

Structural Biology

Structural and enzymatic analyses of a glucosyltransferase Alr3699/HepE involved in *Anabaena* heterocyst envelop polysaccharide biosynthesis

Xue-Ping Wang, Yong-Liang Jiang, Ya-Nan Dai, Wang Cheng, Yuxing Chen¹, and Cong-Zhao Zhou¹

Hefei National Laboratory for Physical Sciences at the Microscale and School of Life Sciences, University of Science and Technology of China, Hefei, Anhui 230027, People's Republic of China

¹To whom correspondence should be addressed: Tel: +86-551-63600406; Fax: +86-551-63600406; e-mail: zcz@ustc.edu.cn (C.Z.Z.); cyxing@ustc.edu.cn (Y.C.)

Received 26 May 2015; Revised 10 December 2015; Accepted 13 December 2015

Abstract

Formation of the heterocyst envelope polysaccharide (HEP) is a key process for cyanobacterial heterocyst differentiation. The maturation of HEP in *Anabaena* sp. strain PCC 7120 is controlled by a gene cluster termed HEP island in addition to an operon *alr3698-alr3699*, which encodes two putative proteins termed Alr3698/HepD and Alr3699/HepE. Here we report the crystal structures of HepE in the apo-form and three complex forms that bind to UDP-glucose (UDPG), UDP&glucose, and UDP, respectively. The overall structure of HepE displays a typical GT-B fold of glycosyltransferases, comprising two separate $\beta/\alpha/\beta$ Rossmann-fold domains that form an inter-domain substrate-binding crevice. Structural analyses combined with enzymatic assays indicate that HepE is a glucosyltransferase using UDPG as a sugar donor. Further site-directed mutageneses enable us to assign the key residues that stabilize the sugar donor and putative acceptor. Based on the comparative structural analyses, we propose a putative catalytic cycle of HepE, which undergoes “open-closed-open” conformational changes upon binding to the substrates and release of products. These findings provide structural and catalytic insights into the first enzyme involved in the HEP biosynthesis pathway.

Key words: cyanobacteria, crystal structure, glycosyltransferase, GT-B fold, heterocyst envelope polysaccharide

Introduction

Cyanobacteria represent a unique phylum of oxygenic phototrophic prokaryotes that are widespread on the earth, participating in global carbon cycle (Bishop 1972; Pelroy and Bassham 1972). In some filamentous cyanobacteria, such as *Anabaena* sp. strain PCC 7120, the normal photosynthetic vegetative cells can differentiate into highly specialized cells, called heterocysts, to fix nitrogen once the combined nitrogen sources are limited (Wolk et al. 1977; Haselkorn 1978). Heterocysts, which are formed along the filament in a semi-regular pattern

separated by ~10 vegetative cells, are morphologically distinct from vegetative cells under light microscope (Nicolaisen et al. 2009). A key feature of heterocysts is the deposition of thick layers of polysaccharide and glycolipid onto the outer-membrane. The double layers impede the entry of oxygen to maintain a micro-oxic intracellular milieu (Walsby 1985; Wolk and Murry 1989), which guarantees the activity of oxygen-sensitive nitrogenases (Fay 1992; Gallon 1992). The inner glycolipid layer functions as the principal barrier against oxygen permeation, whereas the outer polysaccharide layer is required

for the reinforcement of the glycolipid (Wolk and Murry 1989). To date, the fine structure of heterocyst envelope polysaccharide (HEP) remains largely unknown. Only the skeleton structures of HEP in two *Anabaena* strains (*A. variabilis* ATCC 29413 and *A. cylindrica*) and a *Cylindrospermum* strain have been identified, revealing a same repetitive backbone (Cardemil and Wolk 1979;981) that consists of glucose and mannose as the most abundant monosaccharide residues (Wolk and Dunn 1970). In detail, the repetitive unit of HEP backbone is $\beta(1,3)$ -linked tetrasaccharides composed of one mannosyl and three glucosyl residues, and is further decorated with glucose, mannose at internal position of branches at varying ratios in different strains (Cardemil and Wolk 1981). Besides the xylose and galactose residues which have been identified at the termini of branches in all three strains mentioned above, the terminal arabinose is only found in *A. variabilis* ATCC 29413, whereas the terminal mannose is only detected in *A. cylindrical* and *Cylindrospermum* strains. Due to the close genetic relationship (<http://www.kazusa.or.jp/cyano/Anabaena/>), *Anabaena* sp. strain PCC 7120 most likely possesses an HEP structure similar to that of *A. variabilis*.

To date, the *Anabaena* sp. strain PCC 7120 (*Anabaena* for short) represents a heterocyst-forming cyanobacterium, the whole genome of which has been completed (Kaneko 2001). It has been demonstrated that a gene cluster (genes *alr2823–alr2841*) in *Anabaena* is required for the biosynthesis of HEP, and thereafter termed the HEP island (Ehira et al. 2003; Huang et al. 2005), which encodes putative dehydrogenase, epimerase, oxidoreductase, glucosyltransferase (GT) and other hypothetical proteins. Beyond this island, three putative GT genes (*alr3698*, *alr3699* and *all4160*) also participate in HEP formation (Wang et al. 2007). Notably, knockout of *alr3699* resulted in the deficiency of HEP formation during heterocyst differentiation. Furthermore, another gene *all1711* encoding a membrane protein is predicted to be an exporter of glycosides (Lopez-Igual et al. 2012). However, the molecular function and mechanism of these putative proteins remain unclear.

In a frame of systematic investigation of HEP synthesis-related proteins, we first focused on the putative proteins encoded by *alr3698* and *alr3699*, which were termed HepD and HepE, respectively. Bioinformatic analyses against the carbohydrate-active enzymes database (Campbell et al. 1997) suggested that HepE belongs to the GT4 family, in which all structure-known members are retaining enzymes and adopt a GT-B fold (Breton et al. 2012), one of the two major folds/topologies of GTs. Distinct from the GT-A fold that contains two closely abutting Rossmann-fold-like domains, the GT-B fold consists of two separate $\beta/\alpha/\beta$ Rossmann-fold-like domains with the active site located at the inter-domain crevice (Breton et al. 2006). Generally, the conserved C-terminal domain (CTD) of GT-B enzymes provides the major contacts for the donor substrate nucleotide-diphosphate sugar, whereas the N-terminal domain (NTD) recognizing diverse acceptor substrates is variable. To date, the structures of 19 members in the GT4 family have been solved, including two *N*-acetylglucosamine transferases: *Bacillus anthracis* BA1558 (Parsonage et al. 2010) and *Corynebacterium glutamicum* MshA (Vetting et al. 2008), one glucosyltransferase *Escherichia coli* WaaG (Martinez-Fleites et al. 2006) and two mannosyltransferases containing *Mycobacterium tuberculosis* PimA and *C. glutamicum* PimB' (Guerin et al. 2007; Batt et al. 2010). All these members share a relatively conserved donor-binding domain fused with a rather variable acceptor-binding domain. However, the low sequence identity among varying GTs makes it difficult to predict the *bona fide* donor and/or acceptor of HepE.

To elucidate the molecular function of *Anabaena* HepE, we determined four crystal structures: the apo-form and three complex forms

with UDP-glucose (UDPG), UDP-glucose and UDP, respectively. Structure-guided enzymatic assays identified that HepE is a glucosyltransferase using UDPG as the sugar donor. Moreover, these four structures present snapshots of conformational changes during the catalytic cycle, emboldening our understanding of the catalytic mechanism of GTs.

Results

Overall structure of HepE

The structure of apo-HepE was determined at 2.01 Å resolution by single-wavelength anomalous dispersion (SAD) phasing method using the selenomethionine (SeMet)-substituted crystals (Table I). Each asymmetric unit of apo-HepE contains two subunits, which are quite similar to each other with a root-mean-square deviation (RMSD) of 0.48 Å over 364 C α atoms. The two subunits form a buried dimeric interface of 1099 Å² per subunit. In fact, HepE also exists as a dimer in solution, as confirmed by size-exclusion chromatography and bis(sulfosuccinimidyl) suberate (BS³) cross-linking assays (Supplementary data, Figure S1). The dimer adopts a slightly bent rod shape with dimensions of 118 × 30 × 4 Å³ (Figure 1A). Each subunit displays a typical GT-B fold of GTs, consisting of two separate $\beta/\alpha/\beta$ Rossmann-fold domains (NTD and CTD) with the catalytic center located at a deep inter-domain crevice (Figure 1A). The NTD (residues Met1–Asn174 and Val364–Leu378), assumed to bind the acceptor, comprises a seven-stranded parallel β -sheet (β 3, β 2, β 1, β 4, β 5, β 6 and β 7) surrounded by eight α -helices (α 1– α 7 and α 16). The CTD (residues Gly175–Asp363), proposed to bind the nucleotide-sugar donor, has a six-stranded parallel β -sheet (β 10, β 9, β 8, β 11, β 12 and β 13) flanked by eight α -helices (α 8– α 15) on both sides. The C-terminal helix α 16 (Val364–Leu378) runs backward to pack against the NTD, which is a common feature of GT-B fold (Vrielink et al. 1994). The residues (Ile59–Gly72) connecting β 3 and α 3 in addition to the most C-terminal residues (Gly379–Arg382) could not be traced in the electron-density map, indicating their flexibility (Figure 1A).

