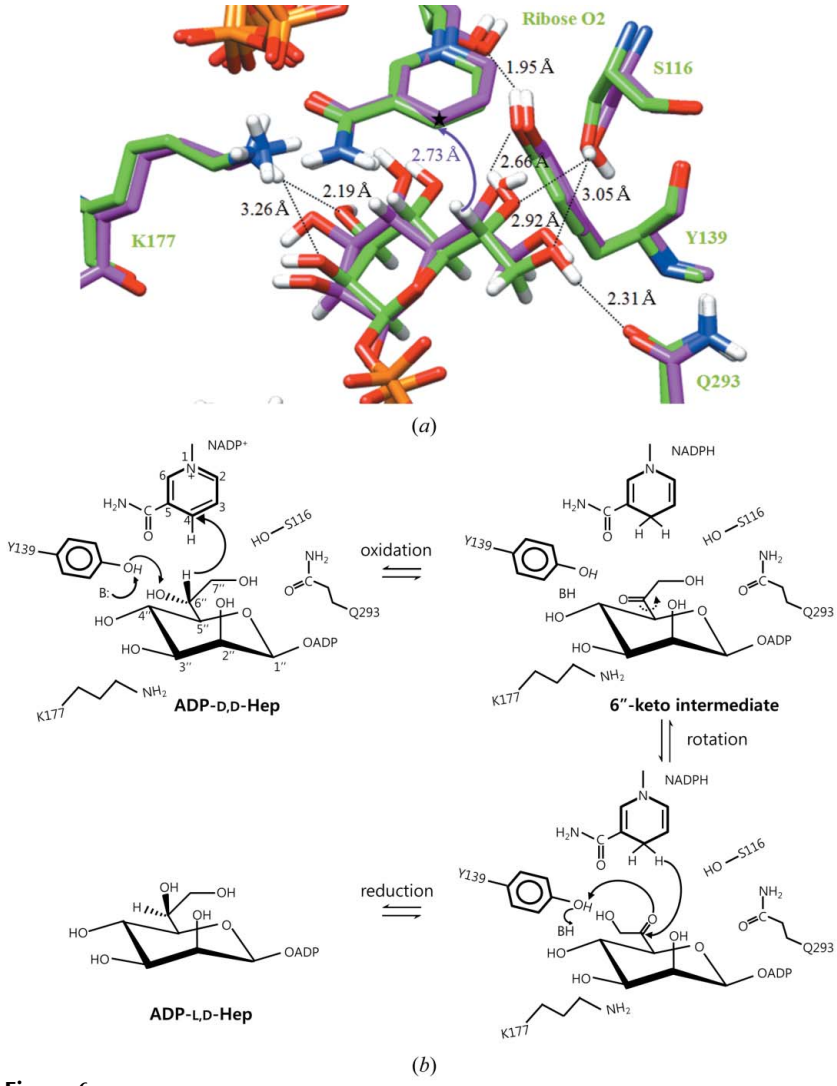


Figure 6
The binding modes of ADP-D,D-Hep and a scheme for a one-base mechanism. (*a*) The *in silico* modelled ADP-D,D-Hep complexed with *Bt*AGME (green) is superimposed on that of *Ec*AGME (pink). The functional residues and NADP$^+$ molecules are represented by ball-and-stick models (blue represents N atoms, red O atoms, green C atoms and scarlet P atoms). Only the hydrogen bonds between ADP-D,D-Hep and *Bt*AGME are indicated by dotted lines. The C4 position of NADP$^+$, where hydride transfer occurs, is indicated by a filled star. (*b*) Scheme for a one-base mechanism catalyzed by *Bt*AGME. The interconversion of ADP-D,D-Hep and ADP-L,D-Hep involves oxidation at C6″ followed by rotation around C5″—C6″ and then reduction at C6″ with the aid of redox reactions of NADP$^+$/NADPH together with the function of Tyr139 in acid/base catalysis. B represents a base, which may be one of the surrounding water molecules.

strongly support the one-base mechanism. Therefore, the conventional role of Tyr139 in hydroxyl-group proton abstraction suggested in the previous study (Deacon *et al.*, 2000; Kowatz *et al.*, 2010) is further corroborated by our *in silico* models.

Based on these observations, the whole catalytic process of *Bt*AGME can be briefly summarized as follows. For forward catalysis, the deprotonation of Tyr139 needs to be accomplished. This can be performed by lowering the pK_a of Tyr139, with the positive charge of Lys143 stabilizing a phenoxide anion. The phenoxide anion of Tyr139 abstracts the proton from the C6″ hydroxyl group and simultaneously $NADP^+$ abstracts hydride from C6″. The resulting ketone intermediate of ADP-D,D-Hep would perturb interactions with *Bt*AGME by altering its geometry. As a result, the ketone intermediate rotates around the C5″—C6″ bond. Finally, epimerization is accomplished by reverse catalysis by NADPH coordinated with protonation of the carbonyl O atom of the ketone intermediate by Tyr139. Therefore, the modelled structures support the one-base mechanism as a major catalytic pathway, in which Tyr139 solely functions as a catalytic acid/base residue (Fig. 6*b*).

