## research papers

Acta Crystallographica Section D Biological Crystallography ISSN 0907-4449

#### 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

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

ADP-L-glycero-D-manno-heptose 6-epimerase (AGME), the product of the *rfaD* gene, is the last enzyme in the heptosebiosynthesis pathway; it converts ADP-D-glycero-D-mannoheptose (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<sup>+</sup>. 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 (BtAGME), a surrogate organism for studying the pathogenicity of melioidosis caused by B. pseudomallei, has been determined. The crystal structure determined with copurified NADP<sup>+</sup> revealed common as well as unique structural properties of the AGME family when compared with UDPgalactose 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 BtAGME with AGME from Escherichia coli indicates that they may recognize their substrate in a 'lock-and-key' fashion. Unique structural features of BtAGME are found in two regions. The first region is the loop between  $\beta 8$  and  $\beta 9$ , affecting the binding affinity of *Bt*AGME for the ADP moiety of ADP-D,D-Hep. The second region is helix  $\alpha$ 8, 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 BtAGME 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<sup>+</sup>. 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,

© 2013 International Union of Crystallography Printed in Singapore – all rights reserved



Received 6 September 2012 Accepted 11 January 2013

#### PDB Reference:

ADP-L-glycero-D-mannoheptose 6-epimerase, 4ej0 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 heptosebiosynthesis 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<sup>+</sup> 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- $\beta$ -D-mannose led to the proposal of a