The dimeric interface is formed between the two NTDs of subunits A and B (Figure 1A), composed of helices α 3, α 4, α 5 and α 6 and their 2-fold symmetrically related counterparts. The interface is mainly stabilized by extensive hydrophobic interactions. Notably, helix α 3 of one subunit extends to the counterpart subunit and makes hydrophobic interactions with helices α 5' and α 6'. Moreover, the side-chain nitrogen atom of Asn56 forms two hydrogen bonds with the main chain oxygen atoms of Gln76' and Leu77', respectively, further fixing the dimeric interface. Multiple-sequence alignment demonstrated that the dimeric interface is highly conserved in the heterocyst differentiation cyanobacteria (Figure 1B). Notably, it was reported that *C. glutamicum* MshA adopts a similar dimeric interface (Vetting et al. 2008).

Structural comparison

Structural homology search using DALI (http://ekhidna.biocenter.helsinki.fi/dali_server/) (Holm and Sander 1993) indicated that HepE is a member of GT-B GTs in the GT4 family, despite with a sequence identity of $\leq 25\%$. The top hits include *B. anthracis* GT BA1558 (PDB code 3MBO, Z-score 37.8, RMSD 2.0 Å over 316 C α atoms) (Parsonage et al. 2010), followed by *C. glutamicum* PimB' (PDB code 3OKA, Z-score 37.2, RMSD 2.4 Å over 324 C α atoms) (Batt et al. 2010), *C. glutamicum* MshA (PDB code 3C48, Z-score 36.9, RMSD 2.0 Å over 317 C α atoms) (Vetting et al. 2008) and *E. coli* $\alpha(1,3)$ glucosyltransferase WaaG (PDB code 2IV7, Z-score 32.1, RMSD 2.3 Å over 293 C α atoms) (Martinez-Fleites et al. 2006). Structural comparison

Table 1. Crystal parameters, data collection and structure refinement

	SeMet-HepE	HepE-UDP	HepE-UDPG	HepE-UDP&glucose
Data collection				
Space group	C222 ₁	C222 ₁	P3 ₂ 21	P3 ₂ 21
Unit cell (Å, °)	78.29, 133.47, 142.81 90.00, 90.00, 90.00	78.88, 133.22, 143.02 90.00, 90.00, 90.00	130.23, 130.23, 157.36 90.00, 90.00, 120.00	130.99, 130.99, 156.33 90.00, 90.00, 120.00
Resolution range (Å)	50.00–2.01	50.00–2.15	50.00–2.39	50.00–2.48
Unique reflections	48,403 (4,914) ^a	41,058 (4,046)	59,901 (5,956)	55,301 (5,470)
Completeness (%)	96.6 (99.0)	99.5 (99.8)	99.2 (97.7)	99.7 (97.9)
$\langle I/\sigma(I) \rangle$	17.6 (3.0)	19.2 (3.5)	13.4 (2.8)	19.4 (3.9)
R_{merge}^b (%)	8.4 (48.1)	8.5 (50.1)	9.8 (47.9)	7.5 (50.8)
Average redundancy	3.3 (3.2)	4.0 (4.0)	3.5 (3.5)	5.7 (5.7)
Structure refinement				
Resolution range (Å)	38.71–2.01	35.78–2.15	33.94–2.39	47.40–2.48
R factor ^c /R-free ^d (%)	19.5/24.6	20.6/24.4	19.4/22.0	22.3/26.7
Number of protein atoms	5560	5570	5575	5618
Number of water atoms	149	152	196	219
RMSD ^e bond lengths (Å)	0.018	0.011	0.012	0.011
RMSD bond angles (°)	1.806	1.334	1.508	1.476
Mean B factors (Å ²)	41.2	42.1	53.3	60.5
Individual B factors (Å ²)				
Protein	41.1	42.4	52.7	60.9
Ligands	51.6	43.4	60.4	65.7
Water	45.5	38.3	47.3	55.9
Ramachandran plot ^f (residues, %)				
Most favored (%)	96.7	97.6	97.1	96.3
Additional allowed (%)	3.3	2.4	2.9	3.7
PDB entry	4XSO	4XSP	4XSR	4XSU

^aThe values in parentheses refer to statistics in the highest bin.

^b $R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, where $I_i(hkl)$ is the intensity of an observation and $\langle I(hkl) \rangle$ is the mean value for its unique reflection; summations are over all reflections.

^cR factor = $\frac{\sum_b |F_o(b) - |F_c(b)||}{\sum_b |F_o(b)|}$, where $|F_o|$ and $|F_c|$ are the observed and calculated structure-factor amplitudes, respectively.

^dR-free was calculated with 5% of the data excluded from the refinement.

^eRMSD from ideal values.

^fCategories were defined by MOLPROBITY.

of HepE against BA1558, with CTDs superimposed first, revealed a quite similar overall structure with the NTDs rotate against each other as rigid bodies (Figure 1C), indicating HepE is most likely a GT-B GT. However, the NTD could not be well superimposed (Figure 1C), suggesting that it is rather variable.

From the structural point of view, the inverting and retaining members in GT-B superfamily differ from each other mainly in two loops: one in CTD that binds to the sugar donor and another in NTD that fixes the sugar acceptor (Zheng et al. 2011). Similar to the typical retaining GTs WaaG (PDB code 2IW1) (Martinez-Fleites et al. 2006) and MshA (PDB code 3C4V) (Vetting et al. 2008), HepE also possesses a relatively long loop (loop $_{\beta 1-\alpha 12}$) that blocks the pyrophosphate and glycosyl moiety of donor in the active crevice (Supplementary data, Figure S2A). In contrast, the corresponding loop in the inverting enzymes, like MurG (PDB code 1NLM) (Hu et al. 2003) and VvGT1 (PDB code 2C1Z) (Offen et al. 2006), is much shorter (Supplementary data, Figure S2B). Moreover, the putative acceptor-binding loop between $\beta 1$ and $\alpha 1$ of HepE also adopts a conformation similar to that of the retaining enzymes (Supplementary data, Figure S2A), but distinct from the inverting ones (Supplementary data, Figure S2B).

Identification of the sugar donor and acceptor

To explore the sugar donor of HepE, we tested its hydrolytic activities towards six common sugar donors, including UDPG, ADP-glucose,

GDP-glucose, UDP-galactose, UDP-N-acetylglucosamine (UDP-GlcNAc) and GDP-mannose. As shown in Figure 2A, HepE hydrolyzed UDPG to UDP at a slow but detectable velocity with K_m , k_{cat} and k_{cat}/K_m values of 0.11 ± 0.02 mM, $6.8 \pm 0.2 \times 10^{-3} \text{ min}^{-1}$ and $0.1 \text{ mM}^{-1} \text{ min}^{-1}$, respectively, whereas it displayed undetectable activities towards the other five sugar donors. These results indicated that UDPG is the preferred sugar donor of HepE, which was further validated by the complex structures of HepE.

To identify the favored acceptor of HepE, we compared the activities of HepE towards UDPG in the presence of previously identified monosaccharide residues of HEP (Cardemil and Wolk 1976), including glucose, mannose, galactose, xylose and arabinose. Notably, addition of mannose increased the activity of HepE towards UDPG by 4-folds (Figure 2B), whereas other monosaccharide residues had no detectable effects. Moreover, we also tested the activities of HepE in the presence of several oligosaccharides (Figure 2C). We found the addition of α (1,3)-linked mannanose [Man α (1,3)Man] or β (1,4)-linked mannosyl-glucose [Man β (1,4)Glc] increased the activity by ~ 5 -folds. Meanwhile, we detected the hydrolytic profiles of HepE towards UDPG in the presence of 0, 5 and 10 mM mannose using evaporative light scattering detector implemented in HPLC. The results indicated that only glucose could be detected in the reaction mix in the absence of mannose (the middle panel of Figure 3A), whereas the addition of mannose resulted in a new peak of longer retention time (the lower panel of Figure 3A), which most likely corresponds to the disaccharide Glc-Man. Notably,

