Title

Structural characterization of a nonhydrolyzing UDP-GlcNAc 2-epimerase from Neisseria meningitidis serogroup A

Journal

Acta Crystallographica Section F: Structural Biology Communications, 76(11)

ISSN

2053-230X

Authors

Hurlburt, Nicholas K
Guan, Jasper
Ong, Hoonsan
et al.

Publication Date

2020-11-01

DOI

10.1107/s2053230x20013680

Peer reviewed

manuscript ir5017 for review

STRUCTURAL BIOLOGY
COMMUNICATIONS

Structural characterization of a non-hydrolyzing UDP-GlcNAc 2-epimerase from Neisseria meningitidis serogroup A

Nicholas Hurlburt, Jasper Guan, Hoonsan Ong, Hai Yu, Xi Chen and Andrew Fisher

Category:

Co-editor:

Dr J. Agirre

York Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York, North Yorkshire YO10 5DD, United Kingdom

Telephone: +44(0)1904328252

Fax: +44(0)1904328266

Email: jon.agirre@york.ac.uk

Submitting author:

Andrew Fisher

Chemistry, University of California, One Sheilds Avenue, Davis, CA, 95616, United States

Telephone: 5307546180

Fax: 5307528995

Email: ajfisher@ucdavis.edu

Structural characterization of a non-hydrolyzing UDP-GlcNAc 2-epimerase from *Neisseria meningitidis* serogroup A

Nicholas K. Hurlburt[1,#], Jasper Guan[1], Hoonsan Ong[1], Hai Yu[1], Xi Chen[1,*], Andrew J. Fisher[1,2,*]

[1]Department of Chemistry, University of California, Davis, CA 95616, USA; [2]Department of Molecular and Cellular Biology, University of California, Davis, CA 95616, USA

Running title:

[#]Present address: Fred Hutchinson Cancer Research Center, Vaccine and Infectious Disease Division, 1100 Fairview Ave. N., Seattle, WA 98109, USA

*Corresponding Authors Andrew J. Fisher: ajfisher@ucdavis.edu; and Xi Chen: xiichen@ucdavis.edu

Keywords:

X-ray crystallography, structure, UDP-GlcNAc, UPD-ManNAc, epimerase, epimerization, Rossmann fold.

Abstract

Bacterial non-hydrolyzing UDP-*N*-acetylglucosamine 2-epimerases catalyze the reversible interconversion of UDP-*N*-acetylglucosamine (UDP-GlcNAc) and UDP-*N*-acetylmannosamine (UDP-ManNAc). UDP-ManNAc is an important intermediate in the biosynthesis of certain cell-surface polysaccharides, including in some pathogenic bacteria such as *Neisseria meningitidis* and *Streptococcus pneumoniae*. We present here two crystal structures of UDP-*N*-acetylglucosamine 2-epimerase from *Neisseria meningitidis* serogroup A (NmSacA). One crystal structure is of the completely substrate-free enzyme, and the other structure contains UDP-GlcNAc substrate bound to the active site. Both structures form dimers as seen in similar epimerases, and substrate binding to the active site induces a large conformational change where two Rossmann-like domains clamp down on the substrate. Unlike other epimerases, NmSacA does not require UDP-GlcNAc to initiate epimerization of UDP-ManNAc although UDP-GlcNAc is an activator. In spite of conserved residues involved in binding the allosteric UDP-GlcNAc, the structures presented here do not contain UDP-GlcNAc bound in the allosteric site, which is observed adjacent to the active site in similar epimerases. The structures presented here provide additional insight to the mechanism and regulation of this critical enzyme and improves our structural understanding of NmSacA's ability to epimerize modified substrates.

1. Introduction

Bacterial capsular polysaccharides (CPSs) are distinct, organized structures found on the surface of a wide range of bacterial species and are important virulence factors that provide protection from various circumstances, ranging from harsh environmental conditions to the immune system of a host (Roberts, 1996, Cress *et al.*, 2014).

Neisseria meningitidis is a Gram-negative bacterium with at least twelve known serogroups which are classified based on their CPS structures. *N. meningitidis* serogroup A was historically responsible for large epidemics of meningitis and septicemia in the meningitis belt countries and still causes life-threatening invasive meningococcal diseases (IMDs) in some countries (Aye *et al.*, 2020, Dutta *et al.*, 2020, Novak *et al.*, 2019). The CPS of Nm serogroup A is a homopolymer of [–6ManNAcα1-PO_4–] and is unique from the other disease-causing Nm serogroups including serogroup X which has a CPS consisting of a [–6GlcNAcα1-PO_4–] homopolymer (Xie *et al.*, 2012) or B, C, W-135, and Y which all contain sialic acid in their CPSs (Fiebig *et al.*, 2014, Jennings *et al.*, 1977). The *cps* operon of Nm serogroup A contains four open reading frames, the first being *sacA*, which encodes an enzyme (NmSacA) that catalyzes the first step in Nm serogroup A CPS biosynthetic pathway (Swartley *et al.*, 1998). This enzyme [EC: 5.1.3.14] catalyzes the conversion of uridine 5'-diphosphate *N*-acetylglucosamine (UDP-GlcNAc) to its C2"-epimer UDP-*N*-acetylmannosamine (UDP-ManNAc) (Zhang *et al.*, 2016, Swartley *et al.*, 1998).

NmSacA was verified to be a non-hydrolyzing UDP-GlcNAc 2-epimerase (Zhang *et al.*, 2016). The mechanism of this type of bacterial epimerases is believed to involve *anti*-elimination of the C2" proton and UDP from UDP-GlcNAc, generating enzyme-bound intermediates of UDP and 2-acetamidoglucal, followed by a subsequent *syn*-addition proton to C2" and UDP to the same face of the double bond, producing the UDP-ManNAc product (Morgan *et al.*, 1997, Tanner, 2002). NmSacA was shown to release 2-acetamidoglucal and UDP which were observed easily when the reactions were carried out for an extended period of time (Zhang *et al.*, 2016).

Bacterial non-hydrolyzing UDP-GlcNAc 2-epimerases are found in both Gram-negative and Gram-positive strains and vary widely in their regulation. The ones from Gram-positive bacteria *Bacillus anthracis* (Velloso *et al.*, 2008), *Bacillus cereus* (Kawamura *et al.*, 1978), and *Staphylococcus aureus* (Mann *et al.*, 2016) as well as the ones from the Gram-negative bacterium *E. coli* O14 K7 H (Morgan *et al.*, 1997) and from archaea *Methanococcus jannaschii* (Chen *et al.*, 2014) were shown to be allosterically regulated, requiring UDP-GlcNAc to catalyze the reversible epimerization of UDP-ManNAc to UDP-GlcNAc (Kawamura *et al.*, 1979, Morgan *et al.*, 1997). The crystal structures of these enzymes

reveal that the allosteric UDP-GlcNAc binds in a conserved site adjacent to the active site, which contains a bound UDP after the GlcNAc hydrolyzed off from UDP-GlcNAc in an extended incubation time for crystal growth (Chen *et al.*, 2014, Velloso *et al.*, 2008). The GlcNAc moiety of the UDP-GlcNAc allosteric effector makes extensive interaction with the pyrophosphate of the UDP in the active site. The allosteric UDP-GlcNAc binding is presumed to optimize the configuration of the active site and exclude solvent.

UDP-GlcNAc was not required for the activity of NmSacA in epimerizing UDP-ManNAc to UDP-GlcNAc, but UDP-GlcNAc did appear to enhance the kinetics of epimerization (Zhang *et al.*, 2016). To better understand the epimerization reaction mechanism of NmSacA, its apparent lack of the requirement of UDP-GlcNAc for reactivity, and to determine the structural basis of its activity on modified substrates, we solved two crystal structures of NmSacA, a ligand-free structure and a structure bound to the ligand UDP-GlcNAc.

2. Materials and methods

2.1 Cloning, expression, and purification

As reported previously (Zhang *et al.*, 2016), the gene that encodes for NmSacA was amplified from the genomic DNA of *N. meningitidis* serogroup A stain M1027 and cloned into a pET22b(+) vector (Novagen) with a C-terminal His_6-tag using the *Nde*I and *Xho*I restriction sites. Plasmids were sequenced to verify correct ligation and were transformed into *E. coli* BL21 (DE3) cells (Invitrogen) for expression.

Plasmid-bearing strains were grown in LB-rich media with ampicillin (100 µg/mL) and grown to a density of $OD_{600\ nm}$ = 0.8. Overexpression was induced with 0.1 mM of isopropyl-1-thio-β-D-galactopyranoside (IPTG) and cultures were allowed to be incubated at 293 K for 20 h with shaking.

After 20 h incubation, cells were pelleted from culture by centrifugation at 4,000 rpm for 2 h. Cell pellet was resuspended in a lysis buffer containing Tris-HCl buffer (100 mM, pH 8.0) and 0.1% Triton X-100. Cells were lysed by treatment with 50 µg/mL of lysozyme and 3 µg/mL DNaseI for 60 min at 310 K with vigorous shaking. Cell lysate was cleared by centrifugation at 12,000 rpm for 30 min. The His_6-tagged protein was purified from cell lysate by Ni^{2+}-NTA column. The column was equilibrated with 10 column volumes (CV) of binding buffer containing Tris-HCl (50 mM, pH 7.5), NaCl (500 mM), and imidazole (5 mM). The cell lysate was loaded to the column which was then washed with 8 CV of binding buffer followed by 10 CV of washing buffer containing Tris-HCl (50 mM, pH 7.5), NaCl (500 mM), and imidazole (40 mM). The protein was eluted from the column using an elution buffer containing Tris-HCl (50 mM, pH 7.5), NaCl (500 mM), and imidazole (200 mM). Elution fractions containing the purified protein were collected, dialyzed, and stored at 277 K.

2.2 Crystallization

For the ligand-free structure, purified NmSacA at 10 mg/mL was crystallized using hanging drop vapor diffusion equilibrated against reservoir of 50% (v/v) PEG-200, phosphate-citrate (100 mM, pH 4.2), and NaCl (200 mM). Crystals were flash-cooled in liquid nitrogen prior to data collection.