BT_4EJ0/1-330	1 · · · · · · · · · · · · · · · · · · ·	DNLT	
EC 1EQ2/1-310	1 · · · · · · · · · · · · · · · · · · ·	DNLK	
HP 35XP/1-330	1 · · · · · · · · · · · · · · · · · · ·	DKFRSNTLFSNNRPSSLG HFKNLI GF	
SC 3VPS/1-321	1 · · · · · · · · · · · · · · · · · · ·	DDLR	
BC 3M2P/1-309	1XSI KLAVIGGIGELGOYVVESIKN, DONTPILLIE	RSIG	YEYRVS
PS 3/ //1/1-343	1MDIYMSRYEELTOOLL. ESPKTWLITGYAGELGSNLLEKLIK, INOVVIGLDI	NESK	TI VSTEOWSBECELEGDI
PA 4500/4 344			EVELOGDI 70
AT 0050/4 070		NE N	
A/_2059/1-3/9	GAMGTINGTDYGAYTYKELEREUTWPSENLKTSTTGAGGFTASHTARRLKH-EGH-TVTASL	DWKC	DEFHLVDLRVM84
SC_1245/1-357	1 MTAQLQSESTSKIVLVIGGAGYIGSHTVVELTE.NGYD.CVVAL	DNLS	IPFYEVDLCDR73
EC_1UDA/1-342	1 · · · · · · · · · · · · · · · · · · ·	NLCN · · · · · · · · · · · · · · · SKRSVLP · · VIERLOGK · · · · H	••••••••••••••••••••••••••••••••••••••
BP_3ENK/1-341	1 · · · · · · · · · · · · · · · · · · ·	DNLV·····GKINGK····T	
AF_3EHE/1-313	1 · · · · · · · · · · · · · · · · · · ·	DNLSA	ARLVKADL52
Conserved Seq.	β1 α1 β2	P P G1	β3 P
	T G x x G a h G		$D \times \times D$
BT 4EJ0/1-330	56 EFVERFA-R-GDFGKVRAVFHEGACSDTMETDGRYMMDNNFRYSRAVLDACLA	- QGA - QFLYASSAAIYGGSS R - FVEEREVEAP	LNVYGYSKFLFDQVIRRVMP - 155
EC 1EQ2/1-310	55 DELIGINAG - EEFODVEALEHEGAXS STT - E - WORKYMMDNNYQYSKELLHYCLE	- REI - PFLYASSAATYGGRT SD FIESREYEKP	LNVYGYSKFLFDEYVRQILP - 156
HP 35XP/1-330	81 LDL RRI EKI . HEDVLEHOAAVS DTT M I NOELVMKTNYOAFLNI LELARS.	KKA, KVIYASSAGVYGNTKAP. NVVG. KNESP	ENVYGESKI CMDEEVI SHS 178
SC 3VPS/4 334	AN LELE	WAVE KVVVGSTCEVVGGAD TI BTRED SPISE	PSPYAASKVGLEMVAGAHOPA 180
BC_3M3B/4 309			NI YAYEKI ACEHIANIYEB 140
BC_3W2F71-309		ANY OF TYAA COSTYCE UP ALD WEEK ELFLE	LENTOVSKERCENTONTTSR- 140
PS_3L01/1-343	83 ROLLIGE. QVMRGVDHVLHQAALGSVPRSTVDPTTNATNITGFLNTLHAARN	AQVQSFTYAASSSTYODHPALPKVEE.NTGNP	LSPYAVIKYVNEIYAQVYAR- 185
PA_15B8/1-341	80 RNLDDCNNACAGVDYVEHQAALGSVPRSTNDPTTSNATNTDGFENMLTAARD.	AKVQSFTYAASSSTYGDHPGLPKVED.TIGKP	LSPYAVTKYVNELYADVFSR- 182
AT_2C59/1-379	85 ENCLEVV TEGVDHVFNLAADMGGMGFIQ. S. NHS VIMYNNTMISFNMIEAARIN	NGIK · RFFYASSACIYPEFKQLETTNVSL · · KESDAWPAEP	QDAYGLEKLATEELCKHYNK - 192
SC_1Z45/1-357	74 KGLEKV ···· FKEYKIDSVIHFAGLK ···· AVG ·· ESTOIPLRYYHNNILGTVVLLELMQQY	YNVS · KFVFSSSATVYGDATRFPNMI · · P · · IPEE · CPLGP	TNPYGHTKYAIENILNDLYNS 180
EC_1UDA/1-342	64 RNEALMT.EILHDHAIDTVIHFAGLKAVGESVOKPLEYYDNNVNGTLRLISAMRAA	ANVK-NFIFSSSATVYGDQPKIPYVESFPTGTP	QSPYGKSKLMVEQILTDLQKA 170
BP_3ENK/1-341	65 SDERALA-RIFDAHPITAAIHFAALKAVGESVAKPIEYYRNNLDSLLSLLRVMREF	RAVK - RIVFSSSATVYGVPE RSPIDET - FPLSA	TNPYGQTKLMAEQILRDVEAA 170
AF_3EHE/1-313	53 AADDIK DYLKGAEEVWHIAANP DVRIGA ENPDEIYRNNVLATYRLLEAMRKA	AGVS · RIVFTSTSTVYGEAK · · · · · · VIPTPEDYP · THP	ISLYGASKLACEALIESYCH - 154
Conserved Seq.	62 84 A P P 63 A A	85 062 YG 63	• • œ4 G4
	DybbHyAA	hhhySSyyhaG	PYYYSKYYDD
	2411114466	1111.55.221.40	FTXXSKXXIID
BT 4F.10/1-330	156 GAKSOLAGERYENVYGP RESHKORMASVAEHNENOERAEG. KVKLEGEYSG	Y GPGEOTRDEVSVEDVAKVNI VEEDH. PE K	SOLENIGTORAOPENDIAA 253
EC 1E02/1-310	157 - EANSOLVAERYENVYAR,, REAHKASMASVAEHINTOLNN, GE. SPKLEE.	GSENEK PDEVYVGDVADVNIWELEN. G	SOLENI GTORAESEOAVAD 248
UP 25VP/1-220		FOEDLEDVIDANVKAM K AO	SOVENUGY SOARSYNELVS 288
HF_33XF/1-330			
30_3VP3/1-321	101 - SVAPEVGTVRFF NVTGP GER PDALVPRLCANLLI RN. ELFVEG	DGEQRRDFTTTTDVVDREVALANRP	- PSVVNFGSGQSLSVNDVTR24/
BC_3M2P/1-309	150 - KKGLCTKNERFAHLYGF	AKREFLYAKDAAKSVIYALKQ.EKV	SGIFNIGSGDALINYEVAN 235
PS_3LU1/1-343	188 . TYGE KTIGERY FNVFGR RUDPNGAYAAV TPKWTAAMEKGD. DVYTNGD	·····GETSRDFCYIDNVIQMNILSALA·KDSA·K····	· · DNTYNVAVGDRTTLNELSG 2/9
PA_1SB8/1-341	183 - CYGFSTIGLRYFNVFGRRQDPNGAYAAVIPKWTSSMIQGD.DVYING	· · · · · DGETSRDFCYLENTVQANLLAATA · GLDA · R · · · · ·	· · NOVYNIAVGGRTSLNQLFF 276
AT_2C59/1-379	193 - DFGIECRIGRFHNINGPFGTWKGGREKAPAAFCRKAQTSTDRFEMWG	· · · · · DGLQTRSFTFIDECVEGVLRLTKS · · D · · · F · · · · ·	DEDV/NIGSDEMV/SMNEMAE204
SC_1Z45/1-357			· · KEPVNIOSDEWVSWNEWAE204
	181 DKKSWKFAILRYFNPIGAHPSGLIGEDPLGIPNNLLPYMAQVAVGRREKLYIFGDDYDS	S · · · RDGTPIRDYIHVVDLAKGHIAALQY · LE · · · AYNENE	GLCREWNL GSGKGSTVFEVYH 293
EC_10DA/1-342	181 DKKSWKFAILRYFNPIGAHPSGLIGEDPLGIPNNLLPYMAQVAVGRREK · · · LYIFGDDYD 171 · QPDWSIALLRYFNPYGAHPSGDMGEDPQGIPNNLMPYIAQVAVG · · · RR·DSLAIFGNDYF	S···RDGTPIRDYIHVVDLAKGHIAALQY·LE···AYNENE PTEDG··TG <mark>VR</mark> DYIHVMDLADGHVVAMEKL <mark>AN··</mark> ·K····	GLCREWNLGSGKGSTVFEVYH293 PGVHIYNLGAGVGNSVLDVVN278
EC_1UDA/1-342 BP_3ENK/1-341	181 DKKSWKKFALLRYFNP IGANPSGLIGEDPLGIPNNLLPYMAQVA/VGRREK···LYIFGDPVDI 171 · OPDWSLALLRYFNPVGAHPSGDMGEDPQGIPNNLMPYIAQVAVG···R·DSLAIFGNDYF 171 · OPSWRVATLRYFNPVGAHESGLIGEDPAGIPNNLMPYVAQVAVOK··LE·KLRVFGSDYPT	S···RDGTPIRDYIHVVDLAKGHIAALQY·LE···A YNENE PTEDG··TG <mark>VRDYIHVMDLADGHVVAMEKLAN··</mark> K···· T···PDGTGVR <mark>DYIH</mark> VVDLARGHIAALDA·LERRDA····	GLCREWNLGSGKGSTVFEVYH293 PGVHIYNLGAGVGNSVLDVVN278 ··SLTVNLGTGRGYSVLEVVR278
EC_1UDA/1-342 BP_3ENK/1-341 AF_3EHE/1-313	181 DKKSWKFALRYFNPIGAHPSGLIGEDPLGIPNNLLPYMAQVA/VGRREK···LYIFGDDYD: 171 · OPDWSLALLRYFNPVGAHPSGLIGEDPAGIPNNLMPYLAQVAVG···RF.DSLAIFGNDYD: 171 · OPSWRVATLRYFNPVGAHESGLIGEDPAGIPNNLMPYVAQVAVOK··LE·KLRVFGSDYD1 155 · TFJDMQAWIYRFNVIGR····RST····HOVIYDFIMKLKR···NPELEILG····	SRÓGTPIR DYIH VVDLAKGHIAALQY.LEA YN ENE PTEDGTG <mark>VR DYIH VMDLADGH VVAME KLAN</mark> K TPDGTGVR DYIH VVDLARGHIAALDA.LERRDA NGGQN ISYII VISDCVDAM LFGLRG,DER	GLCREWNLGSGKGSTVFEVYH 293 PGVHIYNLGGGVGNSVLDVVN 278 - SLTVNLGTGRGYSVLDVVR 278 - VNIFNIGSEDGLKVKRIAE 242
EC_1UDA/1-342 BP_3ENK/1-341 AF_3EHE/1-313 Conserved Seq.	181 DKKSWKFALLRYFNPIGAHPSGLIGEDPLGIPNNLLPYMAQVAJVGRREKLYIFGDPVG 171. OPDWGIALLRYFNPVGAHPSGDMGEDPOGIPNNLLPYMAQVAJVGRRR.DSLAIFGMDY1 171. DPSWRVATLRYFNPVGAHSGLIGEDPAGIPNNLLMPYVAQVAJVGKLE.KLRVFGSDYP1 185. TEIDMGAWIYEFANVIGRRSTHOVIYDFIMKLKRNPELELIG	S ROGTPIRDYIHUVDLAKGHIAALQY.LEAVNENE PTEDG.TGVRDYIHUNDLADGHUVAMEKLANK TPDGTGVRDYIHVVDLARGHIAALDA.LERRDA NGEQMKSVIYIBVVDLARGHIAALDA.LERRDA	GLCREWNL6S0K0STVFEVYH203     POVHIYNL6S0K0STVFEVYH203     POVHIYNL6S0K0STVEVVY1203     • SLTVNL6T0R6YSVLEVVZ78     • VNIFNI6SE00 IKVKRIAE242     NOFN