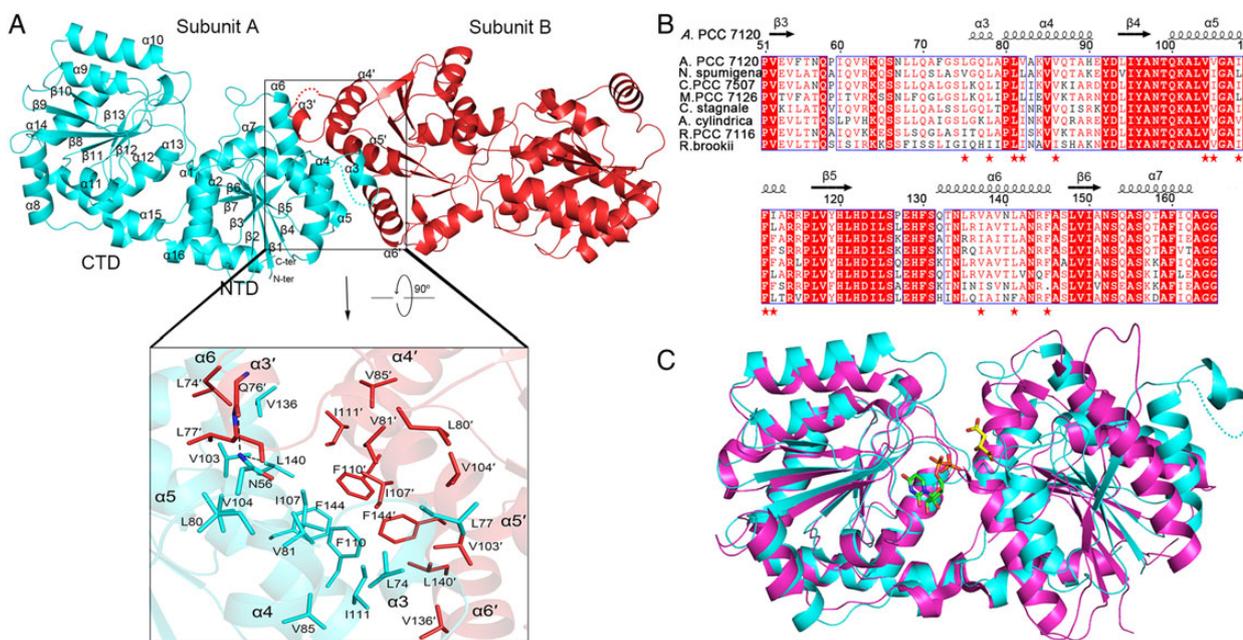


Fig. 1. Overall structure of HepE. **(A)** HepE dimer. The two subunits are colored in cyan (subunit A) and red (subunit B), respectively. The secondary structure elements are labeled in subunit A, and the secondary elements of subunit B at the dimeric interface are labeled with a prime. An enlarged view shows the dimeric interaction. **(B)** Multiple-sequence alignment of HepE and homologs. Residues involved in the dimeric interface are labeled with red stars. **(C)** Comparison of the overall structure between HepE (cyan) and *B. anthracis* GT BA1558 (magenta). The donor (UDP) and acceptor (L-malate) of BA1558 are labelled by green and yellow sticks, respectively.

the yield of Glc-Man is dependent on the concentration of mannose added to the reaction mix. To assign the putative disaccharide product of Glc-Man, we applied the mix of reaction to LC-MS/MS assays. A single MS survey scan gave the m/z values of 179.05, 565.04, 402.99 and 341.11, which correspond to mannose (Mw 180.16), UDPG (Mw 566.27), UDP (Mw 404.16) and Glc-Man (Mw 342.3), respectively (Figure 3B). It confirmed that HepE is able to catalyze the formation of Glc-Man in the presence of mannose. Taken together, HepE is a glucosyltransferase that most likely catalyzes the transfer of a glucose moiety from the donor UDPG to a putative acceptor of mannose or mannose-linked oligosaccharide (*EC* 2.4.1.-).

The donor-binding pattern

To elucidate the binding pattern of sugar donor, we determined three complex structures of HepE bound to UDPG (HepE-UDPG), UDP&glucose (HepE-UDP&glucose) and UDP (HepE-UDP) at 2.39, 2.48 and 2.15 Å, respectively. All three structures are dimers with an almost identical dimeric interface as the apo-form. In the structure of HepE-UDPG, the UDPG molecules fit well in the electron-density maps of both subunits (Figure 4A). The glucosyl moiety of UDPG that is bent-back towards UDP at an angle of $\sim 90^\circ$ adopts a standard chair conformation. This conformation makes the anomeric sugar carbon (carbon 1, C1) exposed for the nucleophilic attack, in agreement with the previous reports (Guerin et al. 2007; Vetting et al. 2008). The UDPG molecule locates at the inter-domain crevice and most UDPG-binding residues are from CTD. In detail, the uracil moiety of UDP is fixed by polar interactions with the main chain atoms of Gly235 and Arg267, in addition to hydrophobic interactions with Phe266 and Ile270. The ribose oxygen atoms O2 and O3 interact with the carboxylate group of highly conserved Glu296 via two hydrogen bonds, respectively, similar to most structure-known members of GT-B superfamily (Hu et al. 2003). The ribose oxygen O3 is further stabilized by the side chain of Arg292. The pyrophosphate moiety interacts with the

main chain amides of Gly14 and Arg292 as well as with the side chains of Arg208, Lys213 (Figure 4A). The O3' atom of glucosyl moiety forms hydrogen bonds with the main chain of Phe290 and the side chain of Glu288, a highly conserved acid residue that has been found to interact with the donor sugar in other GT4 members (Martinez-Fleites et al. 2006; Guerin et al. 2007). The O4' atom forms a hydrogen bond with the main chain of Gly291, whereas the O6' atom engages in strong interactions with the side chains of His121, Asn151 and Asn174 from the NTD. Notably, either mutation of the conserved residue His121 or Glu288 to alanine completely abolished the enzymatic activity of HepE towards UDPG (Figure 4B), indicating their crucial roles in the catalytic process, in agreement with the previous reports (Buschiazzo et al. 2004).

During crystallization in the presence of UDPG and mannose, the donor UDPG is hydrolyzed into the products UDP and glucose, but the mannose molecule is absent from the complex structure of HepE-UDP&glucose. As shown in Figure 4C, the O1' atom of glucose and the closest oxygen atom of β -phosphate has a distance of 3.5 Å. The glucose molecule employs the same conformation and binding residues as the glucosyl moiety in HepE-UDPG, except that the O1' atom is stabilized by the main chain of His121.

In the structure of HepE-UDP, the electron density of UDP in subunit B is untraceable, thus we take subunit A for describing the binding pattern of UDP. The UDP molecule fit well in the electron-density map in subunit A (Figure 4D). The UDP moiety is stabilized in a similar manner as that of HepE-UDPG structure. Most UDP-binding residues, except for the N-terminal residue Gly14, are located at the CTD.

Conformational changes upon substrate binding

Structural comparisons of three complexes of HepE against the apo-form yielded RMSD values of 1.47 Å over 355 C α atoms, 1.35 Å over 362 C α atoms and 0.53 Å over 363 C α atoms, respectively. Despite the significant conformational changes upon the donor binding, the

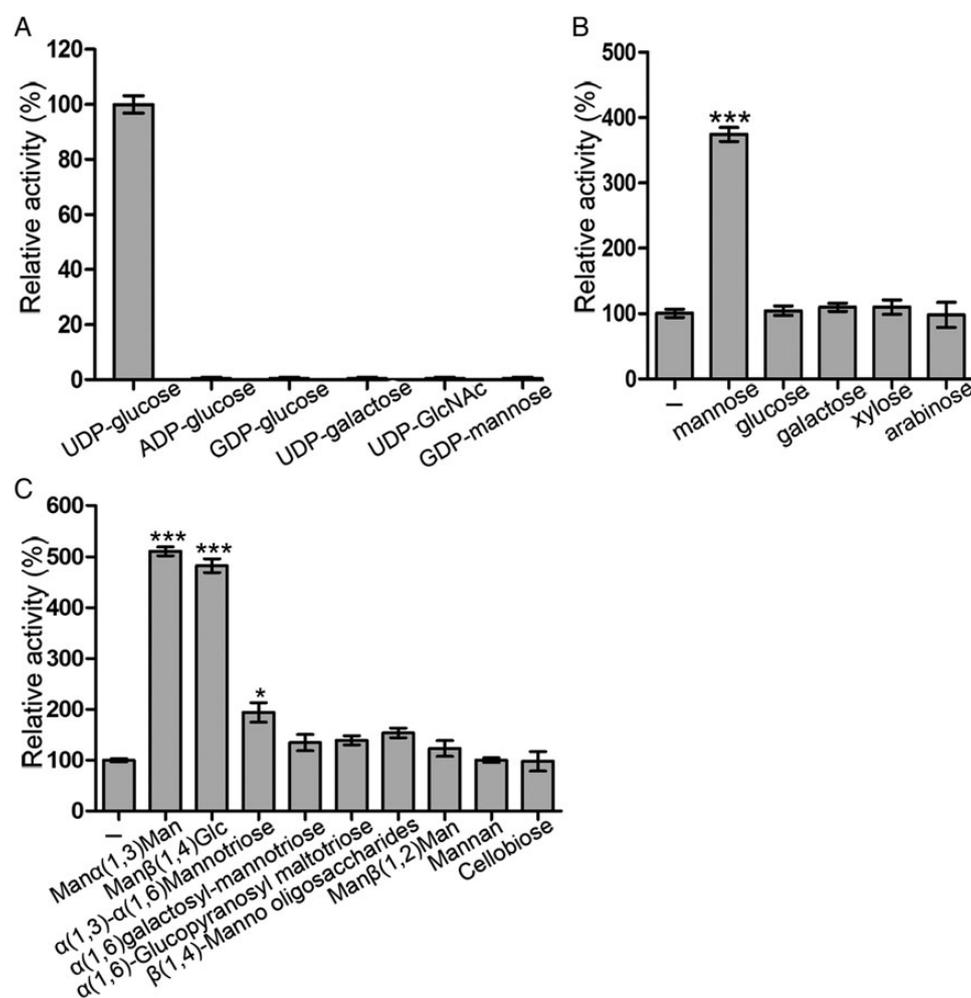


Fig. 2. Enzymatic activity assays. (A) The relative hydrolytic activities of HepE towards six nucleotide-sugar donors (UDP-glucose, ADP-glucose, GDP-glucose, UDP-galactose, UDP-GlcNAc and GDP-mannose). The relative activities of HepE towards UDPG in the presence of (B) some monosaccharide residues and (C) oligosaccharides. The activities of all samples are shown as a percentage to that of the wild-type HepE towards UDPG in the absence of sugar acceptor. Data are presented as the means \pm standard deviations from three independent assays. Statistical significance analyses were performed using a one-way ANOVA. *P*-value of <0.05 and 0.001 are indicated with * and ***, respectively.

