Structural and functional characterization of sialic acid uptake and metabolism in pathogenic bacteria

A Thesis

Submitted to

The University of Trans-Disciplinary Health Sciences and Technology

for the degree of

DOCTOR OF PHILOSOPHY

by

THANUJA GANGISETTY

under the guidance of

Prof. S. Ramaswamy

Bangalore

January, 2020

CERTIFICATE

I certify that this thesis entitled "**Structural and functional characterization of sialic acid uptake and metabolism in pathogenic bacteria**" comprises research work carried out by **Ms. Thanuja Gangisetty** at The Institute for Stem Cell Science and Regenerative Medicine (inStem) under the supervision of **Prof. Ramaswamy Subramanian** during the period 2015-2020 for the degree of Doctor of Philosophy from The University of Trans-Disciplinary Health Sciences and Technology (TDU). The results presented in this thesis have not been submitted previously to TDU or any other University for a Ph. D. or any other degree.

Prof. Mukund Thattai

Head, Academics

Institute for Stem Cell Science and Regenerative Medicine,

GKVK post, Bellary Road,

Bangalore 560065, India.

Date:

Synopsis

Sialic acids are nine-carbon sugar acids, which play an important role in a wide range of biological phenomena. Based on differences in its side chains there are approximately 50 isoforms of this sugar and N-acetylneuraminic acid (Neu5Ac) is the most common sialic acid. N-Glycolylneuraminic acid (Neu5Gc) is another common sialic acid, which differs from Neu5Ac by having an extra hydroxyl group at the fifth position. During evolution, the gene Cytidine Monophosphate-N-Acetylneuraminic acid Hydroxylase (CMAH) that converts Neu5Ac to Neu5Gc lost its activity in humans due to a frameshift mutation. Thus, humans have lost the ability to make Neu5Gc. In general, eukaryotic cells are surrounded by glycocalyx, a sugar jacket composed of proteoglycans, glycosphingolipids, and glycoproteins. Most of these glycans have sialic acid as their outermost sugar that helps in cell adhesion and signaling. Sialic acids serve as an interface between host and commensal or pathogenic microorganisms. In order to evade the host immune response, these bacteria have evolved to express sugars similar to their human host cells (1). During the process of infection, pathogenic bacteria, such as *Haemophilus influenzae* form communities that are termed biofilms, where these bacteria are intimately associated with each other on the exopolymer matrix. A study by Swords et al., shows that lipooligosaccharides (LOS) containing sialic acids promote biofilm formation of *H. influenzae* (2). Given the fact that there is dense array of sugars on the outside of the eukaryotic cells, many pathogenic and symbiotic bacteria have taken advantage of these sugars. Most of the pathogenic bacteria, including *Fusobacterium nucleatum, Pasteurella multocida,* and *Vibrio cholera,* that cause different diseases in humans and animals reside in mucus rich regions and scavenge sialic acids from the host. Further, these bacteria incorporate these sialic acids as the outermost sugars on their LOS/Lipopolysaccharide (LPS) and escape the host immune system via a process called 'molecular mimicry'. In addition, a part of the scavenged sialic acids are utilized by pathogenic bacteria as carbon, nitrogen, and energy sources (3)(4).

Sialic acids released by sialidase activity enter the bacterial periplasmic space through different porins like OmpC, OmpF, and NanC. Further, sialic acids enter the cytosol of the bacteria with the help of a primary active transporter, such as ATP-binding cassette (ABC), or secondary active transporters, such as Major Facilitator Superfamily (MFS), Solute Sodium Symporter (SSS), and Tripartite ATP-independent Periplasmic (TRAP). ABC transporters

utilize the energy derived from ATP hydrolysis for the transport of substrates, whereas the secondary transporters use Na^+ ion gradient for the transport of substrates. In general, ABC transporters contain a substrate-binding subunit in the periplasm, two transmembrane domains, and two nucleotide-binding domains. Whereas MFS, SSS, and TRAP transporters contain only two transmembrane domains for the transport of the substrates. Similar to ABC transporters, TRAP transporters contains a substrate binding protein in the periplasmic space.

In *H. influenzae*, the TRAP transporter that transports sialic acids is encoded by the operon called SiaPQM. Earlier studies on biofilm growth of NTHi shows a defect in the biofilm formation in SiaP or SiaT (SiaQM) deletion mutants, suggesting that sialic acid transport is absolutely necessary for the long-term survival of bacteria in biofilms (5). Likewise, other Gram-negative bacteria, such as *P. multocida, F. nucleatum,* and *V. cholera,* also utilize the TRAP transporters for the transport of sialic acids and the substrate binding protein that scavenges sialic acid in these bacteria is called as SiaP. SiaPs from the TRAP transport system are very well characterized, structurally and functionally.

SiaPs are two-domain proteins and sialic acid binds between both the domains in a well conserved binding pocket. Structural studies show that ligand binding generates conformational changes in the amino acids of the hinge region, which cause the two domains to close upon the substrate. One of the striking features of these proteins is the presence of a conserved Arg at the 127^{th} position that forms salt bridges with C1 carboxyl and C2 hydroxyl groups. Thermodynamic studies show that SiaPs bind to sialic acids with nanomolar affinity, and it is an enthalpically favored process (5–7) .

Haemophilus ducreyi is another Gram-negative bacterium that causes the sexually transmitted disease chancroid. LOS is an important virulence factor in these bacteria, and it is heavily sialylated. *H. ducreyi* uses the ABC transport system for the transport of sialic acid, and it is encoded by the gene SatABCD. Although several periplasmic binding proteins of the ABC transport system that transport different substrates (maltose, molybdate, Vit B12 etc.,) have been well studied, the structural and biophysical properties of SatA binding to sialic acid is not known. The sequence similarity between sialic acid binding proteins of TRAP and ABC transport system is very less; thus, it is important to understand the structural and functional properties of SatA of the ABC transport system.

Some of the sialic acid that is scavenged from the host is incorporated as the outermost sugar on the LOS/LPS of the bacteria by using the enzymes CMP-Neu5Ac synthetase (SiaB) and several sialyltransferases (Lic3B, LsgB, Lic3A, and SiaA). Part of the sialic acid is converted into fructose-6-phosphate by a cluster of enzymes called Nan-Nag gene cluster (NanA, NanK,

NanE – NagA, NagB) enzymes. Fructose-6-phosphate enters into the glycolytic pathway and is converted into carbon and nitrogen sources, which are further utilized by the bacteria. Colonization of sialic acid-rich mucus regions in the host by pathogenic/commensal bacteria emphasize the importance of sialic acid catabolism for in vivo survival of these bacteria (8). Thus, it is important to characterize all these enzymes structurally and functionally, which will be useful for designing drugs against these pathogenic bacteria. The complete overview of the sialic acid pathway in pathogenic bacteria is represented in Figure 1.

N-acetyl mannosamine kinase is the second enzyme in the catabolic pathway, which converts N-acetyl mannosamine to N-acetyl mannosamine-6-phosphate in the presence of ATP. This enzyme belongs to the ROK super family of enzymes, characterized by the presence of an ATP binding motif, a catalytic aspartate residue, and a zinc-binding motif (9). In humans, NanK is involved in the biosynthetic pathway of Neu5Ac and the sequence similarity between human and bacterial NanKs is only approximately 25%. Earlier structural studies have shown that most of the NanKs contain all the signature motifs of the ROK super family of enzymes, except *Fn*-NanK, where the enzyme lacks the Zn-binding motif. Generally, Zn-binding motifs are thought to be useful for the structural integrity and functional properties of the proteins. Although the Zn-binding motif is absent in *Fn*-NanK, its three-dimensional structure is very similar to other previously reported structures. It is also important to characterize the functional properties of *Fn*-NanK in comparison to NanKs from other pathogenic bacteria that contain Zn-binding motif.

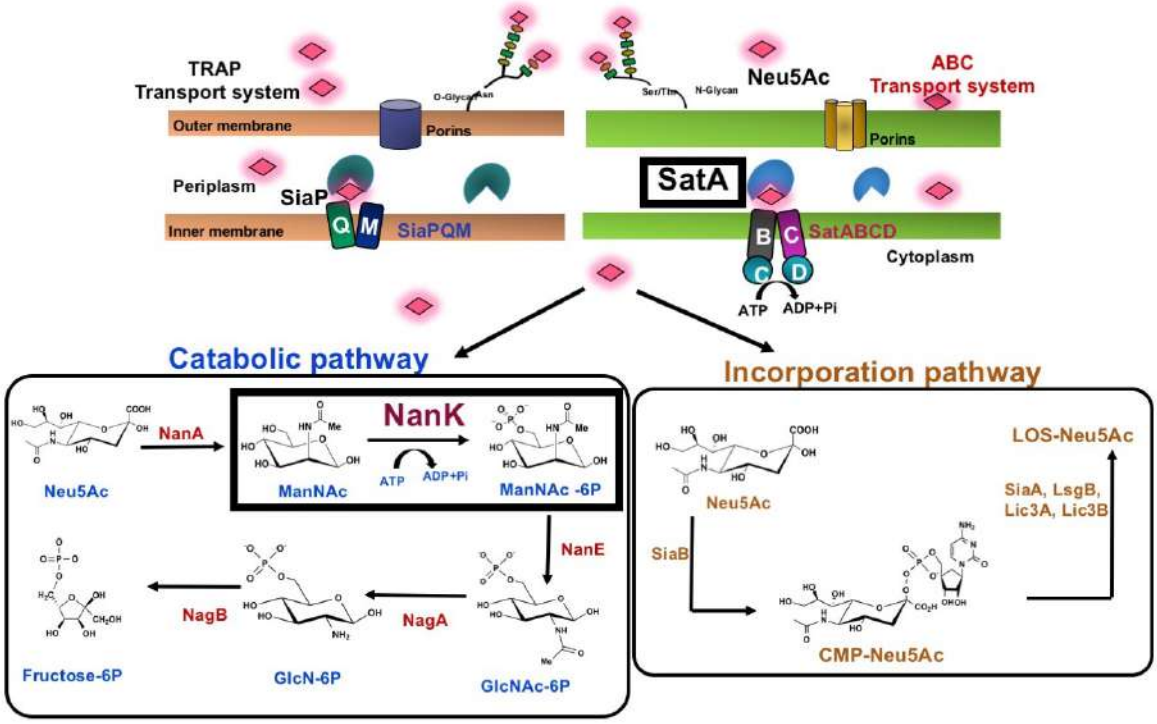


Figure 1: Schematic representation of sialic acid metabolism in bacteria. Summary of the various pathways of sialic acid utilization in bacteria. This figure represents the TRAP and ABC transport systems used by bacteria for scavenging sialic acids. It also represents the enzymes involved in the catabolic and incorporation pathways of sialic acid. Neu5Ac, N-acetyl neuraminic acid; NanA, N-acetylneuraminic acid lyase; ManNAc, N-acetylmannosamine; NanK, N-acetylmannosamine kinase; ManNAc-6-P, N-acetylmannosamine-6-phosphate; NanE, N-acetylmannosamine-6-P epimerase; GlcNAc-6-P, N-acetylglucosamine-6-phospahte; NagA, N-acetylglucosamine-6-phosphate deacetylase; NagB, Glucosamine-6-phosphate deaminase; Fructose-6P, Fructose-6-phosphate. Figure is modified from (12, 13).

The current study titled "**Structural and functional characterization of sialic acid uptake and its metabolism in pathogenic bacteria"** is divided into 5 chapters.

Chapter-I is an outline and review of literature on sialic acids and their metabolism in pathogenic bacteria and eukaryotes. **Chapter–II** describes the experimental procedures used in this study. Some of these experimental procedures are published in the journal "Microbial Biotechnology" titled "Automation aided optimization of cloning, expression and purification of enzymes of the bacterial sialic acid catabolic and sialylation pathways enzymes for structural studies". **Chapter-III** describes the studies related to the structural and functional

characterization of SatA from *Haemophilus ducreyi*. The results of this study were published in the "Journal of Biological Chemistry" in an article titled "Molecular characterization of the interaction of sialic acid with the periplasmic binding protein from *Haemophilus ducreyi*". **Chapter IV** describes the studies related to the structural and functional characterization of N-acetyl mannosamine kinases from pathogenic bacteria. The results of this study have been communicated and is currently under review. The manuscript is titled "To Zinc or not to Zinc– Structure and function of *N*-acetylmannosamine kinases from pathogenic bacteria". **Chapter V** discusses some future directions**.**

Chapter-III: **Structural and functional characterization of sialic acid binding protein of ABC transport system from *Haemophilus ducreyi (Hd*-SatA*).***

The ABC transporter that transports sialic acid in *H. ducreyi* is structurally encoded by a four-gene operon SatABCD. In this operon, SatA encodes for the periplasmic binding protein and SatBCD encodes for integral membrane protein and nucleotide binding domain. In this study, using X-ray crystallography, we structurally characterized *Hd*-SatA, where the unliganded and sialic acid-bound forms (with Neu5Ac and Neu5Gc) of *Hd*-SatA were determined to 2.2, 1.5, and 2.5 Å resolutions, respectively. Structural studies show that, *Hd*-SatA is a three-domain protein and is larger than other known SiaPs of the TRAP transport system. Similar to other SiaPs of the TRAP transport system, sialic acid binds between domains I and II of *Hd*-SatA. Domain III is a small extension of domain I, which does not interact with the sialic acid. Further, its thermodynamic properties were studied using isothermal calorimetry. Thermodynamic studies show that SatA binds to sialic acids with nanomolar binding affinity and is an entropically driven process. Comparison of the binding pockets of SatA and SiaP shows that SiaP has more charged residues in the binding pocket and most of the interactions with sialic acid are mediated by hydrogen bonds, ionic interactions, and salt bridges. In contrast, *Hd*-SatA has only two charged residues, Arg408 and Asp383 in the binding pocket. The other residues in the binding pocket of *Hd*-SatA are Tyr15, Ser27, Gly28, Ala29, Leu105, Pro381, His333, Asp383, and Trp397. Thus, the majority of interactions with sialic acids are mediated by either polar or hydrophobic interactions. The calculation of binding pocket volume in SatA and SiaP by CASTp server shows that upon ligand binding there is a large volume change in SatA. This large volume change contributes to expelling more water molecules, which creates randomness in the system that will further contribute towards the entropic binding effect in SatA. Based on the previous structural and functional studies of SiaP, we presumed that mutating few polar residues in the binding pocket to charged residues would increase enthalpic

contribution of *Hd*-SatA towards sialic acid binding. However, these substitutions did not contribute to the anticipated changes and the mutants behaved similar to the wild type protein. In summary, our biophysical characterization of SatA showed that bacteria have evolved to bind to similar ligands such as sialic acid with almost similar binding affinities with altered binding pocket topologies. Further, our comparative study of structural and functional properties of SiaP and SatA will be useful for future structure-based drug designing against these different pathogenic bacteria.

The studies of this work have been published in the Journal of Biological Chemistry:

Setty TG., Mowers JC., Hobbs AG., Maiya SP., Syed S., Munson RS Jr., Apicella MA., Ramaswamy S. (2018) Molecular characterization of the interaction of sialic acid with the periplasmic binding protein from *Haemophilus ducreyi*. **J Biol Chem.** 293(52): 20073-20084.

Chapter IV: Structural and functional characterization of N-acetyl mannosamine kinases from pathogenic and bacteria

Zn-binding motifs are important for structural integrity and enzymatic properties of proteins. Unlike traditional Zn-binding motifs, the Zn in NanK is coordinated by one histidine and three cysteines. Previous functional studies on *Bacillus subtilis* glucokinase, which belongs to the ROK family, have shown that mutating cysteine residues in the Zn-binding motif to alanine residues results in loss of enzyme activity (14). Our goal was to understand the functional properties of *Fn*-NanK that lacks Zn-binding motif (contains only His159 of the Zn-binding motif) in comparison to other NanKs from *Haemophilus influenzae (Hi), Pasteurella multocida (Pm),* and *Vibrio cholera (Vc)* that contain a Zn-binding motif. To study this, we carried out the kinetic studies of NanKs binding to its substrate ManNAc. Following this, we compared the kinetic parameters of *Fn*-NanK binding to ManNAc with *Hi*-NanK, *Pm*-NanK, and *Vc*-NanK. Interestingly, our kinetic studies did not show any major differences between these two groups (Zn binding and non Zn-binding) of NanKs.

Additionally, we have characterized the thermodynamic properties of these NanKs in binding to its substrates, ManNAc and AMPPNP (an ATP analogue). Isothermal calorimetry (ITC) studies show that *Pm*/*Vc*-NanK binds to ManNAc and AMPPNP in micromolar affinity range and follows an ordered binding pattern indicating that *Pm/Vc*-ManNAc complex formation is necessary for the binding of AMPPNP. In contrast, *Hi/Fn*-NanK also binds to these substrates in micromolar affinity range and follows a non-preferential binding pattern. Overall, these thermodynamic results demonstrate that Zn-binding motif has no role in the order of binding pattern.

Further, our structural studies on *Pm*-NanK and *Hi*-NanK using X-ray crystallography demonstrate that these are two-domain proteins. ManNAc interacts extensively with the N-terminal domain of NanK and is deeply buried inside the pocket between the N- and C- terminal domains. Superposition of the open and closed conformations of NanK structures shows that the N-terminal domain makes a 21^0 rotation towards the C-terminal domain to close upon ManNAc, which further brings the ATP closer to the substrate for phosphorylation. Compared to previous reports, this systematic structural and functional characterization of NanKs shows that although *Fn*-NanK lacks the Zn-binding motif, the presence of histidine alone is sufficient for holding/orienting the ManNAc in the binding pocket.

The structural analysis also demonstrates that Thr131 of NanK is an important and conserved residue in the binding pocket, which forms hydrophobic and hydrogen bonds with **β**-phosphate of the nucleotide in the open and closed conformations. The thermodynamic and kinetic studies of T131V mutant confirms that Thr131 is an important residue and is required for enzyme catalysis.

In summary, our biophysical characterization studies of NanKs demonstrate that even though NanKs have less sequence similarity, the signature motifs in these proteins are well conserved to carry out the phosphorylation step. Further, the biophysical characterization of these enzymes reveals the importance of critical residues that are involved in the interaction with ManNAc. These studies demonstrate the molecular level details of these kinases in the sialic acid catabolic pathway.

Overall, our research findings on sialic acid binding protein from the ABC transport system outlines the molecular level details of these proteins and their binding to sialic acids. Additionally, studying the structural and functional details of SatA from the ABC transport system in comparison to previously studied SiaPs of TRAP transport system will not only divulge the details of their interactions with substrates, but also delineate the strategies used by these pathogenic bacteria to survive in the host system. Our structural and functional characterization of NanKs from different pathogenic bacteria elucidates the role of Zn-binding motif in these kinases at molecular level. In future, our research findings will help in identifying/designing effective small molecule inhibitors against sialic acid binding proteins and kinases of several pathogenic bacteria that can act as potential drug targets.

