A conservative distribution of tridomain NDP-heptose synthetases in actinobacteria

Yue Tang[1*], Wei Tang[1], Min Wang[1,2], Zhilong Zhang[1,2] & Yihua Chen[1,2*]

[1]State Key Laboratory of Microbial Resources and CAS Key Laboratory of Microbial Physiological and Metabolic Engineering, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China;
[2]University of Chinese Academy of Sciences, Beijing 100049, China

Heptoses are important structural components of Gram-negative bacterium cell wall and participate in bacterial colonization, infection, and immune recognition. Current knowledge of NDP-heptose originating from D-sedoheptulose 7-phosphate in Gram-positive bacterium remains limited. Here, *in silico* analysis suggested that the special tridomain NDP-heptose synthetases with isomerase, kinase, and nucleotidyltransferase activities are conservatively distributed in Actinobacteria class of Gram-positive bacterium. Enzymatical characterization of the tridomain proteins from different strains showed that they are involved in ADP-D-*glycero*-β-D-*manno*-heptose biosynthesis despite the unexpected discovery of kinase activities deficient in some proteins. The presence of three types of NDP-heptose synthetases in Gram-positive bacterium suggests that it is also a rich source of heptoses and the heptose moieties may play important roles *in vivo*. Our work updates the understanding of NDP-heptose biosynthesis in Gram-positive bacterium and lays a solid foundation for further physiological function explorations.

tridomain NDP-heptose synthetase, NDP-heptose biosynthesis, widespread distribution, Gram-positive bacterium

INTRODUCTION

Carbohydrates are ubiquitously found in nature and a great abundance of the Earth's biomass is glycosylated (Thibodeaux et al., 2007; Zhang et al., 2020). In addition to the most found hexose and pentose, the versatile biological properties of heptose-containing glycoconjugates have been elaborated simultaneously (Guo et al., 2021). Heptose moieties not only act as building blocks of lipopolysaccharides (LPSs) and capsular polysaccharides (CPSs), which are important for the infection, colonization, and immune recognition of Gram-negative bacteria, but also are components of a variety of natural products with fascinating bioactivities from Gram-positive bacteria (Kosma, 2008;

Zhou et al., 2018; Lu et al., 2014).

We summarized the biosynthesis of diverse heptose moieties recently and showed that most of those moieties are derived from the same precursor, D-sedoheptulose 7-phosphate (S7-P), an intermediate of the pentose phosphate pathway (Guo et al., 2021). In principle, S7-P is converted to NDP-heptose via a four-step process sequentially catalyzed by enzymes with isomerase, kinase, phosphatase, and nucleotidyltransferase activities. To date, two kinds of NDP-heptose, GDP-α-heptose and ADP-β-heptose, derived from S7-P have been confirmed as the initial nucleotide-activated heptose precursors (Valvano et al., 2002). The activated NDP-heptoses are adopted as sugar donors for glycosyl-transfer reactions directly or after appropriate modification (Tang et al., 2018; Thibodeaux et al., 2008).

GDP-activated heptoses are currently proposed as sugar

---

donors for the heptose motifs in both CPS of some Gram-negative bacterium like food-borne pathogen *Campylobacter jejuni* species and S-layer (Surface-layer) glycoprotein of the Gram-positive *Aneurinibacillus thermoaerophilus* DSM 10155 (McCallum et al., 2013; Kneidinger et al., 2001). GDP-D-*glycero-α-D-manno*-heptose is a common intermediate of those GDP-heptoses and its biosynthesis is the same in varied bacteria (Huddleston and Raushel, 2019; Kneidinger et al., 2001). Using *A. thermoaerophilus* DSM 10155 as an example, an isomerase GmhA$_{AT}$ is responsible for the conversion of S7-P to D-*glycero-D-manno*-heptose 7-phosphate, which is then phosphorylated at the anomeric carbon by a kinase HddA and generates the diphosphate sugar, D-*glycero-α-D-manno*-heptose 1,7-bisphosphate; subsequently, a phosphatase GmhB$_α$ that takes *α*-configured heptose 1,7-bisphosphate as a substrate hydrolyzes the 7-phosphate group to form D-*glycero-α-D-manno*-heptose 1-phosphate, which is then activated by the guanylyltransferase HddC to afford GDP-D-*glycero-α-D-manno*-heptose (Figure 1A and B) (Kneidinger et al., 2001).

ADP-activated heptoses have been discovered as precursors for the heptose motifs in the LPS inner cores of Gram-negative bacterium and, sometimes, in their O-antigens (Raetz and Whitfield, 2002). They are also involved in the biosynthesis of a number of heptose-containing natural products from Gram-positive bacteria (Tang et al., 2018; Tang et al., 2020). ADP-D-*glycero-β-D-manno*-heptose is a common intermediate of those ADP-heptoses and its bio-

synthesis has been extensively studied in Gram-negative bacterium (Hu et al., 2019; Kneidinger et al., 2002; Morrison and Tanner, 2007). In *Burkholderia pseudomallei*, the formation mechanism of ADP-D-*glycero-β-D-manno*-heptose is analogous to that of GDP-D-*glycero-α-D-manno*-heptose, which also converts S7-P to D-*glycero-D-manno*-heptose 7-phosphate as the first intermediate, but differs in the following three steps: (i) the kinase HldA catalyzes the phosphorylation of the anomer carbon with a *β*-configuration; (ii) the phosphatase GmhB$_β$ is prone to recognize *β*-configured 1,7-diphosphate heptose; (iii) the nucleotidyltransferase HldC activates D-*glycero-β-D-manno*-heptose 1-phosphate as an ADP-sugar (Figure 1A and B) (Park et al., 2018; Lee et al., 2013; Wang et al., 2010). Notably, in a lot of Gram-negative bacteria like *E. coli*, a bifunctional HldE with a kinase and a nucleotidyltransferase domain replaces HldA and HldC for the C-1 *β*-configured phosphorylation and the last adenylyltransfer step (Kneidinger et al., 2001).

Recently, we discovered the first tridomain NDP-heptose synthesis enzyme SepB with isomerase, kinase, and nucleotidyltransferase activities during elucidating the biosynthesis of septacidins, a group of antifungal antibiotics from the Gram-positive bacterium *Streptomyces fimbriatus* CGMCC 4.1598 (Tang et al., 2018). SepB, together with SepL (a GmhB$_β$ type phosphatase), synthesizes ADP-D-*glycero-β-D-manno*-heptose *via* the same process as that in *B. pseudomallei* and *E. coli*. In this work, we investigated the distribution of the special tridomain NDP-heptose
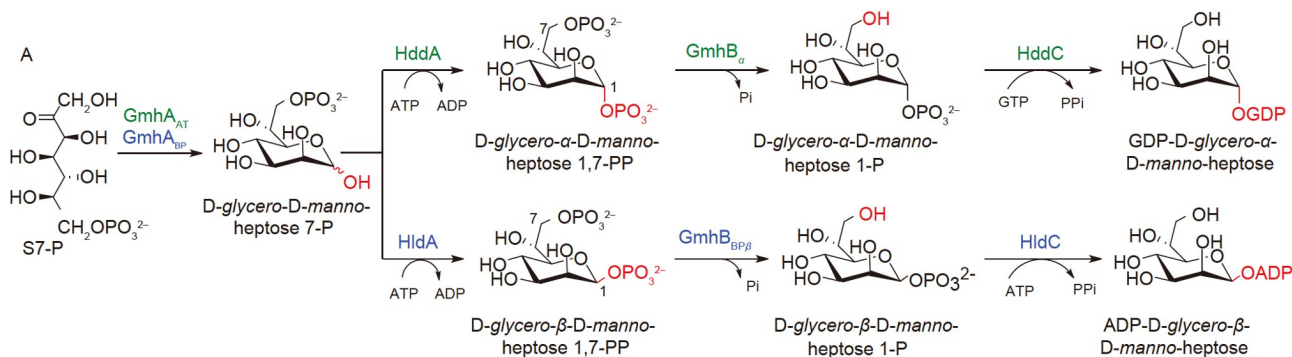
B

