Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

Insights into the binding specificity and catalytic mechanism of *N*-acetylhexosamine 1-phosphate kinases through multiple reaction complexes

Kuei-Chen Wang,[a,b,c] Syue-Yi Lyu,[a] Yu-Chen Liu,[a] Chin-Yuan Chang,[a] Chang-Jer Wu[d] and Tsung-Lin Li[a,b,e]*

[a]Genomics Research Center, Academia Sinica, Taipei 115, Taiwan, [b]Chemical Biology and Molecular Biophysics Program, Taiwan International Graduate Program, Academia Sinica, Taipei 115, Taiwan, [c]Institute of Biochemical Sciences, National Taiwan University, Taipei 115, Taiwan, [d]Department of Food Science, National Taiwan Ocean University, Keelung 202, Taiwan, and [e]Biotechnology Center, National Chung Hsing University, Taichung 402, Taiwan

Correspondence e-mail: tlli@gate.sinica.edu.tw

Received 24 January 2014
Accepted 23 February 2014

PDB references: NahK complexes, 4ocj; 4ock; 4oco; 4ocp; 4ocq; 4ocu; 4ocv

Utilization of *N*-acetylhexosamine in bifidobacteria requires the specific lacto-*N*-biose/galacto-*N*-biose pathway, a pathway differing from the Leloir pathway while establishing symbiosis between humans and bifidobacteria. The gene *lnpB* in the pathway encodes a novel hexosamine kinase NahK, which catalyzes the formation of *N*-acetylhexosamine 1-phosphate (GlcNAc-1P/GalNAc-1P). In this report, seven three-dimensional structures of NahK in complex with GlcNAc, GalNAc, GlcNAc-1P, GlcNAc/AMPPNP and GlcNAc-1P/ADP from both *Bifidobacterium longum* (JCM1217) and *B. infantis* (ATCC15697) were solved at resolutions of 1.5–2.2 Å. NahK is a monomer in solution, and its polypeptide folds in a crescent-like architecture subdivided into two domains by a deep cleft. The NahK structures presented here represent the first multiple reaction complexes of the enzyme. This structural information reveals the molecular basis for the recognition of the given substrates and products, GlcNAc/GalNAc, GlcNAc-1P/GalNAc-1P, ATP/ADP and Mg^{2+}, and provides insights into the catalytic mechanism, enabling NahK and mutants thereof to form a choice of biocatalysts for enzymatic and chemoenzymatic synthesis of carbohydrates.

1. Introduction

The sugars *N*-acetyl-D-glucosamine (GlcNAc) and *N*-acetyl-D-galactosamine (GalNAc) are common primary metabolites that are produced in all living beings (van Heijenoort, 2001; Hart *et al.*, 2007; Lindahl & Höök, 1978). The intestine-colonizing bifidobacteria are considered to be beneficial to human health, but acquire these metabolites from extrinsic sources (Kitaoka, 2012; Fushinobu, 2010). Bifidobacteria feature a unique lacto-*N*-biose (LNB)/galacto-*N*-biose (GNB) pathway to utilize LNB and GNB from human milk oligosaccharides (HMO) and intestinal mucins (Nishimoto & Kitaoka, 2007). The LNB/GNB pathway is encoded by a four-gene *lnpABCD* operon and includes galacto-*N*-biose/lacto-*N*-biose phosphorylase (GLNBP), *N*-acetylhexosamine 1-kinase (NahK), UDP-glucose hexose 1-phosphate uridylyltransferase ($GalT_{BL}$) and UDP-glucose 4-epimerase ($GalE_{BL}$), respectively, as illustrated in Fig. 1 (Nishimoto & Kitaoka, 2007; Kitaoka *et al.*, 2005). In this report, we focus on the structural complexes of *N*-acetylhexosamine 1-kinase (NahK, encoded by the gene *lnpB*), which catalyses the phosphorylation of *N*-acetyl-D-hexosamine. The product GlcNAc-1-phosphate (GlcNAc-1P) or GalNAc-1-phosphate (GalNAc-1P) then undergoes uridylation to generate UDP-*N*-acetyl-hexosamine at the expense of one equivalent of UDP-glucose (UDP-Glc) for subsequent metabolism (Kitaoka, 2012; Nishimoto & Kitaoka, 2007).

© 2014 International Union of Crystallography

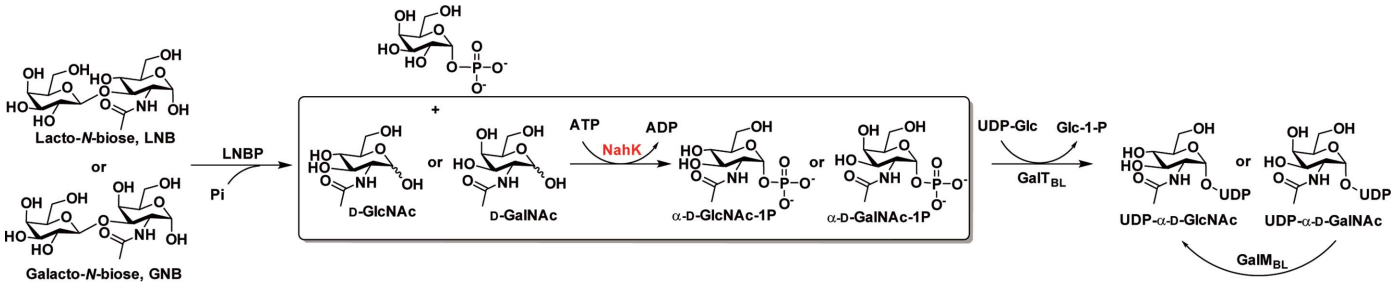


Figure 1
The LNB/GNB pathway in bifidobacteria, in which NahK catalyzes the formation of *N*-acetylhexosamine 1-phosphate (GlcNAc-1P/GalNAc-1P) at the expense of one equivalent of ATP.

The amino-acid sequence of NahK shares only 12% identity with that of choline kinase, a member of the sugar kinase–Hsp70–actin superfamily (Hurley, 1996; Supplementary Fig. S1[1]). All members of the superfamily consist of a characteristic five-stranded β-sheet in their N- and C-terminal domains (Hurley, 1996). The ATP-binding site is often situated in the cleft between the two domains, and the adenosyl moiety of ATP points towards the N-terminal domain. Despite considerable detailed knowledge of the sugar kinase–Hsp70–actin superfamily, correlation of the enzyme structures with their substrate selectivity has been difficult largely owing to the lack of three-dimensional structures of their complexes. This lack of structures is responsible for the difficulty not only in deciphering the regioselectivity and stereoselectivity of the enzymes but also in gaining insights into their catalytic mechanisms at the molecular level. Close members of the sugar kinase family that have had their tertiary structures elucidated include galactokinase (GalK) from human (Thoden *et al.*, 2005) and L-rhamnulose kinase (RhuK) from *Escherichia coli* (Grueninger & Schulz, 2006). Although homology analysis helps to outline general structural topology, the marginal sequence identity between these genes (<10% identity for NahK *versus* GalK or RhuK) does not allow a dependable structure to be predicted, let alone the details of substrate recognition.