References

1. Varki, A. (2017) Biological roles of glycans. *Glycobiology* **27**, 3–49
2. Swords, W. E., Moore, M. L., Godzicki, L., Bukofzer, G., Mitten, M. J., and VonCannon, J. (2004) Sialylation of Lipooligosaccharides Promotes Biofilm Formation by Nontypeable Haemophilus influenzae. *Infect. Immun.* **72**, 106–113
3. Vimr, E. R., Kalivoda, K. A., Deszo, E. L., and Steenbergen, S. M. (2004) Diversity of Microbial Sialic Acid Metabolism. *Microbiol. Mol. Biol. Rev.* **68**, 132–153
4. Vimr, E., Lichtensteiger, C., and Steenbergen, S. (2000) Sialic acid metabolism's dual function in Haemophilus influenzae. *Mol Microbiol* **36**, 1113–1123
5. Johnston, J. W., Coussens, N. P., Allen, S., Houtman, J. C. D., Turner, K. H., Zaleski, A., Ramaswamy, S., Gibson, B. W., and Apicella, M. A. (2008) Characterization of the N-acetyl-5-neuraminic acid-binding site of the extracytoplasmic solute receptor (SiaP) of nontypeable Haemophilus influenzae strain 2019. *J. Biol. Chem.* **283**, 855–865
6. Gangi Setty, T., Cho, C., Govindappa, S., Apicella, M. A., and Ramaswamy, S. (2014) Bacterial periplasmic sialic acid-binding proteins exhibit a conserved binding site. *Acta Crystallogr. D. Biol. Crystallogr.* **70**, 1801–1811
7. Mu, A., Severi, E., Mulligan, C., Watts, A. G., Kelly, D. J., Wilson, K. S., Wilkinson, A. J., and Thomas, G. H. (2006) Conservation of Structure and Mechanism in Primary and Secondary Transporters Exemplified by SiaP , a Sialic Acid Binding Virulence Factor from Haemophilus influenzae *. *J. Biol. Chem.* **281**, 22212–22222
8. Almagro-moreno, S., and Boyd, E. F. (2009) Sialic Acid Catabolism Confers a Competitive Advantage to Pathogenic Vibrio cholerae in the Mouse Intestine. *Infect. Immun.* **77**, 3807–3816
9. Larion, M., Moore, L. B., Thompson, S. M., Miller, B. G., May, R. V, Re, V., Recei, M., and August, V. (2007) Divergent Evolution of Function in the ROK Sugar Kinase Superfamily : Role of Enzyme Loops in Substrate Specificity. , 13564–13572
10. Caing-carlsson, R., Goyal, P., Sharma, A., Ghosh, S., Setty, T. G., North, R. A., Friemann, R., and Ramaswamy, S. (2017) Crystal structure of N -acetylmannosamine kinase from Fusobacterium nucleatum research communications. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **F73**, 356–362
11. Martinez, J., Nguyen, L. D., Hinderlich, S., Zimmer, R., Tauberger, E., Reutter, W., Saenger, W., Fan, H., and Moniot, S. (2012) Crystal structures of N-acetylmannosamine kinase provide insights into enzyme activity and inhibition. *J. Biol. Chem.* **287**, 13656–

13665

12. Severi, E., Hood, D. W., and Thomas, G. H. (2015) Sialic acid utilization by bacterial pathogens. *Microbiology* **153**, 2817–2822
13. Almagro-Moreno, S., and Boyd, E. F. (2009) Insights into the evolution of sialic acid catabolism among bacteria. *BMC Evol. Biol.* **9**, 118
14. Mesak, L. R., Mesak, F. M., and Dahl, M. K. (2004) Bacillus subtilis GlcK activity requires cysteines within a motif that discriminates microbial glucokinases into two lineages. *BMC Microbiol.* **4**, 1–10

DECLARATION

This thesis entitled "**Structural and functional characterization of sialic acid uptake and metabolism in pathogenic bacteria**" is a presentation of my original work. Wherever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature, and acknowledgment of collaborative research and discussions. The results presented in this thesis have not been submitted previously to The University of Trans-Disciplinary Health Sciences & Technology (TDU) or any other University for Ph. D. or any other degree.

The work was done under the supervision of **Prof. Ramaswamy Subramanian** at The Institute for Stem Cell Science and Regenerative Medicine (inStem), Bangalore.

Signature of the Candidate
(Thanuja Gangisetty)

In my capacity as supervisor of the candidate's thesis, I certify that the above statements are true to the best of my knowledge.

Signature of the Supervisor

Prof. Ramaswamy Subramanian

inStem, GKVK post, Bellary Road,
Bangalore 560065, India.

Date:

ACKNOWLEDGEMENTS

I have been fortunate to receive mentorship and guidance from many individuals pre- and post- journey as a graduate student. It is no trivial task to thank all the people who made this PhD thesis possible in a few short words.

First, I would like to thank my supervisor, Prof. S. Ramaswamy (Rams), who accepted me as his PhD student and provided me with support, guidance, and encouragement throughout my graduate studies. I am very grateful to him for the opportunities that he provided me with and I do not think I could have been more successful in any other lab. In particular, I would like to thank him for the freedom he gave me to explore, think, question, and learn. His zeal, enthusiasm, inquisitiveness, and passion are very inspiring, and he taught me to never underestimate myself. I sincerely thank him from the bottom of my heart and I will be obliged to him throughout my life. I am very thankful to Preeti (Mrs. Rams) for hosting lab members during different occasions and for her hospitality.

I am very thankful to my research thesis supervisory committee members, Prof. B. Padmanabhan (NIMHANS) and Dr. Minhaj Sirajuddin (InStem), for the helpful discussions, meetings, comments, and suggestions related to my research. I sincerely and specially thank Dr. Vinothkumar Kutti for his constant support and guidance during the last part of my PhD.

I would like to thank all the present and past Rams lab members for their help, suggestions, guidance, valuable conversations, and humorous chitchat, which made my time enjoyable in the lab. I would like to thank Vinod Nayak, Sucharita, and Sanchari for helping me understand the finer details of crystallography and for their help with structure determination. I want to thank Dr. Sudha for all her suggestions during my biochemical experiments. I thank Dr. Sabari, Dr. Vijayan, Dr. Parveen, Nitish, Dr. JayPrakash, Dr. Laavanya, Dr. Sai Rohit, Partha, Gowtham, Arunabha, Dr. Swagatha, Sowmya, Dr. Shubha Mayya, Dr. Nainesh, and Wendie. I would like to thank Dr. Chetan for his help with the Chem draw figures. I would like to extend my thanks to Vinoth lab members Dr. Lahari, Dr. Megha, Dr. Rahul, Mugdha, Sailesh, and Ananya. I would like to specially thank my dearest and long-term friend and colleague Dr. Sathya, who encouraged me throughout my journey as a student. I would like to convey my special thanks to Dr. Swapna for the critical reading of my thesis.

My heartfelt thanks to the Dolna team (crèche at NCBS/InStem) who made it possible for me to devote my time and attention in the lab. I am very grateful to all the teachers and staff of Dolna for taking care of my kids and ensuring that the kids felt at home.

I would like to thank Mr. Ravi Kumar who helped me with the TDU (Trans Disciplinary University) formalities and thesis submission. I want to thank Rashi, Vishalakshi, and Valsala for their help and support during PhD registration and the report and thesis submission processes. I want to thank Sharat (Travel desk) who helped me during my foreign travel to Soleil (France) for data collection and to Baltimore (USA) to attend the Biophysical society conference. I would like to thank Sunitha, Shobha, Radha, and Shrikant who helped me with all the finance related paperwork during my foreign travel and for the other formalities of the Council of Scientific Industrial Research (CSIR) fellowship.

I gratefully acknowledge the sources of funding support over the course of my PhD work. I received initial support from the Indo-Swedish collaborative grant and later from the Senior Research Fellowship scheme from CSIR. I would like to thank the Department of Science and Technology (DST) for providing me with travel fellowship to attend the Biophysical Society (BPS) Conference at Baltimore, USA. I also thank the InStem TTK travel grant for partially covering the expenses of the BPS conference.

I am very obliged for the infrastructure available at Bangalore Biocluster, without which none of this work would have been possible. In this regard, I greatly appreciate and acknowledge the staff at Kitchen, Instrumentation, IT, Stores, Administration (Purchase and Accounts), Reception and Canteen for the constant support they provided. My special thanks to Pankaj, Chetan, Chakrapani, Divya, Pooja, Ranjith, Manjunath, Allwyn, Prashanth and many others for all the help related to instruments, IT, and purchase.

I am very thankful to both my amma (Kususma Kumari) and daddy (Nageswara Rao), for providing the opportunity to be where I am at right now. Though their understanding of my lab work is limited, their moral support provides me with great strength. Both are wonderful grandparents to my kids, and I owe them huge thanks for their steady support. I also express my thanks to my brother Ramakrishna, sister-in-law Manasa, and all the other family members. My sincere regards to my mother-in-law Savitramma for her unconditional love and moral support during all times.

I owe special thanks to a very special person, my soulmate and husband, Subba Rao for his understanding, love, and patience. I am very grateful to him for all the support and endless encouragement that he provided me with to purse my dream. Despite the responsibilities associated with having two kids, he made my career a priority and I am very thankful to him for my whole life. He is always around me during tough times and I greatly appreciate his contribution and his trust in me. It means the world to me.

Last, I do not know how to thank my two lovely children Kintan and Shalvi. Considering that they were young kids when I started my PhD, it was hard to imagine balancing my role as a mother with my lab work, but their smiles and naughtiness worked as stress busters for me. They are both amazing kids and are the pride and joy of my life. I love them both for their support and patience during their mommy's PhD journey.

Finally, I give all the glory to God for this thesis and thank him for all the support and blessings.

Table of contents

Molecular characterization of the interaction of sialic acid with the periplasmic binding protein from *Haemophilus ducreyi*. *J Biol Chem* (2018) 293(52): 20073-20084.

ANNEXURE I

Bairy S, Gopalan LN, **Setty TG,** Srinivasachari S, Manjunath L, Kumar JP, Guntupalli SR, Bose S, Nayak V, Ghosh S, Sathyanarayanan N, Caing-Carlsson R, Wahlgren WY, Friemann R, Ramaswamy S and Neerathilingam M. Automation aided optimization of cloning, expression and purification of enzymes of the bacterial sialic acid catabolic and sialylation pathways enzymes for structural data. ***Microb Biotechnol*** (2018) Mar; 11(2): 420–428.

ANNEXURE II

Setty TG, Mowers JC, Hobbs AG, Maiya SP, Syed S. Munson RS Jr, Apicella MA and Ramaswamy S. Molecular characterization of the interaction of sialic acid with the periplasmic binding protein from *Haemophilus ducreyi*. ***J Biol Chem*** (2018) 293(52): 20073-20084.

List of Tables

List of figures

List of Abbreviations

Neu5Ac	N-acetyl neuraminic acid
Neu5Gc	N-glycolyl neuraminic acid
ABC	ATP-binding cassette transporters
MFS	major facilitator superfamily
TRAP	Tripartite ATP-independent periplasmic transporter
SSS	Sodium solute symporter
SBP	Substrate binding protein
TMD	Transmembrane domain
NBD	Nucleotide binding domain
SatA	Sialic acid transport A gene
SiaP	Sialic acid binding Protein
Hi	*Haemophilus influenzae*
Pm	*Pasteurella multocida*
Fn	*Fusobacterium nucleatum*
Vc	*Vibrio cholera*
E.coli	*Escherichia coli*
NTHi	Nontypeable Haemophilus influenzae
LOS	Lipooligosaccharide
LPS	Lipopolysaccharide
PCR	Polymerase chain reaction
WT	Wild type
NanK	N-acetyl mannosamine kinase
ITC	Isothermal calorimetry
ManNAc	N-acetylmannosamine
ATP	Adenosine triphosphate
ADP	Adenosine diphosphate

AMPNP	Adenylyl-imidodiphosphate
CMP	Cytidine monophosphate
CMAH	Cytidine Monophosphate-N-acetylneuraminic acid hydroxylase
Å	Angstrom
α	Alpha
β	Beta
γ	Gamma
σ	Sigma
SEM	Scanning electron microscope
Mg^{2+}	Magnesium ion
UDP	Uridine diphosphate
Zn	Zinc
K_d	Dissociate constant
mM	Millimolar
uM	Micromolar
nM	Nanomolar
EDTA	Ethylene diamine tetra acetic acid
SDS	Sodium dodecyl sulphate
ΔH	change in Enthalpy
ΔS	change in entropy

Chapter 1

Introduction

The “central dogma” of life in a cell is the translation of genetic information from DNA to RNA and then to protein. Apart from DNA, RNA, and proteins, the other two major fundamental macromolecules in a cell are lipids and carbohydrates (sugar chains/glycans). Glycans are a small number of monosaccharides that are glycosidically linked to each other and are usually attached to either proteins (glycoprotein) or lipids (glycolipids). Glycans adorn the cell surfaces of various organisms ranging from Archaea to eukaryotes. They play an important role in signal transduction, immunity, structural integrity, inflammation, and pathological interactions. Sialic acids are a class of modified monosaccharides that are often found at the outermost end of glycan chains (1).

1.1 Sialic Acids

Sialic acids are negatively charged nine-carbon keto sugars that are present as outermost sugars on glycoproteins, glycolipids, oligosaccharides, lipooligosaccharides (LOS), and lipopolysaccharides (LPS) of both eukaryotes and prokaryotes. These sugars were first isolated from the bovine submaxillary gland mucin by Blix et al., in the year 1936. They first named these sugars as “Kohlenhydrat I” and later renamed them to “sialic acids” (2). Klenk et al., in 1941 isolated nitrogen containing sugars from the neurons of brains and termed them “neuraminic acid” (3). Subsequently, the Warren laboratory in 1963 showed the absence of these sialic acids in plants (4). Structurally, sialic acids differ from other monosaccharides by having a carboxyl group at the first position and a glycerol chain at sixth position. In nature, around 50 different sialic acid isoforms occur and they differ from each other in their side chains (5). However, they are all grouped under ‘sialic acids’.

1.1.1 Sialic acids and their diversity

N-acetylneuraminic acid (Neu5Ac) is the most common sialic acid. The next three common sialic acids are neuraminic acid (Neu), N-glycolyl neruaminic acid (Neu5Gc), and deaminated neuraminic acid (Kdn). Carboxyl group at position 1 in sialic acids contributes to its pKa of 2. Acetylation of Neu gives rise to Neu5Ac. Neu5Ac and Neu5Gc differ from each other based on their difference in the hydroxyl group at C5 position. In contrast, the amino group at the fifth position in Neu5Gc/Ac is replaced by a hydroxyl group in Kdn (Figure. 1.1) (6).

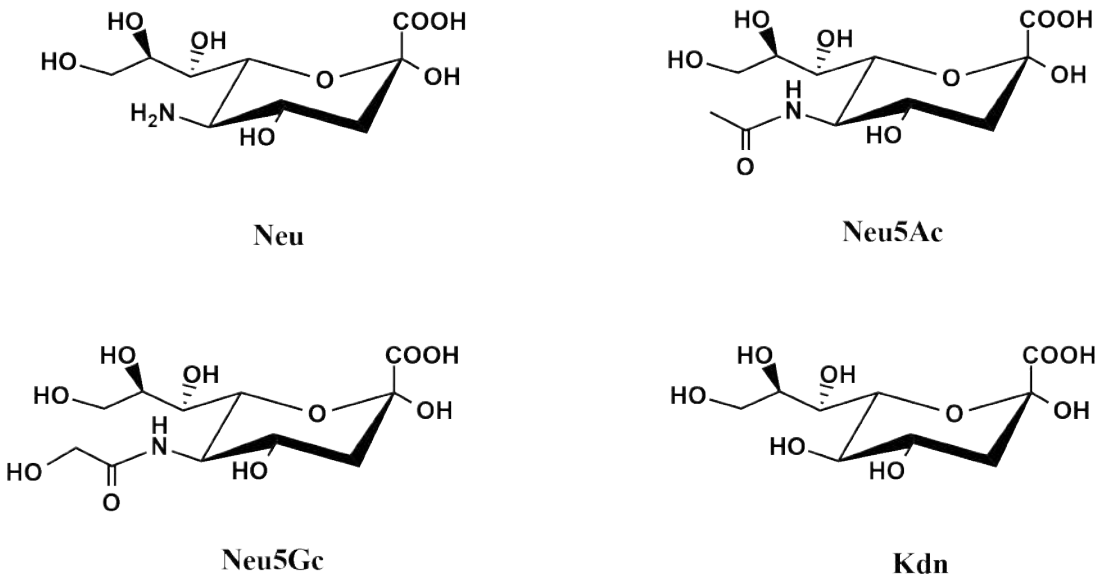


Figure. 1.1. The chemical structure of sialic acid and its analogues. The parent chemical structure of sialic acid family, Neu and its three major derivatives Neu5Ac, Neu5Gc, and Kdn are shown. Modified from Ref. (7).

Apart from the C5 position modifications, all other sialic acid isoforms are generated through modifications in the hydroxyl groups at C6, C7, C8, and/or C9 positions with acetylation, lactylation, sulfation, phosphorylation, or methylation (Figure. 1.2). Further, inter- and intra-lactomization between C1 and C5 carboxyl groups bring diversity in sialic acids. The distribution of sialic acids is different among different organisms (8, 9).

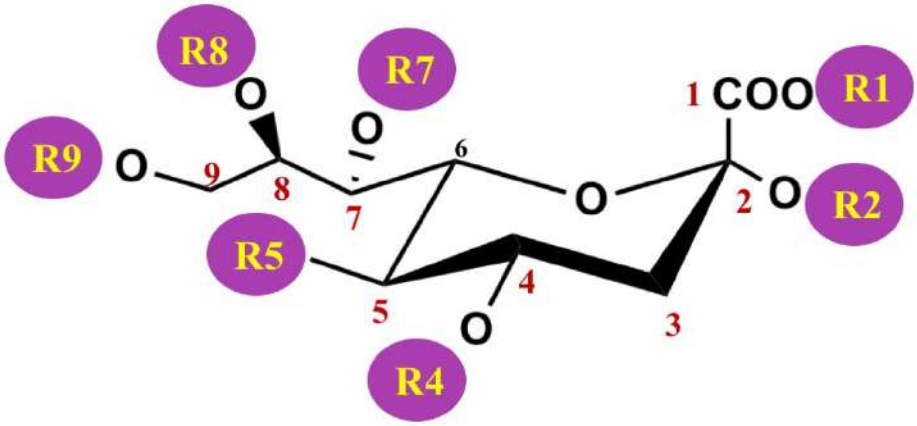


Figure. 1.2. The α configuration of sialic acid (Sia). The figure represents the possible variations at different carbon positions that bring diversity in sialic acids. Modified from Ref. (10).

1.1.2 Neu5Ac and Neu5Gc: Their relation to human evolution

According to Darwin's theory of evolution, all higher order organisms evolved from simplistic ancestors through natural selection, such that there is an increase in the ability to survive, compete, and reproduce. Additionally, the theory of evolution also suggests that "great apes" are the closest evolutionary relatives of the modern man. Consistently, the genetic variation between humans and apes is only 1-2% (11) and is responsible for the morphological and functional differences between humans and chimpanzees. Interestingly, the absence of Neu5Gc in humans is one of the first known differences between humans and the great apes based on a biological readout.

CMP-Neu5Ac gets hydroxylated into CMP-Neu5Gc by CMP-N-acetylneuraminic acid hydroxylase (CMAH) in the presence of an electron transport system consisting of cytochrome b5 and its reducing factors (Figure. 1.3) (12–14).

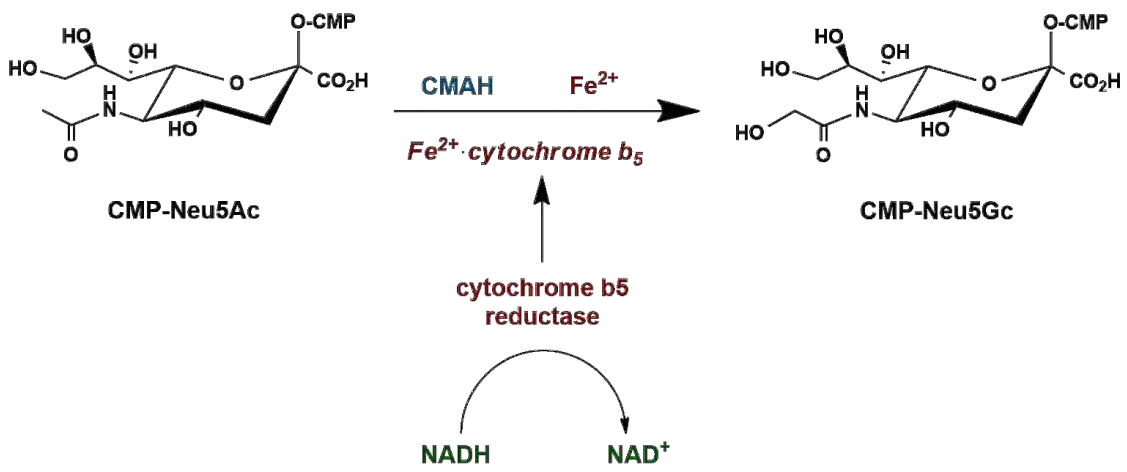


Figure. 1.3. Conversion of CMP-Neu5Ac into CMP-Neu5GC. Figure representing the conversion of CMP-Neu5Ac into CMP-Neu5Gc by CMAH in the presence of cofactor Cytochrome b_5. Modified from Ref. (14).

CMAH is an iron-sulfur protein of Rieske type oxygenase. During evolution, CMAH lost its activity in humans due to a frameshift mutation with the deletion of a 92-base pair exon (Figure. 1.4) (15). Thus, Neu5Gc is the outermost sugar in most mammalian cells, except in humans (16). This was further confirmed by the "serum sickness reaction" induced by infusion of animal serum in humans, which resulted in the generation of antibodies and agglutinated the red blood cells containing Neu5Gc from various animal species like horse, ox, and sheep (17). However, previous studies report detection of low levels of Neu5Gc in human tissues due to

the consumption of dietary products including red meat. Despite mutations in the CMAH gene, previous studies have also shown that Neu5Gc is present in the glycoconjugates of cancer and fetal tissues of humans (18).