EC_1UDA/1-342 BP_3ENK/1-341 AF_3EHE/1-313 Conserved Seq.	181 DKKSWKFALLRYFNPIGAHPSGLIGEDPLGIPNNLLPYMAQVAVGRREK···LYIFGDPVGI 171 · OPDWSLALLRYFNPVGAHPSGLIGEDPLGIPNNLMPYIAQVAVGG···RR·DSLAIFGNDYF 171 · DPSWRVATLRYFNPVGAHESGLIGEDPAGIPNNLMPYIAQVAVG····NE·KLRVFGSDYP1 155 <u>· TF</u> DMDAWIYRFANVIGR·····RST····HGVIYDFIMKLKR···NPEELEILG····· p6 RY Pp7 P● 65 ∞5 R p8 A	S ROGTPIRDYIHUVDLAKGHIAALQY-LEAYNENE PTEDGTGURDYIHVMDLADGHVVAMEKLANK TPOGTGVRDVIHVVDLARGHIAALDA-LERR NGEQNKSYIYISDCVDAMLFGLRG-DER C p9 od	C A E WHILG S G K (Ø S T // E V/H 203     G C A E WHILG S G K (Ø S T // E V/H 203     P O V H I YN LG AG V G N S VL D V V 278     - SL T V N LG T G R G Y S V LE V V R 278     - WHI F N I G S E D G LK V K R I AE 242     P D F N B A D G
EC_1UDA/1-342 BP_3ENK/1-341 AF_3EHE/1-313 Conserved Seq.	181 DKKSWKFAILRYFNPIGAHPSGLIGEDPLGIPNNLPYMAQVAJVGREKLYIFGDYDY   171 - OPD%ILLLRYFNPVGAIPSGDWGEDPGGIPNNLPYLAQVAJVGREKE.KLRYFGNDY1   171 - OPD%WVATLRYFNPVGAHPSGLIGEDPAGIPNNLPYLAQVAJVGKE.KLRYFGNDY1   155 - TFJDMQAWIYFFANVIG   156 - TFJDMQAWIYFFANVIG   157 - RYFNPVGAHPSGLIGEDPAGIPNNLPYLAQVAJVGKLE.KLRYFGSDYP1   158 - TFJDMQAWIYFFANVIG   159 - RYFNNL   150 - RYFNNL   150 - RYFNNL   151 - RYFNNL   152 - TFJDMQAWIYFFANVIG   155 - TFJDMQAWIYFFANVIG	S ROGTPIIRD'YIHVVDLAKGHIAALQY\LEAVNENE PTEDG. TOKRDYIHVVDLAGHVAMEKLANK T PDGTOVRDVIHVVDLARGHIAALDA.LERRDA NGEQHKSVIYISDCVDAMLFGLRG.DER C p0 od	GLCREWILGSSKØSTVE EVYH 29 PGVH I YHLGSGVGNSVLDVVN 278 - SLTVHLGTGRGYSVLEVVR278 - VNIFNIGSEDQILVVKRIAE 242 p10FN p11 Aα7 haNhG
EC_1UDA/1-342 BP_3ENK/1-341 AF_3EHE/1-313 Conserved Seq. BT_4EJ0/1-330	181 DKKSWKFALLRYFNPIØAHPSGLIGEDPLGIPNNLLPYMAQVAJVGRREK···LUIFGDPVGI   171 · OPDWGIALTSODMGEDPOGIPNNLLPYMAQVAJVGRK··RE·DSLAIFØNDY   171 · DPSWRVATLRYFNPVØAHPSGLIGEDPAGIPNNLMPYVAQVAJVGK··LE·KLRVFØNDY   185 · TTDMGAWIYRFANVIG   185 · TTDMGAWISRFANVIG   186 · TTDMGAWISRFANVIG   187 · TTDMGAWIYRFANVIG   188 · TTDMGAWIYRFANVIG   189 · TTDMGAWIYRFANVIG   180 · T	S ROGTP[IRDY1HVVDLAKGHIAALQY.LEA]VNENE PTEDGTOWRDY1HVVDLAKGHIAALQA.LANK TPDGTOWRDV1HVVDLARGHIAALDA.LERR NGEQNKSV[IYISDCVDAMLFGLRG.DER C p0 ad * ****	GLCREWLGSSKKOSTFEWYH 203 PGVH 1YNLGAGVGNSVLDVVN 278 - SLTVNLGTGRGYSVLEVVN 278 - VN IFN IGSEDQIKVKRIAE 242 p10FN p11 Ad7 hanh6 
EC_10DA/1-342 BP_3ENK/1-341 AF_3EHE/1-313 Conserved Seq. BT_4EJ0/1-330 EC_1EC/2(1-340	181 DKKSWKFALLRYFNPIØAHPSGLIGEDPLGIPNNLPYMAQVAJVGREKLYIFGDV9V   171. oPDWSIALLRYFNPVGAHPSGLIGEDPAGIPNNLMPYJAQVAJVGR.REKLYIFGDV9V   171. oPDWSNVATKYFNPVGAHESGLIGEDPAGIPNNLMPYJAQVAJVGKE.K.RYFGSDYP1   155. TFJDMDAWIYRFANVIGRRSTHGVIYDFIMKLKRNPELEILG   06 RY PJ7   07 RY D   08 RYN Nh63   254 TVVNTLRALEGOPALTLAEQUEGOLVEVYPFPDALRGKYQCFTQADDTKLRAQDYDAP.FL   254 TVVNTLRALEGOPALTLAEQUEGOLVEVYPFPDALRGKYQCFTQADDTKLRAQDYDAP.FL   254 TVVNTLRALEGOPALTLAEQUEGOLVEVYPFPDALRGKYQCFTQADDTKLRAQDYDAP.FL	S ROGTPIRDYIHUVDLAKGHIAALQY-LEAYNENE PTEDGTGVRDYIHVVDLAKGHIAALQY-LEK TPOGTQNDVIHVVDLARGHIAALDA-LERRDA C p9 od * **** TVQEGVDRYVRWLFGQL	GL C REWILG S SK KØ STIV E VYN L28     GL C REWILG S SK KØ STIV E VYN 278     - SL TVNLØ TØRØ VSVL E VVN 278     - NI F MI (S SE DØ LKVKR I AE 242     p10 F N     p11 A α7     h a Nh G     - 100%     - 500
EC_1UDA/1-342 BP_3ENK/1-341 AF_3EHE/1-313 Conserved Seq. BT_4EJ0/1-330 EC_1EQ2/1-310 HP_3SYE/1-330	181 DKKSWKFALLRYFNPIØAHPSGLIGEDPLGIPNNLLPYMAQVAJVGRREKLYIFGDV9V   171 - OPD%IALLRYFNPVGAPSGLGBOPOGINGEDPOGIPNNLLPYMAQVAJVGRKE.KLRVFØKDV1   171 - OPSWRVATLRYFNPVGAHSGLIGEDPAGIPNNLLMPYVAQVAJVGKE.KLRVFØKDV1   175 - TFJDMGAWIYGR RSTRSTHØVIVDFIMKLKRNPEELEILG   165 - TFJDMGAWIYGR RSTRSTHØVIVDFIMKLKRNPEELEILG   165 - TFJDMGAWIYGR RSTRSTHØVIVDFIMKLKRNPEELEILG   165 - TTJDMGAWIYGR RSTRSTHØVIVDFIMKLKRNPEELEILG   165 - TTJDMGAWIYGR RSTRSTHØVIVDFIMKLKRNPEELEILG   165 - TTJDMGAWIYGRA RST	S ROGTPIRDYIHVVDLAKGHIAALQY.LEAVNENE PTEDG.TGVRDYIHVVDLARGHIAALQA.LERRDA TPDGTGVRDYIHVVDLARGHIAALDA.LERRDA NGEQNKSVIYISDCVDAMLFGLRG.DER C po e6 * **** TVQEGVDRYVRWLFGQL TVAEGVTEYMAWLMRDA ESQLVVDRUARDA	GLCREWILGSSK 6517 FEV7H 203 PGVH 1YNLGASVK 617 FEV7H 203 PGVH 1YNLGASVK 617 FEV7H 203 PGVH 1YNLGASVK 617 FEV7H 203 PGVH 1YNLGASVK 617 FEV7H 203 VLDVVN 278 VLDVN 278 VLDVN 278 PGVH 204 PGVH 204 PGV
EC_10DA/1-342 BP_3ENK/1-341 AF_3EHE/1-313 Conserved Seq. BT_4EJ0/1-330 EC_1EQ2/1-310 HP_3SXP/1-330	181 DKKSWKFALLRYFNPIØAHPSGLIGEDPLGIPNNLLPYMAQVAVGREKLYIFGDPVG 171 · OPDWSLALLRYFNPVØAHPSGDMGEDPGGIPNNLLMPYLAQVAVGRR.DSLAIFØNDYI 171 · OPSWRVATLRYFNPVØAHPSGDIGEDPAGIPNNLMPYLAQVAVGRR.DSLAIFØNDYI 155 <u>· TF</u> JDMDAWIYRFANVIØAHSGLIGEDPAGIPNNLMPYLAQVAVGKNPEELEILG p0 RY PP7 P● 65 œ5 R p8 A RYNNN60× 254 <u>TVVNTLRAL</u> EGOPALTLAEQVEGGLVEYVPFPDALRGKYQGFTQADDTKLRAAGYDAP.FL 240 <u>ATLAYH</u> KKGQIEYIPFPDKLKGRYQAFTQADDTKLRAAGYDKP.FKT 271 LKEHLGDFKVTYIKNPYAFFQKHTQAHIEPTILDUYTP.LYC	SROGTPIRDYIHVVDLAKGHIAALQY.LEAYNENE PTEDGTGVRDYIHVVDLAKGHIAALQA.LEAYNENE TPOGTOVRDVIHVVDLARGHIAALQA.LERRIDA NGEQNKSYIYISDCVDAMLFGLRG.DER C po od * **** TVQEGVDRYVRWLFGQL TVAEGVTEYNAWLNRDA DESGIKDYLPHIHAIFKQRA	GLC REWILG S SK KOSTIV FEVY H 203 PGVH IYNLG A SVKONSVLDVVN 278 - SLTVNLG 5 G F KOSYSVLDVVN 278 - VNIFNIG 5 G F KOSYSVLEVVN 278 - VNIFNIG 5 G F KOSYSVLEVVN 278 - VNIFNIG 5 G F KOSYSVLEVVN 278 - 100 F N P11 A α7 h a N h G - 59% 310 - 31% 330 - 31% 330