Representative source strains	Isomerase	Kinase	Nucleotidyltransferase	Phosphatase	Final products
A. thermoaerophilus DSM 10155	GmhA$_{AT}$	HddA	HddC	GmhB$_α$	GDP-D-*glycero-α-D-manno*-heptose
B. pseudomallei	GmhA$_{BP}$	HldA	HldC	GmhB$_{BPβ}$	ADP-D-*glycero-β-D-manno*-heptose
E. coli	GmhA	HldE (didomain)		GmhB$_β$	
S. fimbriatus	SepB (tridomain)			SepL	

Figure 1   Comparison of the biosyntheses of GDP-heptose and ADP-heptose. A, Proposed biosynthetic pathways of GDP-D-*glycero-α-D-manno*-heptose and ADP-D-*glycero-β-D-manno*-heptose. B, A summary of the NDP-heptose synthesis enzymes. Enzymes for GDP-D-*glycero-α-D-manno*-heptose and ADP-D-*glycero-β-D-manno*-heptose biosyntheses are indicated with green and blue, respectively. To distinguish the same named enzymes of GmhA and GmhB$_β$ in different organisms, the initials of their bacterial source strains are added as suffixes.

synthesis enzyme and found that they are conservative to the Actinobacteria class of Gram-positive bacterium. Four of those tridomain proteins were characterized *in vitro*, revealing that they are all involved in the biosynthesis of ADP-D-*glycero*-β-D-*manno*-heptose, although three of them were dysfunctional with the kinase activity unexpectedly. We also analyzed the related gene clusters and proposed that they dictate several different types of heptose-containing products, which may participate in various biological processes of actinobacteria.

RESULTS

Distribution of the SepB-like tridomain NDP-heptose synthetases

The special tridomain feature of SepB aroused our interests to investigate its distribution. Blast searches using SepB as a probe at the NCBI nr database (till April 6, 2021) resulted in 85 proteins with sequence identity and coverage more than 40% and 90%, respectively. It is surprising that all the 86 proteins (including SepB) are from 12 different orders of the Gram-positive Actinobacteria class, mainly dispersed in orders Streptomycetales (28/86), Frankiales (18/86), and Micrococcales (16/86) (Table S1 in Supporting Information). A phylogenic tree was then constructed to present a snapshot of the distributions and the phylogenetic relationships of those proteins. As shown in Figure 2A, most orthologs from strains in the same order are grouped together, such as Streptomy-

cetales, Kineosporiales, and Micrococcales; while some orthologous proteins from strains in the same order are split into two distinct branches, such as Frankiales. In addition, a small number of proteins from Streptosporangiales, Pseudonocardiales, and Catenulisporales are embedded in the Streptomycetales clade, indicating a close evolutionary relationship of those proteins with their homologues from Streptomycetales.

Furthermore, a sequence similarity network (SSN) analysis, which has been used to predict protein functions in many examples (Yang et al., 2017), was built for those tridomain proteins. At an *E*-value cutoff of 10^{-180}, the 86 proteins were mainly segregated into 9 main clusters, each containing 3 or more members and the remaining ones dispersed sporadically. In the network topology (Figure 2B), most tridomain proteins from Micrococcales (12/16) are organized together; while those from Streptomycetales (21/27) are separated into two main clusters. Interestingly, the 18 proteins from Frankiales constitute four distinguished clusters, indicating a functional differentiation of those tridomain enzymes.

Analysis of the gene clusters harboring the tridomain *sepB* homologues

ADP-β-heptose and GDP-α-heptose are involved in a variety of important biological processes in Gram-negative bacteria (Guo et al., 2021). In Gram-positive bacteria, we only know that they can participate in septacidin biosynthesis and S-
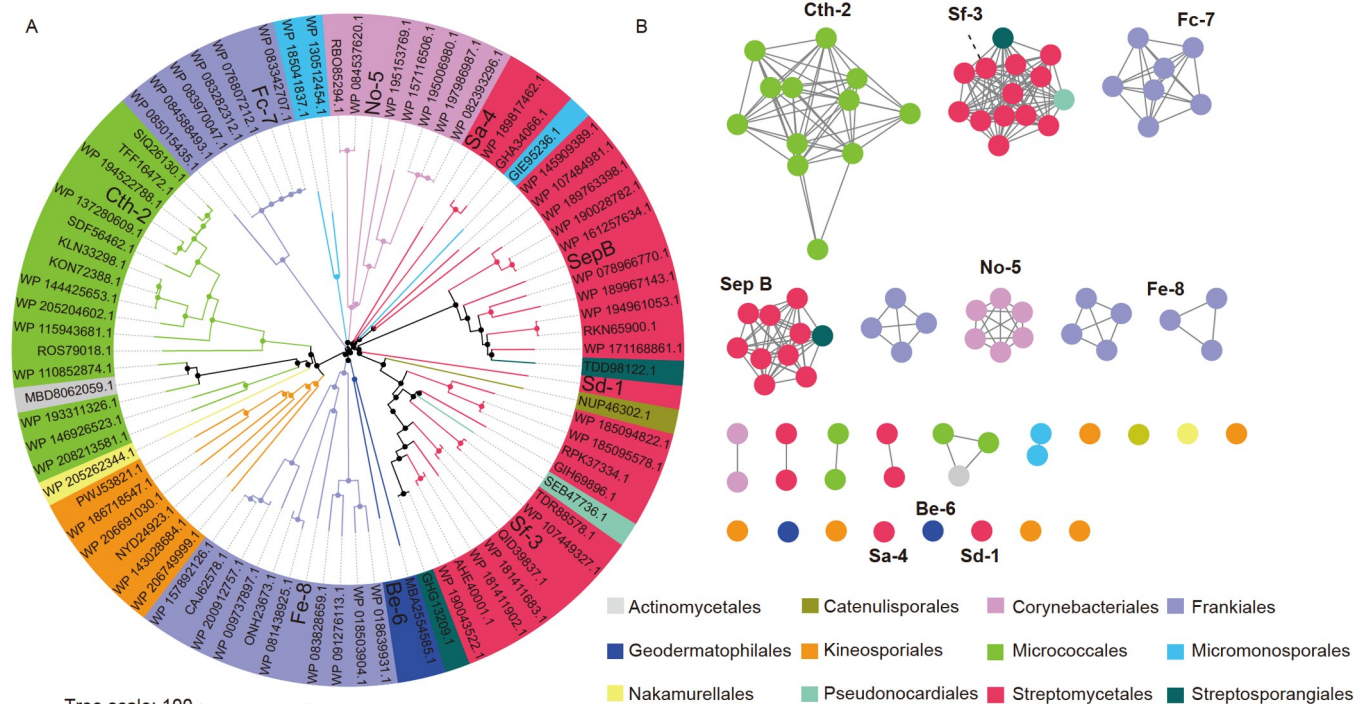