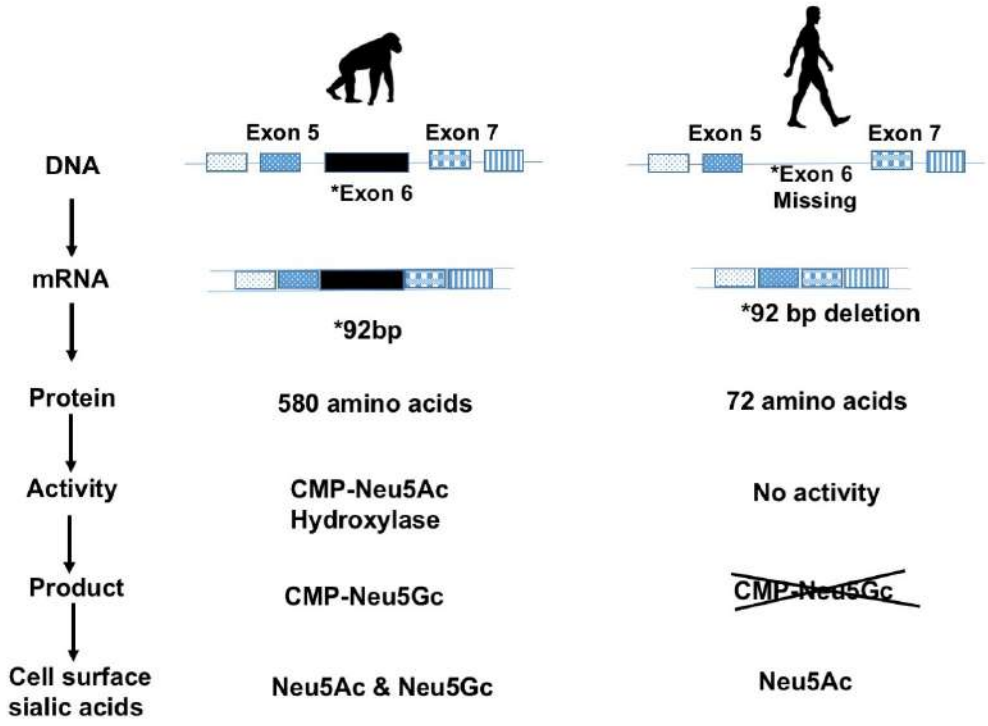


Figure. 1.4. Figure representing the molecular basis of CMAH gene expression in humans and chimps. Due to a mutation in the CMAH gene, 92 basepairs are deleted in exon 6 resulting in a "frameshift". The deleted region contains the amino acid sequence that is critical for the functioning of the enzyme. Thus, the enzyme made in humans is inactive and cannot produce Neu5Gc. Modified from Ref (19)**.**

1.2 Review on sialic acid metabolism in eukaryotes

Sialyation of glycan chains in glycoproteins and glycolipids in humans, vertebrates, and higher invertebrates is very important and is required for several cellular functions like adhesion and signalling. Sialic acids are also essential for the early embryonic development in mice (20). Apart from the cellular functions, these sugars play an important role in the regulation of immune response during pathogenesis and viral infections. The expression of sialic acids is altered in cancerous cells and also during embryogenesis (21). Biosynthesis of glycans takes place in cellular compartments, such as endoplasmic reticulum and Golgi, and involves glycosyl transferases, nucleotide sugar transporters, glycosidases, and glycan-modifying enzymes.

The process starts in the cytosol, either from fructose-6-phosphate formed after glycolysis or from N-acetyl glucosamine derived during nutrient digestion. In the first step, these molecules get amidated irreversibly to glucosamine-6-phosphate (GlcN-6-phosphate) via GlcNAc-6-phosphate synthase. In the second step, GlcN-6-phosphate is converted into GlcNAc-1-phosphate by GlcNAc-6-p-mutase. In the third step, GlcNAc-1-phosphate is activated by UTP and forms UDP-GlcNAc. Subsequently, UDP-GlcNAc is converted into ManNAc (with the release of UDP) and ManNAc gets phosphorylated to ManNAc-6P in the presence of a bifunctional enzyme UDP-GlcNAc 2-epimerase/ManNAc kinase (GNE). The N-terminal region of the GNE encodes for UDP-GlcNAc 2-epimerase and the C-terminal region encodes for ManNAc kinase (22). Further, ManNAc-6P is converted into Neu5Ac-9P through aldol condensation with phosphoenol pyruvate (PEP) and is catalysed by Neu5Ac-9-phosphate synthase. Then, Neu5Ac-9P is dephosphorylated by Neu5Ac-9-phosphate phosphatase to Neu5Ac. The final step of CMP-Neu5Ac synthesis occurs in the nucleus in the presence of CMP-Neu5Ac synthetase (Figure. 1.5). CMP-Neu5Ac is then transported into the Golgi apparatus by CMP-sialic acid transporter and then by several different sialyltransferases. Except humans, in other vertebrates, Neu5Ac is converted into Neu5Gc in the presence of CMAH in the endoplasmic reticulum. During the Neu5Ac biosynthesis, the UDP-GlcNAc 2-epimerase activity of GNE is regulated by feedback inhibition with CMP-Neu5Ac. Previous studies have shown that inactivation of GNE leads to early embryonic lethality in mice. Thus, GNE and sialic acid synthesis are very important during development (20). The heterozygous missense mutation in the allosteric site of GNE leads to a loss in feedback inhibition by CMP-Neu5Ac, which, in turn, results in overproduction of sialic acid, observed in the human disease sialuria (23, 24). Patients with this disease contain quantities of sialic acid in grams in urine and also have increased sialic acid levels in the cytoplasm of cultured fibroblasts (25, 26).

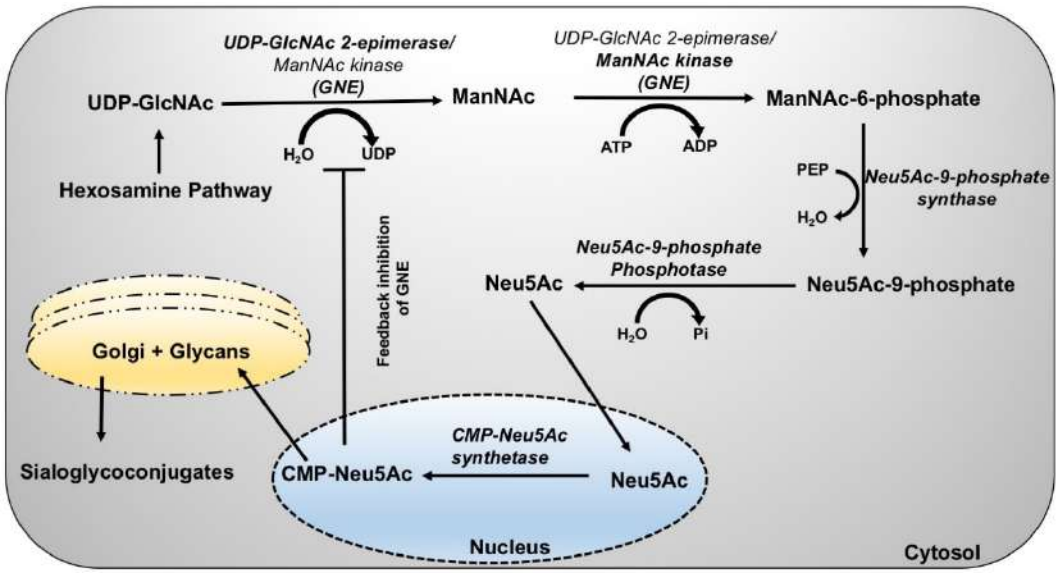


Figure. 1.5. Sialic acid biosynthesis in vertebrates. UDP-GlcNAc is derived from fructose-6-phosphate via the hexosamine pathway. GNE catalyses the first two steps of this pathway. The enzymatic reactions are indicated in bold. The feedback inhibition of UDP-GlcNAc 2-epimerase activity of GNE is controlled by CMP-Neu5Ac, which is formed in the nucleus. The figure is modified from Ref. (23).

1.3 Review on sialic acid significance in bacteria

The human body is inhabited by several symbiotic and pathogenic bacteria. They reside in a wide range of environments and are exposed to different chemical and physical environments. In some environments, bacteria experience sudden osmotic pressure in the host system. To counter the stress, the bacterial surface is covered by a dense array of glycan molecules, which help the bacteria interact with the host metabolic system in addition to escaping the host immune system. Bacteria are classified into Gram-positive and Gram-negative bacteria based on their cell wall composition. Gram-negative bacteria contain both a plasma membrane and an outer membrane. The outer membrane of Gram-negative bacteria contains LPS and LOS, whereas Gram-positive bacteria lack the outer membrane but they contain polysaccharides like teichoic acids (WTA) and lipoteichoic acids (LTA) that are inserted into the cytoplasmic membrane. The peptidoglycan layer between outer and inner membranes is made of N-acetylglucosamine and N-acetylmuramic acid (MurNAc). This layer offers protection towards sudden environmental changes and the peptidoglycan layer is thinner

in Gram-negative bacteria than in Gram-positive bacteria (27). LPS is composed of 3 components: (1) lipid A moiety (endotoxin) (2) inner and outer polysaccharide core and (3) and outer most O-antigen. Lipid A moiety is hydrophobic in nature and serves as an anchor between the bacterial outer membrane and the LPS. It contains 6 acyl chains and provides the rigidity in anchoring to the membrane. Inner polysaccharide core consists of 3-deoxy-D-manno-octulosonic acid (Kdo) and L-glycero-D-manno heptose sugars (Hep). The outer core contains a polymer of saccharides, called O-polysaccharide. Most of the pathogenic bacteria that reside in the mucus-rich regions do not contain these long polysaccharide chains, and they contain LOS, which is a truncated form of LPS (Figure. 1.6). The common sugars in this region are hexoses and hexosamines like glucose, galactose, N-acetyl galactosamine, and N-acetyl glucosamine. Based on the differences in the structure of lipid A moiety, inner core, outer core and polysaccharide chains, different bacteria have different kinds of LPS that confers serum resistance to bacteria.

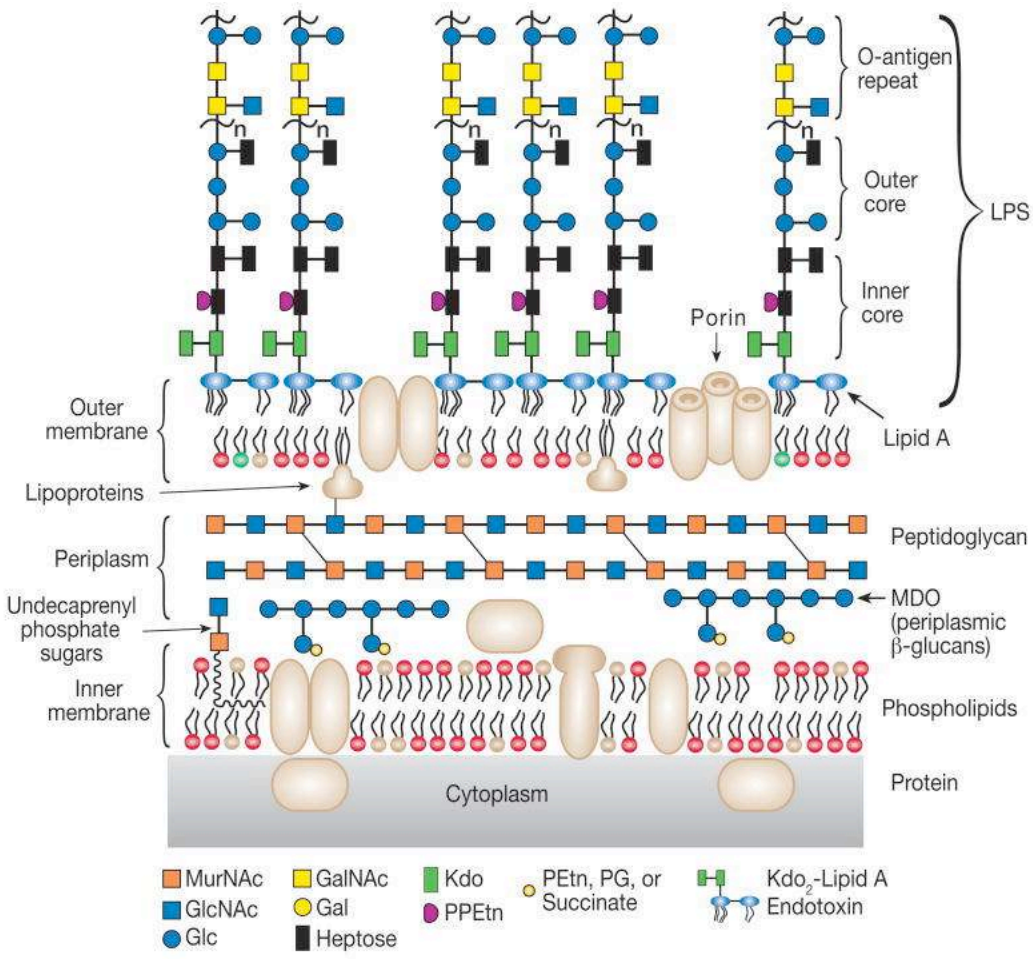


Figure. 1.6. Cell wall of Gram-negative bacteria. Schematic representation of the cell wall of Gram-negative bacteria showing inner membrane, periplasmic peptidoglycan layer, and an outer membrane rich in lipopolysaccharide (LPS). Figure is adapted from Ref. (28).

Bacteria have evolved with different biosynthetic pathways to synthesize these different glycans in a variety of host environments. Glycosylation is one of the most important post-translational modifications and glycans play a crucial role in influencing immune signalling and its activation. In bacteria, glycans are assembled on a lipid-linked precursor, where oligosaccharyl transferase transfers these moieties onto the external surface of the cell through en bloc mechanism (28, 29). The capsular polysaccharides, LPS, and flagellum of bacteria are decorated with glycan molecules, which terminate with negatively charged sialic acids (Figure. 1.7). These molecules represent the first interface between bacteria and the respective mammalian host. Further, by having sialic acids as the outermost sugars on cell wall, bacteria mimic the eukaryotic host cells and escape the host immune system in a mechanism called 'molecular mimicry' (Figure 1.7). In addition, sialic acids also play an important role in biofilm formation.

Glycosylation of TibA in *E. coli* is essential for the dynamics of its folding and also helpful for the adhesion properties of this protein (30). Similarly, pilins (surface hair-like extensions) in *Neisseria gonorrhoeae* and *Pseudomonas aeruginosa* are glycosylated and they are important for their adhesion (Figure 1.7) (31). In case of *Campylobacter Jejuni* and *Heliocobater pylori*, glycosylation of flagellin is important for their motility and also helps them colonize the host (32, 33)(Figure 7).

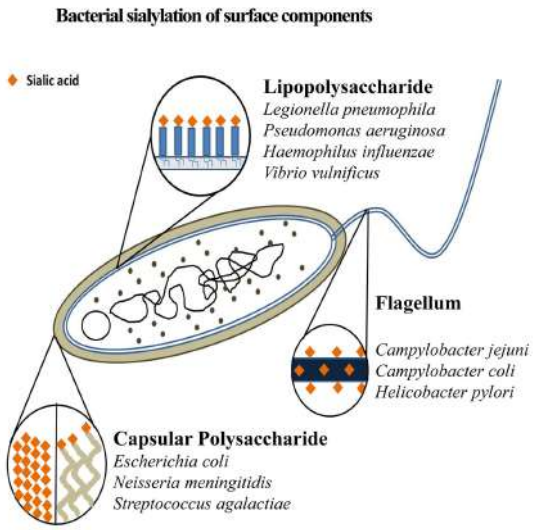


Figure. 1.7. Sialylation of surface components in bacteria. This diagram depicts the different surface structures in bacteria that are known to be decorated with nonulosoinc acids (neuraminic, pseudaminic, or legionaminic acids). Figure also indicates the different bacterial species along with their sialylated surfaces. Figure is adapted from Ref. (34).

Encapsulated strains of *Haemophilus influenzae* causes septicemia and meningitis in kids and non-encapsulated (Nontypeable *Haemophilus influenzae,* NTHi) strains cause otitis media and bronchitis. The LOS of NTHi is a heterogeneous mixture of glycoforms and sialic acids are attached as terminal sugars to N-acetyl hexosamine with an alpha2,6-linkage (Figure 1.7). During the process of infection, these pathogenic bacteria colonize and form biofilms and reside on the exopolymer matrix. The levels of sialylated LOS increases during biofilm formation (35–37). Previous studies have shown that the addition of sialic acid promotes biofilm formation in *H. influenzae* (36). In the continuous-flow model, NTHi2019 has been shown to grow in a medium containing Neu5Ac and its growth is significantly reduced in media lacking sialic acid. NTHi biofilm formation was also previously shown in chinchilla middle ear infection model, microtiter plate biofilm assay, and O'Toole-Kolter microtiter plate assay (Figure. 1.8) (37, 38). Interestingly, mutations in sialyltranferase (SiaA) or mutations in CMP-sialic acid synthetase (SiaB) show defects in biofilm formation and leads to attenuation of bacterial virulence in chinchilla model. Thus, the loss of Neu5Ac in *H. influenzae* makes it to susceptible to death mediated by exposure to human serum.

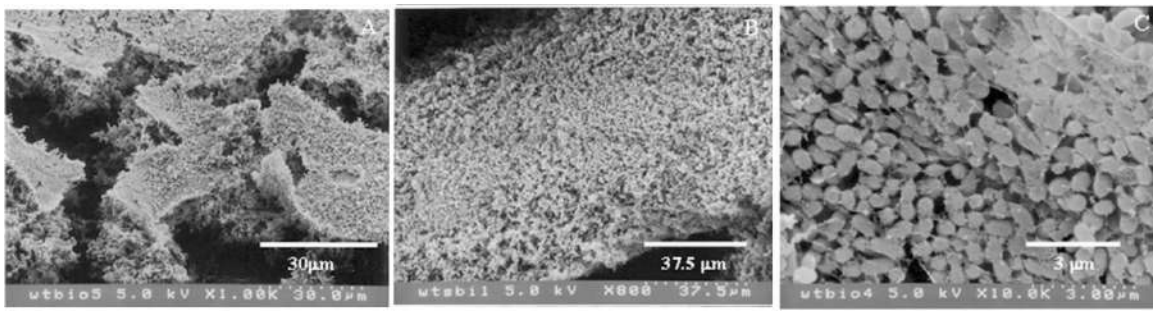


Figure 1.8. SEM micrographs showing the NTHI biofilm formation. Different views of biofilms- (A) top surface, (B) cross-sectional view and (C) higher magnification of cross-sectional view. Figure is adapted from Ref. (37).

1.4 Sialic acid metabolism in bacteria

1.4.1 Sialic acid biosynthesis and its acquisition

Bacteria acquire sialic acids by three different methods: (i) *de novo* biosynthesis (ii) donor scavenging (iii) precursor scavenging (39–41).

(i) De novo biosynthesis: Bacteria, such as *E. coli K1, Neisseria meningitides,* and *C. jejuni* synthesize sialic acid *de novo*. In this pathway, UDP-GlcNAc is converted into ManNAc by the enzyme NeuC and then to Neu5Ac by NeuB in presence of phsopho enol pyruvate (PEP). Further, by the action of CMP-sialic acid synthetase (NeuA), Neu5Ac is converted into CMP-Neu5Ac.

(ii) Donor scavenging: *N. gonorrhoeae* is causative of a sexually transmitted disease in humans and it scavenges CMP-Sialic acid from the host. *N. gonorrhoeae* expresses sialyltransferase that is homologous to the enzyme found in *N. meningitides* and it catalyses the transfer of sialic acid to LOS (41–43).

(iii) Precursor scavenging: Most of the pathogenic bacteria like *H. influenzae, F. nucleatum, P. multocida, and V. cholera* reside in the mucus-rich regions of humans and scavenge sialic acids from the host. Some pathogenic bacteria secrete sialidases to cleave host-derived sialic acids or they scavenge free sialic acids that are released by the sialidases, secreted by the neighbouring bacteria in the niche.