For the substrate-bound structure, NmSacA (13 mg/mL) was crystallized by sitting-drop vapor diffusion against a solution of 22% PEG-5000, sodium citrate/citric acid (100 mM, pH 5.5), and UDP-GlcNAc (10 mM). Crystals were briefly soaked in a solution of 30% ethylene glycol with UDP-GlcNAc (10 mM) in mother liquor and flash-cooled in liquid nitrogen.

2.3 Data collection and structure determination

X-ray diffraction data for the ligand-free structure were collected at SSRL on beamline 7-1 at a wavelength of 1.12709 Å with a 100 µM beam size on a ADSC Q315 detector at a distance of 200 mm. Diffraction data were integrated and scaled with XDS and XSCALE (Kabsch, 2010a, b). The phases were determined by molecular replacement using the structure of a non-hydrolyzing UDP-GlcNAc 2-epimerase from *Bacillus subtilis* (PDB: 4FKZ). X-ray diffraction data for the UDP-GlcNAc-bound structure were collected at SSRL on beamline 7-1 at a wavelength of 1.03317 Å with a beam size of 150 x 100 µm on a ADSC Q315 detector at a distance of 300 mm. Data were collected using an oscillation angle of 0.2° over a 7-sec exposure time and integrated and scaled with XDS and XSCALE (Kabsch, 2010a, b). The phases were solved using molecular replacement using the ligand-free structure. The structures were refined using the program package PHENIX (Adams *et al.*, 2010). The data collection and refinement statistics are summarized in Table 1.

3. Results and discussion

3.1 Crystal structure of NmSacA

The NmSacA monomeric structure consists of two domains each folding into a three-layer (αβα) sandwich architecture of a Rossmann fold (Fig. 1*a*). Domain 1 consists of residues 1–170 and residues 356–371 forming the C-terminal α-helix. Domain 2 spans residues 171-355. Domain 1 is made up of a central seven-stranded parallel β-sheet flanked by seven α-helices with a topology of a dinucleotide-binding Rossmann domain. Domain 2 is comprised of a central six-stranded β-sheet with nine α-helices. Helices α9 to α15 pack against the central β-sheet forming the dinucleotide-binding domain, while helices α7 and α8 form a connecting segment that packs against the C-terminal domain (Fig. 1*a*). The cleft between the two domains defines the substrate binding pocket.

The ligand-free NmSacA structure contains one dimer in the crystallographic asymmetric unit, with both monomers displaying nearly identical conformations (RMSD of 1.00 Å for 350 equivalent α-carbons). The substrate-bound structure contains two dimers in the asymmetric unit, with the UDP-GlcNAc substrate binding to only one monomer of each dimer (Chains A and C) (Fig. 1*b*). The dimer interface is very similar between the two types of dimers, substrate-bound and substrate-free. Three α-helices (α3, α4, and α5) from each monomer of Domain 1 form a six-helix bundle mediating the dimer interaction, and buries a surface area of 1,401 $Å^2$ and 1,368 $Å^2$, for the structures without and with the substrate, respectively (Fig. 1*c*).

3.2 Substrate UDP-GlcNAc bound in active site.

The closed conformation was bound with UDP-GlcNAc in the active site. While the electron density strongly defines the UDP moiety, the density was weaker for the sugar GlcNAc and was refined to have ~60% occupancy, suggesting some hydrolysis occurred in the crystal (Fig. 2*a*). At 100% occupancy, the sugar is surrounded by negative electron density in the F_o-F_c map. With only UDP modeled in the structure, heavy positive electron density extends off the end of the β-phosphate suggesting GlcNAc is present. Lowering the occupancy of the GlcNAc to 60% provided continuous electron density in the $2F_o$-F_c map and no density in the F_o-F_c map.

The UDP-GlcNAc binds at the interface between the two domains. The uracil ring π-stacks with Phe275. The main chain carbonyl of Gln270 hydrogen bonds with N3 of the uracil ring (2.8 Å) and the side chain amide hydrogen bonds with O4 (3.0 Å). Both hydroxyls on the ribose sugar hydrogen bond with the side chain of Glu295 (2.8 Å for both oxygens). The ring-oxygen in the ribose sugar is hydrogen

bonded by Arg10 (2.8 Å), which also bonds to the α-phosphate oxygen (3.0 Å). The sidechain hydroxyl of Ser289 hydrogen bonds with both the α- and β-phosphates (2.7 and 3.1 Å, respectively). The main chain amide nitrogens of Gly290 and Gly291 both interact with the β-phosphate (both 2.9 Å). These two glycine residues are part of the conserved sequence DSGG and initiates the N-terminus of helix α12, suggesting the helix dipole contributes to anchoring the UDP. The closed conformation also orders a loop from His212 to Glu219, which has sparse electron density in the substrate-free open conformation monomer in the crystal. This loop is in close proximity of the substrate binding site with His212 hydrogen bonding to the α- and β-phosphate oxygens (3.3 and 3.0 Å, respectively).

The GlcNAc sugar moiety is bound by a series of hydrogen bonding interactions. Lys15 reaches into the binding pocket and binds to the C4 hydroxyl of GlcNAc (3.2 Å). The C4 hydroxyl is also ligated by the carboxyl group of Glu117 (2.5 Å). The C3 hydroxyl is bound by the carboxyl group of Asp95 (2.8 Å) and the guanidinium group of Arg312 (3.0 Å). Arg135 makes a long hydrogen bond to the acetyl group carbonyl oxygen of the carbohydrate (3.4 Å) (Fig. 2*b*). In summary, most of the contacts to the UDP moiety come from Domain 1 while most of the interacts to the GlcNAc sugar originate from Domain 2.

The annotated catalytic residues of Asp95, Glu117, Glu131, and His212 (Samuel & Tanner, 2004), are all in close proximity to the sugar (Fig. 2*b*). The carboxyl group of Asp95 is 3.3 Å from the GlcNAc C2" carbon suggesting that it serves as the general base to generate the 2-acetamidoglucal intermediate. However, it is uncertain what residue can serve as the general acid for the *syn*-addition protonation of C2". Nevertheless, it is interesting to note that His212 hydrogen-bonds to a β-phosphate oxygen, which in turn is 4.0 Å from the C2" carbon *syn*-face, suggesting that His212 may be a general acid where the proton is shuttle by the β-phosphate in a substrate-assisted fashion.

3.3 UDP-GlcNAc substrate binding triggers a closed conformation.

Substrate binding induces a conformational change where each domain closes down upon the UDP-GlcNAc substrate, which binds at the cleft between the two Rossmann domains. The ligand-free monomers of each dimer (Chains B and D) reside in the open conformation very similar to both monomers of the completely ligand-free structure (RMSD range from 0.374 Å to 1.16 Å). Comparing the two conformations, each domain superimposes well between the substrate-free open structure and the closed UDP-GlcNAc-bound structure with an RMSD of 0.799 Å for Domain 1 and 2.077 Å for Domain 2 (184 and 182 equivalent α-carbons, respectively). However, aligning Domain 1 from each structure reveals that Domain 2 rotates about 29° clamping down on the substrate (calculated by the program DynDom (Hayward & Lee, 2002)) (Fig. 3*a*). The interdomain connecting helix α7, serves as the pivot

point or hinge upon closure, which results in some α-carbon movements over 11 Å distal from the hinge in helix α10.

The closed substrate-bound NmSacA conformation is stabilized by interactions between the UDP-GlcNAc and residues from both domains. The closed state is further reinforced through the creation of new inter-domain ionic interactions. Arg213 from Domain 2 forms a new ionic interaction with Glu131 from Domain 1. Glu311 of Domain 2 also forms a new salt-linkage with Arg135 of Domain 1. Finally, Glu294 from Domain 2 ion-pairs with Lys15 of Domain 1. This substrate-induced closure is also seen in other non-hydrolyzing UDP-GlcNAc 2-epimerases (Chen *et al.*, 2014).

3.4 Na^+ ion ligated in the open conformation

The ligand-free open conformation structure revealed electron density of a potential metal ion with short contacts to mainchain carbonyl oxygens of Pro297, Ala349, and Ile351 as well as two water molecules in a trigonal bipyramidal geometry. Since the crystals were grown in 200 mM of NaCl, we modeled the cation as Na^+ (Fig. 3*a*). The geometry and ligation distances are consistent with a Na^+ ion as confirmed using the "CheckMyMetal" web server (Zheng *et al.*, 2017). The Na^+ ion was observed in both subunits in the asymmetric unit of the completely ligand-free structure as well as the two monomers in the open conformation that do not contain UDP-GlcNAc (Chains B and D) of the substrate-bound crystal form.

This cation binding site is far from the active site (>20Å). It is located in a place where Domain 2 transitions back into Domain 1 and fastens the loop between α15 and α16 to helix α12. Interestingly, the Na^+ ion is absent in the ligand-bound closed conformation. Upon closing, the end of helix α15 unravels a partial turn, shifting the loop connecting helices α15 to α16. Ile351 in the loop swings over and occludes the sodium site. This results in the flipping of Ile351 and moving of Ala349, breaking the trigonal-planer geometry of the main-chain ligations and displacing a water ligand (Fig. 3*b*). The cation does not appear to be involved in catalysis or conformational flexibility as treatment with ethylenediaminetetraacetic acid (EDTA) had little to no effect on catalytic activity (Zhang *et al.*, 2016).

3.5 Comparison with other epimerase structures

NmSacA shares high sequence and structural homology with many epimerase structures in the PDB (Badger *et al.*, 2005, Campbell *et al.*, 2000, Chen *et al.*, 2014, Mann *et al.*, 2016, Velloso *et al.*, 2008) (Fig. 4*a*). The most similar protein is the UDP-GlcNAc 2-epimerase from *E. coli* (UniProtKB: P27828, PDBID: 1VGV) (Badger *et al.*, 2005), which shares 56% identity with NmSacA. The *E. coli* structure also

crystallized with two dimers in the asymmetric unit, where UDP-GlcNAc is bound to only one monomer of each dimer. Although it should be noted that this structure, which is the result of a high-throughput structural genomics consortium, modeled in the UDP-ManNAc epimer in the structure, but labeled it as UDP-GlcNAc.