Figure 2 Bioinformatic analysis of the tridomain NDP-heptose synthetases. A, Phylogenetic analysis of the tridomain NDP-heptose synthetases. B, Sequence similarity network of the tridomain NDP-heptose synthetases. Different colors represent different orders in Actinobacteria class in both (A) and (B).

layer glycoprotein installation (Tang et al., 2018; Kneidinger et al., 2001). Discovery of 85 SepB homologues, which are most likely to catalyze NDP-heptose formation, serves as a good chance to take a glance at the biological meaning of heptose in Gram-positive bacteria. We therefore analyzed the genes in the vicinity of those *sepB* homologues carefully, and categorized most of the gene clusters containing *sepB* homologues (75 of the 85 clusters) into three types (type I to III, Figure 3, Table S1 in Supporting Information). All of them contain genes encoding proteins sharing considerable similarities with SepB and SepL, indicating that an analogous NDP-heptose biosynthetic pathway exists in those organisms. The remaining ten clusters were sorted into the "other" group since the adjacent genes do not belong to the three types (3) or are unavailable (7) (Table S1 in Supporting Information).

For the 75 categorized gene clusters, 13 of them belong to type I, with 11 of them from Streptomycetales (including the *sep* gene cluster), one from Catenulisporales, and one from Streptosporangiales. All of them share very high level of similarities with the *sep* gene cluster, indicating their products are septacidin analogues. There are 27 gene clusters belonging to type II, with 10 from Streptomycetales, eight from Corynebacteriales, seven from Frankiales, and the remaining two from Streptosporangiales and Pseudonocardiales, respectively. Each of the type II gene cluster possesses one glycosyltransferase gene and some genes encoding modification enzymes like cytochrome P450 and radical *S*-adenosyl-L-methionine (SAM) proteins, implying that a group of natural products with decorated heptose moieties are synthesized by the type II gene clusters. Specifically, four or more glycosyltransferase genes are present in the type III gene clusters (36), indicating that the products of those gene clusters are heptose-containing glycan chains. The type III clusters distribute in Micrococcales (16), Frankiales (11), Kineosporiales (3), Micromonosporales (3), Geodermatophilales (1), Nakamurellales (1), and Streptomycetales (1) orders. It is notable that 11 of the 18 clusters from Frankiales belong to type III, and the remaining seven gene clusters belong to type II, which corresponds with the distributions of those gene clusters in the phylogenetic tree (Figure 2A). A striking character of the type III clusters from Frankiales is that each of them contains a gene for the bifunctional HldE homologue, implying that they have two sets of genes for NDP-heptose biosynthesis.

Expression of the SepB homologues in *E. coli*

Taken the phylogeny, SSN, and gene cluster analysis results into consideration, eight representative tridomain proteins were selected for further enzymatic characterization (Figure 2A). According to their strain sources, they were named as Sd-1 (from *Streptomyces* sp. DfronAA-171), Cth-2 (from

Cellulosimicrobium sp. TH-20), Sf-3 (from *Streptomyces* sp. NRRL F-5639), Sa-4 (from *Streptomyces antimycoticus*), No-5 (from *Nocardia otitidiscaviarum* IFM 11049), Be-6 (from *Blastococcus endophyticus* DSM 45413), Fc-7 (from *Frankia* sp. CcI6), and Fe-8 (from *Frankia* sp. EI5c), respectively. The codon optimized genes of the eight proteins were synthesized and expressed as N-His$_6$ tagged proteins in *E. coli* BL21. Only Cth-2, Sa-4, No-5, and Be-6 were obtained as soluble proteins even if various protein expression conditions were tested (Figure 4A). At the preliminary assays for catalytic activity testing, we were plagued by several false-positive results (data not shown), which most likely were caused by the contamination of ADP-heptose synthesis enzymes from the *E. coli* host. To obtain a clean strain for protein expression, a mutant strain *E. coli* BL21 Δ(*gmhA*, *gmhB$_β$*, *hldE*), in which the isomerase gene *gmhA*, the phosphatase gene *gmhB$_β$*, and the bifunctional kinase/adenyltransferase gene *hldE* were sequentially knocked out, was constructed (Figure S1 in Supporting Information). *E. coli* BL21 Δ(*gmhA*, *gmhB$_β$*, *hldE*) was proved to be a reliable host for expressing NDP-heptose synthesis related proteins in the following studies.

Characterization of Be-6 as a tridomain isomerase, kinase, and adenylyltransferase

The high similarities of Cth-2, Sa-4, No-5, and Be-6 with SepB (41%–48% identities) imply they have similar functions as SepB. Therefore, the catalytic activities of the four proteins were first investigated by incubating each of them with a heptose 1,7-biphosphate 7-phosphatase (GmhB$_α$ from *A. thermoaerophilus* DSM 10155 or GmhB$_β$ from *E. coli*), S7-P, ATP (as the phosphate donor for kinase activity), and different NTPs (as the possible sugar acceptor for nucleotidyltransferase activity. Compared with the negative control, a new peak was observed by HPLC only in the Be-6 assays using GmhB$_β$ as the phosphatase and ATP as the sugar acceptor. Addition of the other NTPs did not influence the Be-6 reactions, suggesting that ATP is used as both the phosphate donor and the sugar acceptor in the four-step-relay of Be-6 and GmhB$_β$. The products of the Be-6 assays were identified as ADP-D-*glycero*-*β*-D-*manno*-heptose by comparison with its authentic standard (Figure 4B). Therefore, Be-6 was confirmed as a tridomain protein as SepB with isomerase, kinase, and adenylyltransferase activities.