EC_10DA/1-342 BP_3EN/V1-341 AF_3EHE/1-313 Conserved Seq. BT_4EJ0/1-330 EC_1EQ2/1-310 HP_3SXP/1-330 SC_3VPS/1-321	181 DKKSWKFALLRYFNPIØAHPSØLIGEDPLGIPNNLLPYMAQVAJVGRREKLYIFGDV9V   171 - OPD%IALLRYFNPVØAHPSØLIGEDPGIPNNLLPYMAQVAJVGRKEKE.KLRVFØSDYP1   171 - OPSWRVATLRYFNPVØAHPSØLIGEDPAGIPNNLLMPYVAQVAJVGKE.KLRVFØSDYP1   185 - TFJDMQAWIYRFANVIØRRSTHØVIYDFIMKLKRNPELEILG   186 - RYFNDA   185 - TFJDMQAWIYRFANVIØRRSTHØVIYDFIMKLKRNPELEILG   185 - TFJDMQAWIYRFANVIØRRSTHØVIYDFIMKLKRNPELEILG   185 - TFJDMQAWIYRFANVIØR   185 - TFJDMQAVIYRFANVIØR   185 - TFJDMQAVIYRFANVI   185 - TFJDMQAVIYRFANVI	S ROGTPIRDYIHVVDLAKGHIAALQY.LEAVNENE PTEDG. TGMRDYIHVVDLARGHIAALQA.LERRDA TPDGTGVRDYIHVVDLARGHIAALDA.LERRDA C pp dd * **** TVQEGVDRVVRWLFGQL VGEGVTEVHAWLNRDA DESGIKDVLPHIHAIFKGQRA EGIRLTLEWWQSRDLDD RQRHFGEEGAD	GL CREWILG S SK KØ STIVE EVIT LØD     GL CREWILG S SK KØ STIVE EVVT 278     PGVH I YNLØ A GVØNSVLDVVN 278    SLTVNLØ TØRØ YSVLEVVR278    MIF RÍG SE DØL RVKRIAE 242     p10F N     p11 A d7     h a Nh G    System 100%    Slaven 100%
EC_10DA/1-342 BP_3EN/V/1-341 AF_3EHE/1-313 Conserved Seq. BT_4EJ0/1-330 EC_1EQ2/1-310 HP_3SEP/1-320 SC_3V/PS/1-321 BC_3M2P/1-309 DC_21/PS/1-321	181 DKKSWKFALLRYFNPIØAHPSGLIGEDPLGIPNNLLPYMAQVAJVGRREKLYIFGDYDY   171. OPDWSIALLRYFNPVØAHPSGLIGEDPLGIPNNLLPYMAQVAJVGRKE.KLRVFØKDYD   171. OPDWSIALLRYFNPVØAHPSGLIGEDPAGIPNNLMPYVAQVAJVGKE.KLRVFØKDYD1   171. OPDWSIALLRYFNPVØAHPSGLIGEDPAGIPNNLMPYVAQVAJVGKE.KLRVFØKDYD1   175. <u>TT</u> JDMGAWIYRFANVIGRRSTHØVIYDFIMKLKRNPEELEILG   p0 RY PP7   p0 05 GR   RYNNN63x   254 <u>TVVNTLRAL</u> EGOPALTLAEQVEØLVEYVPFPDALRGKYQGFTQADDTKLRAAGYDAP-FLU   240 ATLAYHKKGQIEYIPFPDKLKGRYQAFTQADLTNLRAAGYDKP-FKT   271 ILKEHLG	S ROGTPIRDYIHVVDLAKGHIAALQY.LEAYNENE PTEDGTOMRDYIHVVDLAKGHIAALQA.LEAYNENE TPOGTOVRDVIHVVDLARGHIAALDA.LERRDA NGEQ <u>NKSVIYISDCVDAMLFGL</u> RG.DER. C p9 ad * **** TVQEGVDRYVRWLFGQL TVAEGVTEYMAWLNRDA. DESGIKDVLPHIHAIFKGQRA. EGIRLTEWWQSRDLDDJRRHFGEEGAD NFATAVEEIHLLIKGLDDVPLWYEGHHHHHH.	GLCREWLGSSKKSTVFEVYFL23 PGVHIYNLGSKKSTVFEVYFL23 PGVHIYNLGSKKSTVFEVYFL23 PGVHIYNLGSGKKSNVLDVVN278 - VNLFNIGSESCULEVYR278 - VNLFNIGSESCULEVYR278 - 100% 330 - 59% 310 - 31% 330 - 22% 321 - 22% 320 - 22% 321 - 22% - 328 - 32% -
EC_10DA/1-342 BP_3EN/4/1-341 AF_3EHE/1-313 Conserved Seq. BT_4EJ0/1-330 EC_1EQ2/1-310 HP_3SXP/1-330 SC_3VPS/1-321 BC_3M2P/1-309 PS_3LU1/1-343	181 DKKSWKFALLRYFNPIØAHPSGLIGEDPLGIPNNLLPYMAQVAJVGRREKLYIFGDYDY   171 - OPD%IALLRYFNPVØAHPSGLIGEDPAGIPNNLMPYLAQVAJVGKE.K.RYFGNDY1   171 - OPD%IALLRYFNPVØAHPSGLIGEDPAGIPNNLMPYLAQVAJVGKE.K.RYFGNDY1   175 - TTJDMQAWIYRFANVIGRRSTHGVIYDFIMKLKRNPELELIG   185 - TTJDMQAWIYRFANVIGRRSTHGVIYDFIMKLKRNPELELIG   186 RY PF7 P 05 d5 R   187 ALLAYFMYNDS P 05 d5 R   188 RYNNNDS ******   249 ATLAYHKKGQIEYIPFPDKLKGRYQGFTGADDTKLRAAGYDAP-FL ******   240 ATLAYHK	SROGTP[IRDYIHVVDLAKGHIAALQY.LEAVNENE PTEDG.TGVRDVIHVVDLARGHIAALQA.LERR]DA TPDGTGVRDVIHVVDLARGHIAALQA.LERR]DA C	C C REWIL G S G K 0 STILE INTEL 204     G C REWIL G S G K 0 STILE INTEL 204     PG VH I YNL G A G V G NS VL D VVN 278     - SL TVNL G T G R G Y SVL E VVN 278     - NI F M I G S E D G I KVKR I A E 242     p10 F N     p11 A d7     h a N h G     - 59% 310     - 22% 321     - 22% 340     - 24%
EC_10DA/1-342 BP_2EN/V/-341 AF_3EHE/1-313 Conserved Seq. BT_4EJ0/1-330 EC_1EQ2/1-310 HP_35XP/1-320 SC_3VPS/1-321 BC_3M2P/1-309 PS_3LU1/1-343 PA_1SB8/1-341	181 DKKSWKFALLRYFNPIØAHPSGLIGEDPLGIPNNLLPYMAQVAJVGREK···LYIFGDV9V   171 - QPU%JALLRYFNPVØAHPSGLIGEDPAGIPNNLLPYMAQVAJVGREK···LE·KLRVFØSDYP   171 - OPSWRVATLRYFNPVØAHPSGLIGEDPAGIPNNLMPYVAQVAJVGK··LE·KLRVFØSDYP   185 - TEJDMGAWIYERANU IGR.···RST.···HØVIYDFIMKLKR.···NPEELEILG   186 - TKJDMGAWIYERANU IGR.···RST.···HØVIYDFIMKLKR.···NPEELEILG   187 - NNh60x   254 - TVVNTLRALEGOPALTLAEQVEGGLVEYVPFPØALRGKYQCFTQADQTKLRAAGYDAP·FL[   240 - ATLAYHK········KGQIEYIPFPØLKLKGRYQAFTQADTTKLRAAGYDAP·FL[   241 - LAATLAYHK····································	S ROGTPIRDYIHVUDLAKGHIAALQY.LEAVNENE PTEDG. TGMRDYIHVUDLARGHIAALQA.LERRDA T PDGTGVRDYIHVUDLARGHIAALDA.LERRDA NGEQNKSVIYISDCUDANLFGLRG.DER. C p9 d5 * **** TVQEGVDRYVRWLFGQL TVAEGVTEYMAWLMRDA DLESGIKDYLPHIHAIFKGQRA GIRLTLEWWQSRDLDDJRQRIFGEGAD NFATAVEEIHLLXRGLDDVPLWYEGHHHHH KIRGGLRLBWWYNFLKG	GL C REWNL G S G K G ST V F EVY H 23 PG V H Y NL G A G V G N S V L D V V N 278 - S L Y NL G T G R G Y S V L E V V N 278 - V N I F N I G S E D G I K V K R I A E 242 p10 F N p11 A d7 h a N h G - 100% 330 - 59% 310 - 22% 321 - 22% 324 - 22% 341
EC_10DA/1-342 BP_2SE/Vi-341 AF_3E/HE/1-313 Conserved Seq. BT_4EJ0/1-330 EC_1EQ2/1-310 HP_35XP/1-330 SC_3VP/5/1-321 BC_3M/2/1-309 PS_3LU/1/1-343 PA_15BB/1-341 AT_2C59/1-379	181 DKKSWKFALLRYFNPIØAHPSGLIGEDPLGIPNNLLPYMAQVAJVGRREKLYIFGDYDY   171. oPD%IJALLRYFNPVGANPSGDNGEDPGGIPNNLMPYVAQVAJVGKLE.KLRYFGSDYP   171. oPD%IVATKYFNPVGANPSGLIGEDPAGIPNNLMPYVAQVAJVGKLE.KLRYFGSDYP   155. TFJDMQAWIYRFANVIGRRSTHGVIYDFIMKLKRNPELELELGG   165. TFJDMQAWIYRFANVIGRRSTHGVIYDFIMKLKRNPELELELGG   166. RYDNNHGX	S ROGTPIRDYIHVVDLAKGHIAALQY.LEAYNENE PTEDG. TGVRDYIHVVDLARGHIAALQA.LERRDA TPOGTGVRDYIHVVDLARGHIAALDA.LERRDA C PO dG * **** TVQEGVDRYVRWLFGQL DLESGIKDYLPHIHAIFKGQRA EDIRLTLEWWQSRDLDDIRQRIFGEEGAD KIREGLRLSMPWYVRFLKG VAGOVALAMPWYIMFLK EKEGLRLSMPWYVRFLKGL	GLC REWILL G S G K G STIVF EVY H 20 PG VH I YHL G A G V G NS VL DVV N 278 - SL TVH LG G G C K G NS VL DVV N 278 - SL TVH LG G G C K G NS VL EVV N 278 - VH I F N G S E D G I KVKR I A E 242 p10F N p11 A σ7 h a N h G 
EC_10DA/1-342 BP_3EN/V1-341 AF_3EHE/1-313 Conserved Seq. BT_4EJ0/1-330 EC_1EQ2/1-310 HP_3SXP/1-320 BC_3M/2P/1-309 PS_3L/U1/1-343 PA_158B/1-341 AT_2C59/1-379 SC_12/4/1-387	181 DKKSWKFALLRYFNPIØAHPSGLIGEDPLGIPNNLLPYMAQVAJVGRREKLYIFGDPVG   171 - OPD%IALLRYFNPVGAHPSGLIGEDPAGIPNNLMPYLAQVAJVGKE.K.RVFGDVD'   171 - OPSWRVATLRYFNPVGAHPSGLIGEDPAGIPNNLMPYVAQVAJVGKE.K.RVFGDVD'   185 - TFJDMGAWIYFANVIGRRSTHOVIYDFIMKLKRNPELEILG   186 RY PJ7   187 NNhhGx   254 TVVNTLRALEGOPALTLAEQVEGGLVEYVPFPOALRGKYQCFTQADCTKLRAAGYDAP-FL   240 ATLAYHKKGQIEYIPFPOALKGRYQAFTQADTKLRAAGYDAP-FL   241 LAYHKKGQIEYIPFPOKLKGRYQAFTQADCTKLRAAGYDAP-FL   242 ILQATSP	S ROGTPIRDYIHVVDLAKGHIAALQY.LEAVNENE PTEDG. T.GVRDYIHVVDLARGHIAALQA.LERRDA TPDGTGVRDYIHVVDLARGHIAALQA.LERRDA C. D9 00 * **** TVQEGVDRYVRWLFGQL TVAEGVTEYMAWLNRDA DESGIKDYLPHIHAIFKGQRA EGIRLTLEWWQSRDLDDIRQRIFGEEGAD WFATAVEEIHLLXRGLDVFUYEGHHHHH KIREGULESDWWYNFLKG VSAGVALAMPWYIMFLKG RLKEGLRITYFWIKEGIEKEKAKGSDVSL VGSSKVVGTQAP	GLCREWRLGSSKKGSTVFEVYFL23 PGVHIYRLGAGVGNSVLDVVN278 - SLTVNLGGRGVGNSVLDVVN278 - VNIFNIGSEDQIKVLEVVR278 - VNIFNIGSEDQIKVLEVVR278 - VNIFNIGSEDQIKVLEVVR278 - 100% 330 - 59% 310 - 31% 330 - 22% 321 - 22% 321 - 22% 341 VQLGSLRAADGKE 21% 379 - 22% 341