The two UDP-GlcNAc-bound chains of NmSacA align with the two substrate-bound chains of 1VGV with RMSDs that range from 0.655 Å to 0.949 Å for 300 α-carbons (Fig. 4*b*). However, the *E. coli* structure actually exhibits a little more of a domain closure upon binding substrate resulting in higher RMSDs for ligand-free superposition comparisons between *E. coli* and *Nm*SacA, which range between 1.15 Å to 1.44 Å for 300 α-carbons.

The only major structure difference between the substrate-bound structures is the disposition of helix α10 and the loop between β9 and α9 above the UDP moiety in Domain 2. The β9-α9 loop in NmSacA contains Arg213, which makes a new inter-domain ionic bond with Glu131 upon binding substrate. In the *E. coli* structure even though Glu121 is conserved, this loop shifts greatly resulting in the guanidinium group of Arg213 to swing over 13 Å to become solvent exposed (Fig. 4*c*). The β9-α9 loop in NmSacA also contains a 3_{10} helix, which is not observed in the *E. coli* structure.

The structures of other epimerases with ligands bound in the active site from *Burkholderia vietnamiensis* (PDBID 5DLD, 48.7% identity), *Bacillus subtilis* (PDBID 4FKZ, 51.3% identity), *Bacillus anthracis* (PDBID 3BEO, 50.6% identity), *Staphylococcus aureus* (PDBID 5ENZ, 47.2% identity), and *Methanocaldococcus jannaschii* (PDBID 4NES, 36.4% identity), all have the conformation of α10 and the β9-α9 loop similar to what is observed in NmSacA. This suggests that the *E. coli* enzyme may be unique in displaying a different conformation near the active site when binding the substrates. Yet the residues implicated in catalysis are in similar positions and orientations to all other epimerases.

3.6 Allosteric site

Most of the non-hydrolyzing UDP-GlcNAc 2-epimerases were shown to require UDP-GlcNAc to initiate the epimerization of UDP-ManNAc (Kawamura *et al.*, 1979, Kawamura *et al.*, 1978), suggesting it possesses a distinct allosteric regulatory site. It has also been demonstrated that UDP-GlcNAc stimulates the enzyme activity resulting in sigmoidal velocity curves with a Hill coefficient of over 2 (Kawamura *et al.*, 1979). However, it appears that NmSacA may be a unique non-hydrolyzing epimerase in that it does not required UDP-GlcNAc to initiate the epimerization of UDP-ManNAc, although UDP-GlcNAc does stimulate the epimerase activity resulting in sigmoidal kinetic curves (Zhang *et al.*, 2016).

Four of the non-hydrolyzing UDP-GlcNAc 2-epimerase structures have been determined with UDP-GlcNAc bound to the allosteric site, all with UDP also bound in the adjacent active site. These epimerases are from *Burkholderia vietnamiensis* (PDBID 5DLD), *Bacillus subtilis* (PDBID 4FKZ), *Bacillus anthracis* (PDBID 3BEO), and *Methanocaldococcus jannaschii* (PDBID 4NES). Overlaying the NmSacA structure with the epimerase from *B. anthracis* reveals that while the overall structures are very similar (RMSD 1.01 Å) the β2-α2 and β3-α3 loops are in a different conformation in *B. anthracis* compared to NmSacA because these loops help shape the allosteric UDP-GlcNAc binding pocket (Fig. 5). These *B. anthracis* loops are in the similar conformation to the three other epimerase structures with bound allosteric effector. Interestingly, all the residues involved in binding allosteric UDP-GlcNAc are conserved in all epimerases with the exception of Glu41 in NmSacA, which is a Gln in many other epimerases including *Bacillus subtilis* (Fig. 4*a*). It is unknown why UDP-GlcNAc is not observed binding to the allosteric site in NmSacA. It could be that the allosteric effector may have a weaker binding affinity in NmSacA since this enzyme does not require UDP-GlcNAc binding to initiate UDP-ManNAc epimerization. Or it could also be due to the low pH of crystal growth, which was hypothesized to prevent allosteric binding in the *E. coli* epimerase (Velloso *et al.*, 2008). The *E. coli* enzyme was crystallized at pH 5.2 and the NmSacA crystals presented here were grown at pH 5.5, which is well below the optimal enzyme activity pH of 8.5 (Zhang *et al.*, 2016).

3.7 Structural basis for tolerance of modified UDP-ManNAc substrates.

NmSacA was shown to epimerize UDP-ManNAc substrates with various modifications at the *N*-acetyl position of carbon 2 (carbon that is epimerized) (Zhang *et al.*, 2016). The mannosamine sugar must still contain the *N*-acetyl group, but some modifications of the acetyl group can serve as substrates. UDP-sugars containing a sugar without the *N*-acetyl group, such as UDP-mannose or UDP-ManNAc derivatives where the *N*-acetyl group is replaced by fluorine, amine, or azide, do not serve as substrates. Only small structural additions to the *N*-acetyl group such as *N*-propyl (with an additional methyl) and *N*-glycolyl (with an additional hydroxyl) groups can be tolerated, while UDP-ManNAc derivativities with larger *N*-acyl groups such as *N*-butyl, *N*-azidoacetyl, *N*-phenylacetyl are not epimerized by NmSacA (Zhang *et al.*, 2016).

The structure of NmSacA complexed with UDP-GlcNAc reported here helps explain its substrate specificity and the tolerance of certain sugar modifications. The oxygen of the acetyl group forms a week hydrogen bond (3.4 Å) to the conserved Arg135 (Fig. 2*b*). This explains the requirement of the N-acyl group, or at least a hydrogen-bonding accepting atom three atoms off the C2" sugar ring. Sugars without

the acetyl group or other small acyl groups are incapable of hydrogen bonding with Arg135, and therefore may not properly position the sugar for the initial C2" proton abstraction. For tolerance of small additions to the acetyl group, the structure revealed that the methyl moiety of the *N*-acetyl group points towards a small pocket defined the loop between β11 and α12 and the loop between β12 and α13. It is interesting to note that the methyl group points towards Gly290 at the start of α12, which is strictly conserved. Inspection of this pocket explains why NmSacA can only tolerate a sugar with the addition of a single hydroxyl or methyl group off the acetyl methyl, because anything larger would clash with the protein (Fig. 6).

4. Conclusions

Two crystal structures of NmSacA were determined, one in the absence of any ligands and one with UDP-GlcNAc bound in the active site. The substrate-bound structure contains two dimers in the asymmetric unit. Each dimer consists of a monomer without a bound substrate and another with the substrate UDP-GlcNAc bound in the active site. The structure reveals a common fold among non-hydrolyzing UDP-GlcNAc 2-epimerases, two domains each with a three-layer αβα-sandwich of a Rossmann fold. As observed in similar epimerases, substrate binding triggers a closure between the two domains, which adds important information to the potential epimerization mechanism. The reaction has been shown to be non-hydrolyzing (Zhang *et al.*, 2016) and the mechanism likely goes through Asp95, which would act as the general base to deprotonate the sugar moiety, but the general acid cannot be discerned from the structure. Nevertheless, conserved His212 may serve as a general acid where the proton is relayed by the β-phosphate in a substrate-assisted fashion.

Orthogonal enzymes share a very similar structure and active site pocket, and conserved residues are nearly identically as NmSacA. The allosteric site of the non-hydrolyzing UDP-GlcNAc 2-epimerases is largely conserved in NmSacA yet nothing was observed to bind to the allosteric site. Structural differences in the β2-α2 and β3-α3 loops explain why no allosteric effector was bound, however all but one of the allosteric-binding residues are conserved in NmSacA. Further studies are required to determine if NmSacA does indeed bind UDP-GlcNAc in an allosteric site. In all, the NmSacA structure and the analysis with structurally homologous enzymes shed additional light on the enzymatic mechanism and an initial understanding of the non-allosteric nature of the enzyme.

Acknowledgements

This work was partially supported by funding from the US National Institutes of Health (NIH) grants [R01GM094523 to X.C. and U01GM125288 to H.Y.]; AJF is partially supported by United States Department of Agriculture National Institute of Food and Agriculture Hatch Grant [CA-D-MCB-2310-H]; NKH was partially supported by NIH training grant T32GM007377. Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the NIH National Institute of General Medical Sciences (NIGMS) (P41GM103393). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIH or NIGMS.

References

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). *Acta Crystallogr D Biol Crystallogr* **66**, 213-221.

Aye, A. M. M., Bai, X., Borrow, R., Bory, S., Carlos, J., Caugant, D. A., Chiou, C. S., Dai, V. T. T., Dinleyici, E. C., Ghimire, P., Handryastuti, S., Heo, J. Y., Jennison, A., Kamiya, H., Tonnii Sia, L., Lucidarme, J., Marshall, H., Putri, N. D., Saha, S., Shao, Z., Sim, J. H. C., Smith, V., Taha, M. K., Van Thanh, P., Thisyakorn, U., Tshering, K., Vazquez, J., Veeraraghavan, B., Yezli, S. & Zhu, B. (2020). *The Journal of infection*.

Badger, J., Sauder, J. M., Adams, J. M., Antonysamy, S., Bain, K., Bergseid, M. G., Buchanan, S. G., Buchanan, M. D., Batiyenko, Y., Christopher, J. A., Emtage, S., Eroshkina, A., Feil, I., Furlong, E. B., Gajiwala, K. S., Gao, X., He, D., Hendle, J., Huber, A., Hoda, K., Kearins, P., Kissinger, C., Laubert, B., Lewis, H. A., Lin, J., Loomis, K., Lorimer, D., Louie, G., Maletic, M., Marsh, C. D., Miller, I., Molinari, J., Muller-Dieckmann, H. J., Newman, J. M., Noland, B. W., Pagarigan, B., Park, F., Peat, T. S., Post, K. W., Radojicic, S., Ramos, A., Romero, R., Rutter, M. E., Sanderson, W. E., Schwinn, K. D., Tresser, J., Winhoven, J., Wright, T. A., Wu, L., Xu, J. & Harris, T. J. (2005). *Proteins* **60**, 787-796.