two individual domains are almost rigid, with an RMSD value of 0.32 \AA over 159 $C\alpha$ atoms for the NTD, and 0.33 \AA over 204 $C\alpha$ atoms for the CTD, respectively. However, we could observe significant induced fit of the inter-domain crevice upon the substrate binding. Given the CTDs superimposed, calculation of the domain motion by DYNDOM server (Hayward and Lee 2002) revealed a 11.3° domain rotation of the NTD relative to the CTD upon UDPG binding (Figure 5A), compared with 9.9° and 2.7° inter-domain motions in HepE-UDP&glucose (Figure 5B) and HepE-UDP (Figure 5C), respectively. The large inter-domain motion upon UDPG binding is mainly due to the interactions of UDPG with both NTD and CTD. In the structure of HepE-UDPG, two significant conformational changes were found as a result of domain rotation. First, the domain motion brings loop $_{\beta 5-\alpha 6}$ and loop $_{\beta 11-\alpha 12}$ to a close proximity, to form the acceptor-binding site at the NTD. The side chain of His128 forms a hydrogen bond with the main chain of Pro287 (Figure 5A), further fixing the conformation of loop $_{\beta 5-\alpha 6}$. Second, loop $_{\beta 1-\alpha 1}$ and loop $_{\beta 9-\alpha 10}$ move towards each other and form a hydrogen bond between the side chain of Lys11 and the main chain of Leu238, making Phe239 flipped towards the substrate-binding pocket (Figure 5A). Thus, loop $_{\beta 9-\alpha 10}$

functions as a lid to cover the sugar donor pocket. These movements lead to a closed conformation, which provide the crosstalk between the donor UDPG at the CTD and the acceptor at the NTD. Similar inter-domain motions have also been detected in other structure-known GT-B members upon binding to sugar donors, from 10° to 25° for MurG (Hu et al. 2003), TDP-epi-vancosaminyltransferase (Mulichak et al. 2003) and glycogen synthase (Buschiazzo et al. 2004) or as large as 97° for MshA (Vetting et al. 2008). Interestingly, compared with HepE-UDPG, loop $_{\beta 9-\alpha 10}$ in HepE-UDP&glucose and its bearing residue Phe239 are flipped outwards, uncovering the active-site pocket for the release of the products (Figure 5B). Notably, the UDP-bound form was obtained by soaking apo-HepE crystals with UDP, in contrast to the other two complex structures that were obtained by cocrystallization. Thus, the relatively slight conformational changes upon binding to UDP might be partly due to crystal packing.

Simulation of the acceptor-binding pattern

We attempted to solve the complex structure of HepE with mannose/mannobiose by either co-crystallization or soaking, but failed.

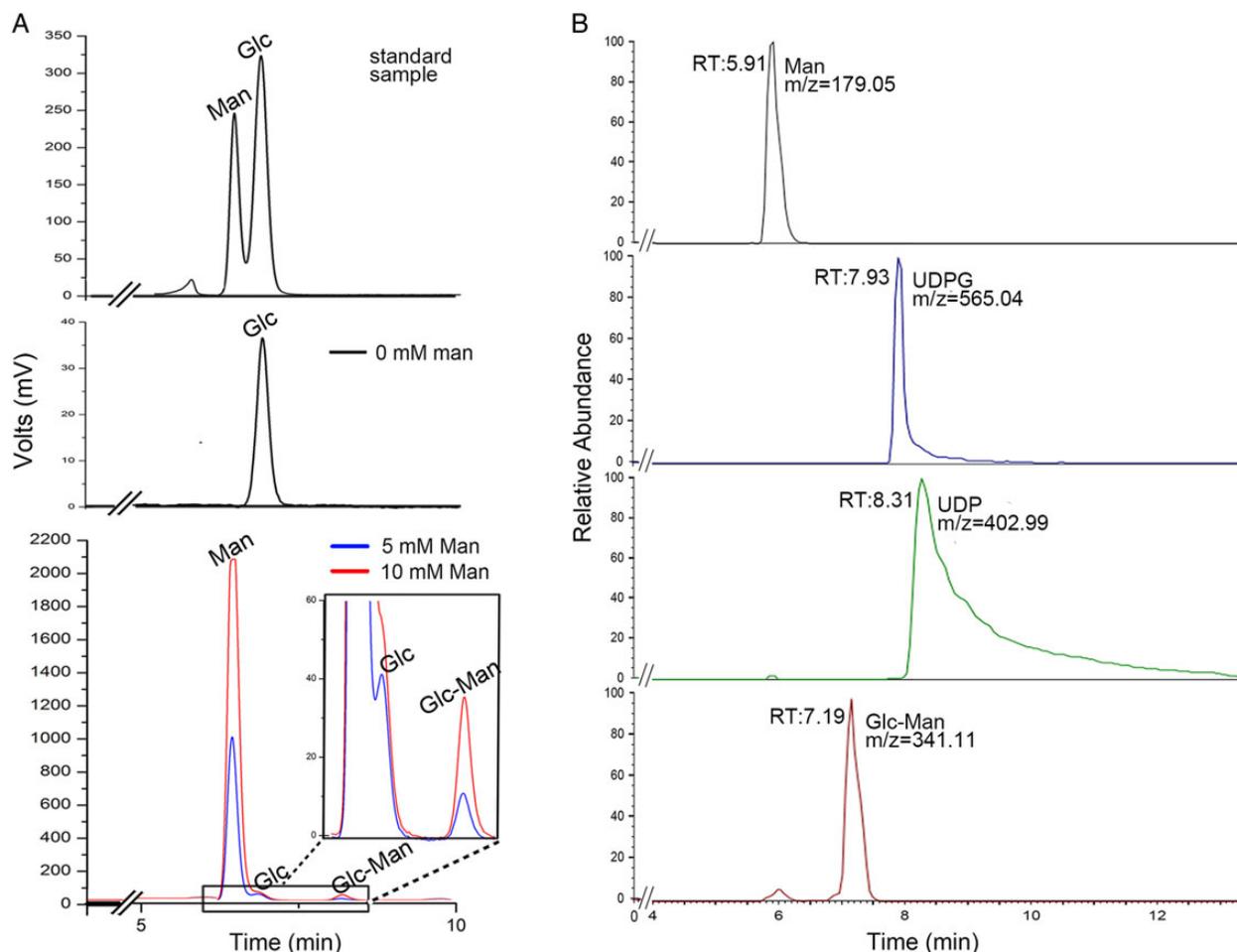


Fig. 3. The production of a putative disaccharide. (A) The hydrolytic profiles of HepE towards UDPG in the absence of mannose (the middle panel) and in the presence of 5 and 10 mM mannose (the lower panel), respectively. The upper panel is the mixed standard samples of mannose and glucose. An enlarged view demonstrated a new peak of longer retention time was produced in the presence of mannose. (B) The chromatograph mass spectrum of the components catalyzed by HepE in the reaction mix.

Alternatively, guided by the previous structures of MshA bound with the acceptor inositol-1-phosphate (PDB code 3C4V) and BA1558 bound with L-malate (PDB code 3MBO), we docked a model of mannose into the acceptor-binding pocket of HepE-UDPG, a mimicking state to bind the approaching acceptor. In the model, the 6-hydroxyl of mannose is ~ 2.8 Å to the exposed phosphoester bond between UDP and glucosyl moiety. This geometric configuration of acceptor mannose is compatible with the requirement of the substrate-assisted S_Ni (internal return) mechanism proposed for the retaining GT-B GTs (Gibson et al. 2002). The hydroxyl groups of mannose are further stabilized by polar interactions with the side chains of Glu16, Asp122 and Arg208, as well as hydrophobic interaction with Phe239 (Figure 6A). Notably, Glu16 in HepE corresponds to Gln25 in MshA, which has been previously demonstrated in acceptor substrate binding (Vetting et al. 2008), suggesting that Glu16 of HepE might also contribute to acceptor recognition. Meanwhile, Arg208 provides a bidentate anchoring interaction between UDPG and mannose, similar to the previous speculations for PimB' (Batt et al. 2010).