EC_1UDA/1-342 BP_2SE/Vi-341 AF_3E/HE/1-313 Conserved Seq. BT_4E/0/1-330 EC_1EQ2/1-310 HP_3SXP/1-330 EC_3VPS/1-321 BC_3M2P/1-309 PS_3LU1/1-343 PA_15SBV1-341 AT_2C591-379 SC_1245/1-367 EC_1UDA/1-342	181 DKKSWKFALLRYFNPIØAHPSGLIGEDPLGIPNNLLPYMAQVAJVGRREKLYIFGDV9V   171. OPDWSILLLRYFNPVØAHPSGLIGEDPGIPNNLLPYMAQVAJVGRKE.KLRVFØBV   171. OPDWSILLLRYFNPVØAHPSGLIGEDPAGIPNNLMPYVAQVAJVGKE.KLRVFØSDYP   185. TTFJDMGAWIYGRARSTH0VIYDFIMKLKRNPEELEILG   p0 RY PP7   p0 05 G   RYNNh6x   254 TVVNTLRALEGOPALTLAEQVEQOLVEYVPFPDALRGKYQCFTQADDTKLRAAGYDAP.FL[   240 ATLAYHK	SRDGTPIRDYIHVVDLAKGHIAALQY.LEAYNENE PTEDGTGVRDVIHVVDLAKGHIAALQA.LERRDAK. TPDGTQNDVIHVVDLARGHIAALDA.LERRDAK. TOGEQ <u>NKSVIYISDCVDAMLFGL</u> RG.DER. C p9 d0 * **** TV <u>AEGVTEYMAWLN</u> RDA LESGIKDTVLPHIHAIFKGQRA DIRTLEWWQSRDLDDJRRHFGEGAD NFATAVEEIHLUXRGLDDVPLWYEGHHHHHH KIREGLRISTYFWIKEQIEKEKAKGSD <u>VSL</u> YGSSKVVGTQAP DVEDSCKDLWKWTTENPFGYQL TDEMAQTWHWQSRHPGYDL	GLCREWRLGSSK 65TVFEVYFL23 PGVH1YRLGSK 65K 65TVFEVYFL23 PGVH1YRLGAGVGNSVLDVVN278 - SLTVNLG5GRGASVLDVVN278 - SLTVNLG5GRGASVLDVN278 - SLTVN278 - SLTVNLG5GRGASVLDVN278 - SLTVN278 - SLTVN278
EC_10DA1-342 BP_2BN/V1-341 AF_2EHE/1-313 Conserved Seq. BT_1EJ0/1-330 EC_1EQ2/1-310 HP_3SXP/1-320 BC_3M/2P/1-321 BC_3M/2P/1-321 BC_3M/2P/1-321 BC_3M/2P/1-321 FA_158BV1-341 AT_2C59H-379 SC_1245/1-341 BP_2BN/V1-341	111 DKKSWKFALLRYFNPIØAHPSGLIGEDPLGIPNNLLPYMAQVAJVGRREKLYIFGDPYG   121 - OPDWRJALLRYFNPØGAJPSGLIGEDPAGIPNNLMPYLAQVAJVGKE.KLRVFGSDYP   121 - OPDWRVATLRYFNPØGAJPSGLIGEDPAGIPNNLMPYLAQVAJVGKE.KLRVFGSDYP   125 - TFJDMGAWIYRFANVIGRRSTHGVIYDFIMKLKRNPELEILG   p0 RY PJ7 P 05 c5 R   p0 RY PJ7 P 05 c5 R   240 ATLAYFNFANVIGRRSTHGVIYDFIMKLKRNPELEILG   254 TVVNTLRALEGOPALTLAEQVEOGLVEYVPFPDALRGKYQCFTQADDTKLRAAGYDAP.FL   240 ATLAYHKKGGIPYIPFPDALKKGRYQAFTQADLTNLRAAGYDAP.FL   241 LAYHKKGGIPYIPFPDALKKGRYQAFTQADLTNLRAAGYDKP.FKI   242 TLAYHKKGGIPYIPFPDALKGRYQAFTQADLTNLRAAGYDKP.FKI   243 TLAYHKKGGIPYIPFPDALKGRYQAFTQADLTNLRAAGYDKP.FKI   244 FLAASH   245 TIVNAFGONKDNLUKNPNANEG   246 TINNAFGONKDNLUKNPNANEG   247 TLKEHGOVS	S RDG TP   RDY I HVVDLAKCH I AALQY.LEAVNENE PTEDG. TG MRDY I HVVDLARCH I AALQA.LERRDA TPDG TG VRDY HVVDLARCH I AALDA.LERRDA C PD 00 * **** TVQEGVDRYVRWLFGQL TVAEGVTEYMAWLNRDA DESGI I KDYLPH I HA I F KGQRA GO I I TLEWWQSRDDDD   RQRI F GEGAD MF ATAVEE I HLLXRGLD V PLWYEGH HHHH MF ATAVEE I HLLXRGLD V PLWYEGH HHHHH K (REGLRSMPWYVRT LKG KASOVALAMPWY I MF LKG K KEGUR I TYFWI KEGI E KEKKAKS DVSL VGSSKVVGTQAP DVEDSCKDLWKWTTENPFGYD DE ERWKADSR F QGYD	GL C REWNL 6 S 6 K 6 ST V F EVY H 20 P 6 V H YNL 6 A 6 V 6 N 5 V L D VV N 278 • S L T V N L 6 T 6 R 6 Y 5 V L E VV R 278 • VN I F N I 6 S E D 0 I KV K R I A E 242 p 10 F N p 11 A a7 h a N h 6 100% 330 59% 310 22% 321 22% 309 V0 L 6 S L R A D 6 K E 21% 376 22% 357 22% 357 22% 341
EC_10DA/1-342 BP_3EN/V1-341 AF_3EHE/1-313 Conserved Seq. EC_1EQ2/1-310 EC_1EQ2/1-310 EC_1EQ2/1-310 EC_3VPS/1-321 BC_3M2P/1-309 PS_3LU1/1-343 PS_3LU1/1-343 PS_12EN/V1-341 AT_2C659/1-371 EC_1UDA/1-342 BP_3EN/V1-341 BP_3EN/V1-341	181 DKKSWKFALLRYFNPIJOAHPSGLIGEDPLGIPNNLLPYMAQVAJVGREKLYIFGDYDY   171 - QPUSILLRYFNPVGAHPSGLIGEDPGIPNNLLPYMAQVAJVGREKLE.KLRVFGODYF   171 - OPSWRVATLRYFNPVGAHPSGLIGEDPAGIPNNLMPYVAQVAJVGKLE.KLRVFGODYF   185 - TEJDMGAWIYEFANVIGRRSTHOVIYDFIMKLKRNPEELEILG   p0 RY PP7   p0 RY PP7   p0 RY PP7   p0 RY PP7   p1 RYNNh60x   xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	S ROGTPIRDYIHVVDLAKGHIAALQY.LEAVNENE PTEDG. T.GVRDYIHVVDLARGHIAALQA.LERRDA NGEQNKSVIYIHVVDLARGHIAALDA.LERRDA NGEQNKSVIYIHVVDLARGHIAALDA.LERRDA NGEQNKSVIYISDCVDAMLFGLRG.DE R C p9 d5 * **** TVQEGVDRVVRWLFGQL  LESGIKDYLPHIHAIFKGQRA GIRLTLEWWQSRDDDJIRQRIFGEGAD NFATAVEEIHLLXRGLDDVPLWYEGHHHHH  K REGLRISDWVYRFLKG VSAGVALAMPWYIMFLKG  RLKEGLRITYFWIKEQIEKEKAKGSDVSI VSAGVALAMPWYIMFLKG  RLKGLRITYFWIKEQIEKEKAKGSDVSI VSAGVALAMPWYIMFLKG  RLKGLRITYFWIKEQIEKEKAKGSDVSI VSAGVALAMPWYIMFLKG  SEGAVENAWGILENEPGYDL 	GLCREWRLG S SK KOSTVFEVYFL23 PGVH1YRLG A SV KOSTVFEVYFL23 + VH1YRLG A SV KOSTVFEVYRL23 + VH1FNLG S SED GLIKVKRIAE 242 β10FN β11 A σ7 h a N h G 
EC_10DA/1-342 BP_2SE/N/-341 AF_3E/HE/1-313 Conserved Seq. BT_4EJ0/1-330 EC_1EQ2/1-310 HP_35XP/1-330 SC_3YP2/1-321 BC_3M2P/1-309 PS_3LU1/1-343 AT_2C59/1-379 SC_1246/1-367 EC_110DA/1-342 BP_3E/N/-1-313 Conserved Seq.	111 DKKSWKFALLRYFNPIØAHPSGLIGEDPLGIPNNLLPYMAQVAJVGRREKLYIFGDYDY   121 - OPDWENJALLRYFNPVGANPSGDWGEOPGGIPNNLMPYLAQVAJVGKE.K.RYFGNDY   121 - OPDWENJALLRYFNPVGANPSGDWGEOPGGIPNNLMPYLAQVAJVGKE.K.RYFGNDY   125 - TTJDMQAWIYFANVIGRRSTHGVIYDFIMKLKRNPELELIG   126 - TTJDMQAWIYFANVIGRRSTHGVIYDFIMKLKRNPELELIG   127 - OPDWENJATIRYFNVGANPSGLIGEOPAGIPNNLMPYLAQVAJVGKLE.K.RYFGSDYP   128 - TTJDMQAWIYFANVIGRRSTHGVIYDFIMKLKRNPELELIG   129 - RYPAT   124 - TLVYNTLRALEGOPALTLAEQVEGGLVEYVPFPDALRGKYQGFTOADDTKLRAAGYDAP.FL   244 - TLVYNTLRALEGOPALTLAEQVEGGLVEYVPFPDKLKGRYQAFTOADDTKLRAAGYDAP.FL   240 - TLAYHK	SROGTP[IRDYIHVVDLAKCHIAALQY.LEAVNENE PTEDG.TG/RD/IHVVDLARGHIAALQA.LERR]DA TPDGTG/RD/IHVVDLARGHIAALDA.LERR]DA C	GLC REWNLG S SK KOSTIVFEVYFL23 PGVH 1YNLG S SK KOSTIVFEVYFL23 PGVH 1YNLG S SK KOSTIVFEVYFL23 PGVH 1YNLG S SK KOSTIVFEVYN 278 - SLTVNLG ST SK KOSTIVFEVYL 278 - SLTVNLG SK SK KOSTIVFEVYL 278 - SLTVNLG SK KOSTIVE - SLTVNLG SK K