Campbell, R. E., Mosimann, S. C., Tanner, M. E. & Strynadka, N. C. (2000). *Biochemistry* **39**, 14993-15001.

Chen, S. C., Huang, C. H., Yang, C. S., Liu, J. S., Kuan, S. M. & Chen, Y. (2014). *Proteins* **82**, 1519-1526.

Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Crystallogr D Biol Crystallogr* **66**, 12-21.

Cress, B. F., Englaender, J. A., He, W., Kasper, D., Linhardt, R. J. & Koffas, M. A. (2014). *FEMS Microbiol Rev* **38**, 660-697.

Dutta, A. K., Swaminathan, S., Abitbol, V., Kolhapure, S. & Sathyanarayanan, S. (2020). *Infectious diseases and therapy*.

Fiebig, T., Freiberger, F., Pinto, V., Romano, M. R., Black, A., Litschko, C., Bethe, A., Yashunsky, D., Adamo, R., Nikolaev, A., Berti, F. & Gerardy-Schahn, R. (2014). *J Biol Chem* **289**, 19395-19407.

Hayward, S. & Lee, R. A. (2002). *Journal of Molecular Graphics and Modelling* **21**, 181-183.

Jennings, H. J., Bhattacharjee, A. K., Bundle, D. R., Kenny, C. P., Martin, A. & Smith, I. C. (1977). *The Journal of infectious diseases* **136 Suppl**, S78-83.

Kabsch, W. (2010a). *Acta Crystallographica Section D* **66**, 125-132.

Kabsch, W. (2010b). *Acta Crystallographica Section D* **66**, 133-144.

Kawamura, T., Ishimoto, N. & Ito, E. (1979). *J Biol Chem* **254**, 8457-8465.

Kawamura, T., Kimura, M., Yamamori, S. & Ito, E. (1978). *J Biol Chem* **253**, 3595-3601.

Mann, P. A., Muller, A., Wolff, K. A., Fischmann, T., Wang, H., Reed, P., Hou, Y., Li, W., Muller, C. E., Xiao, J., Murgolo, N., Sher, X., Mayhood, T., Sheth, P. R., Mirza, A., Labroli, M., Xiao, L., McCoy, M., Gill, C. J., Pinho, M. G., Schneider, T. & Roemer, T. (2016). *PLoS Pathog* **12**, e1005585.

Morgan, P. M., Sala, R. F. & Tanner, M. E. (1997). *J. Amer. Chem. Soc.* **119**, 10269-10277.

Novak, R. T., Ronveaux, O., Bita, A. F., Ake, H. F., Lessa, F. C., Wang, X., Bwaka, A. M. & Fox, L. M. (2019). *The Journal of infectious diseases* **220**, S279-S285.

Roberts, I. S. (1996). *Annu Rev Microbiol* **50**, 285-315.

Samuel, J. & Tanner, M. E. (2004). *Biochim Biophys Acta* **1700**, 85-91.

Swartley, J. S., Liu, L. J., Miller, Y. K., Martin, L. E., Edupuganti, S. & Stephens, D. S. (1998). *J Bacteriol* **180**, 1533-1539.

Tanner, M. E. (2002). *Acc Chem Res* **35**, 237-246.

Velloso, L. M., Bhaskaran, S. S., Schuch, R., Fischetti, V. A. & Stebbins, C. E. (2008). *EMBO Rep* **9**, 199-205.

Xie, O., Bolgiano, B., Gao, F., Lockyer, K., Swann, C., Jones, C., Delrieu, I., Njanpop-Lafourcade, B. M., Tamekloe, T. A., Pollard, A. J. & Norheim, G. (2012). *Vaccine* **30**, 5812-5823.

Zhang, L., Muthana, M. M., Yu, H., McArthur, J. B., Qu, J. & Chen, X. (2016). *Carbohydr Res* **419**, 18-28.

Zheng, H., Cooper, D. R., Porebski, P. J., Shabalin, I. G., Handing, K. B. & Minor, W. (2017). *Acta crystallographica. Section D, Structural biology* **73**, 223-233.

Table 1

Data collection and refinement statistics for NmSacA

Structure	**No substrate**	**With UDP-GlcNAc Substrate**
PDB ID	**6VLB**	**6VLC**
X-ray Source	SSRL 7-1	SSRL 7-1
Wavelength (Å)	1.12709	1.03317
Temperature (K)	100	100
Detector	ADSC Q315	ADSC Q315
Crystal-detector distance(mm)	200	300
Rotation range per image (°)	0.2	0.2
Exposure time per image (s)	10	7
Space group	*C2*	$C222_1$
Unit-cell parameters (Å, °)	A=211.55, b=49.51, c=81.22	a=124.88, b=129.74, c=213.39
	α=γ=90, β=90.3	α=β=γ=90
Resolution (Å)	105.76-1.85 (1.89-1.85)	38.98-2.15 (2.20-2.15)
R_{merge} [a] (%)	7.7 (49.8)	6.3 (56.4)
<I/σ(I)>	10.80 (2.65)	18.68 (2.45)
$CC_{1/2}$ (%)	99.7 (89.0)	99.9 (73.6)
No. of reflections	254,006 (19,336)	348,834 (24,367)
No. of unique reflections	71,340 (5,160)	92,958 (6,787)
Completeness (%)	98.7 (97.0)	98.8 (98.6)
Redundancy	3.56 (3.75)	3.75 (3.59)
Refinement Statistics		
Resolution (Å)	105.76-1.85 (1.89-1.85)	38.98-2.15 (2.20-2.15)
No. of reflections (F>0) used in refinement	67,746 (4,873)	94,030 (2,719)
R_{factor} [b] (%)	18.59	16.79
R_{free} [c] (%)	21.56	21.17
RMS bond length (Å)	0.011	0.012
RMS bond angle (°)	1.22	1.492
Overall B Value ($Å^2$)	36.0	43.3
Ramachandran Plot Statistics[c]		
Residues	371	371
Most favored Region	97.6%	96.5%
Allowed Region	2.3%	2.8%
Disallowed	0.1%	0.7%

[a] R_{merge} = $[\sum_h\sum_i|I_h - I_{hi}|/\sum_h\sum_i I_{hi}]$ where I_h is the mean of I_{hi} observations of reflection *h*. Numbers in parenthesis represent highest resolution shell. [b] R_{factor} and [c] R_{free} = $\sum||F_{obs}| - |F_{calc}|| / \sum|F_{obs}|$ x 100 for 95% of recorded data (R_{factor}) or 5% data (R_{free}). [c] From MolProbity (Chen *et al.*, 2010)

Figure Legends

Figure 1. Structure of NmSacA. (*a*) Overall structure of NmSacA monomer bound to UDP-GlcNAc substrate. The protein is color-coded through the rainbow spectrum starting with blue at the N-terminus and ending with red at the C-terminus. Secondary structure elements used in the manuscript are labeled. The UDP-GlcNAc substrate is show as ball-and-stick with grey-colored carbon atoms. (*b*) Dimeric structure of one dimer in the crystallographic asymmetric unit of the NmSacA substrate-bound structure. Each dimer only displays UDP-GlcNAc substrate binding (Space-filling spheres) in one monomer of each dimer (Chain A, green-colored). The substrate-free monomer is in an open conformation (Chain B, sand-colored). For clarity the other dimer (Chains C and D) is not shown. (*c*) Dimer interaction is mediated by three helices (α3, α4, and α5) of each monomer forming a six-helix bundle at the dimer interface. UDP-GlcNAc substrate is shown as space-filling spheres with green-colored carbon atoms (bound to chain A). Similar dimeric interactions are observed in the substrate-free structure.

Figure 2. UDP-GlcNAc substrate binds at the interface between the two domains of NmSacA and is bound by both aromatic interactions and hydrogen bonding. (*a*) Domain 1 is shown in salmon and domain 2 is shown in green. The electron density map ($2F_O$-F_C) around the UDP-GlcNAc is shown in the blue mesh and is contoured to 1σ. (*b*) Interactions between UDP-GlcNAc (gray-colored carbon atoms) and NmSacA are shown. Hydrogen bonds to the UDP-GlcNAc substrate are shown as yellow dashes. Newly formed inter-domain ionic interactions, by substrate induced enzyme closure, are shown with gray-colored dashes. Ordered water molecules that interact with UDP-GlcNAc are shown as small red spheres. The residues involved in binding are labeled and the catalytic residues are highlighted with the underlined text. (c) Superposition of substrate-bound (green) and ligand-free (brown) NmSacA structures. Domain 1 is aligned (RMSD of 0.799Å), showing the closure of Domain 2 onto the substrate. Domain 2 rotates 29°.

Figure 3. Na^+ ion is coordinated in the open conformation. *(a)* The open conformation coordinates what is assumed to be a Na^+ ion through three main chain carbonyls (Pro297, Ala349, Ile351) and two waters. It is coordinated in the trigonal bipyramidal geometry. The distances of each bond are shown. *(b)* The

Na^+ ion is not found in the closed conformation. Upon substrate-induced conformational change the loop between helices H17 and H18 shifts, breaking the planer geometry of the main-chain coordinating residues and displacing a water and consequently the Na^+ ion.