Here, we present seven three-dimensional structures of NahK including four binary and three ternary structures in complex with GlcNAc, GalNAc, GlcNAc-1P, GlcNAc/AMPPNP (a nonhydrolyzable ATP analogue) and GlcNAc-1P/ADP. These snapshot structures are contributory to the delineation of the full course of catalytic events during phosphoryl transfer by NahK. Because of an observed configurational inversion, a direct inline phosphoryl-transfer mechanism was proposed. Two patches in the sugar-binding cavity with characteristic hydrophobic and hydrophilic attributions were identified, underscoring the substrate specificity of the enzyme. Given that NahK has been utilized as a biocatalyst in carbohydrate synthesis (Zhao *et al.*, 2010; Cai, Guan, Kitaoka *et al.*, 2009; Li *et al.*, 2011; Cai, Guan, Wang *et al.*, 2009), we discuss the implications of these protein structures in protein engineering, in view of the regioselectivity, stereoselectivity and catalytic mechanism of the enzyme. The future design of sugar-1-phosphate kinases as an alternative approach for carbohydrate-related synthesis will be greatly facilitated by these findings.

2. Materials and methods

2.1. Protein expression and purification

The *lnpB* gene was amplified from the genomic DNA of *Bifidobacterium longum* (JCM1217) or *B. infantis* (ATCC15697) and subcloned into the pET-28a(+) expression vector (Novagen). Each clone was transformed into *Escherichia coli* BL21 (DE3) cells for protein expression. Each transformant was grown in LB medium containing 50 µg ml^{-1} kanamycin overnight at 37°C. The overnight culture was inoculated with 50 times the volume of LB medium and grown at 37°C until the OD_{600} reached 0.7. Protein expression was then induced by the addition of isopropyl-β-D-1-thiogalactoside (IPTG) to a final concentration of 0.1 m*M*, followed by incubation for a further 16 h at 16°C. The cells were harvested by centrifugation at 6500 rev min^{-1} for 20 min at 4°C, resuspended in a binding buffer solution [50 m*M* Tris pH 8.0, 500 m*M* NaCl, 10%(*v*/*v*) glycerol, 10 m*M* imidazole] and disrupted using a microfluidizer. All steps were carried out on ice. The cell debris was removed by centrifugation at 20 000 rev min^{-1} for 30 min. The supernatant was loaded into a 2 ml Ni^{+}–NTA agarose resin column (Qiagen) pre-equilibrated with the binding buffer solution. The column was sequentially washed with binding buffer and wash buffer solutions [50 m*M* Tris pH 8.0, 500 m*M* NaCl, 10%(*v*/*v*) glycerol, 50 m*M* imidazole]. The desired protein was then eluted with elution buffer solution [50 m*M* Tris pH 8.0, 500 m*M* NaCl, 10%(*v*/*v*) glycerol, 90 m*M* imidazole]. Gel filtration was performed using an ÄKTA FPLC system equipped with a HiLoad Superdex 200 column (Amersham Bioscience) under isocratic conditions (50 m*M* HEPES pH 7.5) at a flow rate of 1 ml min^{-1}. Protein purity was examined by SDS–PAGE. Protein concentrations were determined using the Bradford assay.

2.2. Crystallization and data collection

NahK complexes were crystallized using the hanging-drop vapour-diffusion method at 20°C. The concentration of NahK

[1] Supporting information has been deposited in the IUCr electronic archive (Reference: MN5056).

Table 1
Data-collection, phasing and refinement statistics.

Values in parentheses are for the highest resolution shell.

	SeMet NahK–GlcNAc								
	Peak	Inflection	Remote	NahK–GalNAc	NahK–GlcNAc-1P	NahK–GlcNAc–AMPPNP	NahK–GlcNAc-1P–ADP	$NahK^{15697}$–GlcNAc†	$NahK^{15697}$–GlcNAc–AMPPNP
PDB code	4ocj			4ocq	4oco	4ock	4ocp	4ocu	4ocv
Data collection									
Space group	$P2_12_12_1$			$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P4_12_12$	$P4_12_12$
Unit-cell parameters									
a (Å)	55.4			55.2	55.6	54.8	55.1	73.5	73.2
b (Å)	79.5			79.6	78.9	79.0	79.2	73.5	73.2
c (Å)	98.1			98.4	97.4	98.0	97.8	166.3	166.6
$\alpha = \beta = \gamma$ (°)	90.0			90.0	90.0	90.0	90.0	90.0	90.0
Wavelength	0.97874	0.97893	0.96353						
Resolution (Å)	30.00–1.60 (1.66–1.60)	30.00–1.60 (1.66–1.60)	30.00–1.60 (1.66–1.60)	30.00–1.90 (1.97–1.90)	30.00–2.15 (2.23–2.15)	30.00–1.72 (1.78–1.72)	30.00–1.94 (2.01–1.94)	30.00–1.90 (1.97–1.90)	30.00–1.47 (1.52–1.47)
R_{merge}	0.086 (0.447)	0.076 (0.483)	0.066 (0.471)	0.081 (0.447)	0.090 (0.478)	0.073 (0.492)	0.118 (0.489)	0.086 (0.449)	0.069 (0.436)
$\langle I/\sigma(I)\rangle$	130.12 (6.00)	26.62 (4.02)	29.64 (3.98)	19.30 (4.06)	21.39 (5.03)	26.69 (3.78)	14.70 (4.74)	36.43 (7.89)	33.75 (7.13)
Completeness (%)	99.9 (99.9)	99.9 (99.9)	99.9 (99.8)	99.6 (99.8)	99.9 (99.9)	100.0 (100.0)	98.5 (100.0)	100.0 (100.0)	99.1 (98.4)
Multiplicity	13.4 (12.8)	6.6 (6.4)	6.5 (6.3)	6.1 (5.8)	7.1 (7.2)	7.1 (6.9)	6.0 (6.4)	14.2 (14.6)	12.1 (12.4)
Refinement									
Resolution (Å)	1.60			1.90	2.15	1.72	1.94	1.90	1.47
No. of reflections	56696			33551	22449	43636	30364	34696	73058
R_{cryst}/R_{free}	0.16/0.18			0.19/0.22	0.18/0.23	0.18/0.20	0.18/0.22	0.18/0.20	0.19/0.20
No. of atoms									
Protein	2776			2775	2776	2776	2776	2810	2800
Ligand/ion	15/—			14/—	19/—	46/—	46/1	14/24	45/19
Water	533			282	192	354	284	305	449
B factors (Å^2)									
Protein	19.76			25.69	29.30	15.72	22.91	20.82	13.85
Ligand/ion	15.08/—			20.79/—	24.06/—	14.90/—	19.71/24.24	16.06/51.64	25.07/39.06
Water	32.88			30.74	31.85	24.78	27.56	29.95	26.83
R.m.s deviations									
Bond lengths (Å)	0.008			0.007	0.011	0.009	0.007	0.011	0.006
Bond angles (°)	1.256			1.035	1.426	1.312	1.165	1.293	1.225

† $NahK^{15697}$: NahK from *B. infantis* (ATCC15697)