#### 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  $\alpha$ -helix, red for  $\beta$ -strand and green for  $3_{10}$ -helix. Green and red characters represent residues comprising the NADP<sup>+</sup>- 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<sup>+</sup>. A blue P indicates residues interacting with the phosphate attached to the ribose moiety of NADP<sup>+</sup>. 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-p.-help 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- $\alpha$ -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;  $\sim$ 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 TGXXGXXG motif for NAD(P)<sup>+</sup> binding (Persson et al., 2003). These are the signature sequences of the eSDR family (Fig. 1). BtAGME 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<sup>1</sup>). Therefore, the elucidation of the unique structural features of BtAGME 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 BtAGME. We have also performed an in silico study of BtAGME and EcAGME 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 BtAGME

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  $25 \,\mu M$  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<sub>6</sub> tag. After transformation into DH5 $\alpha$ , 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<sup>-1</sup> 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<sub>600</sub> of 0.6–0.8 was reached. At this point, expression of *Bt*AGME was induced by isopropyl  $\beta$ -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 7650g (6500 rev min<sup>-1</sup>) 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 mM Tris-HCl pH 8.0, 100 mM NaCl, 10 mM imidazole, 1 mM PMSF, 10  $\mu$ g ml<sup>-1</sup> 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 900g (15 000 rev min<sup>-1</sup>) for 30 min in a high-speed refrigerated centrifuge at 277 K. The supernatant was affinity purified using a HisTrap column on an ÄKTAexplorer system (GE Healthcare, Piscataway, New Jersey, USA). The column was equilibrated with a buffer consisting of 50 mM Tris-HCl pH 8.0, 300 mM NaCl, 10 mM imidazole. The target protein was eluted with a buffer consisting of 50 mM Tris-HCl pH 8.0, 100 mM NaCl with a gradient from 10 to 500 mM imidazole. BtAGME was further purified by ionexchange chromatography using a 5 ml Hi-Trap Q column (GE Healthcare, Piscataway, New Jersey, USA) equilibrated with buffer containing 20 mM 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<sub>6</sub>-tagged BtAGME. The purified protein contained a noncleavable N-terminal His<sub>6</sub> tag followed by five glycine residues. For crystallization, the protein was concentrated to 8.8 mg ml<sup>-1</sup>, 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 mM Tris-HCl pH 8.0.