1.4.2 Sialic acid transport in bacteria

1.4.2.1 Porins: Sialic acid transporters on the outer membrane

Sialic acids released by the sialidases enter into the periplasmic space of bacteria through passive diffusion using porins like OmpC, OpmF, and NanC that are present in the outer membrane. In bacteria, there are general/nonspecific porins and solute-specific porins. Under experimental conditions, *E. coli* uses OpmF and OmpC for the transport of sialic acids. Under low sialic acid conditions, the transport of sialic acids is carried out by specific channels. During these conditions, suitable residues in the pore bind to sialic acids and help the diffusion of these molecules into the periplasmic space. Interestingly, temporal gene expression studies based on microarray analysis in *E. coli* show a five-fold increase in NanC expression in biofilms, when compared to their suspension cultures. NanC is an orthologue of KdgM and is a solute specific porin. This protein contains 12 strand beta barrel structure with an average pore radius of 3.3 A^0. The inner pore of NanC is lined with two tracks of basic residues, which

face each other and help in the translocation of acidic sugars like Neu5Ac (44, 45). *T. forsythia* is a Gram-negative bacterium that causes periodontitis in humans. Recently, it has been reported that NanO and NanU genes form a pair of outer membrane sialic acid transporters, wherein NanU binds to sialic acid (released by sialidases) and is transported to the periplasm by NanO (Figure 1.9). TonB-ExbB-ExbD complex provides energy to NanO for sialic acid transport (46).

Bacteria live in challenging environments, where the availability of nutrients is limited. In these conditions, solutes are transported across the inner membrane from lower to higher concentration, against the concentration gradient. This active transport requires energy and is provided by ATP hydrolysis or electrochemical gradient. Based on the source of energy, there are two types of active transporters in bacteria: (i) primary active transporter (ii) secondary active transporter (Figure. 1.9).

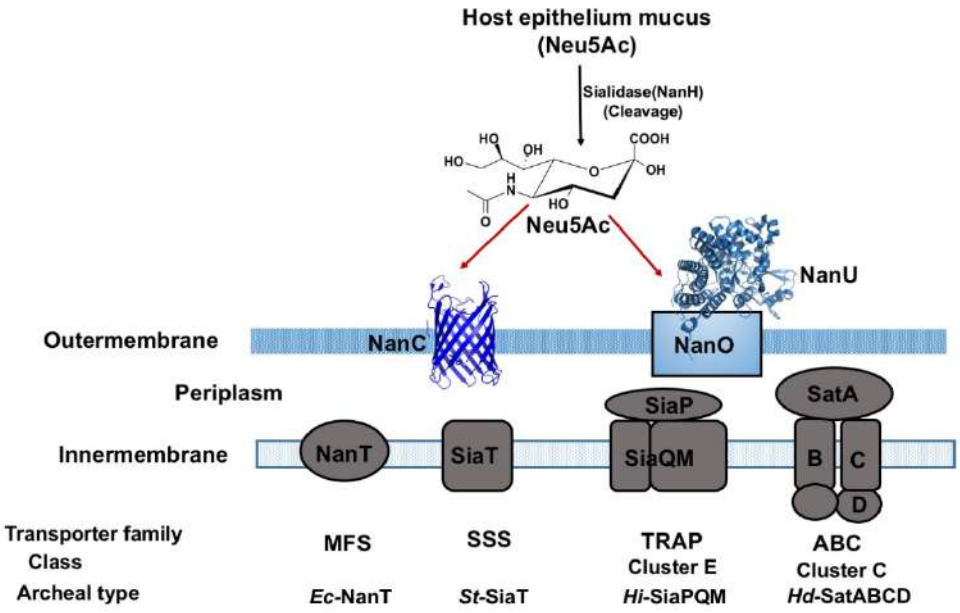


Figure. 1.9. Schematic representation sialic acid transport in bacteria. The first step in sialic acid catabolism is uptake of free sialic acid molecules from the host that are cleaved by a sialidase NanH, localized across the cell wall and cell membrane. Four bacterial sialic acid transporters (MFS, SSS, TRAP, and ABC) in the inner membrane and two in the outer membrane are depicted in the figure. Chemical structure of sialic acid is also represented here. Figure is modified from Ref. (34, 47).

1.4.2.2 Primary active transporters: ABC transporters

ABC class of transporters, called primary active transporters, use the energy driven by ATP hydrolysis to import a wide variety of substrates. ABC transporters are classified into importers and exporters. They import different substrates like sugars, vitamins, amino acids, and large organic compounds across the membrane and they export toxins, drugs, and lipids across the membranes. They are also involved in different cellular processes like signal transduction, protein secretion, antigen presentation, and bacterial pathogenesis. ABC class of importers are only present in prokaryotes, whereas exporters are present in both prokaryotes and eukaryotes, including humans. In bacteria, ABC transporters are generally assembled into four subunits- two transmembrane domains (TMD) and two nucleotide binding domains (NBD). In Gram-negative bacteria, importers also contain a substrate-binding protein in the periplasm that traps the substrate and delivers it to the TMD (48, 49). As these proteins transport a diverse range of substrates, they lack sequence similarity in their TMDs. However, these proteins are grouped into one class based on the presence of a signature motif in the ATP-binding domains: Walker-A motif (P loop) GXXGXGKS/T (X can be varied); and Walker-B motif hhhhD (h stands for hydrophobic) and a linker peptide LSGGQQ/R/KQR (50).

Some of the examples of ABC transport system are as follows:

Example 1: ABC transporter from *E. coli* –maltose transport system:

E. coli uses the maltose transport system MBP-MalFGK$_2$ (MalEFGK$_2$) for the transport of various nutrients like maltose and maltodextrins. It contains an outer membrane channel maltoporin (LamB), periplasmic maltose binding protein (MBP or MalE), two transmembrane domains (MalF and MalG) (make channel for transport of substrates) and two nucleotide binding domains (NBD or MalK$_2$) (51). Maltoporin is a homotrimer of β barrel protein and each barrel contains a pore to transport the substrate from outside. In the periplasm, the substrate binds to the maltose binding protein (MBP), which is composed of two globular domains. It has central β –pleated sheet and is flanked by 3 α -helices. Once the ligand binds to the MBP, it gets trapped between both the domains of MBP, similar to the Venus-fly trap model. Ligand binding induces a conformational change in the protein and the unliganded and the ligand-bound forms are always in equilibrium. Further, ligands establish interactions with the binding proteins by forming hydrogen bonds and van der Waals interactions with the residues in both the domains. The two domains in MBP are connected by 3 hinge regions and upon ligand binding, there is a 35^0 hinge bending and an 8^0 twisting motion between the two

domains (52). The MBP binds to maltose, amylose, cyclic maltodextrins, and its derivatives. Structural analysis of maltose transporters shows that substrate specificity is decided by both the MBP and $MalFGK_2$ (53). Membrane bound receptors bind preferentially to the ligand-bound MBP when compared to the unliganded MBP (54). For the transport of substrates from the periplasm to the cytoplasm, TMDs adopt two conformations- inward facing conformation (inactive, open towards cytoplasmic side, ADP bound form) and outward facing conformation (active, open towards periplasmic side, ATP bound form). In the nucleotide free form, NBDs are separated and their contact is maintained through the C-terminal regulatory domains. In the ATP bound form, the two NBDs dimerize in the presence of two molecules of ATP. This structural information suggests that ATP is required for the dimerization of NBDs. Further, binding of MBP to the TMDs induces a conformational change in the TMDs and this further triggers ATP hydrolysis. This helps in the release of the substrate from MBP to the TMD. In the ADP bound form, the two NBDs are separated, which leads to the release of the substrate into the cytoplasm, followed by bringing the transporter back to the resting state (Figure 1.10, 1.11A, 1.12A) (53–56).

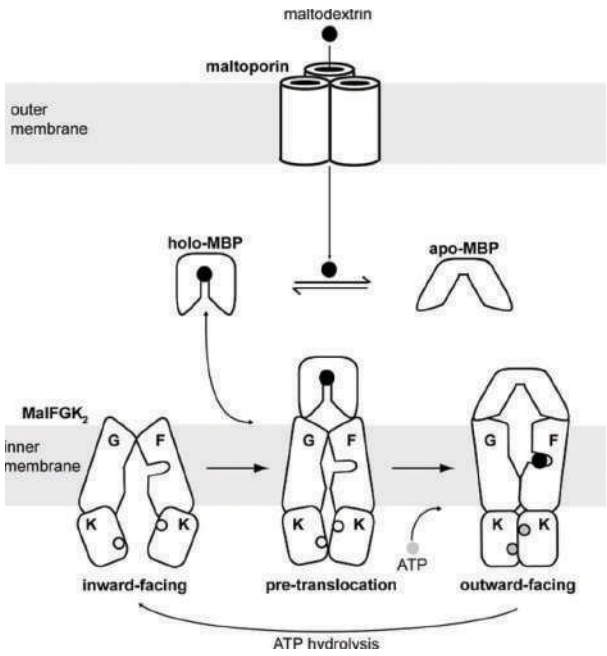


Figure. 1.10. Pictorial representation of maltose transporter showing the transport of substrates in *E .coli.* Figure is adapted from (53)**.**

Example 2: ABC transporter from *Archaeoglobus fulgidus* (*A. fulgidus* $ModB_2C_2$):

Similar to the maltose transporter of *E. coli*, the ABC transporter of *A. fulgidus* has two TMDs with 12 transmembrane segments. The protein has two cytoplasmic NBDs. These

hydrolyse ATP and provide energy for the transport of molybdate. ModA is the substrate binding protein (SBP) in this transporter that binds molybdate and tungstate. ModA also has two domains and they are connected by hinge regions. The complete structure of $ModB_2C_2$ in complex with ModA clearly shows that the mouth of ModA is placed exactly on top of the transmembrane domain $ModB_2$ for the translocation of the substrate. The surface of the ModA contains several charged residues that contribute to its interaction with $ModB_2$. Overall, these structural studies also reveal a requirement of 2 ATP molecules for the transport of every molybdate (Figure 1.11B, 1.12B)(57).

Example 3: ABC transporter from *E. coli* for Vitamin B12 transport (BtuCD and BtuF):

BtuCD is another ABC transporter that transports vitamin B_{12} and consists of two copies of BtuC and BtuD (58). BtuF acts as a periplasmic SBP of this transporter and is a bi-lobed protein, where vitamin B12 binds in a deep cleft formed by these two lobes of the protein. Structural studies show that two surface glutamates of BtuF interact with the surface arginines of BtuCD transporter (Figure 1.11C, 1.12C) (59).

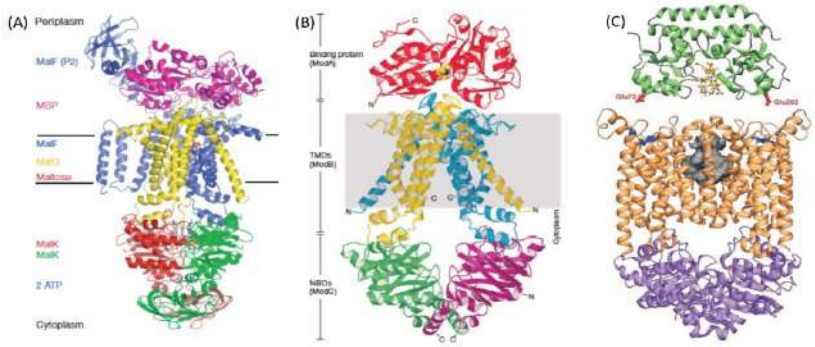


Figure. 1.11. Structure of ABC transporters from bacteria.

(A) Cartoon representation of maltose transport system in *E. coli* showing the transport of maltose. Colour codes: magenta, MBP; blue, MalF; yellow, MalG; red and green, MalK dimer. Figure is adapted from Ref. (55).

(B) Structure of $ModB_2C_2A$ complex from *Archaeoglobus fulgidu*. Colour codes: red, Substrate binding protein ModA; yellow and blue, transmembrane subunits ModB; green and magenta, nucleotide binding domain ModC. Figure is adapted from Ref. (57).

(C) Structure of BtuFCD from *E. coli*. Periplasmic binding protein BtuF is represented in green color; BtuC and BtuD in orange and purple respectively. Figure is adapted from Ref. (59).

The SBPs of these ABC transporters vary in their sizes (from 25-60 Kda) and possess little sequence homology. Thus, based on structural and functional characterization studies, the periplasmic binding proteins of these different bacteria have been classified into different groups. In spite of the low sequence homology, the tertiary structure of different SBPs is very similar, where they contain two globular domains and the ligand binds between the two domains. Structurally, the opening and closing of the domains around the ligand is similar to the Venus flytrap mechanism and is controlled by the hinge regions (Figure 1.12)(60, 61).

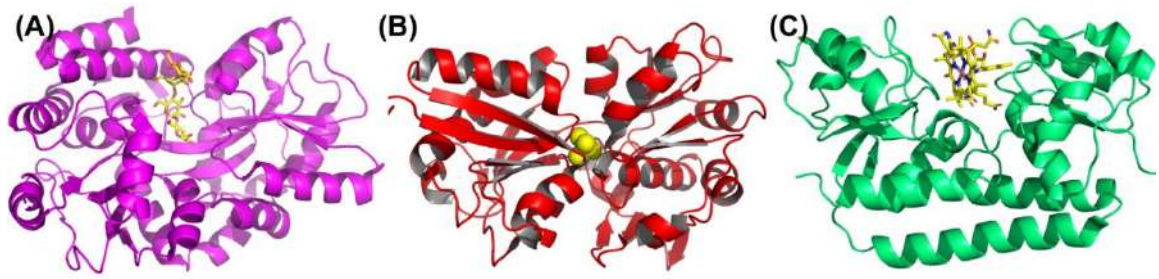


Figure. 1.12. Substrate binding proteins from different ABC transporters. (A) MBP from maltose transport system of *E. coli*: (PDB : 4khz) (B) ModA from molybdate transport system of *Archaeoglobus fulgidu*: (PDB ID: 2ONR) and (C) BtuF from Vit B12 transporter of *E. coli*: (PDB ID: 1N2Z).

1.4.2.3 Secondary active transporters

Secondary active transport is a mechanism where solutes are transported across the membrane due to the difference in electrochemical gradient established by the transport of ions, such as H^+ and Na^+, across the membrane. The three main families in this category are Major Facilitator superfamily (MFS), Sodium Solute Symporter (SSS), and Tripartite ATP-independent periplasmic (TRAP) transporters.

1.4.2.3.1 MFS transporter

The MFS transport system is one of the major transport systems found in bacteria, archaea, and eukaryotes. Unlike ABC transporters, where they transport both small molecules and macromolecules, the MFS transporters transport only small solutes in response to the electrochemical ion gradient. Based on the direction of solute/ion transport, MFS family of transporters are classified into uniporter, symporters, and antiporters (USA). They transport essential nutrients and export toxic substances or end products of the metabolism. The different substrates that are transported by MFS transporters include antibiotics (H^+/tetracycline antiporter), sugars (mammalian glucose carrier and the H^+/lactose permease of *E. coli*), Krebs cycle intermediates (H^+/citrate symporter), and phosphate/phosphate esters (Hexose phosphate antiporter). Most of these transporters contains 12 transmembrane alpha-helical spanners (Figure. 1.13) (62) .

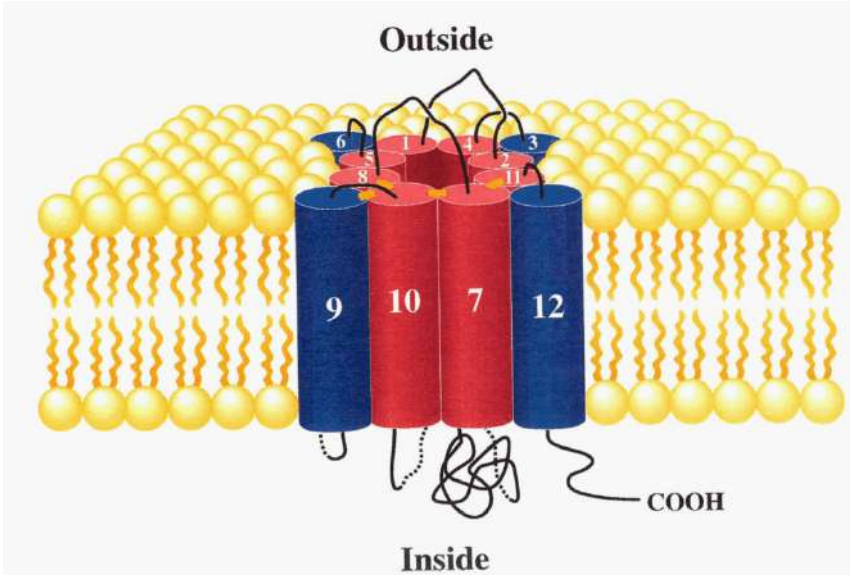


Figure. 1.13: Schematic representation of arrangement of transmembrane domains within the USA superfamily. Figure is adapted from (62)**.**

Based on sequence similarity, Saier Jr. et al., constructed a phylogenic tree and classified the members of the MFS transport system into 17 families. Distinct structurally related compounds are transported by the members of these families (63).

Example 1: NanT, which transports sialic acid, belongs to MFS transport system. NanT was first identified in *E. coli* and its mutation leads to loss of function. Unlike the traditional MFS family members, this protein contains 14 predicted transmembrane helices and an amphiphilic alpha helix that might play an important role in structure and function of the protein (39, 64).

Example 2: In *E. coli*, the LacY uses the energy generated by the proton transfer for the energetically uphill stoichiometric accumulation of galactose against its concentration gradient (65). The monomer of LacY contains N- and C-terminal domains with six transmembrane helices each forming two helical bundles, connected by a long loop between VI and VII helices. The previously reported structure clearly shows the presence β-D-galactopyranosyl-1-thio-β-D-galactopyranoside (TDG) in the large hydrophilic cavity formed by the two domains. The topology of N- and C-terminal domains is similar to most of the other MFS transporters (Figure. 1.14) (65).

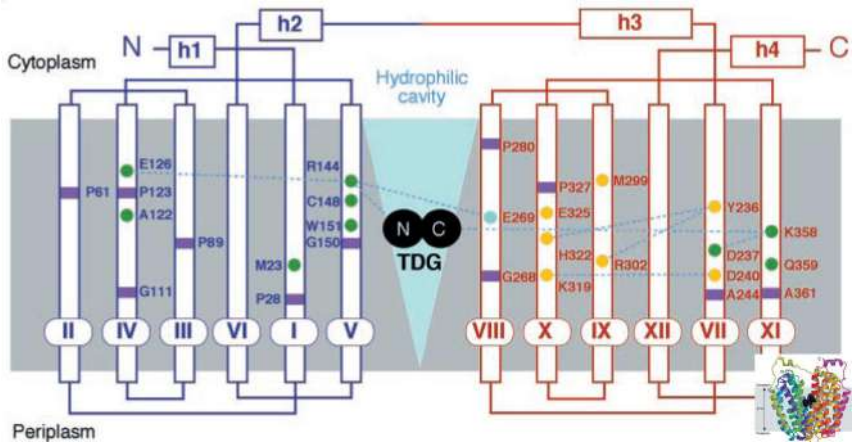


Figure. 1.14: Schematic secondary structure of LacY (Lactose permease from *E. coli*). The inset shows the overall structure of LacY. Figure is adapted from (65).

1.4.2.3.2 SSS transporters

In the SSS transport system, Na+ ion electrochemical gradient is used for the uptake of sialic acids. The Na+/galactose/glucose symporter from *Vibrio parahaemolyticus* (vSGLT) is the first representative of the SSS transport family (66). The SSS transporter in *Drosophila melanogaster* requires both Na+ and Cl^- ions for the transport of neurotransmitters and is called neurotransmitter sodium symporter (NSSs) (67). STM1128, from the human pathogen

Salmonella enterica serovar Typhimurium, is a reversible seconadary transporter that transports sialic acid and its activity is dependent on two Na^+ ions (68). The uropathogen *Proteus mirabilis* transports and catabolizes host derived sialic acids (SiaT) using the SSS transport system. The functional and molecular dynamic studies of SiaT shows that uptake and transport of sialic acid is also controlled by two Na^+ ions (Na2 and Na3). The structural studies show that Na2 occupies a conserved site and the Na3 occupies a unique position, which is present at ~14 Å away from the substrate-binding site and 6.5 Å away form the Na2. Functional studies show that a mutation of Na2 abolishes the transport function, whereas a mutation of Na3 does not show any subtle difference. These results suggests that Na3 might be palying a modulatory role by pre-organizing the binding site, which further increases the binding affinity for the substrate (68, 69).

1.4.2.3.3 TRAP transporter

The tripartite ATP-independent periplasmic transporters (TRAP) are extra cytoplasmic solute receptors (ESRs) that belong to the category of secondary transporters and are present only in bacteria and archaea. Similar to the MFS transport system, electrochemical ion gradient provides the driving force for the transport of solutes across the membrane. Unlike conventional MFS and SSS transport systems, the TRAP transport system contains an extra SBP to trap the substrate in the periplasm and confers unidirectionality to substrate transport (70). TRAP transport system is a three-component system: an extracellular solute receptor (ESR, periplasmic binding protein) and two unequal sized distinct integral membrane proteins (Figure. 1.15). In Gram-negative bacteria, solute binding proteins are present in the periplasm and in Gram-positive bacteria they are anchored to the inner membrane of the bacteria.