To validate our docking model, we performed a series of site-directed mutageneses in combination with activity assays. Comparative activity assays in the presence of mannose indicated that mutation of Glu16, Asp122 and Phe239 to alanine resulted in an apparently

reduced activity of $\sim 20\%$ to that of the wild-type HepE (Figure 6B). In contrast, mutants E16A, D122A and F239A had a comparable activity with that of the wild type if mannose was absent from the reaction mix. These results suggested that Glu16, Asp122 and Phe239 are involved in acceptor recognition. Arg208 contributes to recognizing both donor and acceptor (Figure 6A), as well as stabilizing the UDP leaving group, as observed in the HepE-UDP complex and previously reported structures ((Martinez-Fleites et al. 2006; Vetting et al. 2008). Indeed, mutation of R208A completely abolished the activity of HepE (Figure 6B).

Discussion

HepE is required for the formation of the HEP during *Anabaena* heterocyst differentiation (Wang et al. 2007). However, the molecular function of HepE in the HEP biosynthesis pathway remains unclear. Here we report that HepE is a GT-B fold glucosyltransferase in the GT4 family using UDPG as the sugar donor. Structural comparison against retaining or inverting GTs indicates HepE is most likely a retaining enzyme. Currently, there are two catalytic mechanisms of retaining GTs. The double displacement mechanism which proposes the sugar moiety

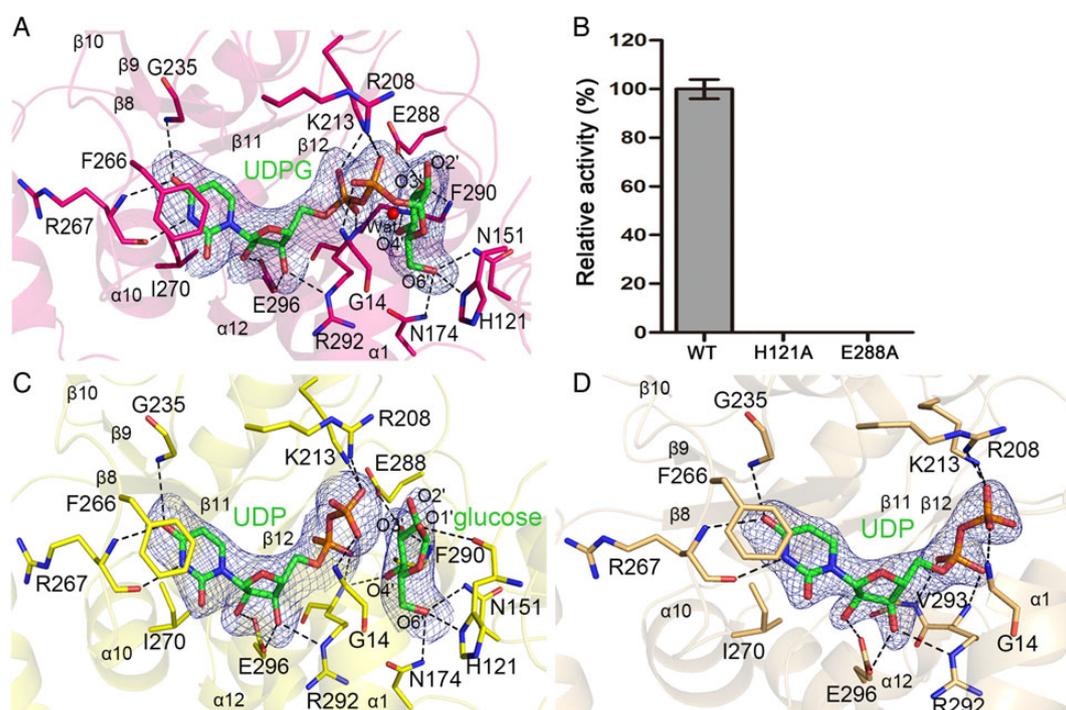


Fig. 4. A stereoview of the donor-binding site. (A) The binding site of UDPG. The complex structure HepE-UDPG is colored in pink. Residues involved in binding donor are displayed as sticks. The polar interactions are indicated by dash lines. A water molecule is shown as a red dot. The $3\sigma F_o - F_c$ electron-density map of UDPG was calculated before modeling the ligand. (B) The relative activities of wild-type HepE and mutants towards UDPG in the presence of mannose. Data are presented as the means \pm standard deviations from three independent assays. Binding sites of (C) UDP&glucose and (D) UDP. The two complex structures HepE-UDP&glucose and HepE-UDP are colored in yellow and wheat, respectively. The $3\sigma F_o - F_c$ electron-density maps of the molecules were calculated before modeling the ligands.

is first transferred to the enzyme on the β -face forming an enzyme-glycosyl intermediate and subsequently transferred to the acceptor on the α -face (Lairson et al. 2008). In contrast, the substrate-assisted S_Ni -like mechanism proposes that the acceptor front-face attacks anomeric carbon and invokes the formation of an oxocarbenium ion-like transition state, followed by the phospho-sugar bond breakage and glycosidic bond formation in a concerted but asynchronous manner (Gibson et al. 2002). In HepE-UDP&glucose, the anomeric carbon of glucose is stabilized on the α -face by the main chain of H121. Moreover, the addition of mannose, mannoside or Man β (1,4)Glc is able to increase the activity of HepE towards UDPG, providing further evidence for a substrate-assisted S_Ni -like reaction mechanism of HepE.

Despite extensive screening of various sugars, we failed assigning the proper acceptor of HepE. Previous reports have suggested that the activity of a retaining GT could be dramatically increased by over 100-folds in the presence of a partner protein (Lu et al. 2005; Shi et al. 2014) or an optimal acceptor (Sindhuwinata et al. 2010). Thus, we propose that the low activity of HepE towards UDPG might be due to the absence of a favored acceptor or an activating partner. As the genes encoding the GT and its activating partner are usually in a same operon, a neighboring putative glycotransferase HepD/Alr3698, which was co-transcribed with HepE (Wang et al. 2007), might be the activating partner of HepE. However, further experiments are needed to verify this hypothesis.

Based on structures of four catalytic states and a docking model of the acceptor, we propose a putative catalytic cycle of HepE (Figure 7). HepE works with the sugar donor binding first and induces a significant inter-domain movement, followed by the acceptor binding, which is similar to other retaining GTs (Boix et al. 2002; Lee et al. 2011).

In the apo-form state, the two domains are somewhat separated with an open inter-domain crevice (a). Upon binding to the donor UDPG, a significant induced fit makes the NTD and CTD move towards each other. In addition, the loop $_{\beta 9-\alpha 10}$ flips towards the active site to form an intact acceptor-binding pocket (b). This induced fit facilitates the binding of the sugar acceptor (c) and promotes the hydrolysis of UDPG (d), accompanying with the formation of glycosidic bond on the same face of the sugar. The reaction would trigger the loop $_{\beta 9-\alpha 10}$ kicked outward, and make the two domains separated again to release the elongated sugar product (e). Finally, upon the release of the second product UDP (Persson et al. 2001), the enzyme is turned-over to the apo-form state and ready for the next catalytic cycle (a). In conclusion, we present here the structural analyses and enzymatic characterization of HepE, the first enzyme involved in *Anabaena* HEP biosynthesis.

Materials and methods

Expression and purification of recombinant HepE and mutants

The full-length coding region of *alr3699* gene was amplified from the genomic DNA extracted from *Anabaena*. The coding region was cloned to a pET28a-derived vector, with an N-terminal 6 \times His tag. The recombinant plasmid was transformed into *E. coli* strain Rosetta (DE3) strain (Novagen), growing at 37°C in LB culture medium (10 g NaCl, 10 g Bacto-Tryptone and 5 g yeast extract per liter), supplemented with 30 μ g/mL kanamycin and 34 μ g/mL chloramphenicol. When the OD $_{600\text{ nm}}$ reached \sim 0.8, the expression of HepE was induced for another