<sup>&</sup>lt;sup>1</sup> 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 sittingdrop 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  $(8.8 \text{ mg ml}^{-1})$  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  $0.5^{\circ}$  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 EcAGME (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 REFMAC5 (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<sup>+</sup> 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

#### 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	$P2_{1}2_{1}2_{1}$
Unit-cell parameters (Å)	a = 158.3, b = 160.9, c = 169.6
Solvent content (%)	57.76
$V_{\rm M} ({\rm \AA}^3{\rm Da}^{-1})$	2.91
Total No. of residues	3300
Total non-H atoms	27040
No. of NADP <sup>+</sup> molecules	10
No. of water molecules	456
Average temperature factors $(Å^2)$	
Protein	50.52
Solvent	41.49
NADP <sup>+</sup>	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
0	

† The data were obtained from a single crystal.

coordinates and structure factors have been deposited in the Protein Data Bank (PDB entry 4ej0).

#### 2.7. Docking study

Docking studies to predict the catalytic mechanism of BtAGME and EcAGME 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 BtAGME and chain D of PDB entry 1eq2 for EcAGME) 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

### research papers

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<sup>+</sup> 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 BtAGME. (a) The crystal structure of BtAGME superimposed on that of EcAGME. Green represents BtAGME and pink represents EcAGME. The NADP<sup>+</sup> 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 EcAGME forms two hydrogen bonds to ADP- $\alpha$ -D-glucose whereas Gln211 of BtAGME forms one hydrogen bond to the modelled ADP- $\alpha$ -D-glucose. The left box represents the decamerization interface (UR2) of BtAGME, 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) BtAGME 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 BtAGME. (c) Diagrams of the decameric form of BtAGME. Left, top view; right, side view. Each subunit is coloured differently. The NADP<sup>+</sup> molecules are represented by stick models. The dimensions of the decamer are represented.

Tables S1a and S1c). 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<sup>+</sup> of *Bt*AGME were almost unchanged compared with those of substrate-bound crystal structures of EcAGME, 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 BtAGME and the first conformer for EcAGME as obtained from their Surflex-Dock runs (Supplementary Tables S1b and S1d). 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 BtAGME complexed with ADP-D,D-Hep and PM0078687 for EcAGME 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-CGGGCAGACGCGC-3' for the forward strand and 5'-GCGCGTCTGCCCGGCCCGGGCC-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 pM 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<sub>2</sub>O. 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  $\mu$ l *Dpn*I was added to the 40  $\mu$ l 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  $\mu$ l of the mixture was used to transform 45  $\mu$ l of XL10-Gold competent cells treated with 2  $\mu$ l  $\beta$ -mercaptoethanol. The transformed *E. coli* XL10-Gold competent cells were grown in 500  $\mu$ l NZY broth and were incubated at 310 K for 1 h. A 100  $\mu$ l aliquot of the NZY broth was spread onto LB plates containing 50 mg ml<sup>-1</sup> 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<sup>-1</sup> were used. A micro cuvette with a volume of 12  $\mu$ l 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 *E*cAGME

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<sup>+</sup>, which co-purified during protein purification. The



#### Figure 3

Electrostatic surface potentials of AGMEs. (a) BtAGME complexed with *in silico* modelled ADP-D,D-Hep. The dimensions of the entrances of its cofactor-binding and substrate-binding pockets are approximately  $16 \times 8$  and  $13 \times 8$  Å, respectively. (b) EcAGME complexed with ADP- $\alpha$ -D-glucose. (c) HpAGME. The flexible loop L $\beta$ 13 $\alpha$ 9 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<sup>+</sup> or NAD<sup>+</sup>, located in the lower cavities are represented by stick models.

final model exhibited good stereochemical geometry and was refined to R and R<sub>free</sub> 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. 2c). The previously solved structures of *EcAGME* have been crystallized in the presence of substrate homologues such as ADP- $\alpha$ -D-glucose (Deacon *et al.*, 2000) or ADP- $\beta$ -Dmannose (Kowatz et al., 2010). However, BtAGME was crystallized in a substrate-free form. Structural analyses of ligand-free BtAGME and ligand-bound EcAGME revealed that their overall architecture is almost conserved, with an r.m.s.d. of 0.59 Å for 281 C<sup> $\alpha$ </sup> atoms (henceforth, structural comparison is between chain A of BtAGME and chain D of EcAGME unless mentioned otherwise). When the NADP<sup>+</sup>- binding pockets of the two structures are compared along the bound NADP<sup>+</sup> molecule, the residues of BtAGME 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- $\alpha$ -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 BtAGME are almost unchanged compared with those of EcAGME. 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 BtAGME and EcAGME 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) BtAGME complexed with modelled ADP-D,D-Hep. (b) E. coli GalE complexed with UDP- $\alpha$ -D-glucose. (c) WbgU from *Pseudomonas aeruginosa* complexed with UDP-*N*-acetylglucosamine. (d) A. thaliana GME complexed with GDP- $\beta$ -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<sup>+</sup> 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	e substrate-binding cavities of AGMEs and GalE homolo	gues.