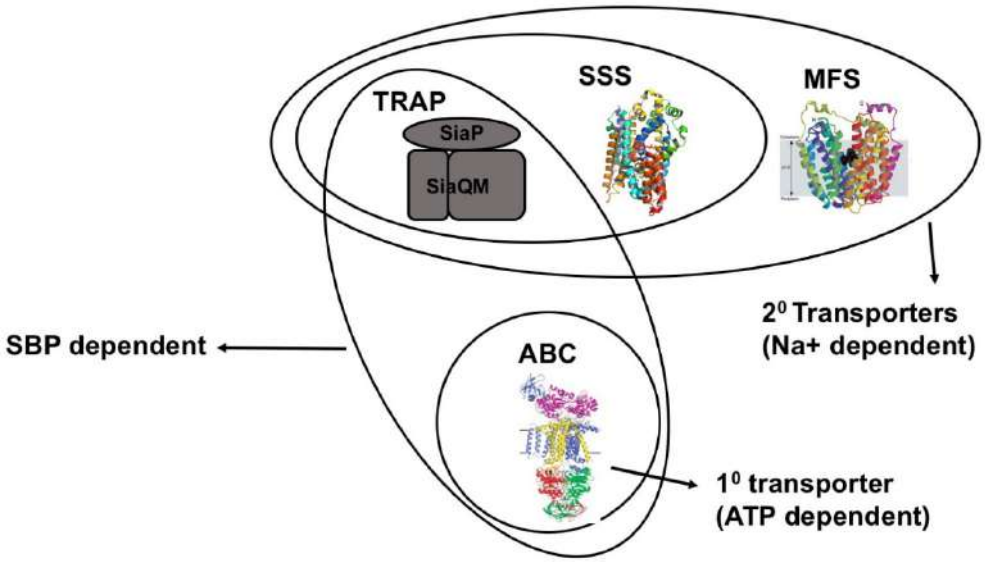


Figure. 1.15. Venn diagram showing the commonalities between different transport systems. Figure is modified from Ref (47)**.**

TRAP transporters are first identified in *Rhodobacter capsulatus*, encoded by dctPQM genes and they transport C4-dicarboxylates (71). Sialic acid transport system in *H. influenzae* is a classic example of the TRAP transport system. Similar kind of sialic acid transport system is also identified in *F. nuleatum*, *V. cholera,* and *P. multocida* that cause different diseases in humans and animals. Overall, sialic acid transport in all these bacteria is driven by an electrochemical ion gradient but not ATP hydrolysis.

1.4.3 Sialic acid utilization and their role in the pathogenicity in humans and animals by different bacteria such as *Haemophilus influenzae*, *Fusobacterium nucleatum*, *Pasteurella multocida,* and *Vibrio cholera*

Nontypeable (non-encapsulated) form of *H. influenzae* (NTHi) colonizes in the nasopharynx. In this bacterium, sialic acids are incorporated as the outermost sugar on their LOS. NTHi is incapable of synthesizing sialic acids and scavenges it from the host using TRAP transport system, which is encoded by genes HI0146 and HI0147. These genes correspond to a periplasmic sialic acid binding protein (SiaP, HI0146) and two integral membrane proteins (SiaQM, HI0147) that are of unequal size, together termed SiaPQM. Previous studies have shown that deletion of SiaT (SiaQM) gene in the NTHi affects the sialic acid incorporation on LOS. Further, the strains are more susceptible to host complement-mediated killing that leads to reduced viability in the growth of biofilms. This was further proven by continuous flow cell

assay, where NTHi2019 and 2019SiaT mutants were tested for their ability to form biofilms. Cell viability staining of these strains showed that 2019SiaT mutant displayed higher proportion of cell death (Figure. 1.16.) (72).

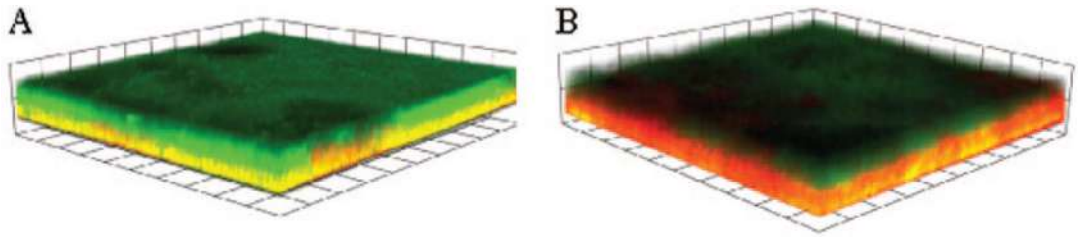


Figure. 1.16. Biofilm formation by wild-type 2019 and 2019SiaT mutant. 2019 and 2019SiaT strains were grown for 2 days in supplemented RPMI diluted to 1:10 with PBS, then stained with LIVE/DEAD Bac light stain, and examined using confocal microscopy. Three-dimensional representations show that biofilm of (A) 2019 has more viable cells (green) and (B) 2019SiaT mutant has more nonviable cells (red). Figure is adapted from (72).

Similarly, Johnston et al., performed the continuous flow cell assays for mutants of *H. influenzae* 2019SiaP and 2019SiaPSiaT to check if the phenotype of biofilm formation is similar to SiaT mutant. Their experiments suggest that sialic acid is not necessary for biofilm formation; however, sialic acid transport by SiaP and SiaT is definitely essential for the long-term viability of cells in the biofilms (Figure. 1.17) (73).

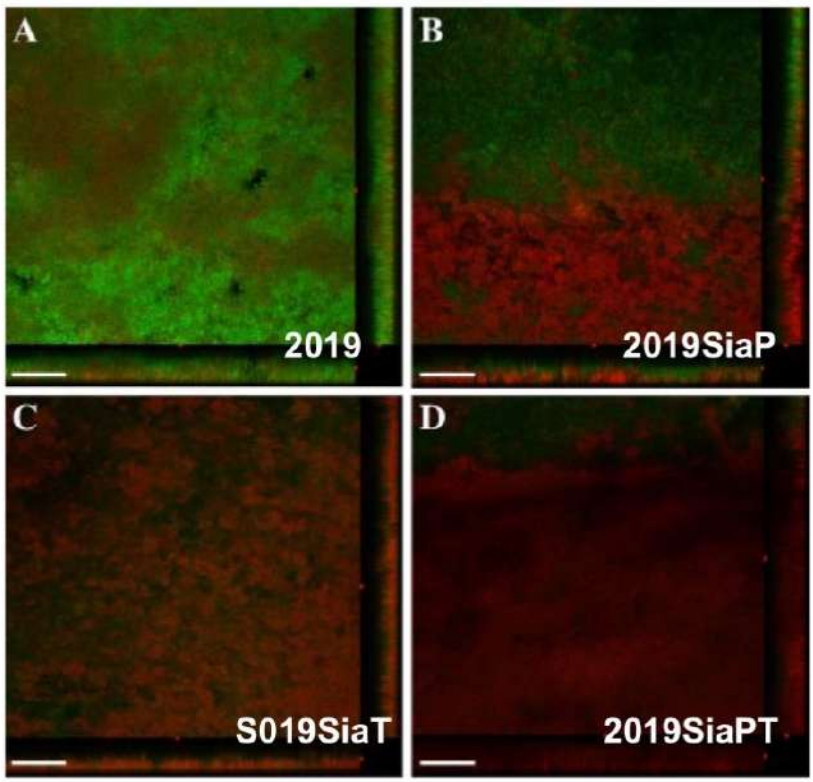


Figure 1.17. Biofilm formation of wild type and sialic acid transporter mutant *H. influenzae*. Strains were grown for 2 days in supplemented RPMI, diluted to 1:10 with PBS, stained with the LIVE/DEAD BacLight viability stain, and examined using confocal microscopy. Mutant strains from B-D show significant number of dead cells compared to wild type (A). Figure is adapted from (73).

Pasteurella multocida is a Gram-negative bacterium and causes several diseases in cattle, buffalo, sheep, poultry, horses, and camels. It is a causative agent of fowl cholera in birds, bovine hemorrhagic septicaemia, and atropic rhinitis in swine. It is also responsible for infection in humans resulting from cat and dog bites. Endotoxins are shown to be the major virulence factors in haemorrhagic septicaemia in the buffalo (74). The entire genome of *P. multocida* Pm70 has been sequenced and shows the presence of genes involved in sialyl transfer and also putative TRAP transporter, similar to *H. influenzae* (75–77). Previous signature tagged mutagenesis studies show that mutations in TRAP transporter greatly decreases the virulence of *P. multocida* in septicaemic mouse model (77) .

F. nucleatum is another Gram-negative periodontal pathogen that cause gingivitis and periodontitis. Genome analysis of *F. nucleatum* W1481 showed the presence of TRAP-like transporter genes (74).

Vibrio cholerae (SiaPQM, VC1777-VC1779) that causes diseases in animals and human lives in mucus-rich regions and does not synthesize sialic acid. This bacteria has a similar TRAP transport system to scavenge and transport sialic acid from the host (78, 79). Growth studies of *V. cholerae* in the presence of different sugars and the measurement of Neu5Ac incorporation on LOS using ELISA-based experiments show that the bacteria can incorporate sialic acid on LOS (Figure. 1.18) (80).

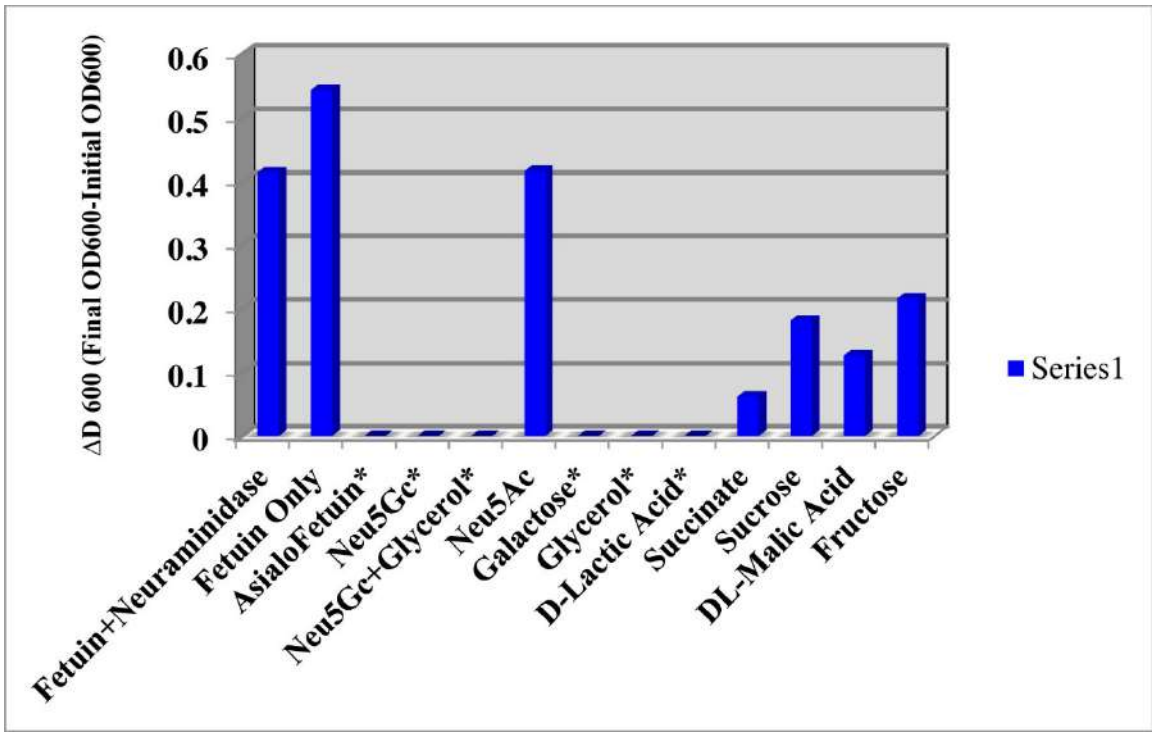


Figure. 1.18. Growth of *V. cholerae* in presence of different sugars. *Vibrio cholerae* O1 biovar eltor str. N16961 was grown in M9 minimal medium supplemented with different sugars, incubated at 37°C. Optical densities at A_{600} were measured at various time points using Spectronic 20 Genesys™ (Spectronic). The growth of bacteria is clearly observed in Neu5Ac. In contrast, bacterial growth is not observed in the media containing Neu5Gc. Figure is adapted from Ref. (80).

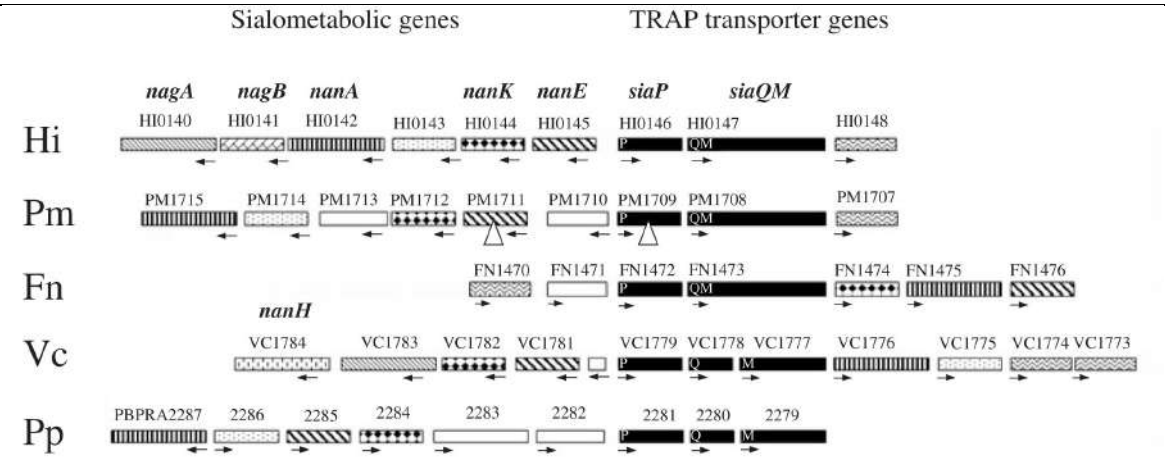


Figure. 1.19. Schematic representation of sialic acid transporter (solid black) and metabolic genes (boxes with various patterns) from various pathogenic bacteria like *H. influenzae (Hi)*; *P. multocida (Pm)*; *F. nucleatum (Fn)*; *V. cholera (Vc)* and *Photobacterium profundum (Pp).* Figure is adapted from Ref (81)**.**

Interestingly, the genes for TRAP transporters are closely associated with sialic acid metabolic genes in all these pathogenic bacteria (Figure. 1.19). Bioinformatic and experimental evidence show that sialic acids are used by these bacteria for molecular mimicry, nutrition, and cell signalling. Given the fact that sialic acids are important for virulence and survival, the inhibition of sialic acid binding proteins that scavenge sialic acid is a potential antimicrobial target. Thus, it is very important to understand the structural and functional properties of these proteins.

1.5 Structural and functional characterization of sialic acid binding protein (SiaP) from *H. influenzae, P. multocida, F. nucleatum and V. cholera*

SiaP of *H. influenzae* (referred from here as *Hi*-SiaP) is a classic example of sialic acid binding proteins from TRAP transport system and was initially characterized by two different groups, Muller et al. (82) and Johnston et al. (73). Muller et al., solved the first structure of *Hi*-SiaP in complex with the sialic acid analogue 2,3-didehydro-2-deoxy-N-acetyneruaminic acid (Neu5Ac2en), wherein the protein is a two-domain protein. DALI search shows that SiaP structure is similar to structures of glycine-betaine ESR proteins, LysR-type transcription factors, and eukaryotic glutamate receptors that belong to the ABC transport system. Superposition of open and closed conformations show that there is a rigid body rotation after ligand binding and Dyndom analysis shows about 25-31^0 rotation in the hinge region. Later Johnston et al., solved the structure of SiaP in complex with Neu5Ac. The main difference

between Neu5Ac2en and Neu5Ac is the presence of a double bond between C2 and C3, which introduces partial planarity in the ring. Studies from our laboratory aided in the structural and functional characterization of the SiaP from other pathogenic bacteria such as *F. nuleatum, P. multocida,* and *V. cholerae* (80). The sequence identities of *Pm*-SiaP, *Fn*-SiaP, and *Vc*-SiaP to *Hi*-SiaP are 73%, 65%, and 47%, respectively (Figure 1.20). Though these proteins show distinct phylogenetic origins (they belong to two phyla and three orders), they have a well-conserved three-dimensional structure (Figure. 1.20) (80).

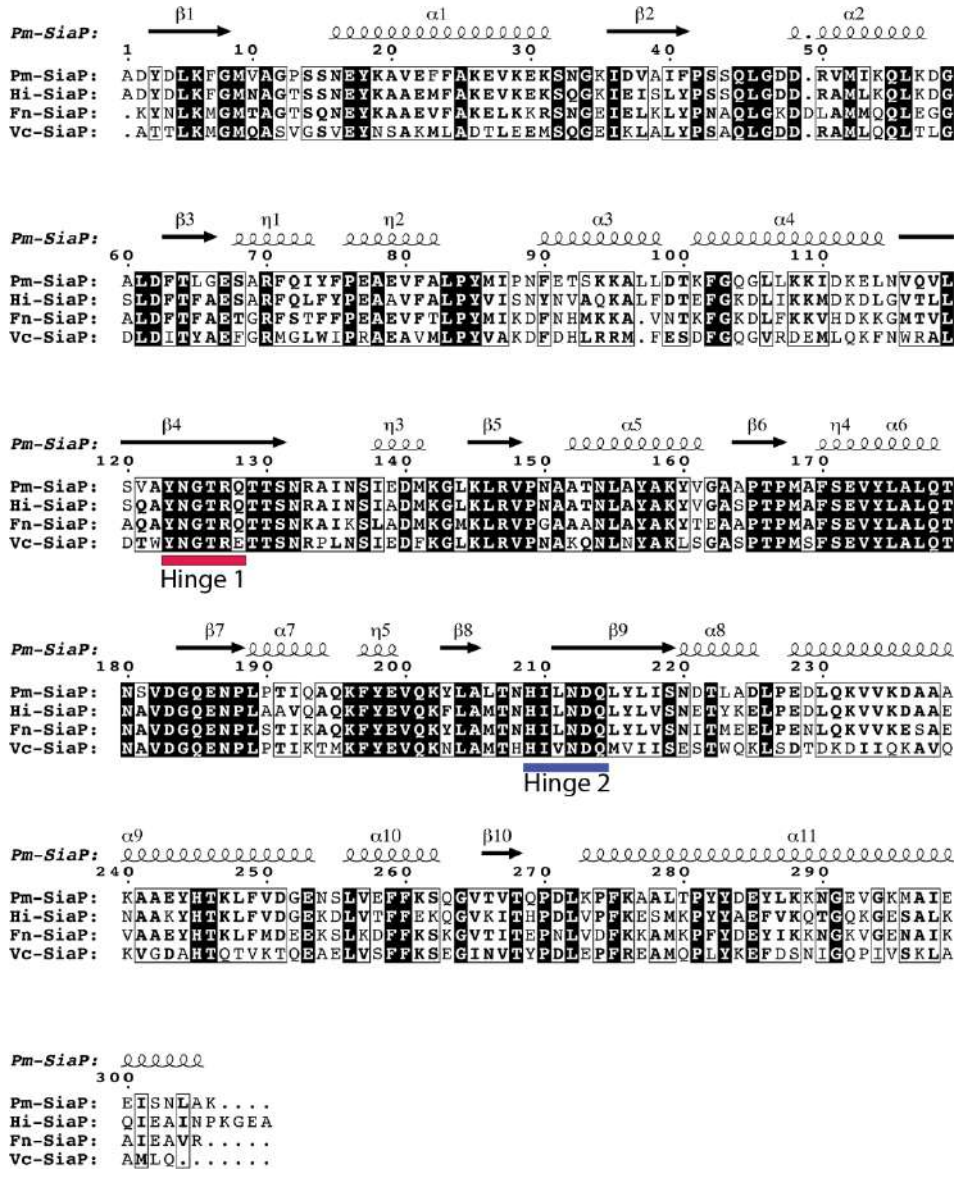


Figure 1.20. Sequence alignment of sialic acid binding proteins from four different pathogenic bacteria using *ClustalW2* and *ESPript.* *Figure is adapted from* (80)**.**

The common features of SiaP from all these bacteria are as follows:
(a) SiaPs are two-domain proteins, (b) ligand binds between the N- and C-terminal domains, (c) two domains close upon ligand binding, similar to the Venus flytrap mechanism (Figure. 1.21), (d) amino acids in the hinge regions are responsible for the opening and closing of the domains, and (e) carboxylate group at C1 position of sialic acid interacts with Arg127 and Arg147. Arg127 is a well-conserved residue in all the SiaP proteins and interacts with both C1 carboxyl group and C2 hydroxyl group of the Neu5Ac by forming salt bridges.