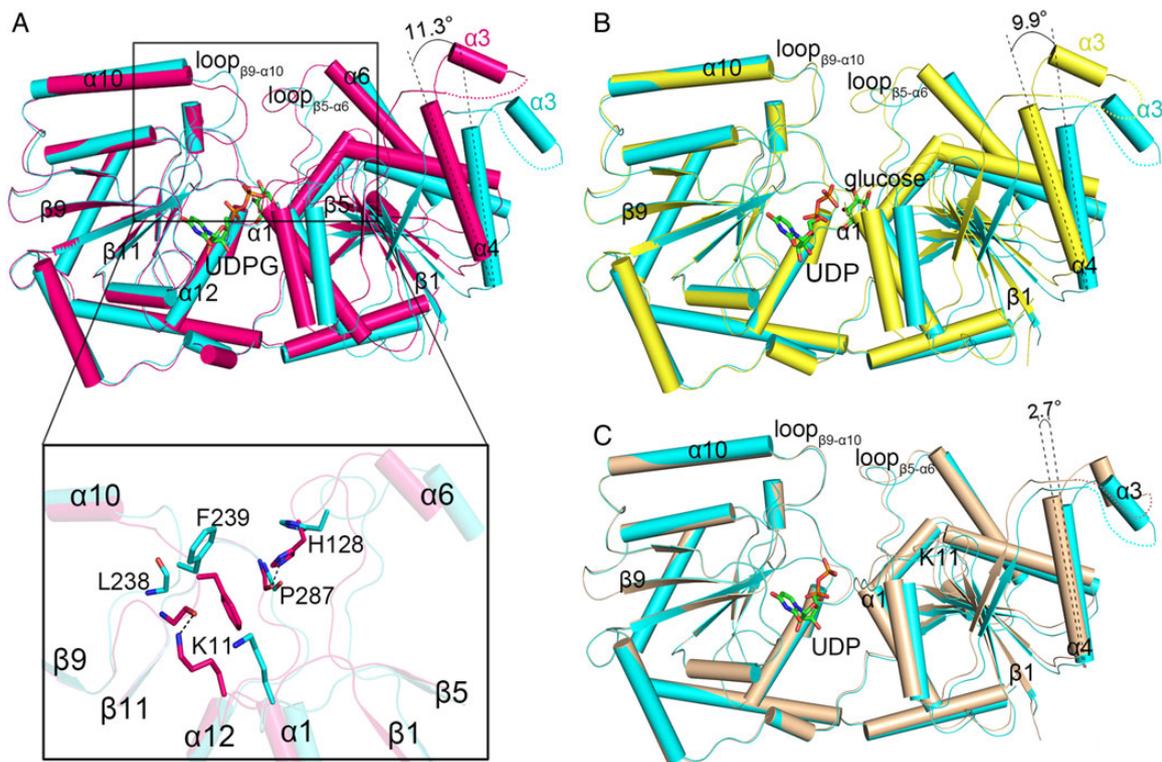


Fig. 5. Conformational changes upon substrate binding. Structural comparisons of apo-HepE (cyan) with (A) HepE-UDPG (pink), (B) HepE-UDP&glucose (yellow) and (C) HepE-UDP (wheat). The movement of NTD relative to CTD is shown by the rotated angle of helix α . An enlarged view shows the conformational changes of loop_{P1- α 1}, loop_{P5- α 6} and loop_{P9- α 10} of the apo-form and HepE-UDPG given the CTDs superimposed. Residues in loops are shown as sticks with hydrogen bonds in dash lines.

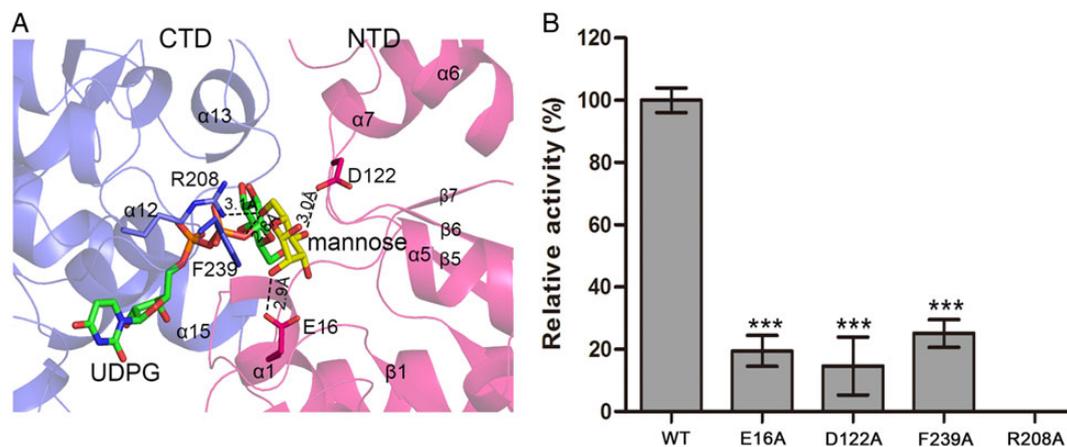


Fig. 6. Simulation and validation of the acceptor-binding pattern. (A) A docked model of mannose in the inter-domain crevice. Mannose is shown as yellow sticks, and is docked to the inter-domain crevice of HepE-UDPG complex structure with NTD and CTD colored in pink and blue, respectively. Residues Glu16, Asp122, Arg208 and Phe239 fixing the docked mannose are shown as sticks with hydrogen bonds in black dashes. (B) The relative activities of wild-type HepE and mutants towards UDPG in the presence of mannose. Data are presented as the means \pm standard deviations from three independent assays. Statistical significance analyses were performed using a one-way ANOVA. *** indicates a P -value of <0.001 .

4 h at 37°C by adding 0.2 mM isopropyl- β -D-thiogalactopyranoside. Cells were harvested by centrifugation at 6000 \times g for 10 min, then resuspended in 40 mL lysis buffer (20 mM Tris-Cl, pH 8.0, 100 mM NaCl, 5%, v/v, glycerol). The suspension was sonicated on ice for a total time of 30 min and the lysate was centrifuged (12,000 \times g,

30 min, 4°C). The supernatant was loaded onto a nickel-nitrilotriacetic acid column (GE healthcare) pre-equilibrated with the binding buffer (20 mM Tris-Cl, pH 8.0, 100 mM NaCl). The column was subsequently washed with gradient imidazole. The target protein was eluted with 500 mM imidazole, and further loaded onto a HiLoad 16/60

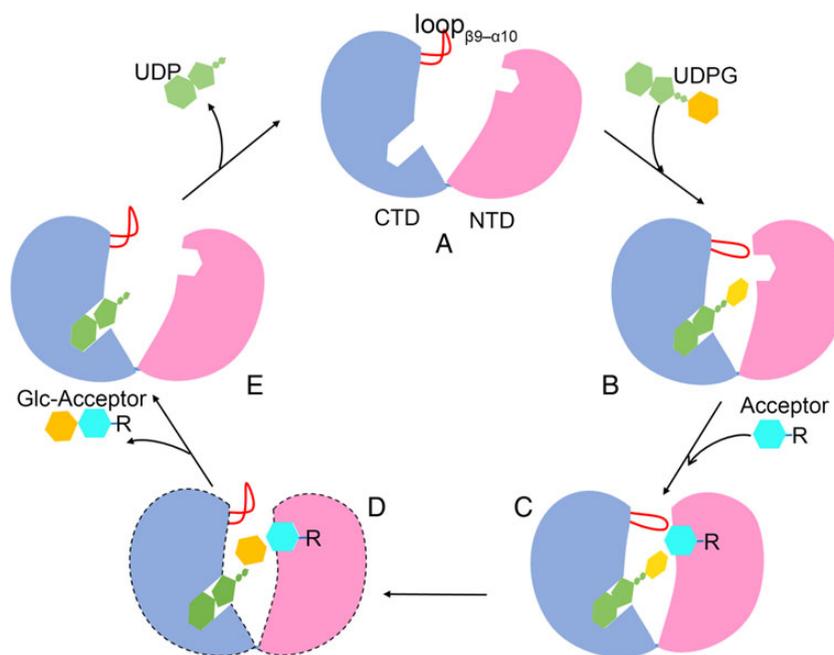


Fig. 7. A schematic diagram of the putative catalytic cycle. (A–E) The five consecutive catalytic states. The NTD and CTD are colored in pink and blue, respectively.

Superdex™ 200 column (GE Healthcare) pre-equilibrated with 20 mM Tris–Cl, pH 8.0, 100 mM NaCl, 5% (v/v) glycerol. Fractions containing the target protein were pooled and concentrated to 10 mg/mL for crystallization. Protein samples for enzymatic activity assays were stored at -80°C . Protein purity was evaluated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS–PAGE).

For the expression of SeMet-labeled HepE, the bacteria were grown to an $\text{OD}_{600\text{ nm}}$ of ~ 0.6 , then harvested and washed twice with the M9 medium. The cells were incubated in SeMet medium (M9 medium with 50 mg/L L-SeMet and the other essential amino acids at 50 mg/L) for 0.5 h and induced with 0.2 mM isopropyl- β -D-thiogalactopyranoside for another 4 h at 37°C . Protein expression and purification were carried out as the procedure for the native protein.

Site-directed mutagenesis was performed using the QuickChange site-directed mutagenesis kit (Stratagene, La Jolla, CA) with the plasmid encoding the wild-type HepE as a template. The mutant proteins were expressed, purified and stored in the same manner as the wild-type protein.

Crystallization, data collection and processing

Crystallization of HepE was performed using a Mosquito robot (TTP Labtech) in 96-well plates (Greiner) at 16°C . SeMet-substituted protein crystals were optimized from drops containing 1 μL 10 mg/mL protein with 1 μL reservoir solution (1.0 M ammonium citrate tribasic, 0.1 M Bis–Tris propane, pH 7.0). The crystals of HepE–UDP were obtained by soaking the apo–HepE crystals with 20 mM UDP. HepE–UDPG complex crystals were obtained by cocrystallization with 10 mM UDPG in 1.5 M Li_2SO_4 , 0.1 M HEPES, pH 7.5, whereas HepE–UDP&glucose complex crystals were grown in the same condition by cocrystallization with 10 mM UDPG and 20 mM mannose simultaneously. All Crystals were transferred to the cryoprotectant (reservoir solution supplemented with 30%, v/v, glycerol) and flash-frozen with liquid nitrogen. The diffraction data were collected at 100 K in a liquid nitrogen stream using beamline 17 U with a

Q315r CCD (ADSC, MARresearch, Germany) at the Shanghai Synchrotron Radiation Facility (SSRF). All diffraction data were indexed, integrated and scaled with HKL2000 (Otwinowski and Minor 1997).