Species (PDB code)	Protein	Substrate	Z-score†	R.m.s.d.‡	SeqID§ (%)	SeqSim¶ (%)	Cov†† (%)	Cavity volume (Å <sup>3</sup> )	Molecular volume $(\text{\AA}^3)$	Cavity/ molecule	Related PDB entries
<i>B. thailandensis</i> (4ei0)	AGME	ADP-D.D-Hep	8.03	0	100	100	100	416.0	213.5	1.9	
E. $coli$ (2x6t)		ADP- $\beta$ -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)	e	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- $\beta$ -L-galactose or GDP- $\alpha$ -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- $\alpha$ -D-glucose		7.13	2.72	22	37	93	830.8	181.5	4.6		
A. fulgidus (3ehe) N/A		N/A	6.81	2.92	22	41	85	N/A	N/A	N/A	
E. coli (11rk)		UDP-N-acetylglucosamine	7.02	2.97	21	34	94	756.5	237.4	3.2	1lrj, 1kvs, 1kvq, 1kvr, 1kvu, 1kvt

 $\dagger$  Measure of the statistical significance of the result relative to an alignment of random structures.  $\ddagger$  R.m.s.d. value of the alignment. § Percentage sequence identity in the alignment.  $\P$  Percentage sequence similarity in the alignment  $\dagger$  The coverage of aligned residues in chain 1.

dimensions of  $12 \times 8$  Å and  $13 \times 8$  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 BtAGME. 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 EcAGME and AGME from Helicobacter pylori (HpAGME). The remainder are UDPgalactose 4-epimerase (GalE) homologues such as GalEs, NAD-dependent NDP-hexosamine 5,6-dehydratase, UDP-Nacetylglucosamine 4-epimerases (WbgUs/WbpPs) and GDPmannose-3',5'-epimerases (GMEs). Since AGMEs and GalE homologues belong to the eSDR family, their overall structures are quite similar to each other (Fig. 2b), with several highly conserved eSDR signature motifs (Persson et al., 2003). In addition, there is one modified motif <sup>165</sup>RYhNhhGx<sup>172</sup> instead of the known motif h(KR)xxNGP that is common to the eSDR family and one new motif <sup>237</sup>haNhGA<sup>241</sup> 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<sup>+</sup> and GalE homologues prefer NAD<sup>+</sup> as a cofactor. Therefore, the residues Asn33, Lys39 and Lys54 that attract the phosphate moiety of NADP<sup>+</sup> 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. 4a and 4b). The spacious room required for the GalE epimerization mechanism is inferable from the 180° rotation step of the carbohydrate moiety of UDP- $\alpha$ -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  $\sim$ 700 Å<sup>3</sup>, which is three times larger than the volume of the carbohydrate moieties of their substrates (Fig. 4c). In contrast, that of GME from Arabidopsis thaliana (AtGME), which adopts a ring-flip mechanism that requires a smaller volume compared with that of a ring-rotation mechanism (Major et al., 2005) occupies  $\sim 460 \text{ Å}^3$  (Fig. 4d). Since the volumes of the two AGMEs are smaller than that of AtGME, 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 2b).

#### 3.3. Structural comparison with HpAGME

The crystal structure of AGME from *H. pylori* (*Hp*AGME) has recently been determined (Shaik et al., 2011). A PSI-BLAST search showed that HpAGME comprises a speciesspecific 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 BtAGME and EcAGME, respectively, its structural similarity is quite high, with an r.m.s.d. of 0.93 Å for 213 C<sup> $\alpha$ </sup> atoms with *Bt*AGME and of 0.92 Å for 208 C<sup> $\alpha$ </sup> atoms with *Ec*AGME. However, there are several local conformations of HpAGME that deviate severely from those of the other two AGMEs. Prominent differences are observed in the three loops LN, L $\beta$ 2G1 and L $\beta$ 13 $\alpha$ 9 of HpAGME (Fig. 5a 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. L $\beta$ 2G1 of HpAGME does not have the conserved asparagine and lysine (Asn33 and Lys54 in BtAGME), the role of which is to bind to the phosphate moiety of NADP<sup>+</sup> in the cofactor-binding site. Instead, the side chain of Ser58 of HpAGME occupies the position of the phosphate moiety of  $NADP^+$  (Fig. 5b). Consequently,  $NAD^+$  (or NAD) was detected in the crystal structure of HpAGME. L $\beta$ 13 $\alpha$ 9, which covers the substrate-binding pocket of HpAGME, is also very flexible owing to the absence of the <sup>283</sup>PFP<sup>285</sup> motif stabilizing this loop region in BtAGME (Fig. 4a and Supplementary Fig. S1). As a result, the space for the adenosine portion of ADP- $\alpha$ -D-glucose is totally exposed in the crystal structure of HpAGME (Fig. 3c). These species-specific sequence and structural properties indicate that the substrate-binding affinity and epimerization mechanism of HpAGME may differ from those of BtAGME and EcAGME.

#### 3.4. Structural properties of two unique regions

The crystal structure of BtAGME provides information on the structural properties of two unique regions that are only found in Burkholderia species. The first region is the loop between  $\beta 8$  and  $\beta 9$  (residues 200–210) located next to the adenine moiety of ADP- $\alpha$ -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 EcAGME. Consequently, the binding affinity of ADP-D,D-Hep for BtAGME is expected to be slightly decreased compared with that for EcAGME. 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 substratebinding 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  $\alpha$ -helix  $\alpha$ 8 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<sup>2</sup> 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 BtAGME and HpAGME. Green represents BtAGME and gold represents HpAGME. The NADP<sup>+</sup> and NAD<sup>+</sup> molecules are represented by stick models (blue represents N atoms, red O atoms, green C atoms and scarlet P atoms). Secondarystructure elements of BtAGME are labelled. (b) Cofactor-binding pockets. NADP<sup>+</sup> and NAD<sup>+</sup> molecules are represented by stick models. The highly conserved residues that interact with the phosphate moiety of NADP<sup>+</sup> 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 onebase 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- $\beta$ -D-mannose (Kowatz et al., 2010) and ADP- $\alpha$ -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 BtAGME and EcAGME 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 BtAGME and 14.4 for *EcAGME* (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. 6a). 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<sup>+</sup>/NADPH. Therefore, the correct positioning of the proton at C6" in ADP-D,D-Hep relative to the C4 atom of

NADP<sup>+</sup> 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



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<sup>+</sup> 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<sup>+</sup>, 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<sup>+</sup>/NADPH together with the function of Tyr139 in acid/base catalysis. B represents a base, which may be one of the surrounding water molecules.

the C7" hydroxymethyl group in the substrate mimics ADP- $\beta$ -D-mannose and ADP- $\alpha$ -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- $\beta$ -D-mannose rather than to that of ADP- $\alpha$ -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<sup>+</sup> at a distance of 2.7 Å

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 BtAGME 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<sup>+</sup> abstracts hydride from C6". The resulting ketone intermediate of ADP-D,D-Hep would perturb interactions with BtAGME 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. 6b).

#### 4. Concluding remarks

In conclusion, the AGME family has structural features in common with GalE homologues. Nevertheless, AGMEs prefer NADP<sup>+</sup> as a cofactor instead of NAD<sup>+</sup>. Both enzymes have distinctive substrate-binding cavities that are suitable for their epimerase mechanism. BtAGME has unique structural properties that are not found in other bacterial AGMEs. The first unique region of BtAGME 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 BtAGME and ligandbound EcAGME 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 BtAGME and EcAGME 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), BtAGME 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. D66, 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. D68, 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.* D67, 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. D66, 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.* D57, 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.* D67, 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. D67, 235-242.