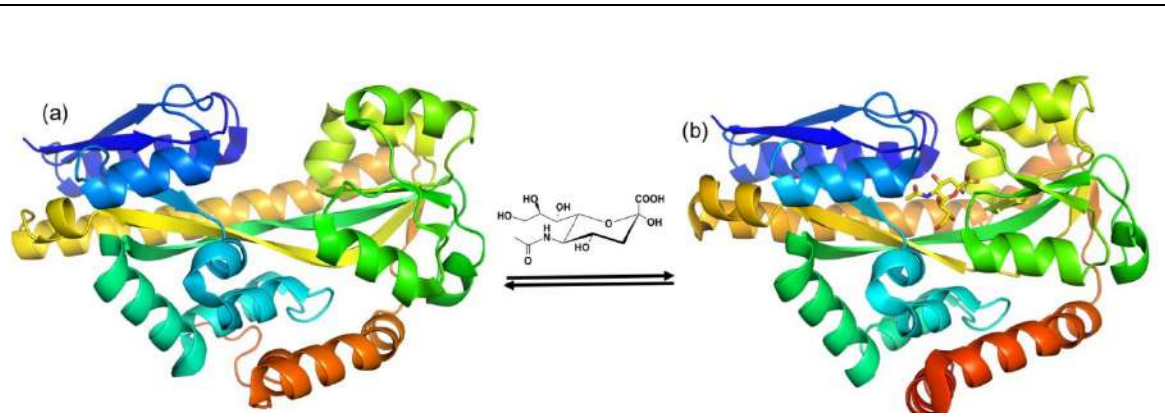


Figure 1.21. Cartoon representation of sialic acid binding protein (SiaP). (a) Unliganded form of *Vc*-SiaP and (b) ligand-bound *Fn*-SiaP in the presence of Neu5Ac. The structures are colored from blue (n-terminus) to red (c-terminus). Neu5Ac is shown as stick representation. Figure is modified from Ref. (80)

An interaction between O-10 of Neu5Ac and Asn-10 of SiaP is also seen in both the ligand-bound structures of *Hi*-SiaP, but this residue is not conserved in the other SiaPs. Except *Hi*-SiaP:Neu5Ac2en, all the other SiaP structures reveal an interaction between the O2 of Neu5Ac and Asn187; and also between O2 and Arg127. Asp49 interaction with O7 is conserved in all the SiaP structures that are bound to Neu5Ac (the numbering corresponds to *Hi*-SiaP). Except few minor changes in the residues, these interactions show that *Hi*-SiaP, *Fn*-SiaP, *Pm*-SiaP, and *Vc*-SiaP have well-conserved ligand binding pockets (Figure. 1.22).

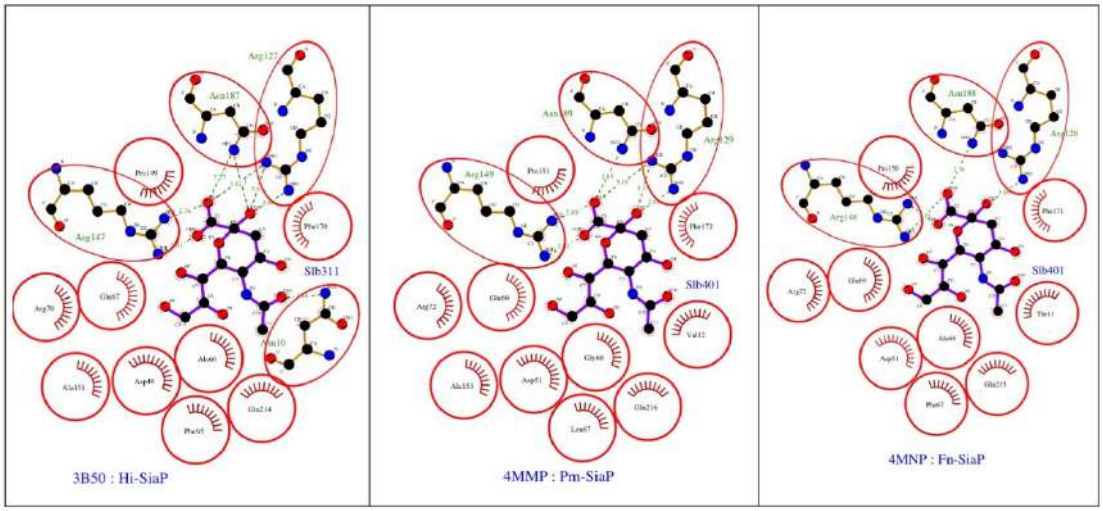


Figure 1.22. Ligplot representation showing the sialic acid binding pockets. *H. influenzae* (3B50), *P. multocida* (4MMP) and *V. cholera* (4MNP).

Structural comparison of the open and closed conformations of SiaPs and sequence analysis shows two broader conserved hinge regions. These are referred to as hinge 1 (YNGTRQ; residues from 123-128 in *Hi*-SiaP) and hinge 2 (HILND; residues from 209-213 in *Hi*-SiaP). Examination of the superimposed open and closed structures of *Hi*-SiaP and Pm-SiaP reveals that upon Neu5Ac binding, the residues Arg129 and His211 (numbering corresponds to *Pm*-SiaP) from the hinges operate by changing the hydrogen-bond structure through making and breaking a series of hydrogen bonds, which might help with the opening and closing of these two domains (Figure. 1.23).

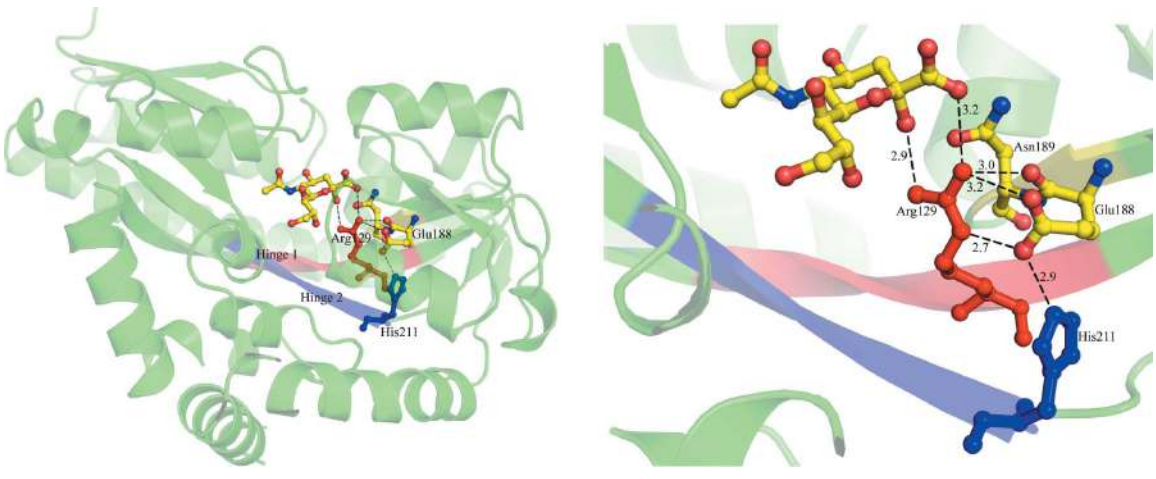


Figure 1.23. Cartoon representation showing hinge regions from SiaP. (a) Cartoon representation of *Fn*-SiaP showing the hinge regions, colored in blue and red. (b) Close-up view of *Pm*-SiaP showing the binding pocket and conserved residues. The dotted lines show the residues interacting through hydrogen/ionic bonding and the distances are labelled in Å (80).

In the Neu5Ac bound structures, Arg127 interacts with the C1 carboxylate group and the C2 hydroxyl group of SiaP by forming salt bridges and this gives high affinity to the ligand. In case of Neu5Ac2en bound structure, there is no salt bridge formation with the SiaP because of the absence of the hydroxyl group. Consistently, binding affinity studies of SiaP to sialyl amide (amide group at C1 position) by tyrosine fluorescence spectroscopy shows an approximate Kd of 243±28 µM, which is approximately 2000 fold lower than the affinity towards Neu5Ac (120 nM). Further, changes in the other groups on sialic acid show a difference in binding affinities towards SiaP. Interestingly, 4-acetylaminocyclohexane carboxylic acid contains only a carboxyl group at the first position and an amido group at the fifth position shows no binding affinity to SiaP. The binding studies of different sialic acid analogues with SiaP show that the high binding affinity requires other groups on sialic acids apart from the carboxyl group and the amido group at the fifth position (Figure. 1.24) (82).

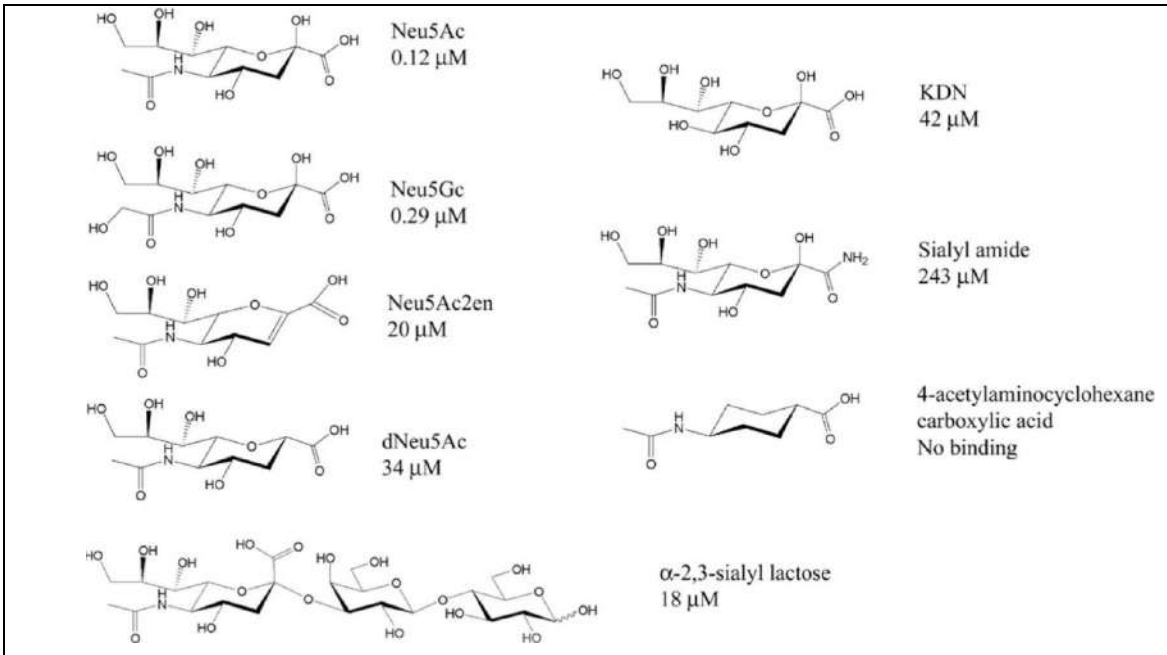


Figure 1.24. Chemical structures and K_d values of sialic acid and its analogues binding to *Hi*-SiaP protein (81, 82)**.**

The binding affinity studies using isothermal calorimetry show that Neu5Ac and Neu5Gc bind to SiaPs at a 1:1 stoichiometric ratio (Figure 1.25). *Hi*-SiaP, *Fn*-SiaP, and *Pm*-SiaP bind to Neu5Ac and Neu5Gc with approximately 20-50 nM binding affinity, whereas *Vc*-SiaP binds to Neu5Ac with a binding affinity of approximately 300 nM, which is 10X weaker than other SiaPs. At present, the structural reasons contributing to the low affinity of *Vc*-SiaP towards sialic acids are not known. Thermodynamic studies show that all these are enthalpically driven interactions (Table 1).

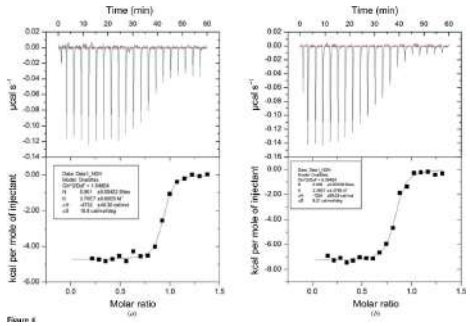


Figure 1.25. Thermodynamic studies of *Pm*-SiaP. Isothermal calorimetric titration isotherms of *Pm*-SiaP binding to (a) Neu5Ac and (b) Neu5Gc. Figure is adapted from Ref. (80).

SiaP *Protein*	***Ligand***	***K_d (nM)***	***ΔH (cal/mol)***	***-TΔS (cal/mol)***
F. nucleatum	Neu5Ac	45.5	-9165	-849.3
	Neu5Gc	45.7	-12074	2059.4
P. multocida	Neu5Ac	19.7	-6560	-3948.2
	Neu5Gc	30.7	-7469	-2777.3
V. cholerae	Neu5Ac	306	-7175	-508.9
	Neu5Gc	1090	-5200	-1127.6
****H. influenzae***	Neu5Ac	28	-15300	5000

Table 1.1. Binding affinities of SiaPs to Neu5Ac and Neu5Gc and their entropic and enthalpic contributions.

All these reports suggest that bacteria have clearly evolved with multiple methods to scavenge and import different molecules and nutrients from the environment using these different transport systems.

Haemophilus ducreyi is a Gram-negative coccobacillus that belongs to the pasteurellaceae family and causes the sexually transmitted disease chancroid. The putative virulence factors of *H. ducreyi* include haemolytic toxin and cytolethal-distending toxin. It has been shown that LOS of *H. ducreyi* also acts as putative virulence factor and the glycan moieties terminate with N-acetyllactosamine and sialyl-N-acetyllactosamine and they are heavily sialylated (83). It has been shown that *H. ducreyi* uses a novel ABC-transport system for the transport of sialic acid and is encoded by a gene SatABCD (84). In this operon, SatA encodes for the periplasmic binding protein that binds sialic acid and SatBCD encodes for the integral membrane protein with nucleotide binding domain. The sequence similarity between SatA and other SiaPs that bind to sialic acids is approximately 20-25%. **Though several other periplasmic binding proteins of thebABC transport system that transport various substrates have been studied before, the structure of SatA and the mechanism of its binding to sialic acids is unknown.**

After internalization of sialic acid by the different transporters, part of the ligand gets incorporated onto LOS for molecular mimicry and the other part is catabolized into carbon and nitrogen sources. At present, there is no evidence to show how bacteria maintain equilibrium between the incorporation and catabolism processes of sialic acid.

1.6 Incorporation of sialic acid onto LOS/LPS

During the process of incorporation, sialic acid is, first, activated by CMP-sialic acid synthetase SiaB to form CMP-Neu5Ac. Confocal and electron microscopy analyses of SiaB mutants show significant reduction in biofilm production (37). Further, different linkage specific sialyltransferases add sialic acids to different acceptors of LPS. Lic3A is a sialyltransferase that adds Neu5Ac to lactosyl acceptor of LPS in an alpha-2,3-linkage (85). Lic3B is another sialyltransferase, a close homologue of Lic3A. It has both alpha-2,3 and alpha-2,8 silalyltranferase activities and the disialylated species are more resistant to being killed by human serum than the mono-sialylated species (86). SiaA and LsgB are two other additional sialyltranfereses, responsible for LPS sialylation. Mutations in SiaA affects the glycoforms containing N-acetylhexosamine and mutants of LsgB affect the glycoforms

containing N-acetyllactosamine (87). It has been previously shown that sialylation of *H. influenzae* LPS is a critical virulence factor in chinchillas otitis media. Consistently, sialylation deficient mutants show greater attenuation during infection in the chinchilla model. Further, it was also shown that sialylation of LPS is important for biofilm formation (37, 40, 86, 88–90).

1.7 Catabolism of sialic acids in pathogenic bacteria

Catabolism of Neu5Ac is an important survival mechanism for both commensal and pathogenic bacteria that reside in the mucus-rich regions of eukaryotic hosts. In the catabolic pathway, N-acetylneuraminic acid lysate (NanA) converts Neu5Ac to N-acetyl mannosamine (ManNAc) by cleaving pyruvate. ManNAc is then phosphorylated into ManNAc-6P by the action of N-acetyl mannosamine kinase (NanK). Further, ManNAc-6P gets epimerized into GlcNAc-6-P by the action of N-acetylmannosamine-6-phosphate 2-epimerase (NanE). GlcNAc-6-P is then deacetylated and converted into GlcN-6P by the action of N-acetylglucosamine-6-phosphate deacetylase (NagA). GlcN-6P is subsequently deaminated by glucosamine-6-phosphate deaminase (NagB), yielding fructose-6-phosphate and ammonia. Fructose-6-phosphate then enters the glycolytic pathway and is utilized as carbon, nitrogen, and energy sources. All these genes in the catabolic pathway are called the Nan-Nag gene cluster of enzymes. The entire pathway of sialic acids in Gram-negative bacteria is pictorially represented in Figure 1.26.

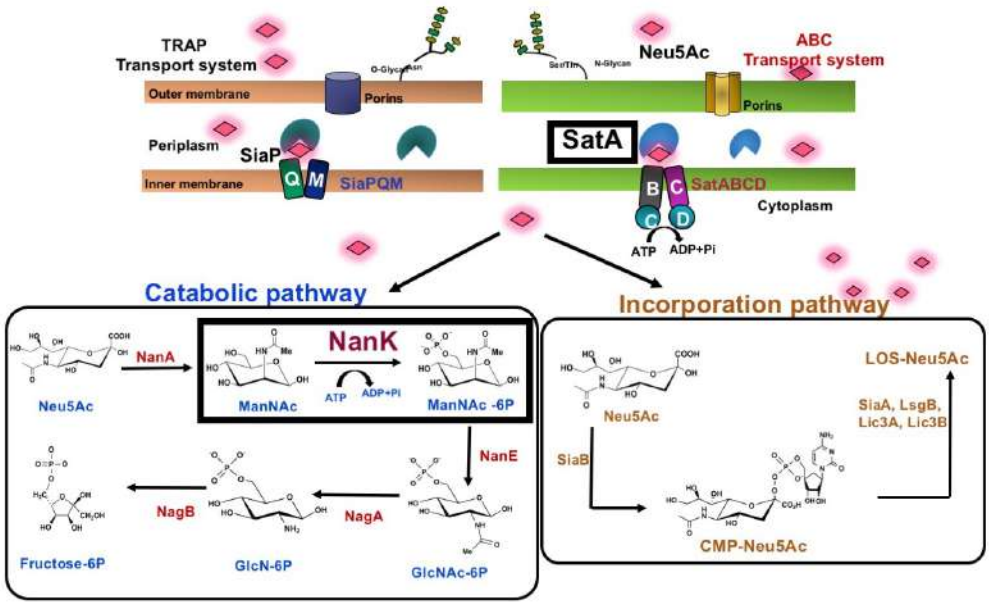


Figure 1.26. Schematic representation of sialic acid metabolism in bacteria. Summary of various pathways that utilize sialic acids in bacteria. Figure represents the TRAP and ABC transport system used by bacteria for scavenging sialic acids and also represents the enzymes involved in the catabolic and incorporation pathways of sialic acid. Neu5Ac, N-acetyl

neuraminic acid; NanA, N-acetylneuraminic acid lyase; ManNAc, N-acetylmannosamine; NanK, N-acetylmannosamine kinase; ManNAc-6-P, N-acetylmannosamine-6-phosphate; NanE, N-acetylmannosamine-6-P epimerase; GlcNAc-6-P, N-acetylglucosamine-6-phospahte; NagA, N-acetylglucosamine-6-phosphate deacetylase; NagB, Glucosamine-6-phosphate deaminase; fructose-6P, fructose-6-phosphate. Figure is modified from (40, 91).