The kinase domains of Cth-2, Sa-4, and No-5 are dysfunctional

It was unexpected that no product was detected in the assays of Cth-2, Sa-4, and No-5 regardless they were combined with GmhB$_α$ or GmhB$_β$. One possibility is that not all three domains in those proteins are functional. Multiple sequence
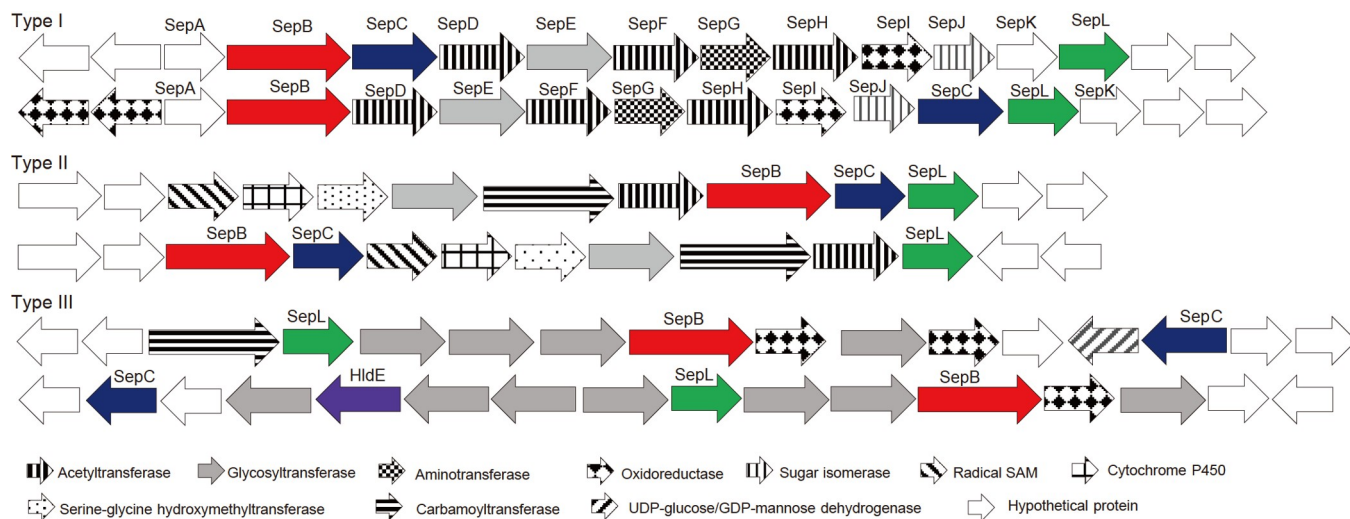
Figure 3 Three types of tridomain NDP-heptose synthetase containing gene clusters.
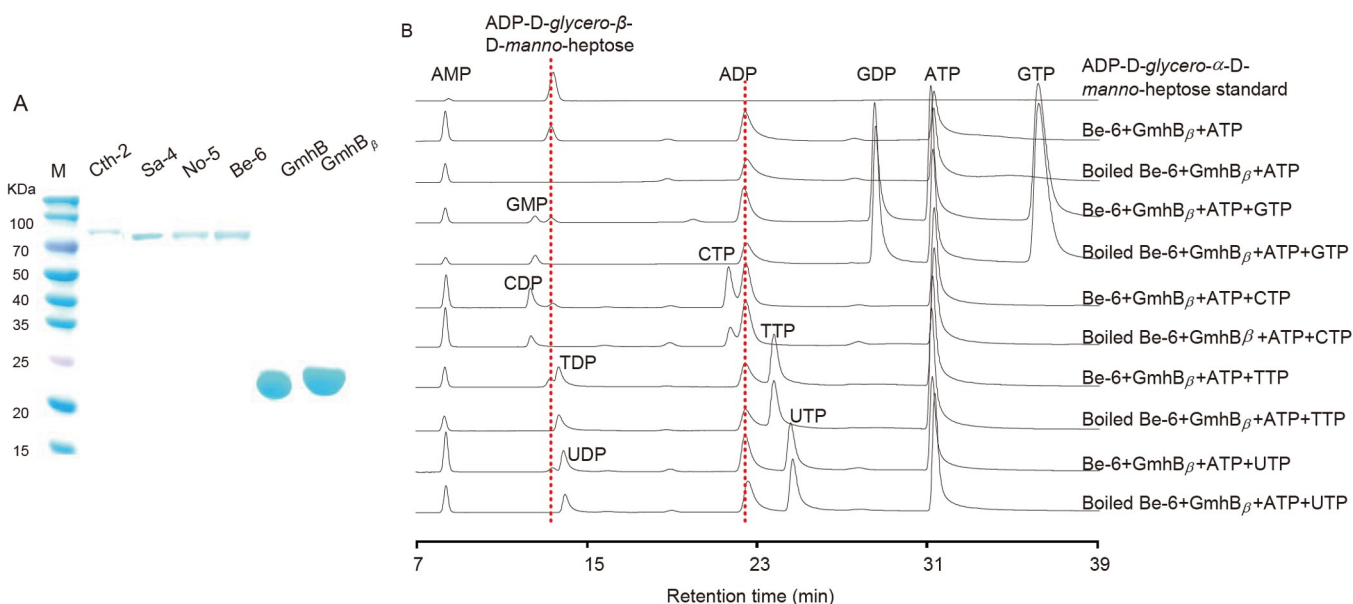

Figure 4 *In vitro* characterization of Be-6. A, SDS-PAGE analysis of N-His$_6$ tagged Cth-2, Sa-4, No-5, Be-6, GmhB$_\alpha$, and GmhB$_\beta$. Lane M, protein marker. B, HPLC profiles of Be-6+GmhB$_\beta$ assays with different NTPs using S7-P as a substrate.

alignments suggested that the N-terminal isomerase domains of them share comparable similarities (24%–35% identities) with isomerases GmhA and GmhA$_{AT}$, the two isozymes involved in the biosynthesis of ADP-β-heptose and GDP-α-heptose, respectively; while their kinase and nucleotidyl-transferase domains are more like those from the ADP-β-heptose pathway (23%–43% identities with HldA, HldC, and the bifunctional HldE) than HddA and HddC from the GDP-α-heptose pathway (less than 14% identities) (Table S2 in Supporting Information). Therefore, the conserved active site residues for the isomerase, kinase, and nucleotidyl-transferase domains of Cth-2, Sa-4, and No-5 were analyzed by comparison with those in GmhA, HldA, and HldC, re-

spectively (Taylor et al., 2008; Park et al., 2018; Lee et al., 2013). However, no mutation or deletion on those residues was found giving no clue to the dysfunction of the three proteins (Figure 5A).