Figure 4. Comparison of bacterial UDP-GlcNAc-2-epimerases. (*a*). Sequence alignment of all known bacterial UDP-GlcNAc-2-epimerase structures in PDB. Secondary-structure elements defined by NmSacA structure are drawn above the sequence alignment. Red boxes indicate residues conserved in all epimerases. The putative catalytic residues are designated by a blue asterisk (*) below the sequence, residues that bind UDP-GlcNAc substrate are designated by magenta triangle (Δ), and residues observed to bind the UDP-GlcNAc allosteric effector in other epimerase structures are represented by an orange cross (+). (*b*) Superposition of NmSacA epimerase from *N. meningitidis* (green) onto *E. coli* (salmon) (PDBID: 1vgv). Superposition of the substrate-bound monomers with UDP-GlcNAc bound (shown as sticks). The RMSD ranges between 0.655 Å to 0.949 Å (300 α-carbons) in comparing the two substrate-bound monomers from each structure. (*c*) Closeup view of the active site revealing the major structural difference between the two epimerases seen in the shift of helix α10 and the β9-α9 loop. Residues are labeled in their respective colors for each epimerase. Yellow dashed lines represent interactions.

Figure 5. Allosteric UDP-GlcNAc binding site. Shown is superposition of NmSacA (green) onto UDP-GlcNAc 2-epimerase from *B. anthracis* (PDBID 3BEO) (sand color), whose structure was determined with UDP-GlcNAc binding to the allosteric site (brown-colored carbon atoms), and UDP bound to active site. Shown are side chains that interact with the allosteric UDP-GlcNAc in the *B. anthracis* structure, with the corresponding NmSacA residues labeled in green. Yellow dashed lines show interactions to allosteric UDP-GlcNAc in *B. anthracis*. All residues are conserved in binding UDP-GlcNAc except Glu41, which is Gln in *B. anthracis*. Also shown is the UDP-GlcNAc bound in the NmSacA active site (light-green colored carbon atoms).

Figure 6. Slice through a space-filling representation of the NmSacA structure bound with UDP-GlcNAc bound in the active site. Protein is shown in green-colored surface with the UDP-GlcNAc shown in sticks with gray-colored carbon atoms. The C2" atom is labeled along with the small pocket off the *N*-acetyl group (red circle). The structure explains how the NmSacA enzyme can accept small modifications of the *N*-acetyl methyl group in UDP-GlcNAc analogs.

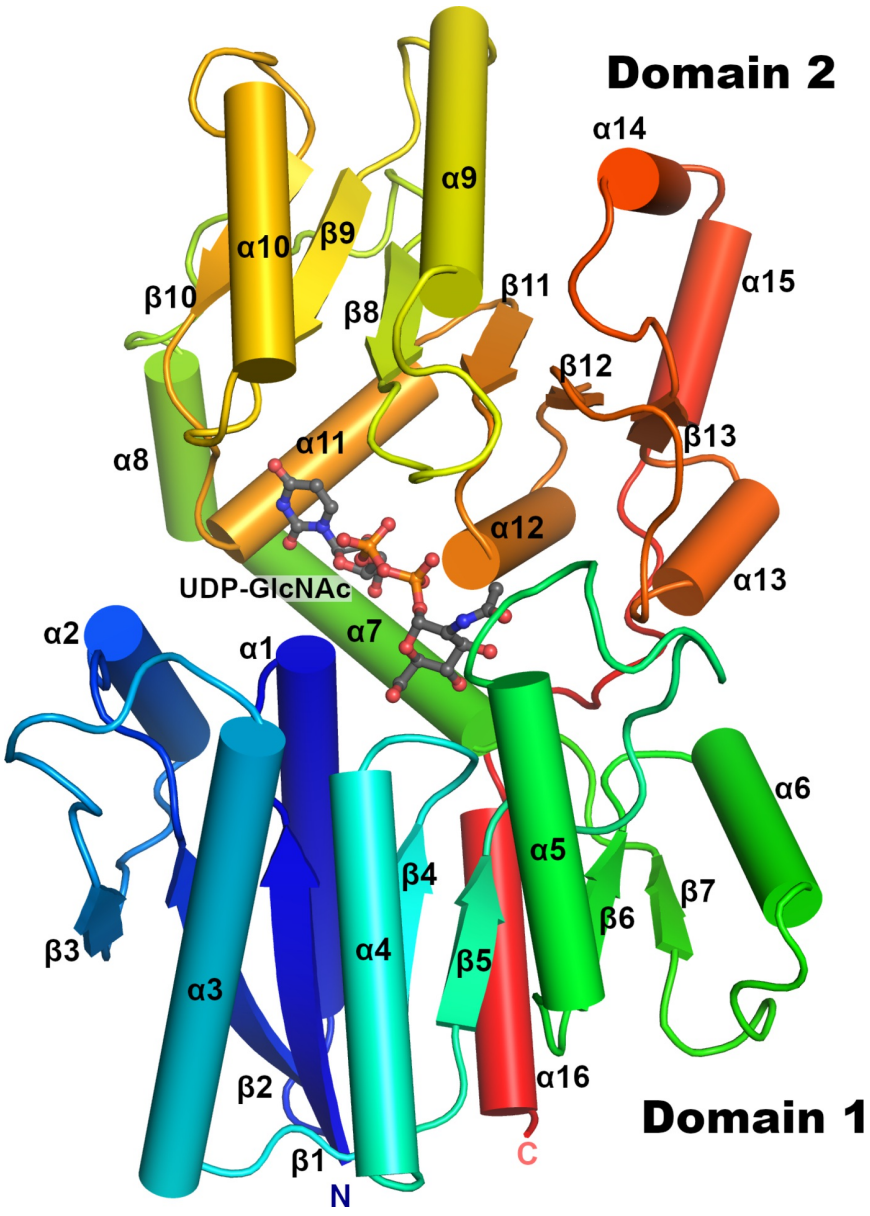
Domain 2
α14
α9
α10
β9
β10
β8
β11
α15
β12
α11
α8
β13
α12
α13
UDP-GlcNAc
α7
α2
α1
α6
α5
β4
β7
β6
β3
α4
β5
α3
α16
β2
Domain 1
β1
C
N