(JCM1217) was 20 mg ml^{-1} in 50 m*M* HEPES pH 7.5. Crystals of the NahK–GlcNAc, NahK–GalNAc and NahK–GlcNAc-1P complexes were obtained by pre-incubation of SeMet-substituted/native NahK with GlcNAc (20-fold excess), GalNAc (20-fold excess) or GlcNAc-1P (tenfold excess), and the protein solutions were then mixed with a crystallization buffer solution consisting of 20%(*w*/*v*) PEG 3350, 0.1 *M* bis-tris pH 5.5. Crystals were obtained within 10–14 d. The crystals of NahK–GlcNAc–AMPPNP were obtained by soaking NahK–GlcNAc crystals with 10 m*M* AMPPNP and 20 m*M* $MgCl_2$ for 15 min. Similarly, the crystals of NahK–GlcNAc-1P–ADP were obtained by soaking NahK–GlcNAc-1P crystals with 5 m*M* ADP and 10 m*M* $MgCl_2$ for 30 min. Likewise, the crystals of the $NahK^{ATCC15697}$–GlcNAc complex were obtained in a crystallization buffer solution consisting of 1.5 *M* Li_2SO_4, 0.1 *M* sodium acetate pH 5.5 within 2 d. The crystals of the $NahK^{ATCC15697}$–GlcNAc–AMPPNP complex were obtained by soaking $NahK^{ATCC15697}$–GlcNAc complex crystals with 5 m*M* AMPPNP and 10 m*M* $MgCl_2$ for 1.5 h. The crystals were transferred to cryoprotectant solution and flash-cooled in liquid nitrogen prior to data collection. The cryoprotectant solution for NahK (JCM1217) consisted of 35%(*w*/*v*) PEG 3350, 0.1 *M* bis-tris pH 5.5, while the cryoprotectant solution for NahK (ATCC15697) consisted of 20%(*v*/*v*) ethylene glycol dissolved in the crystallization buffer solution. X-ray diffraction data sets were collected using an ADSC Quantum-315 or Quantum-210 CCD detector on beamlines 13B1 and 13C1 of the National Synchrotron Radiation Research Center, Taiwan and beamlines 12B2 and 44XU of Spring-8, Japan. Data were indexed and scaled with the *HKL*-2000 package (Otwinowski & Minor, 1997).

2.3. Structure determination and refinement

The multiple-wavelength anomalous diffraction (MAD) method was used to obtain phase information, and *CRANK* (Ness *et al.*, 2004) was used to find the phase solution. The *CRANK* pipeline started with substructure detection and ended with model building, including procedures of substructure detection by *AFRO*/*CRUNCH*2 (de Graaff *et al.*, 2001), substructure refinement by *BP*3 (Pannu & Read, 2004), hand determination and density modification by *SOLOMON* (Abrahams & Leslie, 1996) and model building by *Buccaneer* (Cowtan, 2006, 2008). Phase extension yielded electron-density maps where a polypeptide model was built with *Coot* (Emsley *et al.*, 2010; Emsley & Cowtan, 2004). The model was further refined with *REFMAC* (Murshudov *et al.*, 2011). Other native complexes were solved by the molecular-replacement

(MR) method using SeMet-substituted NahK as the model. Figures were generated using *PyMOL* (http://www.pymol.org). Detailed data-collection and refinement statistics are given in Table 1.

2.4. Site-directed mutagenesis

Site-directed mutagenesis was performed using QuikChange (Stratagene). The wild-type NahK (JCM1217) was used as the template for the single mutation. For the double mutation, the single mutant was used as the template. All mutations were confirmed by DNA sequencing. NahK mutants were expressed and purified with the same protocol as used for the wild-type protein.

2.5. Enzymatic activity assay

Typical enzymatic reactions were carried out in a 100 µl reaction mixture consisting of 100 m*M* Tris buffer pH 8.0, 2 m*M* of the tested substrate, 2.4 m*M* ATP, 1 m*M* $MgCl_2$ and 0.1 mg NahK or the tested NahK mutant. The reactions were incubated at 37°C for 4 h and stopped by boiling the mixtures for 10 min, followed by centrifugation at 14 500 rev min^{-1} for 20 min. The supernatants were analyzed for the production of sugar-1P by high-performance anion-exchange chromatography coupled with a pulsed amperometric detection system (HPAEC-PAD, Dionex ICS-3000 ion-chromatography system) and installed with a Carbopac PA-1 analytical column (4 × 250 mm) at a flow rate of 1 ml min^{-1}. The mobile phase consisted of *A* (100 m*M* NaOH) and *B* (1 *M* sodium acetate with 100 m*M* NaOH). A two-stage elution gradient was programmed as 97% *A* for 10 min, a linear gradient to 90% *A* in 5 min followed by 90% *A* for 10 min, a second linear gradient to 50% *A* in 10 min and then back to 100% *A* in 2 min followed by 100% *A* for a further 8 min.

Assays for synthesis of UDP-sugars were performed in a 200 µl reaction mixture consisting of 0.2 mg NahK, 0.04 mg GlmU (*E. coli* K12), 5 m*M* mono-sugar, 6 m*M* ATP, 6 m*M* UTP, 1 U pyrophosphatase, 10 m*M* $MgCl_2$ and 100 m*M* Tris pH 8.0 at 37°C overnight. Reactions were quenched by boiling the mixtures for 10 min, followed by centrifugation at 14 500 rev min^{-1} for 20 min. The supernatants were subjected to ion-exchange chromatography using an Agilent 1100 series HPLC equipped with a SOURCE 15Q 4.6/100 PE column (GE Healthcare) at a flow rate of 1 ml min^{-1}. The UV wavelength was set at 280 nm. A linear gradient of 0.3–80% 1 *M* ammonium formate against H_2O for 30 min was performed.

2.6. Isothermal titration calorimetry (ITC) analysis

ITC experiments (Huang *et al.*, 2009) were carried out at 25°C using an iTC200 microcalorimeter (MicroCal, Northampton, Massachusetts, USA). The buffer solution for ITC analysis was 50 m*M* HEPES pH 7.5. Proteins were prepared as above. Each titration contained 0.1 m*M* protein in the sample cell (250 µl) and 1 or 2 m*M* testing substrate loaded in an injection syringe (40 µl). A typical titration experiment consisted of 18 injections of 2 µl over a duration of 4 s; each injection was separated by 120 s. For ATP, the titration consisted of 30 injections of 1.2 µl at 120 s intervals. The reference cell was filled with H_2O throughout all experiments. The cell stirring speed was 1000 rev min^{-1}. Raw data were collected and fitted by a nonlinear least-squares method to a single-site binding model using the *Origin* 7.0 software provided by MicroCal.

2.7. Analytical ultracentrifugation analysis

Sedimentation-velocity experiments were performed with a Beckman-Coulter XL-I analytical ultracentrifuge. Samples and buffers (50 m*M* HEPES pH7.5) were loaded in 12 mm standard double-sector Epon charcoal-filled centrepieces and mounted in an An-60 Ti rotor. Protein samples (1 mg ml^{-1}, 400 µl each) were loaded into the cell. Sedimentation-velocity experiments were performed at a rotor speed of 42 000 rev min^{-1} at 20°C. The signals of samples were monitored at 280 nm and collected every 3 min for 6 h. The raw data of experiments were analyzed using the *SedFit* software (http://www.analyticalultracentrifugation.com).

3. Results and discussion

3.1. Structure determination

The *lnpB* gene encoding NahK of 359 amino acids was amplified from the genomic DNA of *B. longum* (JCM1217) or *B. infantis* (ATCC15697) (Sela *et al.*, 2008; Nishimoto & Kitaoka, 2007). Each recombinant NahK was expressed in *E. coli* BL21 (DE3) cells as an N-terminally His_6-tagged protein and was purified sequentially by Ni^{2+}-affinity and gel-filtration chromatography (Supplementary Fig. S2*b*). NahK was first co-crystallized in the presence of GlcNAc. Although NahK is structurally related to the sugar kinase–Hsp70–actin superfamily, the phase problem could not be solved using the molecular-replacement (MR) method. Selenomethionine-labelled NahK was then crystallized, a crystal of which diffracted to 1.6 Å resolution in space group $P2_12_12_1$. The Se peak wavelength was used to obtain the initial phase by multiple-wavelength anomalous dispersion (MAD). The Se anomalous peaks were used to build the initial model (*CRANK*; Ness *et al.*, 2004). The R_{cryst} and R_{free} were 16.1 and 18.3%, respectively. The final model displays correct stereochemistry, with 99.5% of the residues in the allowed region of the Ramachandran plot, and served as a search model for the other ligand-complexed data sets using the MR routine. We obtained an additional six structures of NahK with GlcNAc, GalNAc and GlcNAc-1P in binary complexes and GlcNAc/AMPPNP and GlcNAc-1P/ADP in ternary complexes, although crystals of the apo form remained unobtainable. The resolutions of these complexes were in the range 1.5–2.2 Å. Data-collection and refinement statistics are summarized in Table 1. Since the NahKs from both *B. longum* (JCM1217) and *B. infantis* (ATCC15697) are homologues with high sequence identity (91.53%) and structural similarity [root-mean-square deviation (r.m.s.d.) of 0.37 Å over 352 C^{α} atoms; Supple-

mentary Fig. S4], the NahK referred to hereafter is that from *B. longum* (JCM1217) unless otherwise specified