Structure determination and refinement

The crystal structure of HepE was determined using the SAD phasing method (Brodersen et al. 2000) from a single SeMet-substituted protein crystal to a maximum resolution of 2.01 Å. The SHELXD program (Sheldrick 2008) implemented in IPCAS was used to locate the heavy atoms and the phase was calculated by OASIS (Hao et al. 2000) and further improved with the programs RESOLVE and Buccaneer (Terwilliger and Berendzen 1999; Cowtan 2006). Electron-density maps showed clear features of secondary structural elements. Automatic model building was carried out using Autobuild in PHENIX (Adams et al. 2010). Afterwards, the initial model was subjected to the molecular replacement against the native data of three complex forms using MOLREP (Read 2001). All structures were refined using the maximum likelihood method implemented in REFMAC5 (Murshudov et al. 1997) as part of CCP4i (Bailey 1994) program suite and Phenix.refine in PHENIX program (Headd et al. 2012), then were rebuilt interactively using the program COOT (Emsley and Cowtan 2004). All the final models were evaluated with the programs MOLPROBITY (Davis et al. 2007) and PROCHECK (Laskowski et al. 1993). The data collection and structure refinement parameters were listed in Table I. All structure figures were prepared with PyMOL (<http://www.pymol.org/>).

Activity assays

The activities of HepE were assayed by high-performance liquid chromatography (HPLC). Standard conditions were 20 mM Tris–Cl, pH 8.0, 100 mM NaCl, 1 mM MgCl_2 with 1 mM UDPG (Sigma) and 20 mM mannose (Sigma). All assays were performed at 37°C for

1 h in a 50 μ L reaction system. Except for the enzyme, all components were mixed in the cuvette and allowed to equilibrate for 2 min. Reactions were initiated by the addition of enzyme, terminated by heating at 95°C for 10 min, the reactions without enzyme served as a control. Then the mixtures were centrifuged at 10,000 \times g for 10 min. Ten microliters supernatant were applied to the HPLC system (Agilent 1200 Series). The column (Eclipse XDB-C18 column, 4.6 \times 250 mm, Agilent) was equilibrated with 100 mM Na₂HPO₄/NaH₂PO₄, pH 6.5, 10 mM tetrabutylammonium bromide (TBAB, Sangon). The components were separated at a velocity of 1 mL/min. The yield of UDP was monitored at 254 nm and was assigned based on the retention time of the standards. The yields of sugar molecules are analyzed by carbohydrate column using evaporative light scattering detector implemented in HPLC. All the measurements were done in triplicate.

Chemical cross-linking

HepE was purified as before and desalted to 1 \times PBS, pH 8.0, buffer. Chemical cross-linking of purified HepE was performed using BS³, which is a homobifunctional sulfo-*N*-hydroxysuccinimide ester analog with a spacer arm length of 11.4 Å (Pierce). Briefly, BS³ was diluted from 25 mM dimethyl sulfoxide dissolved stocks to 0.2 mM with PBS. Fifty microliters recombinant protein (0.5 mg/mL) were incubated with 0.2 mM BS³ at 25°C for 30 min. The reaction was quenched by the addition of 20 mM Tris-Cl, pH 8.0. The sample without BS³ served as a control. Then the samples were loaded to a 12% SDS-PAGE. After electrophoresis, the gel was stained with Coomassie Brilliant Blue G-250. Molecular mass markers for SDS-PAGE were purchased from Thermo Scientific (Wilmington, DE): β -galactosidase, bovine serum albumin, ovalbumin, lactate dehydrogenase, REase Bsp98I, β -lactoglobulin, lysozyme, which have a molecular weight of 116, 66.2, 45.0, 35.0, 25.0, 18.4 and 14.4 kDa, respectively.

Identification of the disaccharide product by LC-MS/MS

Targeted SIM (single ion monitoring) and target MS2 analysis was performed with a Q Exactive Orbitrap (Thermo, CA) instrument using an ESI source with negative ion mode. The reaction of HepE was terminated (95°C for 10 min) and centrifugation (10,000 \times g for 10 min), then the supernatant was loaded to the methanol at a final concentration of 80%, and vacuum dried using Eppendorf Concentrator 5301. The sample resolved in 50 μ L methanol was diluted 100 times. One microliter supernatant was loaded to normal phase chromatography column, then the sample was eluted to orbitrap mass spectrometer with 90% IPA (isopropanol) and 10% ACN (acetonitrile) as eluent from 1 to 99% within 10 min. The stationary phase was 60% ACN with 10 mM ammonium acetate. The spray voltage was set to 2800 V. Data with mass range *m/z* 150–2000 were acquired at negative ion mode using data-dependent MSMS acquisition. The full scan and fragment spectra were collected with a resolution of 70,000 and 17,500, respectively. The source parameters include capillary temperature, 320°C; heater temperature, 300°C; sheath gas flow rate, 35 arb; auxiliary gas flow rate, 10 arb. Mass spectra were analyzed using xcalibur software.

Docking calculations

The program SYBYL-X 2.0 (Jain 2007) was used for mannose docking calculation. Ligands were docked automatically into an acceptor-binding site using a protomol-based method and an empirically derived scoring function. Protomol is a computational representation of the intended binding site to which putative ligands are aligned.

Surflex-Dock automatic protomol was prepared using UDPG-complexed structure of HepE, with a threshold value of 0.5 and a Bloat of 0 Å. All hydrogen atoms were added to define the correct configuration and tautomeric states. The Powell energy minimization algorithm was used for the structure energy minimization. For Surflex-Dock protocol, the search grid was expanded in 6 Å. Twenty conformations were used for each fragment and the maximum number of rotatable bonds per molecule was 100. Results were analyzed using SYBYL program with a minimum RMSD of 0.05 Å and figures were prepared with the program PyMoL.

Supplementary data

Supplementary data for this article are available online at <http://glycob.oxfordjournals.org/>.

Funding

This work was supported by the National Natural Science Foundation of China (31370757), the Specialized Research Fund for the Doctoral Program of Higher Education (20133402110023), the Ministry of Science and Technology of China (2012CB911000 and 2014CB910100) and the Hefei Science Center of CAS (2015SRG-HSC045 and 2015SRG-HSC046).

Acknowledgements

We thank the staff at the SSRF for technical assistance. We also thank the Innovation Center for Cell Signaling Network, the Core Facility Center for Life Sciences of University of Science and Technology of China and the Center of Biomedical Analysis of Tsinghua University.

Conflict of interest statement

The authors declare no competing financial interests.

Abbreviations

BS³, bis(sulfosuccinimidyl) suberate; HEP, heterocyst envelope polysaccharide; HPLC, high-performance liquid chromatography; RMSD, root-mean-square deviation; SAD, single-wavelength anomalous dispersion; SeMet, selenomethionine; UDPG, UDP-glucose.

References

- Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, et al. 2010. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D*. 66:213–221.
- Bailey S. 1994. The Ccp4 Suite – Programs for protein crystallography. *Acta Crystallogr D*. 50:760–763.
- Batt SM, Jabeen T, Mishra AK, Veerapen N, Krumbach K, Eggeling L, Besra GS, Futterer K. 2010. Acceptor substrate discrimination in phosphatidylinositol mannoside synthesis: Structural and mutational analysis of mannose transferase *Corynebacterium glutamicum* PimB'. *J Biol Chem*. 285:37741–37752.
- Bishop NI. 1972. Photosynthesis: The electron transport system of green plants. *Annu Rev Biochem*. 40:197–226.
- Boix E, Zhang Y, Swaminathan GJ, Brew K, Acharya KR. 2002. Structural basis of ordered binding of donor and acceptor substrates to the retaining glycosyltransferase α -1,3 galactosyltransferase. *J Biol Chem*. 277:28310–8.