Figure 1a

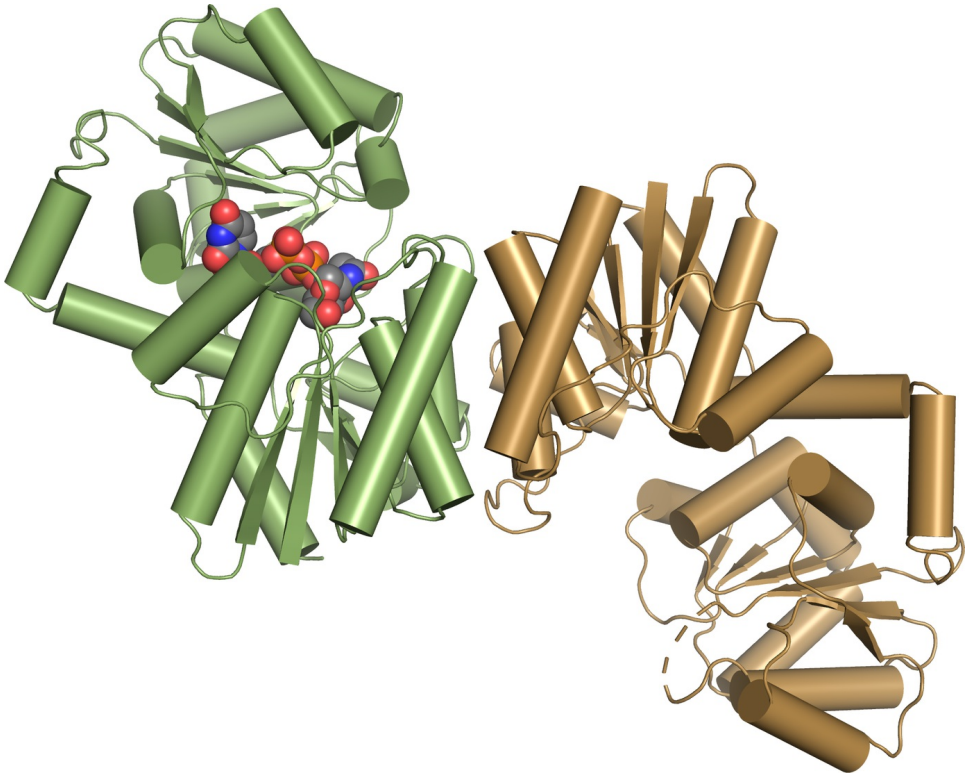

Figure 1b

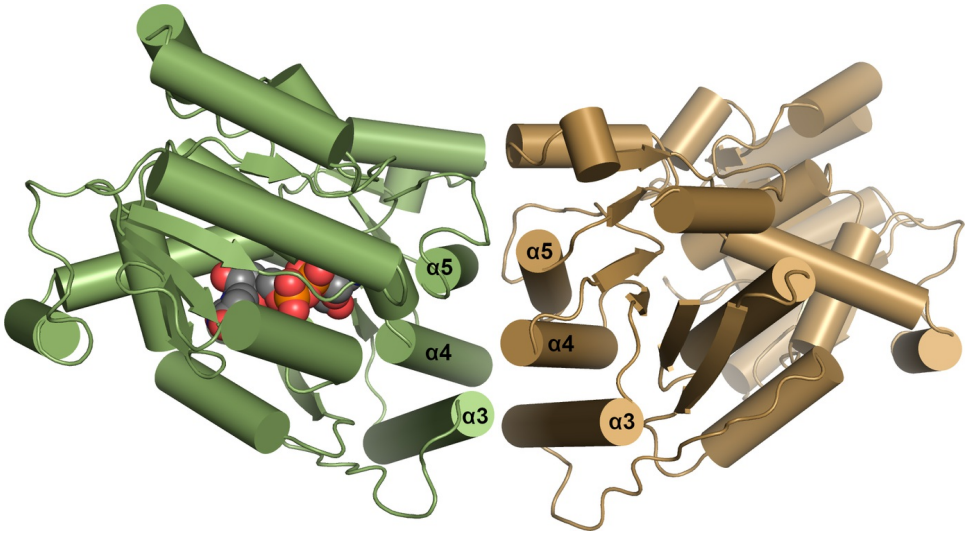

Figure 1c

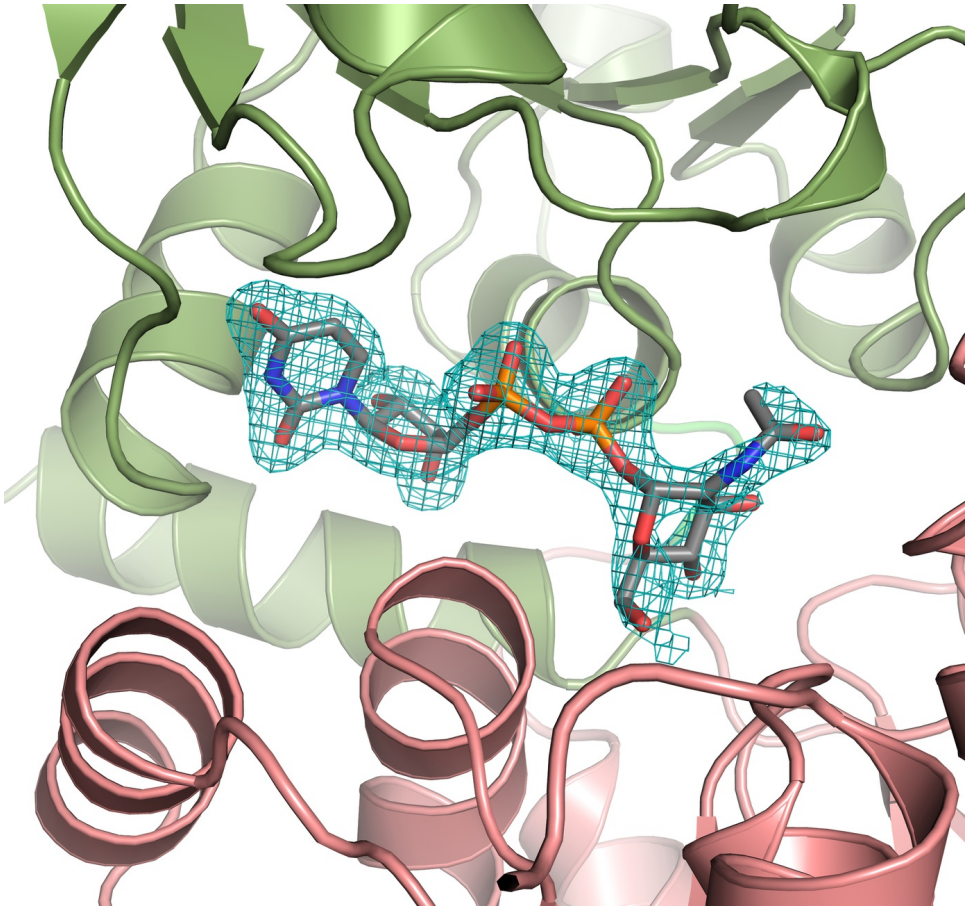

Figure 2a

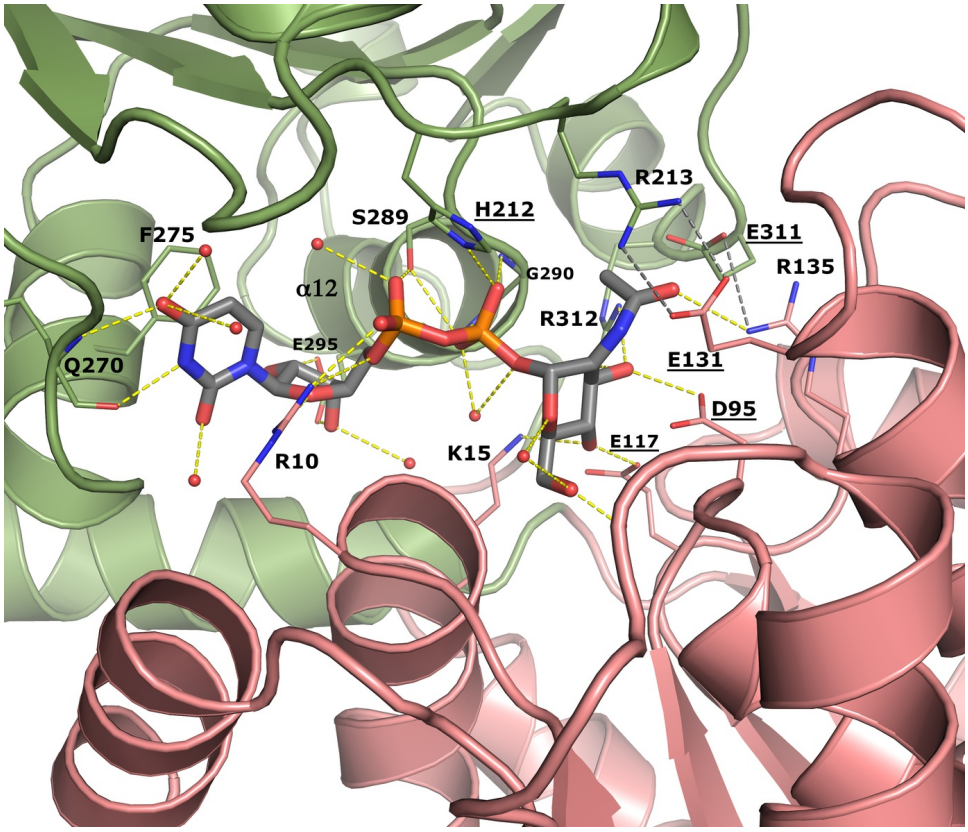
R213
S289
H212
E311
F275
G290
R135
α12
R312
E295
E131
Q270
D95
E117
K15
R10