3.2. Overall structure

NahK is a monomer in solution on the basis of gel-filtration and analytical ultracentrifugation analyses (molecular mass 42.5 kDa; Supplementary Fig. S2). Consistently, each asymmetric unit contains one NahK molecule, suggesting that the monomer is the most biologically relevant state. Like other structures in the sugar kinase–Hsp70–actin superfamily, NahK has a crescent-like architecture subdivided into two domains by a deep cleft (Fig. 2 and Supplementary Fig. S3). The N- and C-terminal domains of NahK comprise residues 1–103 and 117–359, respectively. The N-terminal domain contains a characteristic five-stranded antiparallel β-sheet (β1–β5) flanked by two α-helices (α1–α2). A large helix α2 (residues 58–75) is inserted between β3 and β4. The C-terminal domain (residues 117–359) is composed of two subdomains. The core subdomain consists of three helices (α3, α7 and α8) along with a Greek-key motif (β6–β9), in which β6 and β7 run antiparallel to β9 and β8, respectively. The helical subdomain contains five helices: helix α4 to helix α6 and a helix–turn–helix motif (α9–α10). The Brenner motif ($H^{206}XDXXXXN^{213}$; Brenner, 1987), a characteristic motif in phosphoryl transferases, is located in loop β6–β7.

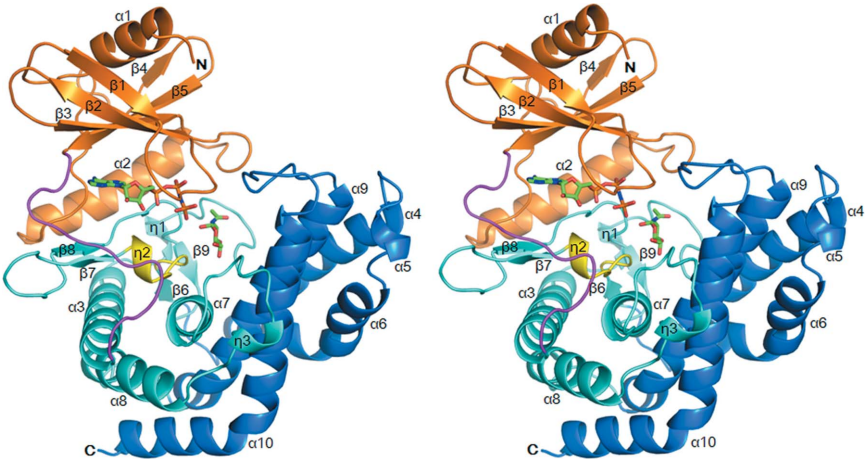


Figure 2
Stereoview of the ternary NahK–GlcNAc–AMPPNP complex. The N-terminal domain is coloured orange, the linker between the N-terminal and C-terminal domains is coloured magenta, the core subdomain of the C-terminal region is coloured cyan and the helical subdomain is coloured marine blue. The Brenner motif is coloured yellow. GlcNAc and AMPPNP are shown as green sticks.

3.3. Structural comparison

Structural analysis using the *DALI* server (Holm & Sander, 1995) revealed that NahK is a structural homologue of human choline kinase (hCK; PDB entry 2ckq; Malito *et al.*, 2006) and aminoglycoside 2″-phosphotransferase IIIa [APH(2″)-IIIa] of *Entercoccus gallinarum* (PDB entry 3tdv; Smith *et al.*, 2012), although both hCK and APH(2″)-IIIa share low sequence identity with NahK (∼12%). Both hCK and APH(2″)-IIIa are members of the 'atypical kinase' enzyme family which have a common catalytic core that is dressed with varied conserved motifs (Scheeff & Bourne, 2005). Superimposition of the structure of NahK with that of hCK or APH(2″)-IIIa shows root-mean-square deviations (r.m.s.d.) of 4.70 and 4.45 Å for 304 and 280 C^{α} atoms, respectively, indicating that NahK preserves the general atypical kinase fold. Despite such high overall structural similarities, the topologies of the substrate-binding sites differ considerably from one another. The most noticeable difference lies on the segment that bridges the N-terminal and C-terminal