To probe each domain's function of Cth-2, Sa-4, and No-5, the isomerase and the different kinase and nucleotidyl-transferase proteins from the ADP-β-heptose and GDP-α-heptose pathways were synthesized, expressed as N-His$_6$ tagged or GST-tagged proteins in *E. coli*, and purified for domain compensation assays (Figure 5B). In total, five proteins were prepared, including the isomerase GmhA, the β-kinase HldA and its cognate adenylyltransferase HldC, and the α-kinase HddA and its cognate guanylyltransferase HddC
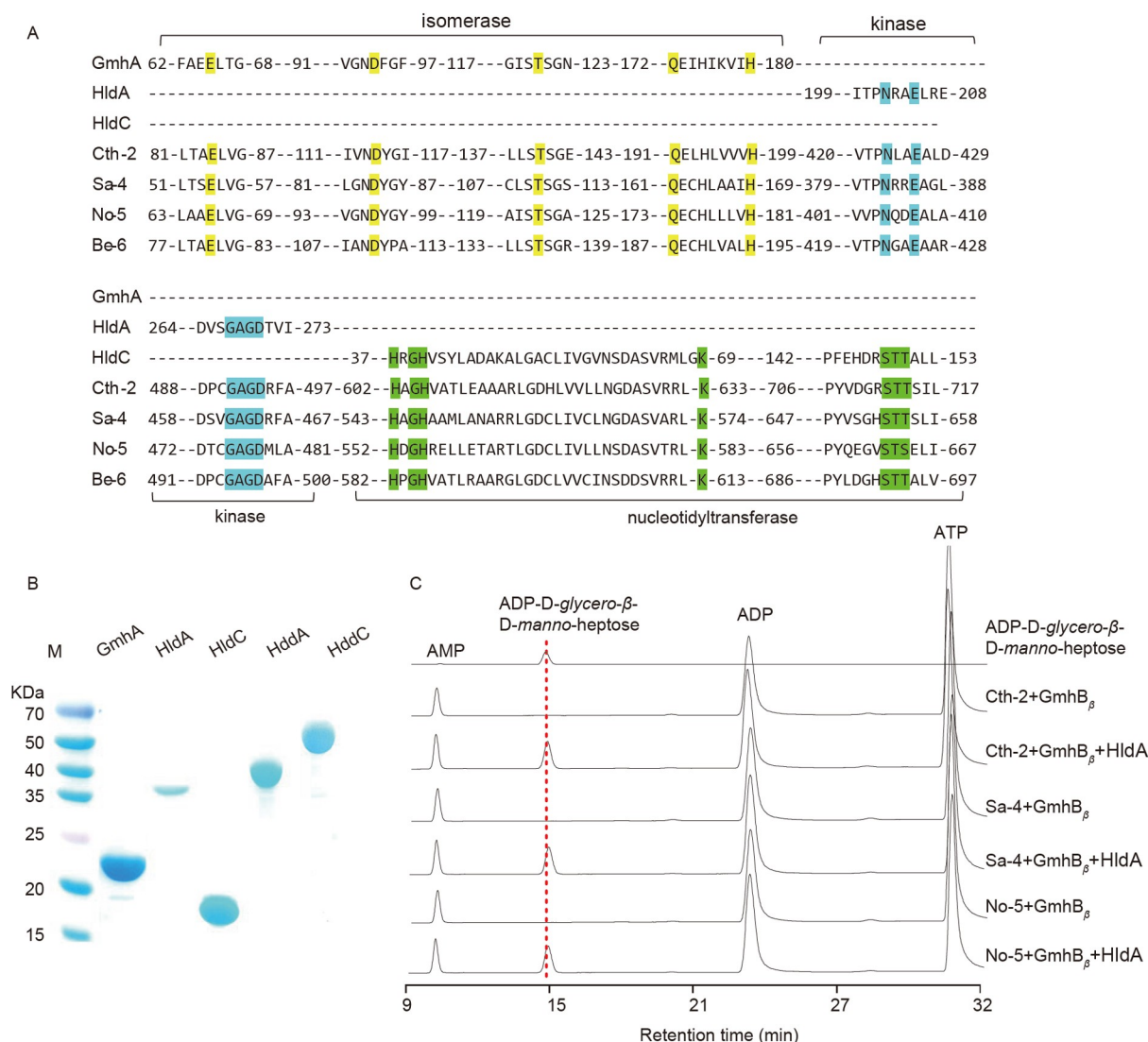
Figure 5 Enzymatical characterizations of Cth-2, Sa-4, and No-5. A, Sequence alignments of conserved amino acids in tridomain NDP-synthetases with GmhA, HldA, and HldC. The active site residues in isomerase, kinase, and nucleotidyltransferase are indicated with yellow, blue, and green, respectively. B, SDS-PAGE analysis of recombinant proteins. M, protein marker; GmhA, HldA, HldC, and HddA are *N*-His$_6$ tagged proteins; HddC, GST-tagged HddC. C, HPLC profiles of enzymatic assays of Cth-2, Sa-4, and No-5.

(Figure 5B). Two types of compensation assays were conducted: (i) for the ADP-β-heptose assay, each of the tridomain protein was incubated with GmhB$_\beta$ and an additional single-function protein (GmhA, HldA, or HldC) using S7-P and ATP as the substrates; (ii) for the GDP-α-heptose assay, each of the tridomain protein was incubated with GmhB$_\alpha$ and an additional single-function protein (GmhA, HddA, or HddC) using S7-P, ATP (phosphate donor) and GTP (sugar acceptor) as the substrates. HPLC analyses revealed that no new product was detected in the GDP-α-heptose assays, and production of ADP-D-*glycero*-β-D-*manno*-heptose was observed only when HldA, the kinase that forms D-*glycero*-β-D-*manno*-heptose 1,7-biphosphate, was added to the ADP-β-heptose assay (Figure 5C). Taken together, the results suggested that Cth-2, Sa-4, and No-5 keep their isomerase and

nucleotidyltransferase activities, but lose kinase activity during evolution.

DISCUSSION

As is well known, Gram-negative bacterium is a rich source of heptose motifs, which can participate in bacterial colonization, infection, and immune recognition (Hu et al., 2019). Both ADP- and GDP-heptoses that participate the installation of polysaccharides like CPS and LPS are biosynthesized via a similar process involving four enzymes, an isomerase, a kinase, a phosphatase, and a nucleotidyltransferase (Valvano et al., 2002). Given the very important role of LPS for Gram-negative bacteria, two of the enzymes synthesizing ADP-

heptose, which is the sugar donor of LPS inner core that usually contains 2 to 4 heptose moieties (Cloutier et al., 2018), were fused together during evolution to generate the didomain kinase-adenylyltransferase HldE. The discovery of a tridomain SepB from septacidin biosynthesis in *Streptomyces* is unexpected, which indicates that heptoses take a conserved and important role in Gram-positive bacterium, which is not recognized as a rich source of heptose before. In this study, the bioinformatic analysis of the genome sequenced bacteria revealed that 86 tridomain NDP-heptose synthetases are conservatively distributed in Actinobacteria, indicating that the tridomain protein was generated by protein fusion not very long ago and therefore spread only in this class till now. Worthy of note, we could not exclude that the gene of this fused tridomain enzyme may also distribute in some other organisms whose genome sequences are not available. Moreover, the didomain HldE homologues were observed in clusters from Frankiales. Taken the single domain GDP-heptose synthetase from *A. thermoaerophilus* into account, all three forms of NDP-heptose synthetases are present in Gram-positive bacteria, suggesting that the roles of heptoses in Gram-positive bacteria are worthy of attention.