4. Concluding remarks

In conclusion, the AGME family has structural features in common with GalE homologues. Nevertheless, AGMEs prefer $NADP^+$ as a cofactor instead of NAD^+. Both enzymes have distinctive substrate-binding cavities that are suitable for their epimerase mechanism. *Bt*AGME has unique structural properties that are not found in other bacterial AGMEs. The first unique region of *Bt*AGME seems to slightly influence its binding affinity for ADP-D,D-Hep. The second region, which is involved in decamerization at low pH, confers resistance of the wild type to acid-induced denaturation. A comparison of the crystal structures of ligand-free *Bt*AGME and ligand-bound *Ec*AGME indicated that there are no structural changes in the active site and cofactor-binding site, implying that substrate binding does not induce a conformational change. The *in silico* models of ADP-D,D-Hep complexed with *Bt*AGME and *Ec*AGME support the one-base mechanism as a major catalytic pathway, in which Tyr139 solely functions as a catalytic acid/base residue. In addition, we propose that the molecular role of the conserved Gln293 in the AGME family is to hold ADP-D,D-Hep. Since mutants lacking the *rfaD* gene are known to be susceptible to antibiotics (Coleman, 1983), *Bt*AGME should be a good target for developing epimerase inhibitors as novel antibiotic adjuvants against melioidosis.

We are grateful to Dr Heenam Stanley Kim at Korea University for kindly providing us with *B. thailandensis* E264 genomic DNA and also to the staff at Pohang Light Source and the Australian Synchrotron for their assistance with synchrotron data collection. We greatly appreciate the comments of Dr Rosalind Kim at the Lawrence Berkeley National Laboratory. This work was supported by Basic Science Research Program (2012-0003936) and the National Core Research Center program (2012-0000952) through the National Research Foundation of Korea grant funded by the Korean Government (MEST). Some of the instruments used in this work were funded by the Ewha Global Top5 Grant 2011 of Ewha Womans University. M-SK, AL, SY, JP and DL were supported by the Brain Korea 21 (BK21) Project.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* D**68**, 352–367.
Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.
Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* D**67**, 271–281.
Coleman, W. G. (1983). *J. Biol. Chem.* **258**, 1985–1990.
Deacon, A. M., Ni, Y. S., Coleman, W. G. & Ealick, S. E. (2000). *Structure*, **8**, 453–462.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Heumann, D., Glauser, M. P. & Calandra, T. (1998). *Curr. Opin. Microbiol.* **1**, 49–55.
Humphrey, W., Dalke, A. & Schulten, K. (1996). *J. Mol. Graph.* **14**, 33–38.
Ishiyama, N., Creuzenet, C., Lam, J. S. & Berghuis, A. M. (2004). *J. Biol. Chem.* **279**, 22635–22642.
Jain, A. N. (2007). *J. Comput. Aided Mol. Des.* **21**, 281–306.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Kellenberger, E., Rodrigo, J., Muller, P. & Rognan, D. (2004). *Proteins*, **57**, 225–242.
Kim, S.-H. *et al.* (2005). *J. Struct. Funct. Genomics*, **6**, 63–70.
Kissinger, C. R., Gehlhaar, D. K., Smith, B. A. & Bouzida, D. (2001). *Acta Cryst.* D**57**, 1474–1479.
Kowatz, T., Morrison, J. P., Tanner, M. E. & Naismith, J. H. (2010). *Protein Sci.* **19**, 1337–1343.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Major, L. L., Wolucka, B. A. & Naismith, J. H. (2005). *J. Am. Chem. Soc.* **127**, 18309–18320.
Morrison, J. P., Read, J. A., Coleman, W. G. & Tanner, M. E. (2005). *Biochemistry*, **44**, 5907–5915.
Morrison, J. P. & Tanner, M. E. (2007). *Biochemistry*, **46**, 3916–3924.
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.* D**67**, 355–367.
Nikaido, H. & Vaara, M. (1985). *Microbiol. Rev.* **49**, 1–32.
Persson, B., Kallberg, Y., Oppermann, U. & Jörnvall, H. (2003). *Chem. Biol. Interact.* **143–144**, 271–278.
Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.
Raetz, C. R. & Whitfield, C. (2002). *Annu. Rev. Biochem.* **71**, 635–700.
Schnaitman, C. A. & Klena, J. D. (1993). *Microbiol. Rev.* **57**, 655–682.
Shaik, M. M., Zanotti, G. & Cendron, L. (2011). *Biochim. Biophys. Acta*, **1814**, 1641–1647.
Thoden, J. B., Frey, P. A. & Holden, H. M. (1996). *Protein Sci.* **5**, 2149–2161.
Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. (2009). *Bioinformatics*, **25**, 1189–1191.
Wiersinga, W. J., van der Poll, T., White, N. J., Day, N. P. & Peacock, S. J. (2006). *Nature Rev. Microbiol.* **4**, 272–282.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.