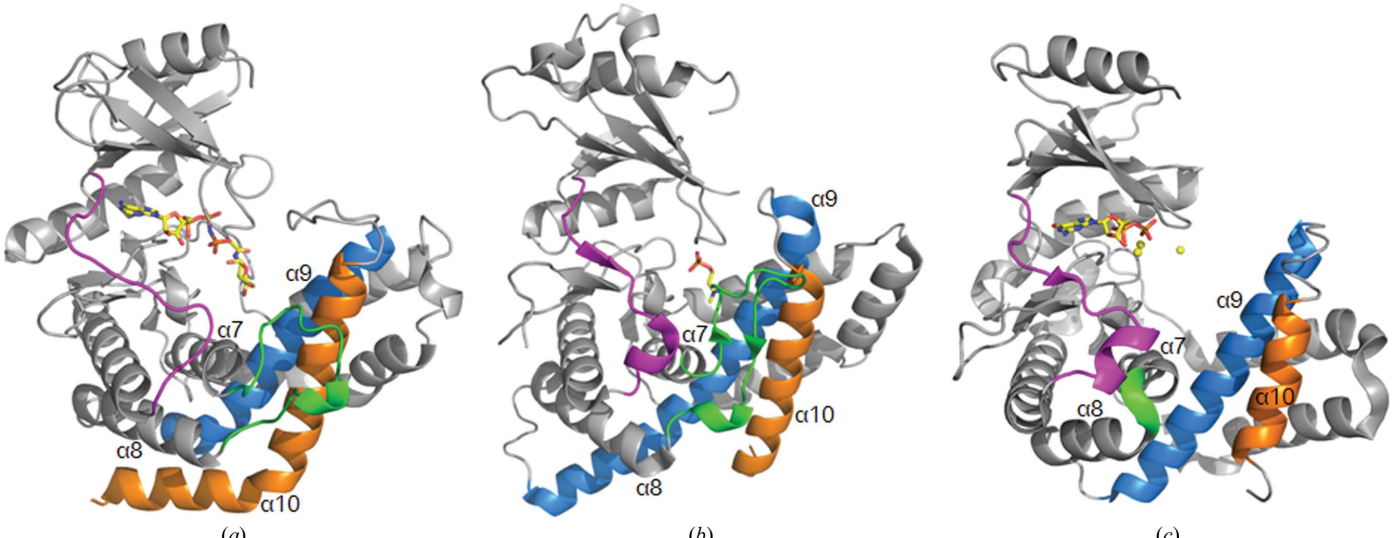
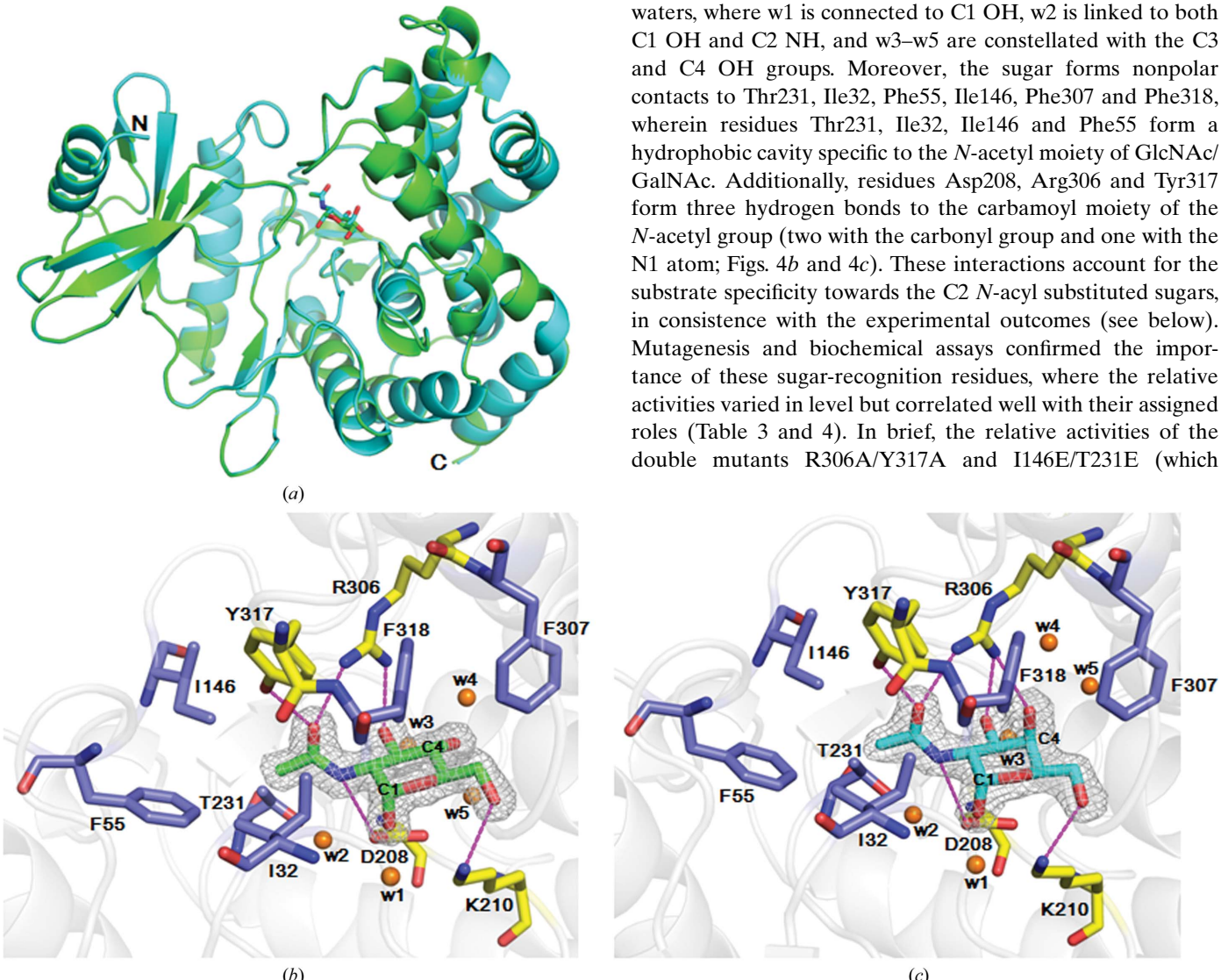


Figure 3
Structural comparison of NahK, hCK and APH(2″)-IIIa. (*a*) NahK; (*b*) hCK; (*c*) APH(2″)IIIa. The major structural differences among these three enzymes are highlighted in magenta, green, marine and orange; the conserved structures are shown in grey.

domains. It is a long loop in NahK, while it is structured as a short β-strand and an α-helix in both hCK and APH(2″)-IIIa (magenta in Fig. 3). The regions between α7 and α8 are also quite dissimilar from one another (green in Fig. 3). Helices α9 and α10 of NahK are unexpectedly a short and a long helix, respectively, when compared with those of hCK and APH(2″)-IIIa (orange and marine in Fig. 3). These differences reflect the considerably diverse substrate selectivity in the atypical kinases.

3.4. Sugar-binding site

The NahK–GlcNAc and NahK–GalNAc complexes are virtually identical (r.m.s.d. of 0.10 Å for 352 C^{α} atoms), suggesting that the binding of different sugars does not provoke a discernible conformational change (Fig. 4*a*). In the complexes, GlcNAc/GalNAc was positioned and identified using ($F_o - F_c$) difference electron-density maps (Figs. 4*b* and 4*c*). The density is well defined, showing clearly the conformation and stereochemistry of these sugars. Since the average *B* factors of the substrates matched that of the surrounding residues, the site was considered fully occupied, suggesting that the binding is specific for α-configuration sugars. GlcNAc/GalNAc resides in a cavity moulded by the core subdomain and the helical subdomain in the C-terminal domain (Fig. 2). When compared with GlcNAc, GalNAc finds sufficient space for the C4 hydroxyl group that interacts with the terminal guanidino group of Arg306 through an extra hydrogen bond (Fig. 4*c*). GalNAc serves as an alternative substrate for NahK with an equivalent conversion rate and binding affinity to GlcNAc (Tables 2 and 3). Both sugars form hydrogen bonds to the side chains of Asp208, Lys210, Arg306 and Tyr317 (Figs. 4*b* and 4*c*). Five water molecules w1–w5 are likely to be bound waters, where w1 is connected to C1 OH, w2 is linked to both C1 OH and C2 NH, and w3–w5 are constellated with the C3 and C4 OH groups. Moreover, the sugar forms nonpolar contacts to Thr231, Ile32, Phe55, Ile146, Phe307 and Phe318, wherein residues Thr231, Ile32, Ile146 and Phe55 form a hydrophobic cavity specific to the *N*-acetyl moiety of GlcNAc/GalNAc. Additionally, residues Asp208, Arg306 and Tyr317 form three hydrogen bonds to the carbamoyl moiety of the *N*-acetyl group (two with the carbonyl group and one with the N1 atom; Figs. 4*b* and 4*c*). These interactions account for the substrate specificity towards the C2 *N*-acyl substituted sugars, in consistence with the experimental outcomes (see below). Mutagenesis and biochemical assays confirmed the importance of these sugar-recognition residues, where the relative activities varied in level but correlated well with their assigned roles (Table 3 and 4). In brief, the relative activities of the double mutants R306A/Y317A and I146E/T231E (which



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

Figure 4
Sugar-binding site of NahK. (*a*) Superimposition of the NahK–GlcNAc (green) and NahK–GalNAc complexes (r.m.s.d. of 0.1 Å over 352 C^{α} atoms). NahK and GlcNAc/GalNAc are shown as ribbon diagrams and stick models, respectively. (*b*, *c*) The GlcNAc/GalNAc-complexed sugar-binding site. GlcNAc and GalNAc are shown as green and cyan stick models, respectively. The $2F_o - F_c$ electron-density map for GlcNAc/GalNAc is carved at 1.5σ. The residues that interact with the sugar through hydrogen bonds and hydrophobic forces are highlighted as yellow and purple sticks, respectively. Water molecules are shown as orange balls. Broken lines represent hydrogen bonds.

Table 2
Summary of ITC analysis of NahK and mutants.