Figure 2b

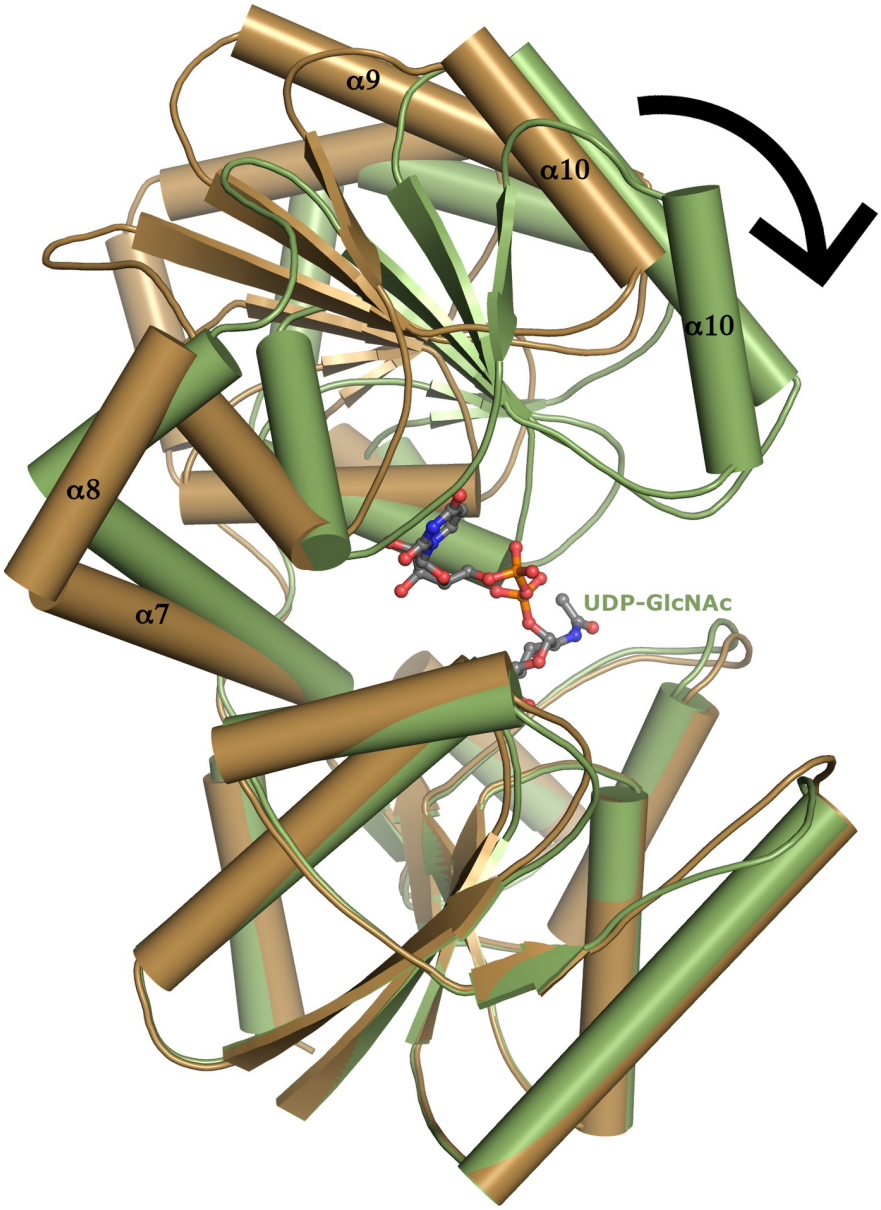

Figure 2c

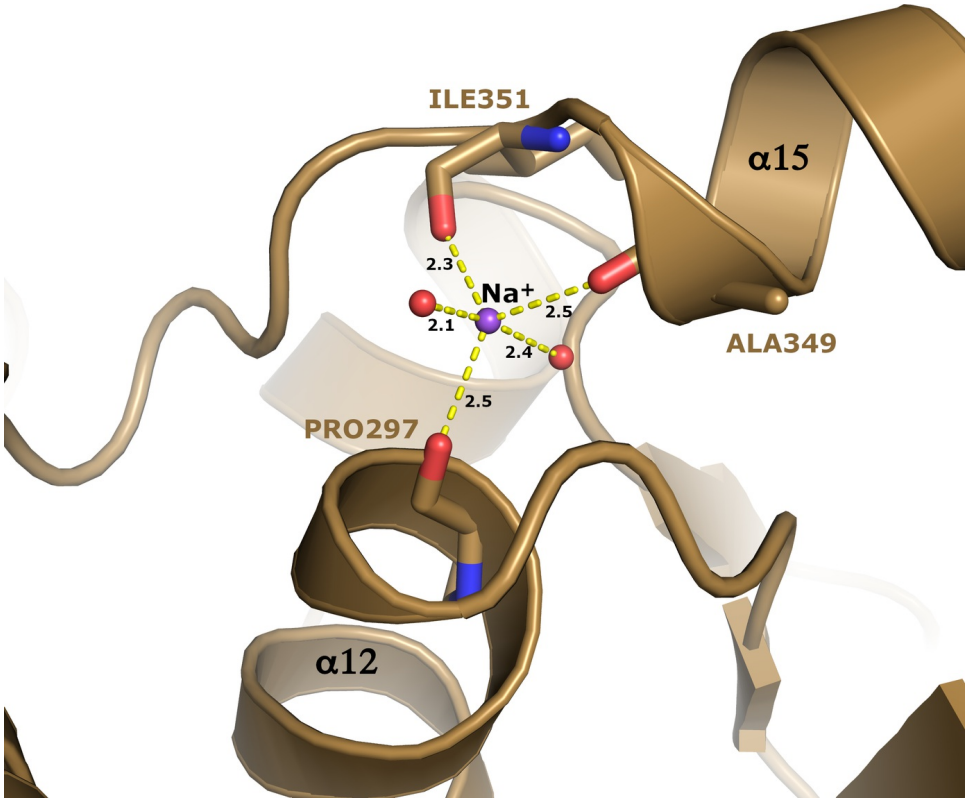

Figure 3a

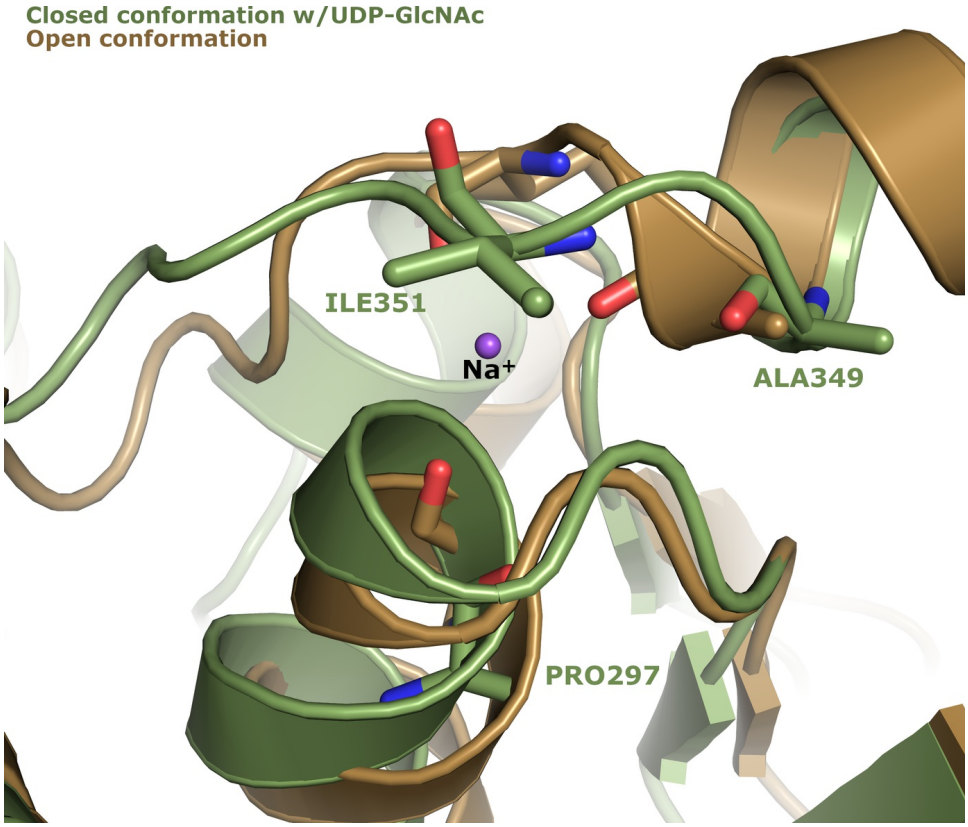
Closed conformation w/UDP-GlcNAc
Open conformation
ILE351
Na+
ALA349
PRO297

Figure 3b

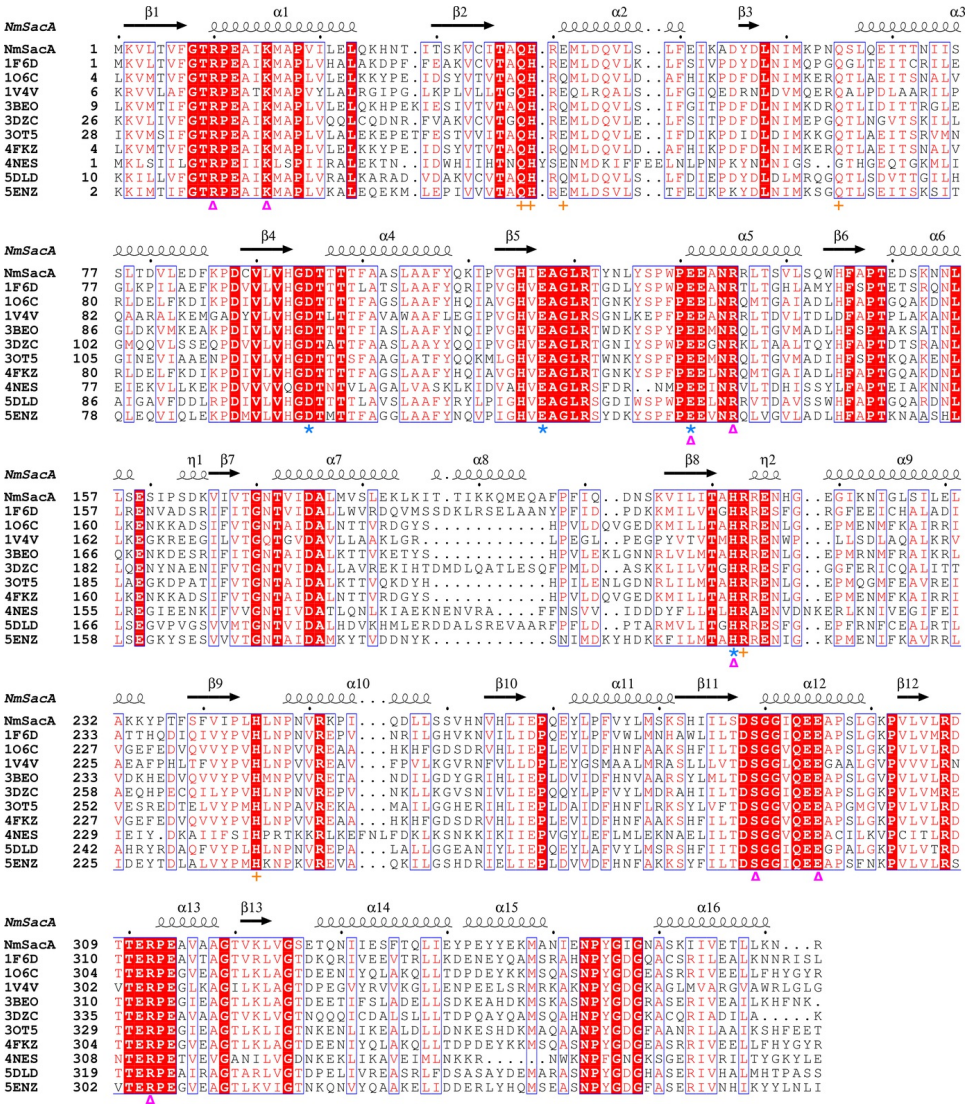
NmSacA
NmSacA
1F6D
1O6C
1V4V
3BEO
3DZC
3OT5
4FKZ
4NES
5DLD
5ENZ