Previous reports have suggested that the Nan-Nag gene cluster of *V. cholera* is very helpful for initial bacterial colonization. Similarly, growth studies have shown that *V. cholerae* grown on sialic acid display an increase in expression of enzymes nanH (scavenging), dctP, nanA and nanK that are present in the Vibrio pathogenicity island (VPI-2) (78). Inactivation of the catabolic pathway genes in VPI-2 shows reduced ability of *V. cholerae* to colonize in the intestine of the infant mouse model. Overall, these studies suggest that bacteria have evolved to utilize the sugars present in the mucus-rich environment and these findings also highlight the significance of sialic acid catabolism in bacterial pathogenesis (78). **Hence, it is important to characterize these enzymes involved in both the incorporation and catabolic pathways, structurally and functionally. These studies will help us understand the molecular basis of their mode of action and aid in the development of drugs against these pathogenic bacteria.**

1.8 Phosphorylation by kinases

Phosphorylation by kinases is important in cellular signalling in both prokaryotes and eukaryotes. During the process of phosphorylation, ligand binding causes a conformational change in the enzyme, which helps the kinase trap the substrate in its active site and positions the nucleotide to facilitate the transfer of **γ** - phosphate from the ATP to its substrate in the presence of a divalent ion Mg^{2+}. Even though most kinases have a similar mechanism of phosphorylating the substrate, they control a wide variety of intracellular signaling pathways including sugar metabolism, cell cycle, gene regulation, and several other cellular processes (92–94).

1.9 N-acetyl mannosamine kinases

NanK is the second enzyme in the nan-nag cluster of enzymes and it phosphorylates ManNAc to ManNAc-6P. NanKs belong to ROK (bacterial **R**epressors, uncharacterized **O**pen reading frames and sugar **K**inases) superfamily of enzymes. The members of this family are diverse and found in prokaryotes as well as in eukaryotes. ROK family of enzymes contain conserved motifs such as (i) DXDGT, located in the ATP binding site at the N-terminus of the protein, (ii) EXGH motif, and (iii) catalytic aspartate residue. Excluding *Fn*-NanK, all the other bacterial NanKs contain a Zn-binding motif, composed of one histidine and three cysteines (HC3-type zinc binding motif).

In pathogenic bacteria, NanK is involved in the catabolic pathway of Neu5Ac, whereas in humans, interestingly, NanK is involved in the synthetic pathway of Neu5Ac. In humans, UDP-GlcNAc 2-epimerase/ManNAc kinase (GNE) consists of two functional domains, N- and C-terminal domains. The N-terminal part encodes for UDP-GlcNAc 2-epimerase and the C-terminal part encodes for ManNAc kinase. Previously, structures of NanK from human, *E. coli,* and *F. nucleatum* were reported by different groups, both in the unliganded and the ligand-bound forms in complex with ManNAc and nucleotide (95–97)(Figure 1.27).

Crystal structures of these NanKs show that ManNAc and ATP bind between the two domains and the Zn-binding motif plays an important role in holding ManNAc in the binding pocket (Figure. 1.27). Further, Martinez et al., characterized the functional properties of hMNK towards its substrates using kinetic and thermodynamic studies. Due to difference in the hydrogen bonding network, the kinetic studies of hMNK shows higher affinity towards ManNAc (K_m = 95 µM) and low affinity towards ADP (K_m= 4.4 mM). Further, the binding studies of hMNK show a K_d of 2.5µM for ManNAc (96) .

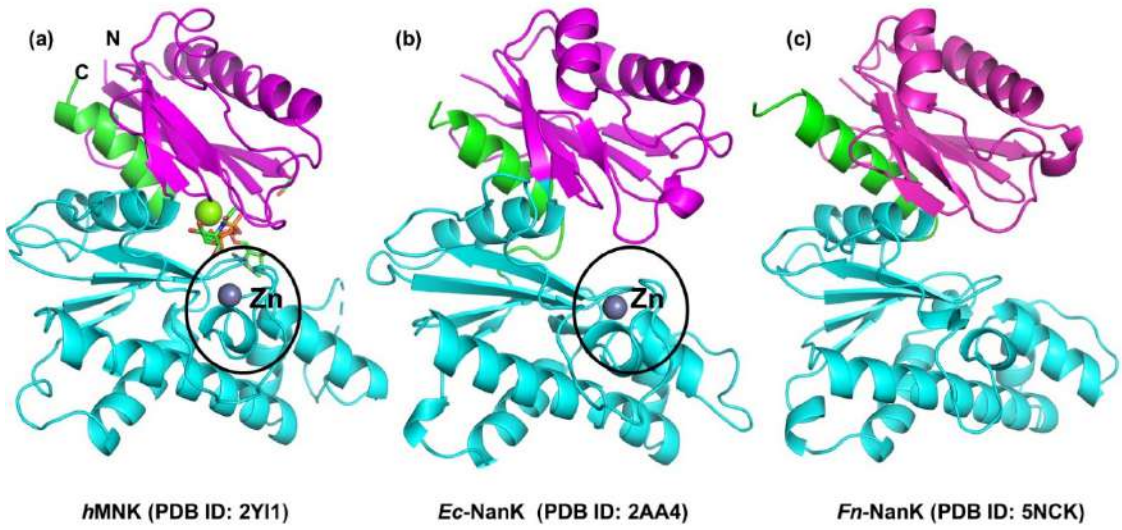


Figure 1.27. Cartoon representation showing the structures of NanKs. (a) Human (b) *E. coli* (c). *F. nucleatum.* Figure shows the absence of Zn- and Zn-binding motif in the NanK from *F. nucleatum*.

1.9.1 Mechanism of phosphate transfer by kinases

The following mechanism of phosphorylation by hMNK is hypothesized by Martinez et al., Similar to other hexokinases, NanK also transfers the phosphate from ATP to the 6-hydroxyl group of ManNAc via trigonal-bipyramidal intermediate, which finally results in the inversion of phosphate configuration. Mutational and enzymatic studies of hMNK showed Asp-517 functions as a catalytic aspartate residue. The general hypothesized mechanism for NanK phosphorylation is that –Asp517 acts as general base and the O6 atom of ManNAc gets deprotonated leading to an increase in its nucleophilicity. Further, the nucleophilic attack by the O6 atom in ManNAc leads to the transfer of the **γ** -phosphate of ATP, which further leads to phosphorylation of the substrate (Figure. 1.28) (96).

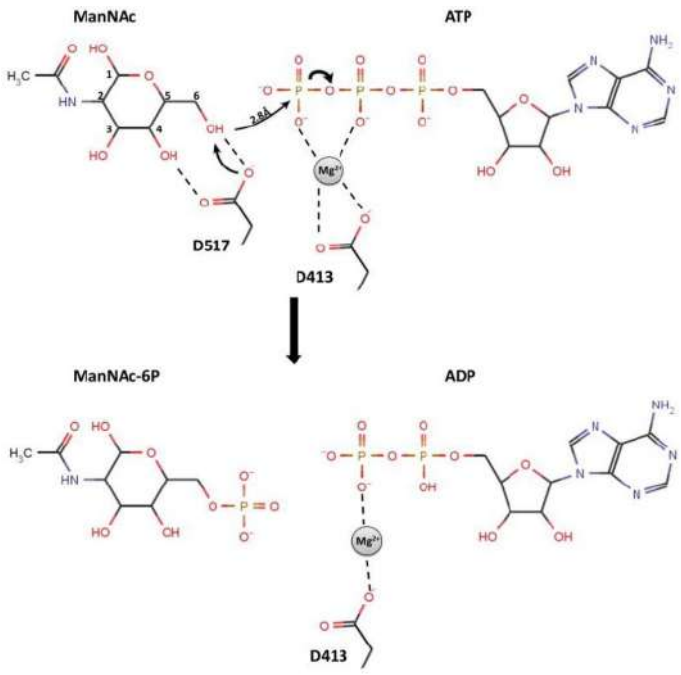


Figure 1.28. Schematic representation of phosphorylation of ManNAc to ManNAc-6P by hMNK. Figure is adapted from (96).

The structure of *Fn*-NanK was previously studied using X-ray crystallography. Unlike other ROK kinases, *Fn*-NanK does not have a conserved Zn-binding motif (Figure 1.27) and only contains the histidine residue in the motif (97). Structural studies show that although the Zn-binding motif is absent in *Fn*-NanK, the three-dimensional structure of the protein is well conserved. Moreover, it was hypothesized that the histidine will play an important role in holding the ManNAc substrate in position. However, the effect of the absence of the Zn-binding motif on the functional properties has not yet been characterized. **Thus, a comparition of the molecular characterization of NanKs from other pathogenic bacteria that contain Zn-binding motif with that of NanK from *F. nucleatum* will help in understanding the significance of this motif in this group of kinases.**

1.10 Aims of this investigation:

The overall goal of the present doctoral thesis is to study the structural and functional characterization of sialic acid uptake and its metabolism in pathogenic bacteria. The main objectives of this thesis are:

(1) Structural and functional characterization of the sialic acid binding protein (SatA) of the ABC transport system from *Haemophilus ducreyi*,

(2) Design, clone, and express the genes from *H. influenzae, P. multocida, V. cholera,* and *F. nucleatum* that are involved in the incorporation and catabolic pathways of sialic acid

and

(3) Structural and functional characterization of N-acetyl mannosamine kinases from pathogenic bacteria – with a special insight into the Zn-binding motif.

1.11 References

1. Varki, A. (2017) Biological roles of glycans. *Glycobiology* **27**, 3–49
2. Blix, G. (1936) Über die Eohlenhydratgruppen des Submaxillarismucins. *Hoppe. Seylers. Z. Physiol. Chem.* **240**, 43–54
3. Klenk, E. (1941) Neuraminsäure, das Spaltprodukt eines neuen Gehirnlipoids. *Hoppe. Seylers. Z. Physiol. Chem.* **268**, 50–58
4. Warren, L. (1963) The Distribution of silaic acids in nature*. *Comput. Biochem. Physiol* **10**, 153–171
5. Varki A, S. R. (2009) *Chapter 14 Sialic Acids*,
6. Inoue, S., and Kitajima, K. (2006) KDN (Deaminated neuraminic acid): Dreamful past and exciting future of the newest member of the sialic acid family. *Glycoconj. J.* **23**, 277–290
7. Schauer, R., and Kamerling, J. P. (2018) *Exploration of the Sialic Acid World*, 1st Ed., Elsevier Inc.
8. Angata, T., and Varki, A. (2002) Chemical Diversity in the Sialic Acids and Related r -Keto Acids : An Evolutionary Perspective. *Chem. Rev.* **102**, 439–469
9. Schauer, R. (2006) Sialic acids: fascinating sugars in higher animals and man. *Zoology* **107**, 49–64
10. Varki, A., and Schauer, R. (2009) Chapter 14 Sialic Acids. , 1–17
11. Goodman, M. (1999) MOLECULAR EVOLUTION '99. The Genomic Record of Humankind's Evolutionary Roots. *Amer. J. Hum. Genet.* **64**, 31–39
12. Takematsu, H., Kawano, T., Koyama, S., Kozutsumi, Y., Suzuki, A., and Kawasaki, T. (1994) Reaction mechanism underlying CMP-N-acetylneuraminic acid hydroxylation in mouse liver: Formation of a ternary complex of cytochrome b5, CMP-N-acetylneuraminic acid, and a hydroxylation enzyme. *J. Biochem.* **115**, 381–386
13. Takehiro Kawano, Susumu Koyama, Hiromu Takemastsu, Yasunori Kozutsumi, Hiroshi Kawasaki, Seiichi Kawashima, Toshisuke Kawasaki, A. suzuki (1995) Molecular cloning of cytidine monophosphate-N-acetylneruaminic acid hydorxylase. *J. Biol. Chem.* **270**, 16458–16463
14. Takahashi, T., Kawagishi, S., Funahashi, H., and Hayashi, N. (2015) Production and Purification of Secretory Simian Cytidine Monophosphate- N -acetylneuraminic Acid Hydroxylase Using Baculovirus-Protein Expression System. *Biol. pharm. Bull* **38**, 1220–1226
15. Atsushi Irie, Susumu Koyama, Yasunori Kozutsumi, Toshisuke Kawasaki, and A. S. (1998) The Molecular Basis for the Absence of N -Glycolylneuraminic Acid in Humans *. *J. Biol. Chem.* **273**, 15866–15871
16. Chou, H. H., Takematsu, H., Diaz, S., Iber, J., Nickerson, E., Wright, K. L., Muchmore, E. A., Nelson, D. L., Warren, S. T., and Varki, A. (1998) A mutation in human CMP-sialic acid hydroxylase occurred after the Homo-Pan divergence. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 11751–6
17. Higashi, H., Naiki, M., Matuo, S., and Okouchi, K. (1977) Antigen of "serum sickness" type of heterophile antibodies in human sera: Identification as gangliosides with N-glycolylneuraminic acid. *Biochem. Biophys. Res. Commun.* **79**, 388–395
18. Malykh, Y. N., Schauer, R., and Shaw, L. (2001) N -Glycolylneuraminic acid in human tumours *. **83**
19. Varki, A. (2001) Loss of N-glycolylneuraminic acid in humans: Mechanisms, consequences, and implications for hominid evolution. *Am. J. Phys. Anthropol.* **116**, 54–69

20. Schwarzkopf, M., Knobeloch, K.-P., Rohde, E., Hinderlich, S., Wiechens, N., Lucka, L., Horak, I., Reutter, W., and Horstkorte, R. (2002) Sialylation is essential for early development in mice. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 5267–70
21. Varki, N. M., and Varki, A. (2007) Diversity in cell surface sialic acid presentations: Implications for biology and disease. *Lab. Investig.* **87**, 851–857
22. Effertz, K., Hinderlich, S., and Reutter, W. (1999) Selective Loss of either the Epimerase or Kinase Activity of due to Site-directed Mutagenesis Based on Sequence Alignments *. *J. Biol. Chem.* **274**, 28771–28778
23. Hinderlich, S., Weidemann, W., Yardeni, T., Horstkorte, R., and Huizing, M. (2015) UDP-GlcNAc 2-epimerase/ManNAc kinase (GNE): A master regulator of sialic acid synthesis. *Top. Curr. Chem.* **366**, 97–138
24. Reinke, S. O., Lehmer, G., and Reutter, W. (2009) Regulation and pathophysiological implications of UDP-GlcNAc 2-epimerase / ManNAc kinase (GNE) as the key enzyme of sialic acid biosynthesis. *Biol chem* **390**, 591–599
25. Seppala, R., Tietze, F., Krasnewich, D., Weiss, P., Ashwell, G., Barsh, G., Thomas, G. H., Packman, S., and Gahl, W. A. (1991) Sialic acid metabolism in sialuria fibroblasts. *J. Biol. Chem.* **266**, 7456–7461
26. Leroy, J. G., Seppala, R., Huizing, M., Dacremont, G., De Simpel, H., Van Coster, R. N., Orvisky, E., Krasnewich, D. M., and Gahl, W. A. (2002) Dominant Inheritance of Sialuria, an Inborn Error of Feedback Inhibition. *Am. J. Hum. Genet.* **68**, 1419–1427
27. Auer, G. K., and Weibel, D. B. (2017) Bacterial Cell Mechanics. *Biochemistry* **56**, 3710–3724
28. Esko, J. D., Doering, T. L., and Raetz, C. R. H. (2009) *Chapter 20 Eubacteria and Archaea*,
29. Tan, F. Y. Y., Tang, C. M., and Exley, R. M. (2015) Sugar coating : bacterial protein glycosylation and host – microbe interactions. *Trends Biochem. Sci.* **40**, 342–350
30. Côté, J. P., Charbonneau, M. È., and Mourez, M. (2013) Glycosylation of the Escherichia coli TibA self-associating autotransporter influences the conformation and the functionality of the protein. *PLoS One* **8**, 1–9
31. Stimson, E., Virji, M., Makepeace, K., Dell, A., Morris, H. R., Payne, G., Saunders, J. R., Jennings, M. P., Barker, S., Panico, M., Blench, I., and Moxon, E. R. (1995) Meningococcal pilin: a glycoprotein substituted with digalactosyl 2,4-diacetamido-2,4,6-trideoxyhexose. *Mol. Microbiol.* **17**, 1201–1214
32. Thibault, P., Logan, S. M., Kelly, J. F., Brisson, J. R., Ewing, C. P., Trust, T. J., and Guerry, P. (2001) Identification of the Carbohydrate Moieties and Glycosylation Motifs in Campylobacter jejuni Flagellin. *J. Biol. Chem.* **276**, 34862–34870
33. Schirm, M., Soo, E. C., Aubry, A. J., Austin, J., Thibault, P., and Logan, S. M. (2003) Structural, genetic and functional characterization of the flagellin glycosylation process in Helicobacter pylori. *Mol. Microbiol.* **48**, 1579–1592
34. Haines-menges, B. L., Whitaker, W. B., Lubin, J. B., and Boyd, E. F. (2014) Host Sialic Acids : A Delicacy for the Pathogen with Discerning Taste. *Microbiol. Spectr.*, 1–17
35. Figueira, M. A., Ram, S., Goldstein, R., Hood, D. W., Moxon, E. R., and Pelton, S. I. (2007) Role of complement in defense of the middle ear revealed by restoring the virulence of nontypeable Haemophilus influenzae siaB mutants. *Infect. Immun.* **75**, 325–333
36. Swords, W. E., Moore, M. L., Godzicki, L., Bukofzer, G., Mitten, M. J., and VonCannon, J. (2004) Sialylation of Lipooligosaccharides Promotes Biofilm Formation by Nontypeable Haemophilus influenzae. *Infect. Immun.* **72**, 106–113
37. Greiner, L. L., Watanabe, H., Phillips, N. J., Shao, J., Morgan, A., Zaleski, A., Gibson, B. W., and Apicella, M. A. (2004) Nontypeable Haemophilus influenzae strain 2019

produces a biofilm containing N-acetylneuraminic acid that may mimic sialylated O-linked glycans. *Infect Immun* **72**, 4249–4260
38. Ehrlich, G. D., Veeh, R., Wang, X., Costerton, J. W., Hayes, J. D., Daigle, B. J., Ehrlich, M. D., and Post, J. C. (2002) Mucosal Biofilm Formation on Middle-Ear Mucosa in the Chinchilla Model of Otitis Media. *JAMA* **287**, 1710–1715
39. Vimr, E. R., Kalivoda, K. A., Deszo, E. L., and Steenbergen, S. M. (2004) Diversity of Microbial Sialic Acid Metabolism. *Microbiol. Mol. Biol. Rev.* **68**, 132–153
40. Severi, E., Hood, D. W., and Thomas, G. H. (2015) Sialic acid utilization by bacterial pathogens. *Microbiology* **153**, 2817–2822
41. Li, Y., and Chen, X. (2012) Sialic acid metabolism and sialyltransferases: natural functions and applications. *Appl. Microbiol. Biotechnol.* **94**, 887–905
42. Gulati, S., Schoenhofen, I. C., Whitfield, D. M., Cox, A. D., Li, J., St. Michael, F., Vinogradov, E. V., Stupak, J., Zheng, B., Ohnishi, M., Unemo, M., Lewis, L. A., Taylor, R. E., Landig, C. S., Diaz, S., Reed, G. W., Varki, A., Rice, P. A., and Ram, S. (2015) Utilizing CMP-Sialic Acid Analogs to Unravel Neisseria gonorrhoeae Lipooligosaccharide-Mediated Complement Resistance and Design Novel Therapeutics. *PLoS Pathog.* **11**, 1–26
43. Lin, L. Y.-C., Rakic, B., Chiu, C. P. C., Lameignere, E., Wakarchuk, W. W., Withers, S. G., and Strynadka, N. C. J. (2011) Structure and mechanism of the lipooligosaccharide sialyltransferase from Neisseria meningitidis. *J. Biol. Chem.* **286**, 37237–37248
44. Vimr, E. R. (2013) Unified Theory of Bacterial Sialometabolism: How and Why Bacteria Metabolize Host Sialic Acids. *ISRN Microbiol.* **2013**
45. Wirth, C., Condemine, G., Boiteux, C., Bernèche, S., Schirmer, T., and Peneff, C. M. (2009) NanC Crystal Structure, a Model for Outer-Membrane Channels of the Acidic Sugar-Specific KdgM Porin Family. *J. Mol. Biol.* **394**, 718–731
46. Roy, S., Douglas, C. W. I., and Stafford, G. P. (2010) A novel sialic acid utilization and uptake system in the periodontal pathogen Tannerella forsythia. *J. Bacteriol.* **192**, 2285–2293
47. Thomas, G. H. (2016) Sialic acid acquisition in bacteria – one substrate , many transporters. *Biochem. Soc. Trans.* **44**, 760–765
48. Locher, K. P. (2004) Structure and mechanism of ABC transporters. *Curr. Opin. Struct. Biol.* **14**, 426–431
49. Rees, D. C., Johnson, E., and Lewinson, O. (2009) ABC transporters: The power to change. *Nat. Rev. Mol. Cell Biol.* **10**, 218–227
50. Schneider, E., and Hunke, S. (2002) ATP-binding-cassette (ABC) transport systems: Functional and structural aspects of the ATP-hydrolyzing subunits/domains. *FEMS Microbiol. Rev.* **22**, 1–20
51. Hiroshi, N. (1994) Maltose transport system of Escherichia coli: An ABC-type transporter. *FEBS Lett.* **346**, 55–58
52. Sharff, A. J., Rodseth, L. E., Spurlino, J. C., and Quiocho, F. A. (1992) Crystallographic Evidence of a Large Ligand-Induced Hinge-Twist Motion between the Two Domains of the Maltodextrin Binding Protein Involved in Active transport and chemotaxix. *Biochemistry* **31**, 10657–10663
53. Oldham, M. L., Chen, S., and Chen, J. (2013) Structural basis for substrate specificity in the Escherichia coli maltose transport system. *Proc. Natl. Acad. Sci.* **110**, 18132–18137
54. Quiocho, F. A. (1990) Atomic Structures of Periplasmic Binding Proteins and the High-Affinity Active Transport Systems in Bacteria. *Philos. Trans. R. Soc. B Biol. Sci.* **326**, 341–352