Subsequent analysis of the genes in vicinity to the tridomain NDP-heptose synthetase genes gave some clues about their roles. Among the 78 clusters whose adjacent genes are available, 75 could be categorized into three types, with type I gene clusters encoding septacidin analogues, type II gene clusters dictating natural products with decorated heptose moieties, and type III gene clusters responsible for the installation of glycan chains. Although the products of type II and type III clusters are unknown, the products of type I clusters, septacidin and its analogues, display a variety of bioactivities including antifungal, antiproliferation, and pain-relief (Tang et al., 2018). Gene fusion is frequently observed in the biosynthetic gene clusters of natural products, and the representative examples are type I polyketides synthases that produce many important drugs like erythromycin and lovastatin and non-ribosomal peptide synthetases that synthesize drugs like daptomycin and vancomycin (Fischbach and Walsh, 2006). Therefore, it may not be surprising that the NDP-heptose synthetase genes are fused in the clusters of natural products with significant activities (e.g., antifungal) for environment adaption of the hosts.

Enzymatical characterization of four selected tridomain proteins, Be-6, Cth-2, No-5, and Sa-4, showed that all of them are involved in the biosynthesis of ADP-D-*glycero-β*-D-*manno*-heptose, while the kinase domains of the latter three proteins are dysfunctional unexpectedly. There are several possibilities of the kinase domain inactivation. One possibility is that the dysfunction of the kinase domain is a result of gene fusion. Alignment analysis revealed that the active

site residues of the isomerase, kinase, and nucleotidyltransferase domains are all conserved in Cth-2, No-5, and Sa-4, suggesting that the improper linkers of those fused domains might render them inappropriate conformations and influence their functions. Another possibility is that at least one enzyme has the same kinase activity, which can compensate the dysfunctional kinase domain, and the unnecessary kinase domain of the tridomain protein is inactivated during evolution. This assumption is supported by the observation that all the 11 type III gene clusters from Frankiales contain a bifunctional HldE homologue with kinase and nucleotidyltransferase domains.

In conclusion, it was showed that, as Gram-negative bacterium, Gram-positive bacterium is also a rich source of heptoses that are derived from S7-P. Moreover, Gram-positive bacteria contain all of the three types of NDP-heptose synthetases, with the tridomain NDP-heptose synthetases only distribute in the Actinobacteria class. All the characterized tridomain proteins are involved in ADP-heptose biosynthesis, albeit some of them are dysfunctional in their kinase domains. Our work updated the current knowledge of NDP-heptose biosynthesis in Gram-positive bacterium and set a stage for further work towards heptose-containing natural products mining and their physiological function explorations.

MATERIALS AND METHODS

Bacterial strains and plasmids

Bacterial strains and plasmids used in this study are listed in Table S3 in Supporting Information.

DNA manipulation and sequence analysis

DNA synthesis and sequencing were carried out in Generay Co. (Shanghai, China) and Biosune Co. (Shanghai, China), respectively. Polymerase chain reactions (PCRs) were performed with PrimeSTAR HS DNA polymerase (Takara, Japan) or Taq DNA polymerase (TransGene, Beijing, China). Restriction enzyme digestions, ligations and transformations were carried out following the general methods. All PCR primers used in this study are listed in Table S4 in Supporting Information. A BLASTP search was used to predict protein functions (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Protein sequence alignments were performed using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/).

Bioinformatic analysis of the SepB homologues

SepB (Accession No.: AVM80385.1) was used as a probe to scan its homologous sequences from the NCBI nr database (till April 1, 2020) with online BLASTP search (https://

blast.ncbi.nlm.nih.gov/Blast.cgi) and the algorithm parameter expect threshold (*E*-value) was set as <10^{-5}. To screen out tridomain sequences, sequence identity and coverage were set >40% and 90%, respectively. For phylogenetic analysis, MEGA6 was used to generate the phylogenetic tree using the neighbor-joining algorithm which was displayed and edited with iTOL (Interactive Tree Of Life) (https://itol.embl.de/). The network topology (*E*-value<10^{-180}) was generated with the Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST) (https://efi.igb.illinois.edu/efi-est/), which was visualized in Cytoscape using the organic layout.

Construction of the *E. coli* mutant strain

The *E. coli* BL21 Δ(*gmhA*, *gmhB$_\beta$*, *hldE*) mutant strain was constructed by in-frame gene deletion via λ RED-mediated recombination coupled with I-SceI cleavage (Lin et al., 2014). The isomerase gene *gmhA* in the genome of *E. coli* BL21 was firstly deleted. In brief, the upstream and downstream arm of *gmhA* were amplified from *E. coli* BL21 genome using primer pairs AUP-1/AUP-2 and ADW-1/ADW-2, respectively, with the 3′ flank of the upstream arm of *gmhA* containing a 20-bp overlapping sequence (AD) of the 5′ flank of the downstream arm of *gmhA*. The tetracycline resistance gene (*tetA*) with I-sceI recognition site on its both sides was amplified from pTKS/CS using primer pairs Atet-1/Atet-2. The *tetA* cassette flanking with the upstream and downstream arm of *gmhA* respectively was obtained by two rounds of overlap PCR using primer pairs AUP-1/Atet-2 and AUP-1/ADW-2 sequentially. The target fragment was then electroporated into *E. coli* BL21-RED (*E. coli* BL21 containing pTKRED) with 0.1 mmol L^{-1} isopropyl-*β*-thiogalactoside (IPTG) for inducing the expression of the recombinase. The intermediate strain with tetracycline resistance was selected and confirmed by PCR using primer pair AV-1/AV-2. Subsequently, 10 mmol L^{-1} L-arabinose was added to induce I-SceI endonuclease expression for cleaving the *tetA* and facilitate the recombination between ADs. After PCR verification with primer pair AV-1/AV-2, the correct recombination strain named *E. coli* BL21 Δ*gmhA* was used to construct the in-frame deletion mutant strain of the phosphatase *gmhB$_\beta$*.

The *E. coli* BL21 Δ(*gmhA*, *gmhB$_\beta$*) was constructed using the same method as that for *gmhA* deletion with primers listed in Table S4 in Supporting Information. The PCR confirmed recombination strain was then used to delete the bifunctional gene *hldE* (kinase/adenylyltransferase) and the detailed method was similar to that for *gmhB$_\beta$* deletion with primers in Table S4 in Supporting Information. After PCR verification, the desired mutant *E. coli* BL21 Δ(*gmhA*, *gmhB$_\beta$*, *hldE*) was used for protein expression in the following studies.

Protein expression and purification

The 2.2-kb *cth-2* gene was synthesized and inserted into the *Nde* I/*Bam*H I sites of pET28a to afford pET28a-Cth-2, which was electroporated into *E. coli* BL21 Δ(*gmhA*, *gmhB$_\beta$*, *hldE*). A single transformant of *E. coli* BL21 Δ(*gmhA*, *gmhB*, *hldE*)/pET28a-Cth-2 was inoculated into LB with 50 μg mL^{-1} kanamycin and cultured overnight at 37°C, 220 r min^{-1}. The overnight culture was used to inoculate LB medium with 50 μg mL^{-1} kanamycin at 1:100 dilution and incubated at 37°C, 220 r min^{-1} until A_{600} reached 0.6. Expression of Cth-2 was then induced by the addition of IPTG at a final concentration of 0.1 mmol L^{-1} and cultured at 16°C, 180 r min^{-1} for a further 18 h.