Figure 4a

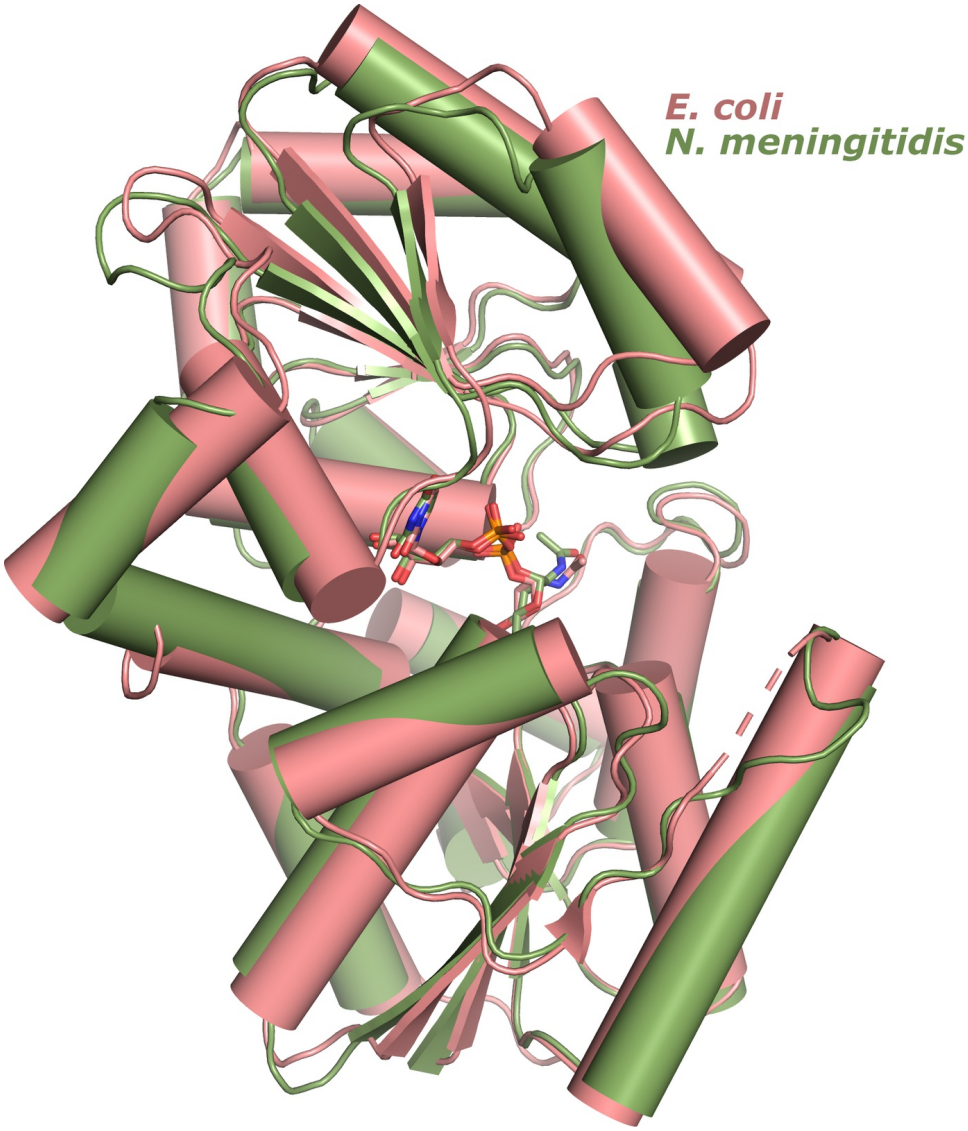

Figure 4b

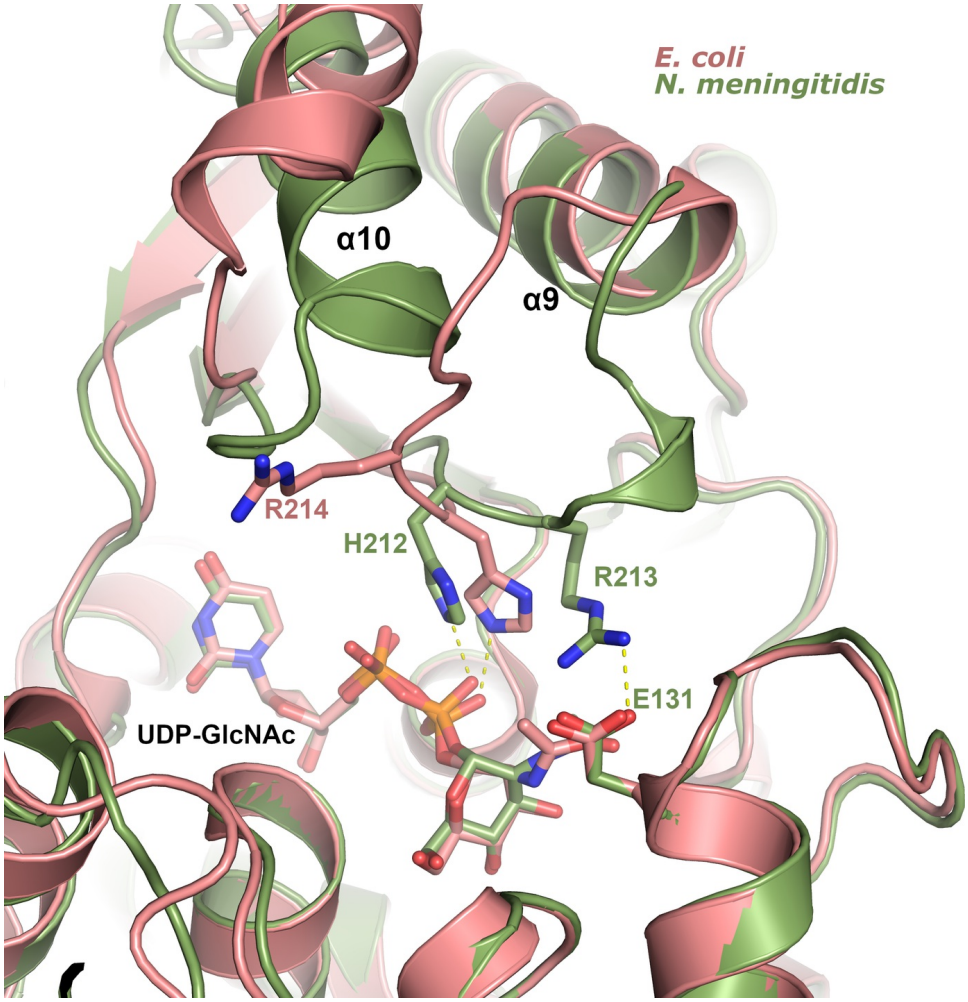

Figure 4c

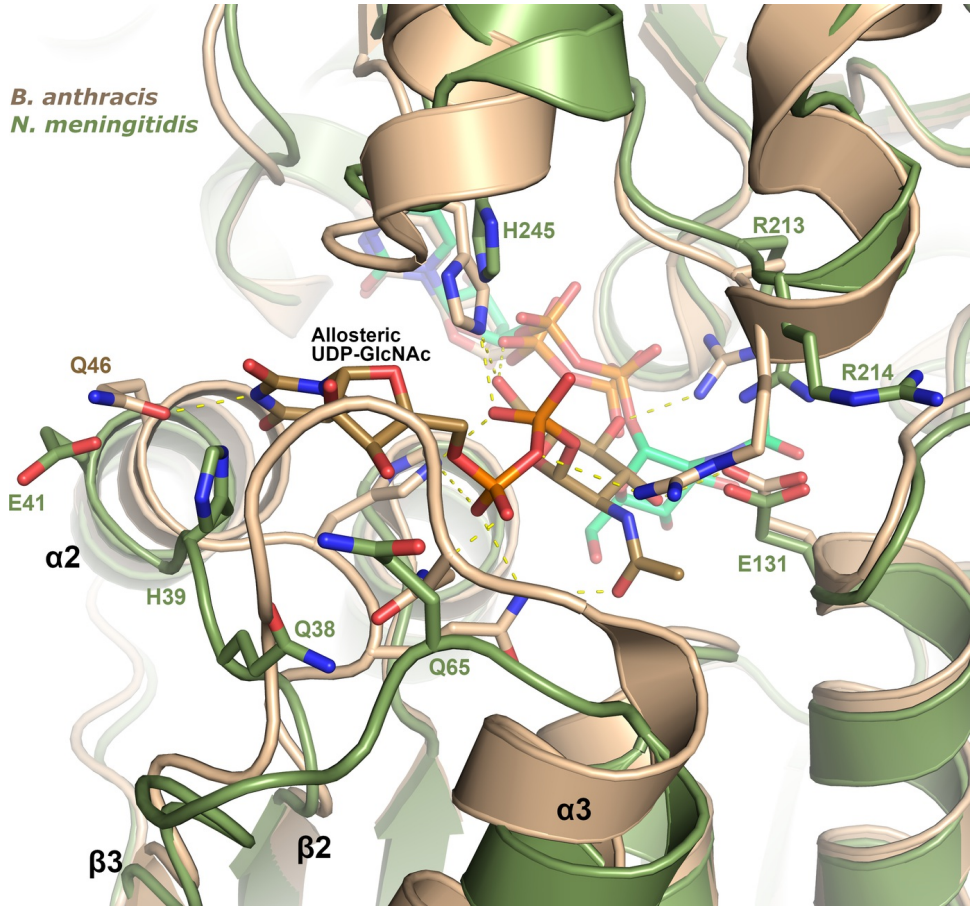
B. anthracis
N. meningitidis
H245
R213
Allosteric
UDP-GlcNAc
Q46
R214
E41
α2
E131
H39
Q38
Q65
α3
β3
β2

Figure 5

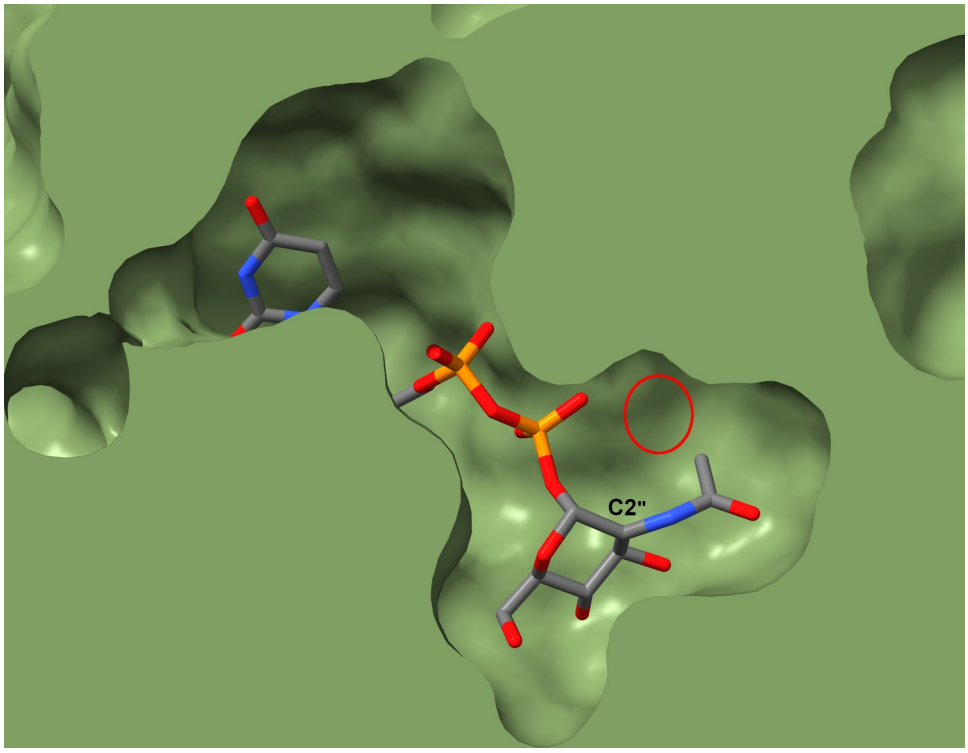

Figure 6