- Breton C, Fournel-Gigleux S, Palcic MM. 2012. Recent structures, evolution and mechanisms of glycosyltransferases. *Curr Opin Struct Biol.* 22:540–549.
- Breton C, Snajdrova L, Jeanneau C, Koca J, Imberty A. 2006. Structures and mechanisms of glycosyltransferases. *Glycobiology.* 16:29–37.
- Brodersen DE, de la Fortelle E, Vornhein C, Bricogne G, Nyborg J, Kjeldgaard M. 2000. Applications of single-wavelength anomalous dispersion at high and atomic resolution. *Acta Crystallogr D.* 56:431–441.
- Buschiazzo A, Ugalde JE, Guerin ME, Shepard W, Ugalde RA, Alzari PM. 2004. Crystal structure of glycogen synthase: homologous enzymes catalyze glycogen synthase and degradation. *EMBO J.* 23:3196–3205.
- Campbell JA, Davies GJ, Bulone V, Henrissat B. 1997. A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *Biochem J.* 326(Pt 3):929–939.
- Cardemil L, Wolk CP. 1976. The polysaccharides from heterocyst and spore envelopes of a blue-green alga. *J Biol Chem.* 251:2967–2975.
- Cardemil L, Wolk CP. 1979. The polysaccharides from heterocyst and spore envelopes of a blue-green alga. *J Biol Chem.* 254:736–741.
- Cardemil L, Wolk CP. 1981. Polysaccharides from the envelopes of heterocysts and spores of the blue-green algae *Anabaena variabilis* and *Cylindrospermum licheniforme*. *J Phycol.* 17:234–240.
- Cowtan K. 2006. The buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D.* 62:1002–1011.
- Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, Snoeyink J, Richardson JS, et al. 2007. MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* 35:W375–W383.
- Ehira S, Ohmori M, Sato N. 2003. Genome wide expression analysis of the responses to nitrogen deprivation in the heterocyst forming cyanobacterium *Anabaena* sp. strain PCC 7120. *DNA Res.* 10:97–113.
- Emsley P, Cowtan K. 2004. Coot: Model-building tools for molecular graphics. *Acta Crystallogr D.* 60:2126–2132.
- Fay P. 1992. Oxygen relations of nitrogen fixation in cyanobacteria. *Micro Rev.* 56(2):340–373.
- Gallon JR. 1992. Reconciling the incompatible N₂ fixation and O₂. *New Phytologist.* 122:571–609.
- Gibson RP, Turkenburg JP, Charnock SJ, Lloyd R, Davies GJ. 2002. Insights into trehalose synthesis provided by the structure of the retaining glucosyltransferase OtsA. *Chem Biol.* 9:1337–1346.
- Guerin ME, Kordulakova J, Schaeffer F, Svetlikova Z, Buschiazzo A, Giganti D, Gicquel B, Mikusova K, Jackson M, Alzari PM. 2007. Molecular recognition and interfacial catalysis by the essential phosphatidylinositol mannosyltransferase PimA from *Mycobacteria*. *J Biol Chem.* 282:20705–20714.
- Hao Q, Gu YX, Zheng CD, Fan HF. 2000. OASIS: A computer program for breaking phase ambiguity in one-wavelength anomalous scattering or single isomorphous substitution (replacement) data. *J Appl Crystallogr.* 33:980–981.
- Haselkorn R. 1978. Heterocysts. *Ann Rev Plant Physiol.* 29:319–344.
- Hayward S, Lee RA. 2002. Improvements in the analysis of domain motions in proteins from conformational change: DynDom version 1.50. *J Mol Graph Model.* 21:181–183.
- Headd JJ, Echols N, Afonine PV, Grosse-Kunstleve RW, Chen VB, Moriarty NW, Richardson DC, Richardson JS, Adams PD. 2012. Use of knowledge-based restraints in phenix.refine to improve macromolecular refinement at low resolution. *Acta Crystallogr D.* 68:381–390.
- Holm L, Sander C. 1993. Protein structure comparison by alignment of distance matrices. *J Mol Biol.* 233:123–138.
- Hu YN, Chen L, Ha S, Gross B, Falcone B, Walker D, Mokhtarzadeh M, Walker S. 2003. Crystal structure of the MurG: UDP-GlcNAc complex reveals common structural principles of a superfamily of glycosyltransferases. *Proc Natl Acad Sci USA.* 100:845–849.
- Huang G, Fan Q, Lechno-Yossef S, Wojciuch E, Wolk CP, Kaneko T, Tabata S. 2005. Clustered genes required for the synthesis of heterocyst envelope polysaccharide in *Anabaena* sp. strain PCC 7120. *J Bacteriol.* 187:1114–1123.
- Jain AN. 2007. Surfex-Dock 2.1: Robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. *J Comput Aid Mol Des.* 21:281–306.
- Kaneko T. 2001. Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120. *DNA Res.* 8:205–213.
- Lairson LL, Henrissat B, Davies GJ, Withers SG. 2008. Glycosyltransferases: Structures, functions, and mechanisms. *Annu Rev Biochem.* 77:521–555.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. Procheck – A program to check the stereochemical quality of protein structures. *J Appl Crystallogr.* 26:283–291.
- Lee SS, Hong SY, Errey JC, Izumi A, Davies GJ, Davis BG. 2011. Mechanistic evidence for a front-side, S_N2-type reaction in a retaining glycosyltransferase. *Nat Chem Biol.* 7:631–638.
- Lopez-Igual R, Lechno-Yossef S, Fan Q, Herrero A, Flores E, Wolk CP. 2012. A major facilitator superfamily protein, HepP, is involved in formation of the heterocyst envelope polysaccharide in the cyanobacterium *Anabaena* sp. strain PCC 7120. *J Bacteriol.* 194:4677–4687.
- Lu W, Leimkuhler C, Gatto GJ Jr, Kruger RG, Oberthur M, Kahne D, Walsh CT. 2005. AknT is an activating protein for the glycosyltransferase AknS in L-aminooxyoxysugar transfer to the aglycone of aclacinomycin A. *Chem Biol.* 12:527–534.
- Martinez-Fleites C, Proctor M, Roberts S, Bolam DN, Gilbert HJ, Davies GJ. 2006. Insights into the synthesis of lipopolysaccharide and antibiotics through the structures of two retaining glycosyltransferases from family GT4. *Chem Biol.* 13:1143–1152.
- Mulchak AM, Losey HC, Lu W, Wawrzak Z, Walsh CT, Garavito RM. 2003. Structure of the TDP-epi-vancosaminyltransferase GtfA from the chloroeremomycin biosynthetic pathway. *Proc Natl Acad Sci USA.* 100:9238–9243.
- Murshudov GN, Vagin AA, Dodson EJ. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D.* 53:240–255.
- Nicolaisen K, Hahn A, Schleiff E. 2009. The cell wall in heterocyst formation by *Anabaena* sp. PCC 7120. *J Basic Microbiol.* 49:5–24.
- Offen W, Martinez-Fleites C, Yang M, Kiat-Lim E, Davis BG, Tarling CA, Ford CM, Bowles DJ, Davies GJ. 2006. Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. *EMBO J.* 25:1396–1405.
- Otwinowski Z, Minor W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol.* 276:307–326.
- Parsonage D, Newton GL, Holder RC, Wallace BD, Paige C, Hamilton CJ, Dos Santos PC, Redinbo MR, Reid SD, Claiborne A. 2010. Characterization of the N-acetyl-alpha-D-glucosaminyl-L-malate synthase and deacetylase functions for bacillithiol biosynthesis in *Bacillus anthracis*. *Biochemistry.* 49:8398–8414.
- Pelroy RA, Bassham JA. 1972. Photosynthetic and dark carbon metabolism in unicellular blue-green-algae. *Archiv Fur Mikrobiologie.* 86:25–38.
- Persson K, Ly HD, Dieckelmann M, Wakarchuk WW, Withers SG, Strynadka NCJ. 2001. Crystal structure of the retaining galactosyltransferase LgtC from *Neisseria meningitidis* in complex with donor and acceptor sugar analogs. *Nat Struct Biol.* 8:166–175.
- Read RJ. 2001. Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallogr D.* 57:1373–1382.
- Sheldrick GM. 2008. A short history of SHELX. *Acta Crystallogr A.* 64:112–122.
- Shi WW, Jiang YL, Zhu F, Yang YH, Shao QY, Yang HB, Ren YM, Wu H, Chen Y, Zhou CZ. 2014. Structure of a novel O-linked N-acetyl-D-glucosamine (O-GlcNAc) transferase, GtfA, reveals insights into the glycosylation of pneumococcal serine-rich repeat adhesins. *J Biol Chem.* 289:20898–20907.
- Sindhuwinita N, Munoz E, Munoz FJ, Palcic MM, Peters H, Peters T. 2010. Binding of an acceptor substrate analog enhances the enzymatic activity of human blood group B galactosyltransferase. *Glycobiology.* 20:718–723.

- Terwilliger TC, Berendzen J. 1999. Automated MAD and MIR structure solution. *Acta Crystallogr D*. 55:849–861.
- Vetting MW, Frantom PA, Blanchard JS. 2008. Structural and enzymatic analysis of MshA from *Corynebacterium glutamicum*: Substrate-assisted catalysis. *J Biol Chem*. 283:15834–15844.
- Vrielink A, Rüger W, Driessen HP, Freemont PS. 1994. Crystal structure of the DNA modifying enzyme β -glucosyltransferase in the presence and absence of the substrate uridine diphosphoglucose. *EMBO J*. 13:3413–3422.
- Walsby AE. 1985. The permeability of heterocysts to the gases nitrogen and oxygen. *Biological Sci*. 226:345–366.
- Wang Y, Lechno-Yossef S, Gong Y, Fan Q, Wolk CP, Xu X. 2007. Predicted glucosyltransferase genes located outside the HEP island are required for formation of heterocyst envelope polysaccharide in *Anabaena* sp. strain PCC 7120. *J Bacteriol*. 189:5372–5378.
- Wolk CP, Dunn JH. 1970. Composition of the cellular envelopes of *Anabaena cylindrica*. *J Bacteriol*. 103:153–158.
- Wolk CP, Meeks JC, Lockau W, Currier TC, Haury JF, Thomas J. 1977. Nitrogen-metabolism, genetics and development of heterocyst-forming blue-green-algae. *J Phycol*. 13:74.
- Wolk CP, Murry MA. 1989. Evidence that the barrier to the penetration of oxygen into heterocysts depends upon two layers of the cell envelope. *Arch Microbiol*. 151:469–474.
- Zheng Y, Anderson S, Zhang Y, Garavito RM. 2011. The structure of sucrose synthase-1 from *Arabidopsis thaliana* and its functional implications. *J Biol Chem*. 286:36108–36118.