Cth-2 was purified with Ni-NTA affinity column at 4°C following the manufacturer's instructions. After harvested by centrifugation, the cells were re-suspended in binding buffer (20 mmol L^{-1} Tris-HCl, 500 mmol L^{-1} NaCl, 5 mmol L^{-1} imidazole, pH 7.9). After sonication, cell debris was removed by centrifugation at 16,000×*g* for 30 min. The supernatant containing His$_6$-tagged proteins was loaded onto the Ni-NTA affinity column, which had been equilibrated with binding buffer, then washed with washing buffer (20 mmol L^{-1} Tris-HCl, 500 mmol L^{-1} NaCl, 60 mmol L^{-1} imidazole, pH 7.9) followed by elution buffer (20 mmol L^{-1} Tris-HCl, 500 mmol L^{-1} NaCl, 500 mmol L^{-1} imidazole, pH 7.9). The fractions containing Cth-2 were pooled and desalted with PD-10 columns (GE Healthcare, USA) and concentrated by ultracentrifugation with an Amicon Ultra-4 10K centrifugal filter device (Merck Millipore, USA). The purified protein was stored in 20 mmol L^{-1} HEPES buffer (pH 8.0) with 100 mmol L^{-1} NaC1 and 20% glycerol at −80°C. Protein concentration was measured by Bradford assay using bovine serum albumin as a standard.

Similarly, the 2.0-kb *sa-4*, the 2.0-kb *no-5*, the 2.1-kb *be-6*, 0.9-kb *hldA*, the 0.5-kb *hldC*, the 1.0-kb *hddA*, and the 0.5-kb *gmhB$_\alpha$* genes were also synthesized and inserted into the *Nde* I/*Bam*H I sites of pET28a to afford pET28a-Sa-4, pET28a-No-5, pET28a-Be-6, pET28a-HldA, pET28a-HldC, pET28a-HddA, and pET28a-GmhB$_\alpha$, respectively. Plasmids pET28a-GmhA and pET28a-GmhB$_\beta$ were constructed in our previous work (Tang et al., 2018). Protein expression and purification of these proteins were performed using the same procedures as those for Cth-2.

The 0.7-kb *hddC* gene was synthesized and inserted into the *Bam*H I/*Xho* I sites of pGEX-6P-1 to afford pGEX-6P-1-HddC. Protein expression of HddC was performed using the same procedures as those for Cth-2 except kanamycin is replaced by ampicillin (100 μg mL^{-1}). To purify the GST-tagged fusion protein, the cells were harvested by centrifugation and re-suspended in lysis buffer (140 mmol L^{-1} NaCl, 2.7 mmol L^{-1} KCl, 10 mmol L^{-1} Na$_2$HPO4, 1.8 mmol L^{-1} KH$_2$PO4, pH 7.3). After sonication and cen-

trifugation, the supernatant was loaded onto a glutathione affinity resin column equilibrated with lysis buffer and washed twice with the same buffer followed by elution buffer (50 mmol L^{-1} Tis-HCl, 10 mmol L^{-1} reduced glutathione, pH 8.0). The desired fraction was pooled, desalted, and concentrated as those for Cth-2.

Enzymatic assays of Be-6

The catalytic activity of Be-6 was studied in a hybrid assay together with the phosphatase GmhB$_\alpha$ or GmhB$_\beta$. The enzymatic reaction of Be-6+GmhB$_\alpha$ was carried out in a 50 μL volume mixture containing 20 mmol L^{-1} HEPES buffer (pH 8.0), 100 mmol L^{-1} NaCl, 2 mmol L^{-1} MgCl$_2$, 2 mmol L^{-1} KCl, 2 mmol L^{-1} ATP, 0.2 mmol L^{-1} S7-P, 5 μmol L^{-1} Be-6, and 5 μmol L^{-1} GmhB$_\alpha$ at 30°C for 3 h. To determine the proper sugar acceptor of Be-6, 2 mmol L^{-1} GTP, CTP, UTP, or TTP was added into the hybrid reactions together with 2 mmol L^{-1} ATP. The enzymatic assays of Be-6+GmhB$_\beta$ were set as well as those of Be-6+GmhB$_\alpha$ except that GmhB$_\alpha$ was replaced with GmhB$_\beta$. All the reactions were quenched by vibrating vigorously with an equal volume of chloroform. After centrifugation, the chloroform layer was removed and 10 μL of aqueous sample was subjected to HPLC analysis.

Enzymatic assays of Cth-2, Sa-4, and No-5

The enzymatic assays of Cth-2, Sa-4, and No-5 were set similarly and here takes Cth-2 for an example. When combined with phosphatase GmhB$_\alpha$ or GmhB$_\beta$, the enzymatic reactions of Cth-2 were carried out as those of Be-6 +GmhB$_\alpha$ or Be-6+GmhB$_\beta$ except that Be-6 was replaced by Cth-2. The ADP-β-heptose compensation experiments of Cth-2 were performed in a 50 μL volume mixture containing 20 mmol L^{-1} HEPES buffer (pH 8.0), 100 mmol L^{-1} NaCl, 2 mmol L^{-1} MgCl$_2$, 2 mmol L^{-1} KCl, 2 mmol L^{-1} ATP, 0.2 mmol L^{-1} S7-P, 5 μmol L^{-1} Cth-2, 5 μmol L^{-1} GmhB$_\beta$, and 5 μmol L^{-1} sing-functional enzyme (GmhA, HldA or HldC) at 30°C for 3 h. The GDP-α-heptose compensation experiments of Cth-2 were performed in a 50 μL volume mixture containing 20 mmol L^{-1} HEPES buffer (pH 8.0), 100 mmol L^{-1} NaCl, 2 mmol L^{-1} MgCl$_2$, 2 mmol L^{-1} KCl, 2 mmol L^{-1} ATP, 2 mmol L^{-1} GTP, 0.2 mmol L^{-1} S7-P, 5 μmol L^{-1} Cth-2, 5 μmol L^{-1} GmhB$_\alpha$, 5 μmol L^{-1} sing-functional enzyme (GmhA, HddA or HddC) at 30°C for 3 h. The enzymatic assays of Sa-4 and No-5 were performed the same as those of Cth-2 except that Cth-2 was replaced by Sa-4 or No-5, respectively. All the reactions were quenched by vibrating vigorously with an equal volume of chloroform. After centrifugation, the chloroform layer was removed and 10 μL aqueous sample was subjected to HPLC analysis.

Spectroscopic analysis

HPLC analysis was performed with a Dionex CarboPac^{TM} PA1 BioLC^{TM} column (4×250 mm, Thermo Fisher Scientific, USA) on a Shimadzu HPLC system (Shimadzu, Japan). The column was developed with solvent A (ddH$_2$O) and solvent B (1.0 mol L^{-1} CH$_3$COONH$_4$) at a flow rate of 1.0 mL min^{-1}. The percentage of solvent B was changed using the following gradient: 0–2 min, 25%; 2–7 min, 25%–45%; 7–20 min, 45%–65%; 20–22 min, 65%–90%; 22–33 min, 90%; 34–40 min, 25%. The detection wavelength was 254 nm (Tang et al., 2020).