ND, not detected.

	n	K_d (μM)	ΔH (cal mol^{-1})	ΔS (cal mol K^{-1})
WT				
GlcNAc	1.1 ± 0.0	20.4 ± 1.5	−4019 ± 106.0	8.0
GalNAc	1.0 ± 0.0	24.7 ± 1.5	−6805 ± 177.4	−1.7
GlcNAc-1P	ND	ND	ND	ND
ATP	0.9 ± 0.0	0.4 ± 0.0	−19580 ± 128.9	−36.5
AMPPNP	1.2 ± 0.0	8.3 ± 0.5	−7349 ± 185.0	−1.4
ADP	0.9 ± 0.0	0.8 ± 0.1	−10600 ± 75.0	−7.9
AMP	0.8 ± 0.1	165.0 ± 14.5	−10170 ± 1493.0	−17.0
Mannose	ND	ND	ND	ND
Glucosamine	ND	ND	ND	ND
Galactosamine	ND	ND	ND	ND
Mannosamine	ND	ND	ND	ND
Glucose	ND	ND	ND	ND
Galctose	ND	ND	ND	ND
ManNAc	ND	ND	ND	ND
R306A				
GlcNAc	1.0 ± 0.1	22.4 ± 5.9	−1479 ± 130.5	16.5
GalNAc	ND	ND	ND	ND
Y317A				
GlcNAc	1.1 ± 0.1	65.2 ± 9.8	−3260 ± 192.1	8.3
GalNAc	1.1 ± 0.0	36.5 ± 6.1	−2362 ± 107.0	12.4
R306A/Y317A				
GlcNAc	ND	ND	ND	ND
GalNAc	ND	ND	ND	ND

disrupt hydrogen bonds and hydrophobic interactions, respectively) *versus* GlcNAc/GalNAc fell to 18.8/17.6% and 25.0/12.5% relative to that of the wild type, respectively. Analogously, the binding affinity for the double mutant R306A/Y317A *versus* GlcNAc/GalNAc was too weak to measure.

3.5. Nucleotide-binding site

The holo NahK complex contained AMPPNP and GlcNAc, as observed in the ($F_o - F_c$) difference electron-density maps (Fig. 5*a*). The density is well defined, showing clearly the conformations of both AMPPNP and GlcNAc. The bound AMPPNP is located in a slot between the N-terminal and C-terminal domains (Fig. 2). The adenine moiety adopts an *anti* conformation and binds in a nonpolar crevice lined by Leu80, Ile105, Leu215 and Ile227 on one side and Ile46, Tyr102 and Phe104 on the other (Fig. 5*b*). The N6 atom on the adenine ring makes a hydrogen bond to the backbone carbonyl O atom of Lys103. There are no other direct hydrogen bonds between the adenine ring and protein side chains. The ribose moiety assumes a C2 *exo*–C3 *endo* conformation in association with two water molecules through hydrogen bonds (Fig. 5*b*). The triphosphoryl moiety interacts with several polar side chains of surrounding residues *via* a constellation of water-mediated hydrogen bonds: the α-phosphoryl group interacts with Gln48, equivalent to Lys52 in APH(9)-Ia, which is highly conserved in all APH and ePKs for binding and orienting the phosphoryl group (Burk *et al.*, 2001; Fong *et al.*, 2010; Thompson *et al.*, 1998); the β-phosphoryl group interacts with Ile32 and Asn33, of which the latter is equivalent to Asn122 in hCK and is also highly conserved in the protein families of CKs and ePKs (Malito *et al.*, 2006); and the γ-phosphoryl group interacts with the backbone amide group of His31 (Fig. 5*b*). An anticipated divalent ion (Mg^{2+}) does not exist in the holo complex; however, this enzyme is indispensable, as no enzyme reaction took place when Mg^{2+} was excluded from the reaction solutions. In contrast, the electron density shows a bulge in the NahK–GlcNAc-1P–ADP complex, which was identified to be a divalent cation chelating to Asn213, Asp228 and the α- and β-phosphoryl groups of ADP (see below).

Table 3
Substrate specificity of NahK.

ND, not detected.

Substrate	Relative activity (%)
GlcNAc	100.0
GalNAc	102.1
ManNAc	4.3
Glucosamine	12.9
Galactosamine	8.6
Mannosamine	10.0
Glucose	5.7
Galactose	8.6
Mannose	21.4
Glucuronic acid	ND
Galacturonic acid	ND
6-Azidoglucose	ND
6-GlcNAc	ND
3-Glucosamine	ND
4-Glucosamine	ND
6-Glucosamine	ND

Table 4
Relative enzymatic activities of NahK and mutants towards GlcNAc/GalNAc.

Mutant	GlcNAc	GalNAc
WT	100.0	100.0
R306A	98.8	88.2
R306E	13.6	16.5
Y317A	103.8	98.8
Y317F	98.8	102.4
R306A/Y317A	18.8	17.6
R306E/Y317F	10.6	18.8
T231E	63.9	70.0
I146E	96.9	90.0
I146E/T231E	25.0	12.5
D208A	10.0	2.4
D208L	8.6	4.8
K210A	71.4	39.8
K210L	28.6	9.6
K210E	7.1	2.4
D208A/K210L	3.1	2.0
D208A/K210A	22.9	20.5
Q48A	84.9	71.8

3.6. Catalytic mechanism

From the four protein complexes NahK–GlcNAc, NahK–GlcNAc–AMPPNP, NahK–GlcNAc-1P–ADP and NahK–GlcNAc-1P, the phosphoryl-transfer reaction can be dissected into a continuum of four reaction states (Fig. 6*a*). The phosphoryl-accepting C1 OH of GlcNAc/GalNAc points towards the γ-phosphoryl group of AMPPNP at a distance of 3.5 Å, as

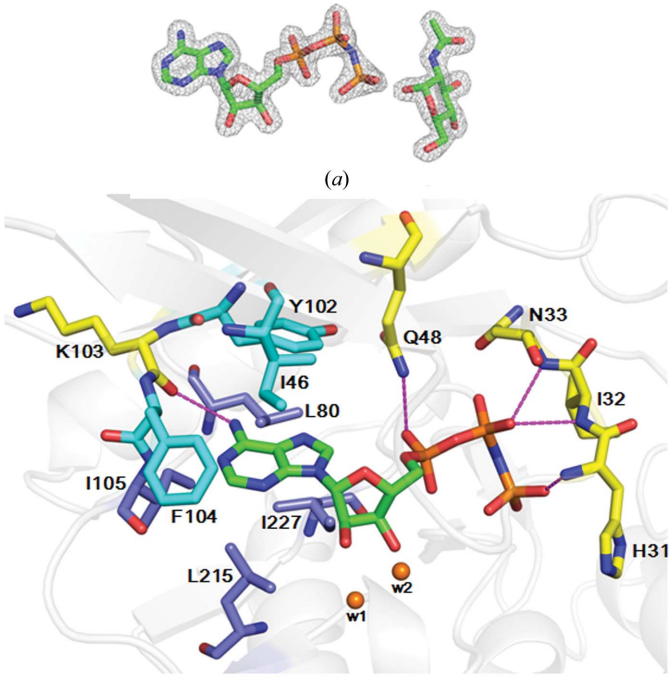


Figure 5
Nucleotide-binding site of NahK. (*a*) The $2F_o - F_c$ electron-density maps of AMPPNP and GlcNAc are contoured at 1.5σ. (*b*) Close-up view of AMPPNP in association with its binding-site residues. The residues that interact with AMPPNP through hydrogen bonds and hydrophobic forces are shown as yellow and cyan/purple sticks, respectively. Water molecules are shown as orange balls. Broken lines represent hydrogen bonds.