References

Cloutier, M., Muru, K., Ravicoularamin, G., and Gauthier, C. (2018). Polysaccharides from *Burkholderia* species as targets for vaccine development, immunomodulation and chemical synthesis. Nat Prod Rep 35, 1251–1293.

Fischbach, M.A., and Walsh, C.T. (2006). Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. Chem Rev 106, 3468–3496.

Guo, Z., Tang, Y., Tang, W., and Chen, Y. (2021). Heptose-containing bacterial natural products: Structures, bioactivities, and biosyntheses. Nat Prod Rep, , doi: 10.1039/d0np00075b.

Hu, X., Yang, C., Wang, P.G., and Zhang, G.L. (2019). ADP-heptose: A new innate immune modulator. Carbohydrate Res 473, 123–128.

Huddleston, J.P., and Raushel, F.M. (2019). Biosynthesis of GDP-D-*glycero*-α-D-*manno*-heptose for the capsular polysaccharide of *Campylobacter jejuni*. Biochemistry 58, 3893–3902.

Kneidinger, B., Graninger, M., Puchberger, M., Kosma, P., and Messner, P. (2001). Biosynthesis of nucleotide-activated D-*glycero*-D-*manno*-heptose. J Biol Chem 276, 20935–20944.

Kneidinger, B., Marolda, C., Graninger, M., Zamyatina, A., McArthur, F., Kosma, P., Valvano, M.A., and Messner, P. (2002). Biosynthesis pathway of ADP-L-*glycero*-β-D-*manno*-heptose in *Escherichia coli*. J Bacteriol 184, 363–369.

Kosma, P. (2008). Occurrence, synthesis and biosynthesis of bacterial heptoses. COC 12, 1021–1039.

Lee, T.W., Verhey, T.B., Antiperovitch, P.A., Atamanyuk, D., Desroy, N., Oliveira, C., Denis, A., Gerusz, V., Drocourt, E., Loutet, S.A., et al. (2013). Structural–functional studies of *Burkholderia cenocepacia* D-*glycero*-β-D-*manno*-heptose 7-phosphate kinase (HldA) and characterization of inhibitors with antibiotic adjuvant and antivirulence properties. J Med Chem 56, 1405–1417.

Lin, Z., Xu, Z., Li, Y., Wang, Z., Chen, T., and Zhao, X. (2014). Metabolic engineering of *Escherichia coli* for the production of riboflavin. Microb Cell Fact 13, 104.

Lu, Q., Yao, Q., Xu, Y., Li, L., Li, S., Liu, Y., Gao, W., Niu, M., Sharon, M., Ben-Nissan, G., et al. (2014). An iron-containing dodecameric heptosyltransferase family modifies bacterial autotransporters in pathogenesis. Cell Host Microbe 16, 351–363.

McCallum, M., Shaw, G.S., and Creuzenet, C. (2013). Comparison of predicted epimerases and reductases of the *Campylobacter jejuni* D-

altro- and L-*gluco*-heptose synthesis pathways. J Biol Chem 288, 19569–19580.

Morrison, J.P., and Tanner, M.E. (2007). A two-base mechanism for *Escherichia coli* ADP-L-*glycero*-D-*manno*-heptose 6-epimerase. Biochemistry 46, 3916–3924.

Park, J., Kim, H., Kim, S., Lee, D., Kim, M.S., and Shin, D.H. (2018). Crystal structure of D-*glycero*-*β*-D-*manno*-heptose-L-phosphate adenylyltransferase from *Burkholderia pseudomallei*. Proteins 86, 124–131.

Raetz, C.R.H., and Whitfield, C. (2002). Lipopolysaccharide endotoxins. Annu Rev Biochem 71, 635–700.

Tang, W., Guo, Z., Cao, Z., Wang, M., Li, P., Meng, X., Zhao, X., Xie, Z., Wang, W., Zhou, A., et al. (2018). D-Sedoheptulose-7-phosphate is a common precursor for the heptoses of septacidin and hygromycin B. Proc Natl Acad Sci USA 115, 2818–2823.

Tang, W., Li, P., Chen, M., Guo, Z., and Chen, Y. (2020). Characterization of SepE and SepF for the N^6-glycosylated adenine structure formation in septacidin biosynthesis. Org Lett 22, 5251–5254.

Tang, Y., Wang, M., Qin, H., An, X., Guo, Z., Zhu, G., Zhang, L., and Chen, Y. (2020). Deciphering the biosynthesis of TDP-β-L-oleandrose in avermectin. J Nat Prod 83, 3199–3206.

Taylor, P.L., Blakely, K.M., de Leon, G.P., Walker, J.R., McArthur, F., Evdokimova, E., Zhang, K., Valvano, M.A., Wright, G.D., and Junop, M.S. (2008). Structure and function of sedoheptulose-7-phosphate isomerase, a critical enzyme for lipopolysaccharide biosynthesis and a target for antibiotic adjuvants. J Biol Chem 283, 2835–2845.

Thibodeaux, C.J., Melançon, C.E., and Liu, H. (2007). Unusual sugar biosynthesis and natural product glycodiversification. Nature 446, 1008–1016.

Thibodeaux, C. ., Melançon Iii, C. ., and Liu, H. (2008). Natural-product sugar biosynthesis and enzymatic glycodiversification. Angew Chem Int Ed 47, 9814–9859.

Valvano, M.A., Messner, P., and Kosma, P. (2002). Novel pathways for biosynthesis of nucleotide-activated *glycero-manno*-heptose precursors of bacterial glycoproteins and cell surface polysaccharides. Microbiology 148, 1979–1989.

Wang, L., Huang, H., Nguyen, H.H., Allen, K.N., Mariano, P.S., and Dunaway-Mariano, D. (2010). Divergence of biochemical function in the HAD superfamily: D-*glycero*-D-*manno*-heptose-1,7-bisphosphate phosphatase (GmhB). Biochemistry 49, 1072–1081.

Yang, Y.L., Zhang, S., Ma, K., Xu, Y., Tao, Q., Chen, Y., Chen, J., Guo, S., Ren, J., Wang, W., et al. (2017). Discovery and characterization of a new family of diterpene cyclases in bacteria and fungi. Angew Chem Int Ed 56, 4749–4752.

Zhang, J., Fan, P.H., Lin, G.M., Chang, W.C., and Liu, H.W. (2020). Recent progress in unusual carbohydrate-containing natural products bio-synthesis. Comprehensive natural products III, pp. 336–392.

Zhou, P., She, Y., Dong, N., Li, P., He, H., Borio, A., Wu, Q., Lu, S., Ding, X., Cao, Y., et al. (2018). Alpha-kinase 1 is a cytosolic innate immune receptor for bacterial ADP-heptose. Nature 561, 122–126.

SUPPORTING INFORMATION

The supporting information is available online at https://10.1007/s11427-021-2000-2. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.