shown in Fig. 6(*b*). This geometry aligns a short path for a nucleophilic attack on the γ-P atom in a bond-breaking/bond-formation manner of transfer (an S_N2-like reaction). The carboxyl group of residue Asp208 within hydrogen-bond distance (2.7 Å) of C1 OH of GlcNAc/GalNAc is likely to serve as the active-site general base. On the opposite side, the γ-phosphoryl group is anchored by the Lys210 terminal amine (3.5 Å, electrostatic force), the backbone amide of His31 (2.6 Å, hydrogen bond) and the terminal carboxylate of Asp228 (2.7 Å), providing the best overlap between the lone pair of the nucleophile and the O—P σ^* antibonding orbital for direct inline transfer. The reaction would proceed through a trigonal bipyramidal transition, resulting in an inversion of configuration. In this way the transition state involves three O atoms perched on the corners of a trigonal plane which contains the ester O atom of the β-phosphoryl group and the C1 O atom of GlcNAc/GalNAc. The terminal amine of Lys210 may stabilize the resulting oxyanion of the γ-phosphoryl group and facilitate collapse of the transition state. The divalent Mg^{2+} ion then takes over to neutralize the growing negative charge on the β-phosphoryl group of ADP, so that the product can be loosened and released (Fig. 6*c*). In contrast, residues Asp208 and Asp228 may exert the terminal repulsive forces (2.5 and 3.3 Å, respectively) to help release the phosphorylated GlcNAc/GalNAc as shown in the NahK–GlcNAc-1P complexes (Fig. 6*d*). This mechanistic notion was supported by mutagenic and biochemical experiments, in which the relative activities of the mutants D208A and D208L declined

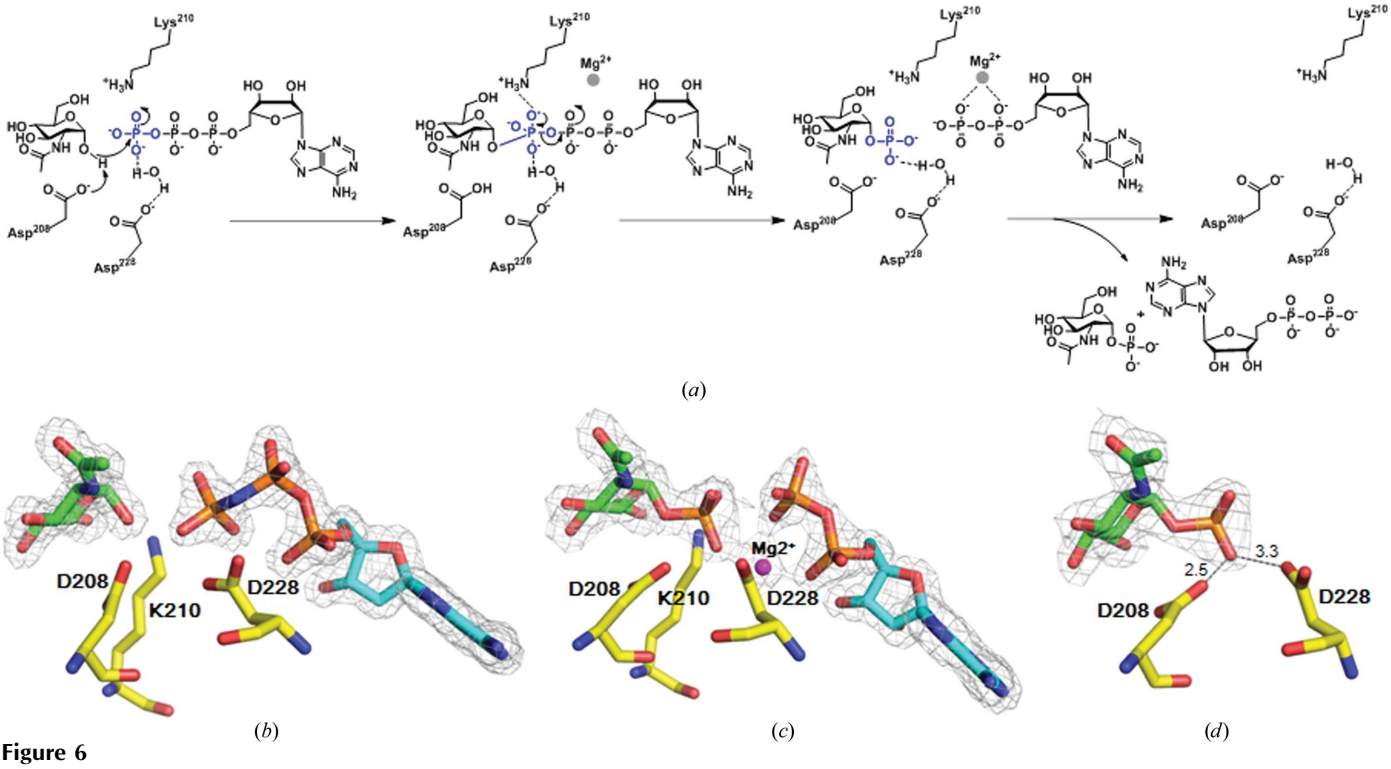


Figure 6
Catalytic mechanism of NahK. (*a*) An S_N2-like mechanism was proposed for the NahK-mediated phosphoryl-transfer reaction, in which Asp208 serves as the general base, Asp228 positions ATP and Lys210 stabilizes the transition. (*b–d*) Snapshots of the phosphoryl-transfer reaction in three different states: (*b*) the initial state; (*c*) the production state; (*d*) the leaving state. The key residues, GlcNAc, AMPPNP and phosphoryl groups are shown as yellow, green, cyan and orange sticks, respectively. The Mg^{2+} ion is shown as a magenta ball. The $2F_o - F_c$ electron-density maps are contoured at 1.5σ.

significantly (10 and 8.6%, respectively, relative to the WT), supporting its role as the general base (Table 4). The relative activities of the K210E (changing the electrostatics) and K210A (disrupting the association) mutants decreased to 7.1 and 71.4% relative to that of the WT, supporting stabilization of the resulting oxyanion of the γ-phosphoryl group by the terminal amine group of Lys210.

3.7. Reaction order

The dissociation constants (K_d) for the substrates and products *versus* NahK were determined using isothermal titration calorimetry (ITC) as shown in Table 2. Both ATP and ADP bind to NahK at a micromolar level; the former is twofold stronger than the latter, consistent with both the experimental and empirical results. In contrast, the K_d for the GlcNAc/GlcNAc-1P pair differs significantly: the K_d for GlcNAc-1P was not measurable, suggesting that the low affinity of GlcNAc-1P is likely to be owing to the introduction of repulsive forces against residues Asp208 and Asp228 (see above). Interestingly, we still obtained the NahK–GlcNAc-1P complex. Given that this complex was crystallized at pH 5.5, the terminal carboxylate of Asp208 is likely to be associated with a neutral OH instead of the negatively charged O^- in the phosphoryl group through a low energy-barrier hydrogen bond (2.5 Å) in addition to being limited to the rigid crystal lattice (Fig. 6*d*). Given also the fact that apo NahK crystals were not obtainable despite tens of thousands of screening conditions, apo NahK is likely to undergo an induced-fit conformational change in the presence of GlcNAc/GalNAc. Overall, ATP may initially bind to the enzyme. When GlcNAc/GalNAc enters NahK, the protein adopts an active conformation to align a better reaction path for the phosphory-transfer reaction. Upon completion of the reaction, GlcNAc-1P/GalNAc-1P is likely to leave prior to ADP to enable the enzyme to readopt the initial state for the next reaction cycle.

3.8. Substrate specificity

From our biochemical examinations, it was concluded that NahK has a high ability to phosphorylate GlcNAc and GalNAc, a low ability to phosphorylate mannose, glucosamine, galactosamine, mannosamine, glucose, galactose and ManNAc, and no ability to phosphorylate glucuronic acid, galacturonic acid, 6-azido glucose, 6-GlcNAc and 3-, 4- or 6-glucosamine (Table 3 and Supplementary Fig. S5*a*). The substrate specificity determined by the biochemical analysis is consistent with the results from ITC (Table 2). The sugar-1P products (*e.g.* GlcNAc-1P and GalNAc-1P) can subsequently be transformed by uridyltransferase (*e.g.* GlmU in *E. coli*; Guan *et al.*, 2009; Mengin-Lecreulx & van Heijenoort, 1993) to the corresponding UDP-sugars for advanced applications (Supplementary Fig. S5*b*). Recent reports have indicated that GlcNAc derivatives with C6, both C2 and C6 or both C2 and C3 modifications are tolerable substrates of NahK (Li *et al.*, 2011; Cai, Guan, Kitaoka *et al.*, 2009). Given the topology of the sugar-binding site of NahK, these modified sugars, not surprisingly, are mainly limited to 2-*N*-acyl sugars/derivatives. Given the fact that the polyhydroxylated nature of usual/unusual saccharides in oligosaccharide synthesis requires elaborate protecting chemistry to ensure an appropriate spatial and temporal control in the subcomponent-coupling stage, the NahK complexes reported here should provide the most complete structural information to facilitate expansion of the substrate dimension for the enzyme-driven carbohydrate synthetic approach.

4. Conclusions

The assimilation of *N*-acetylhexosamine *via* the LNB/GNB pathway is a clear manifestation of the symbiosis between humans and bifidobacteria. The three-dimensional structures of NahK from both *B. longum* (JCM1217) and *B. infantis* (ATCC15697) solved at resolutions of 1.5–2.2 Å represent the first multiple reaction-state complexes of the *N*-acetyl hexosamine kinase. These structures revealed the molecular basis for the recognition of the substrates and products GlcNAc/GalNAc, GlcNAc-1P/GalNAc-1P, ATP/ADP and Mg^{2+} and allowed further insights to be gained into the enzyme mechanism, which will enable rational enzyme engineering for NahK to become a choice biocatalyst in the enzymatic and chemoenzymatic synthesis of carbohydrates.

This work was funded by the National Science Council (NSC) of Taiwan (100-2311-B-001-018-MY3), the National Health Research Institute (NHRI-EX101-10118NC) and Academia Sinica (intramural funding). Portions of this research were carried out at the National Synchrotron Radiation Research Center (NSRRC), a national user facility supported by the National Science Council of Taiwan, ROC. The Synchrotron Radiation Protein Crystallography Facility is supported by the National Research Program for Genomic Medicine. We thank both NSRRC of Taiwan and SPring-8 of Japan for beamtime allocations

References

Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* D**52**, 30–42.
Brenner, S. (1987). *Nature (London)*, **329**, 21.
Burk, D. L., Hon, W. C., Leung, A. K.-W. & Berghuis, A. M. (2001). *Biochemistry*, **40**, 8756–8764.
Cai, L., Guan, W., Kitaoka, M., Shen, J., Xia, C., Chen, W. & Wang, P. G. (2009). *Chem. Commun.*, pp. 2944–2946.
Cai, L., Guan, W., Wang, W., Zhao, W., Kitaoka, M., Shen, J., O'Neil, C. & Wang, P. G. (2009). *Bioorg. Med. Chem. Lett.* **19**, 5433–5435.
Cowtan, K. (2006). *Acta Cryst.* D**62**, 1002–1011.
Cowtan, K. (2008). *Acta Cryst.* D**64**, 83–89.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Fong, D. H., Lemke, C. T., Hwang, J., Xiong, B. & Berghuis, A. M. (2010). *J. Biol. Chem.* **285**, 9545–9555.
Fushinobu, S. (2010). *Biosci. Biotechnol. Biochem.* **74**, 2374–2384.
Graaff, R. A. G. de, Hilge, M., van der Plas, J. L. & Abrahams, J. P. (2001). *Acta Cryst.* D**57**, 1857–1862.
Grueninger, D. & Schulz, G. E. (2006). *J. Mol. Biol.* **359**, 787–797.
Guan, W., Cai, L., Fang, J., Wu, B. & Wang, P. G. (2009). *Chem. Commun.*, pp. 6976–6978.
Hart, G. W., Housley, M. P. & Slawson, C. (2007). *Nature (London)*, **446**, 1017–1022.

Heijenoort, J. van (2001). *Glycobiology*, **11**, 25R–36R.
Holm, L. & Sander, C. (1995). *Trends Biochem. Sci.* **20**, 478–480.
Huang, Y.-T., Lyu, S.-Y., Chuang, P.-H., Hsu, N.-S., Li, Y.-S., Chan, H.-C., Huang, C.-J., Liu, Y.-C., Wu, C.-J., Yang, W.-B. & Li, T.-L. (2009). *Chembiochem*, **10**, 2480–2487.
Hurley, J. H. (1996). *Annu. Rev. Biophys. Biomol. Struct.* **25**, 137–162.
Kitaoka, M. (2012). *Adv. Nutr.* **3**, 422S–429S.
Kitaoka, M., Tian, J. & Nishimoto, M. (2005). *Appl. Environ. Microbiol.* **71**, 3158–3162.
Li, Y., Yu, H., Chen, Y., Lau, K., Cai, L., Cao, H., Tiwari, V. K., Qu, J., Thon, V., Wang, P. G. & Chen, X. (2011). *Molecules*, **16**, 6396–6407.
Lindahl, U. & Höök, M. (1978). *Annu. Rev. Biochem.* **47**, 385–417.
Malito, E., Sekulic, N., Too, W. C., Konrad, M. & Lavie, A. (2006). *J. Mol. Biol.* **364**, 136–151.
Mengin-Lecreulx, D. & van Heijenoort, J. (1993). *J. Bacteriol.* **175**, 6150–6157.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Ness, S. R., de Graaff, R. A., Abrahams, J. P. & Pannu, N. S. (2004). *Structure*, **12**, 1753–1761.
Nishimoto, M. & Kitaoka, M. (2007). *Appl. Environ. Microbiol.* **73**, 6444–6449.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pannu, N. S. & Read, R. J. (2004). *Acta Cryst.* D**60**, 22–27.
Scheeff, E. D. & Bourne, P. E. (2005). *PLoS Comput. Biol.* **1**, e49.
Sela, D. A., Chapman, J., Adeuya, A., Kim, J. H., Chen, F., Whitehead, T. R., Lapidus, A., Rokhsar, D. S., Lebrilla, C. B., German, J. B., Price, N. P., Richardson, P. M. & Mills, D. A. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 18964–18969.
Smith, C. A., Toth, M., Frase, H., Byrnes, L. J. & Vakulenko, S. B. (2012). *J. Biol. Chem.* **287**, 12893–12903.
Thoden, J. B., Timson, D. J., Reece, R. J. & Holden, H. M. (2005). *J. Biol. Chem.* **280**, 9662–9670.
Thompson, P. R., Hughes, D. W., Cianciotto, N. P. & Wright, G. D. (1998). *J. Biol. Chem.* **273**, 14788–14795.
Zhao, G., Guan, W., Cai, L. & Wang, P. G. (2010). *Nature Protoc.* **5**, 636–646.