55. Oldham, M. L., Khare, D., Quiocho, F. A., Davidson, A. L., and Chen, J. (2007) Crystal structure of a catalytic intermediate of the maltose transporter. *Nature* **450**, 515–521
56. Oldham, M. L., and Chen, J. (2011) Crystal Structure of the Maltose Transporter in a Pretranslocation Intermediate State. **332**, 1202–1206
57. Hollenstein, K., Frei, D. C., and Locher, K. P. (2007) Structure of an ABC transporter in complex with its binding protein. *Nature* **446**, 213–216
58. Locher, K. P., Lee, A. T., and Rees, D. C. (2002) The E. coli BtuCD structure: A framework for ABC transporter architecture and mechanism. *Science (80-.).* **296**, 1091–1098
59. Borths, E. L., Locher, K. P., Lee, A. T., and Rees, D. C. (2002) The structure of Escherichia coli BtuF and binding to its cognate ATP binding cassette transporter. *Proc. Natl. Acad. Sci.* **99**, 16642–16647
60. Tam, R., and Saier, M. H. (1993) Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiol. Rev.* **57**, 320–346
61. Berntsson, R. P., Smits, S. H. J., Schmitt, L., Slotboom, D., and Poolman, B. (2010) A structural classification of substrate-binding proteins. *FEBS Lett.* **584**, 2606–2617
62. Goswitz, V. C., and Brooker, R. J. (1995) Structural features of the uniporter/symporter/antiporter superfamily. *Protein Sci.* **4**, 534–537
63. Pao, S. S., Paulsen, I. A. N. T., and Saier, M. H. (2009) Major Facilitator Superfamily. *Microbiol. Mol. Biol. Rev.* **62**, 1–25
64. Martinez, J., Steenbergen, S., and Vimr, E. (1995) Derived structure of the putative sialic acid transporter from Escherichia coli predicts a novel sugar permease domain. *J Bacteriol* **177**, 6005–6010
65. Abramson, J., Smirnova, I., Kasho, V., Verner, G., Kaback, H. R., and Iwata, S. (2003) Structure and Mechanism of the Lactose Permease of Escherichia coli. *Science (80-.).* **301**, 1064–1069
66. Salem Faham, Akira Watanabe, Gabriel Mercado Besserer, Duilio Cascio, A., and Specht, Bruce A. Hirayama, Ernest M. Wright, and J. A. (2013) The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na+ /sugar symport. *Science (80-.).* **321**, 810–814
67. Aravind Penmatsa, Kevin H. Wang, and E. G. (2014) X-ray structure of the dopamine transporter in complex with tricyclic antidepressant. *Nature* **503**, 85–90
68. Severi, E., Hosie, A. H. F., Hawkhead, J. A., and Thomas, G. H. (2010) Characterization of a novel sialic acid transporter of the sodium solute symporter (SSS) family and invivo comparison with known bacterial sialic acid transporters. *FEMS Microbiol. Lett.* **304**, 47–54
69. Wahlgren, W. Y., Dunevall, E., North, R. A., Paz, A., Scalise, M., Bisignano, P., Bengtsson-Palme, J., Goyal, P., Claesson, E., Caing-Carlsson, R., Andersson, R., Beis, K., Nilsson, U. J., Farewell, A., Pochini, L., Indiveri, C., Grabe, M., Dobson, R. C. J., Abramson, J., Ramaswamy, S., and Friemann, R. (2018) Substrate-bound outward-open structure of a Na+-coupled sialic acid symporter reveals a new Na+ site. *Nat. Commun.* **9**, 1–14
70. Mulligan, C., Geertsma, E. R., Severi, E., Kelly, D. J., Poolman, B., and Thomas, G. H. (2009) The substrate-binding protein imposes directionality on an electrochemical sodium gradient-driven TRAP transporter. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 1778–1783
71. Forward, J. A., Behrendt, M. C., Wyborn, N. R., Cross, R., and Kelly, D. J. (1997) TRAP transporters: a new family of periplasmic solute transport systems encoded by the dctPQM genes of Rhodobacter capsulatus and by homologs in diverse gram-negative bacteria. *J Bacteriol* **179**, 5482–5493

72. Allen, S., Zaleski, A., Johnston, J. W., Gibson, B. W., and Apicella, M. A. (2005) Novel sialic acid transporter of Haemophilus influenzae. *Infect. Immun.* **73**, 5291–5300
73. Johnston, J. W., Coussens, N. P., Allen, S., Houtman, J. C. D., Turner, K. H., Zaleski, A., Ramaswamy, S., Gibson, B. W., and Apicella, M. A. (2008) Characterization of the N-acetyl-5-neuraminic acid-binding site of the extracytoplasmic solute receptor (SiaP) of nontypeable Haemophilus influenzae strain 2019. *J. Biol. Chem.* **283**, 855–865
74. Ang, M. Y., Dutta, A., Wee, W. Y., Dymock, D., Paterson, I. C., and Choo, S. W. (2016) Comparative genome analysis of fusobacterium nucleatum. *Genome Biol. Evol.* **8**, 2928–2938
75. May, B. J., Zhang, Q., Li, L. L., Paustian, M. L., Whittam, T. S., and Kapur, V. (2001) Complete Genom Sequence of Pasteurella multocida, Pm70. *PNAS* **98**, 3460–3465
76. Steenbergen, S. M., Lichtensteiger, C. A., Caughlan, R., Garfinkle, J., Fuller, T. E., and Vimr, E. R. (2005) Sialic acid metabolism and systemic pasteurellosis. *Infect. Immun.* **73**, 1284–1294
77. Fuller, T. E., Kennedy, M. J., and Lowery, D. E. (2000) Identification of Pasteurella multocida virulence genes in a septicemic mouse model using signature-tagged mutagenesis. *Microb. Pathog.* **29**, 25–38
78. Almagro-moreno, S., and Boyd, E. F. (2009) Sialic Acid Catabolism Confers a Competitive Advantage to Pathogenic Vibrio cholerae in the Mouse Intestine. *Infect. Immun.* **77**, 3807–3816
79. Chowdhury, N., Norris, J., McAlister, E., Lau, S. Y. K., Thomas, G. H., and Boyd, E. F. (2012) The VC1777-VC1779 proteins are members of a sialic acid-specific subfamily of TRAP transporters (SiaPQM) and constitute the sole route of sialic acid uptake in the human pathogen Vibrio cholerae. *Microbiology* **158**, 2158–2167
80. Setty, T. G., Cho, C., Govindappa, S., Apicella, M. A., and Ramaswamy, S. (2014) Bacterial periplasmic sialic acid-binding proteins exhibit a conserved binding site. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **70**, 1801–1811
81. Severi, E., Randle, G., Kivlin, P., Whitfield, K., Young, R., Moxon, R., Kelly, D., Hood, D., and Thomas, G. H. (2005) Sialic acid transport in Haemophilus influenzae is essential for lipopolysaccharide sialylation and serum resistance and is dependent on a novel tripartite ATP-independent periplasmic transporter. *Mol. Microbiol.* **58**, 1173–1185
82. Mu, A., Severi, E., Mulligan, C., Watts, A. G., Kelly, D. J., Wilson, K. S., Wilkinson, A. J., and Thomas, G. H. (2006) Conservation of Structure and Mechanism in Primary and Secondary Transporters Exemplified by SiaP , a Sialic Acid Binding Virulence Factor from Haemophilus influenzae *. *J. Biol. Chem.* **281**, 22212–22222
83. Melaugh, W., Campagnari, A. A., and Gibson, B. W. (1996) The lipooligosaccharides of Haemophilus ducreyi are highly sialylated. *J. Bacteriol.* **178**, 564–570
84. Post, D. M. B., Mungur, R., Gibson, B. W., and Munson, R. S. (2005) Identification of a novel sialic acid transporter in Haemophilus ducreyi. *Infect. Immun.* **73**, 6727–6735
85. Hood, D. W., Cox, A. D., Gilbert, M., Makepeace, K., Walsh, S., Deadman, M. E., Cody, A., Martin, A., Mansson, M., Schweda, E. K. H., Brisson, J.-R. R., Richards, J. C., Moxon, E. R., and Wakarchuk, W. W. (2001) Identification of a lipopolysaccharide alpha-2,3-sialyltransferase from Haemophilus influenzae. *Mol Microbiol* **39**, 341–350
86. Fox, K. L., Cox, A. D., Gilbert, M., Wakarchuk, W. W., Li, J., Makepeace, K., Richards, J. C., Moxon, E. R., and Hood, D. W. (2006) Identification of a bifunctional lipopolysaccharide sialyltransferase in Haemophilus influenzae: incorporation of disialic acid. *J. Biol. Chem.* **281**, 40024–40032
87. Jones, P. A., Samuels, N. M., Phillips, N. J., Munson, R. S., Bozue, J. A., Arseneau, J. A., Nichols, W. A., Zaleski, A., Gibson, B. W., and Apicella, M. A. (2002) Haemophilus

influenzae type b strain A2 has multiple sialyltransferases involved in lipooligosaccharide sialylation. *J. Biol. Chem.* **277**, 14598–14611
88. Swords, W. E., Moore, M. L., Godzicki, L., Bukofzer, G., Mitten, M. J., and VonCannon, J. (2004) Sialylation of lipooligosaccharides promotes biofilm formation by nontypeable Haemophilus influenzae. *Infect. Immun.* **72**, 106–113
89. Bouchet, V., Hood, D. W., Li, J., Brisson, J. R., Randle, G. A., Martin, A., Li, Z., Goldstein, R., Schweda, E. K., Pelton, S. I., Richards, J. C., and Moxon, E. R. (2003) Host-derived sialic acid is incorporated into Haemophilus influenzae lipopolysaccharide and is a major virulence factor in experimental otitis media. *Proc Natl Acad Sci U S A* **100**, 8898–8903
90. Jurcisek, J., Greiner, L., Watanabe, H., Zaleski, A., Apicella, M. A., and Bakaletz, L. O. (2005) Role of sialic acid and complex carbohydrate biosynthesis in biofilm formation by nontypeable Haemophilus influenzae in the chinchilla middle ear. *Infect. Immun.* **73**, 3210–3218
91. Almagro-Moreno, S., and Boyd, E. F. (2009) Insights into the evolution of sialic acid catabolism among bacteria. *BMC Evol. Biol.* **9**, 118
92. Matte, A., Tari, L. W., and Delbaere, L. T. J. (1998) How do kinases transfer phosphoryl groups ? *Structure* **6**, 413–419
93. L.N., J. (1993) The Effects of Phosphorylation on the Structure and Function of Proteins. *Annu. Rev. Biophys. Biomol. Struct.* **22**, 199–232
94. Hunter, T. (1995) Protein kinases and phosphatases: The Yin and Yang of protein phosphorylation and signaling. *Cell* **80**, 225–236
95. Tong, Y., Tempel, W., Nedyalkova, L., MacKenzie, F., and Park, H. W. (2009) Crystal structure of the N-acetylmannosamine kinase domain of GNE. *PLoS One* **4**, 1–9
96. Martinez, J., Nguyen, L. D., Hinderlich, S., Zimmer, R., Tauberger, E., Reutter, W., Saenger, W., Fan, H., and Moniot, S. (2012) Crystal structures of N-acetylmannosamine kinase provide insights into enzyme activity and inhibition. *J. Biol. Chem.* **287**, 13656–13665
97. Caing-carlsson, R., Goyal, P., Sharma, A., Ghosh, S., Setty, T. G., North, R. A., Friemann, R., and Ramaswamy, S. (2017) Crystal structure of N -acetylmannosamine kinase from Fusobacterium nucleatum research communications. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **F73**, 356–362

Chapter 2

Materials and Methods

2.1 General reagents

2.1.1 Chemicals and reagents

The chemicals and reagents used for experiments were purchased from Sigma-Aldrich, Amresco, Himedia, and ThermoFischer Scientific. Molecular biology reagents such as dNTPs (for PCR), DNA polymerases, and restriction endonucleases (for cloning) were from New England Biolabs and Fermentas Life Sciences. Gel loading dye and molecular weight markers (for electrophoresis) were from Bio-Rad laboratories. The plasmid purification and gel extraction kits were from Qiagen, Himedia, and Macherey-Nagel (Clonetech, Takara). Recombinant genes for protein expression were purchased from GenScript and GenArt (Invitrogen). Protease inhibitor cocktail tablets (without EDTA) were from Sigma-Aldrich. Crystallization screens were from Qiagen and Hampton Research.

2.1.2 Biological reagents

The *E. coli* DH5α and BL21 (DE3*) competent cells were purchased from Novagen.

2.1.3 General materials

The Ni-NTA beads (for the affinity chromatography purification) were from Invitrogen. The empty columns used to pack nickel beads were from Bio-Rad laboratories. All the other prepacked columns such as ion exchange and gel filtration were purchased from GE Health care. The 96 well and 364 well plates (used for crystallization trials and kinetic assays) were from Nunc and Corning. For freezing crystals, the loops, canes etc., were purchased from Hampton Research.

2.2 General methods

2.2.1 Bioinformatic analysis

The catabolic and incorporation pathway gene sequences for different bacteria were extracted using BLASTP software. ClustalW software was used for sequencing analysis.

2.2.2 Sterilization of culture media

For protein expression, bacterial culture media was prepared in Milli-Q water and autoclaved at 121^0C for 15 min. Protein purification buffers were filtered using either 0.22 µm or 0.45 µm Millipore filters.

2.3 Molecular biology techniques

2.3.1 Polymerase chain reaction (PCR)

PCR was carried out using Bio-Rad C1000 Thermal cycler using either Taq or Phusion polymerase. Following components were mixed in a 1.5 ml tube kept on ice.

Component	**50µl reaction**
10X reaction buffer	5 µl
2 mM dNTPs	5 µl
2 µM Forward primer	2.5 µl
2 µM Reverse primer	2.5 µl
Template DNA	Variable
Taq DNA polymerase	Based on manufacturer's specifications
Nuclease-free water	To 50 µl

Further, the components were mixed gently and transferred onto the thermocycler, where the thermal block was pre-heated to 95^0C.

Step	Name of the step	Temperature	Time
1.	Denaturation	95^0C	10 minutes
2.	Denaturation	95^0C	30 sec
	Annealing	Melting temperature-5^0C	30 sec
	Extension	72^0C	1 minute/kb
	Go to step 2 for 30 times		
3.	Final extension	72^0C	10 minutes
4.	Hold	10^0C	For ever

2.3.2 Restriction digestion

Restriction digestion was carried out for obtaining the gene of interest and a plasmid for cloning. Reaction buffer (10x), plasmid/PCR product, restriction enzyme 1 and 2 were mixed according to the manufacturer's protocol and incubated at 37^0C for 1-2 hours.

2.3.3 Agarose gel electrophoresis

After PCR and restriction digestion, DNA was visualized using agarose gel. Based on the size of the DNA, different percentage agarose gels were used. DNA samples were loaded

on the gel along with the DNA ladder. Then, the electrophoresis was carried out in a buffer tank containing 1XTBE buffer and ethidium bromide at 120 Volts for 35 minutes. Ethidium bromide stained DNA gels were visualized using an ultraviolet (UV) transilluminator. Based on the required product, the DNA fragments were excised from the gel and purified using gel extraction kit.

2.3.4 Ligation

Ligation is a procedure where two fragments of DNA are joined using T4 DNA ligase. After digestion and gel extraction, the insert and vector were mixed in the 5:1 ratio along with 10X T4 ligase buffer, T4 DNA ligase, and milli-Q water. The reaction was incubated at room temperature for 3-4 hours and then transformed into *E. coli* DH5α cells.

2.3.5 Plasmid extraction (Miniprep)

The colonies obtained after transformation were inoculated in the LB media containing respective antibiotic and grown at 37^0C for overnight. Next day, the cells were pelleted and subjected to plasmid extraction using manufacturer's protocol. Finally, the DNA was eluted in nuclease-free water. The clones were confirmed by characteristic restriction digestion and confirmed using DNA sequencing.

2.3.6 Site directed mutagenesis

Site directed mutagenesis of *Hd*-SatA and NanK was carried out by PCR using the QuickChange site directed mutagenesis protocol. All primer sequences to NanKs are listed in section III.

2.3.7 DNA Sequencing

After plasmid extraction, the clones were sequenced at the NCBS sequencing facility. The gene sequences related to *Hd*-SatA and NanKs are listed in sections I, II, and III.

2.4 Microbiology

2.4.1 Bacterial strains

For the research work conducted in this thesis, *E .coli* DH5α was used for DNA amplification and *E .coli* BL21 (DE3*) cells for protein expression.

2.4.2 Bacterial culture media

For DNA extraction and protein expression, bacterial cultures were grown in Luria Bertani (LB) broth (purchased from Himedia).

2.4.3 Antibiotics

The following antibiotics were used in the present study.

- Ampicillin - at a final concentration of 100 µg/ml
- Kanamycin - at a final concentration of 30 µg/ml

2.4.4 Competent cell preparation

A protocol from 'Molecular Cloning: A laboratory manual by Sambrook' was followed to prepare the chemical competent *E. coli* cells. Bacterial cells were inoculated from a frozen stock into 5ml of LB medium and grown overnight at 37^0C, 200 rpm. The following day, cells were reinoculated into 50 ml of LB medium to a final concentration of 1% and the culture was allowed to grow at 37^0C in a shaker until the absorbance (600nm) reached 0.35-0.4 OD. The cells were immediately placed on ice and harvested by centrifugation at 2000 rpm for 15 minutes at 4^0C. Supernatant was discarded and the pellet was resuspended gently in sterile ice-cold buffer containing 80 mM $MgCl_2$ and 20 mM $CaCl_2$. Cells were kept on ice for 30 minutes. Further, cells were harvested by centrifugation at 1000 rpm for 15 minutes at 4^0C. Supernatant was discarded and the cell pellet was resuspended in 1 ml ice cold buffer containing 100 mM $CaCl_2$ and 15% glycerol. Cells were aliquoted into sterile 1.5 ml microcentrifuge tubes and flash frozen in liquid nitrogen. These frozen cells were stored at -80^0C.

2.4.5 Bacterial transformation

For bacterial transformation, the frozen competent *E. coli* cells were thawed on ice. About 1-5 µl of DNA (approximately equivalent to 100 ng) was added to the 50 µl of thawed cells and mixed gently by flicking the bottom of the tube. These cells were placed on ice for 30 min and given a heat shock at 42^0C for 40 sec. The cells were immediately placed back on ice for 2 min and then suspended in 1 ml of LB media. Following this, the cells were allowed to grow at 37^0C in a shaking incubator for 45 min-1 hours. After the outgrowth, cells were harvested at 8000 rpm and the supernatant was discarded. Finally, the cells were resuspended in 100 µl of LB and then plated on LB plate containing appropriate antibiotic. These LB plates were incubated at 37^0C overnight.

2.5 Protein purification

Hd-SatA and NanK proteins for this thesis work were expressed in BL21 DE3* cells and purified using Ni-NTA affinity chromatography and ion exchange and size exclusion chromatography techniques. The details of the protein purification procedure and the buffers

used for purification are mentioned in the materials and methods section in Chapters III and IV. Protein concentration was measured using a spectroscopic method by measuring the absorbance at A280. The molecular weight and extinction coefficient of the protein were estimated from http://protcalc.sourceforge.net/. Further concentration of the protein was calculated based on the Beer-Lambert law.

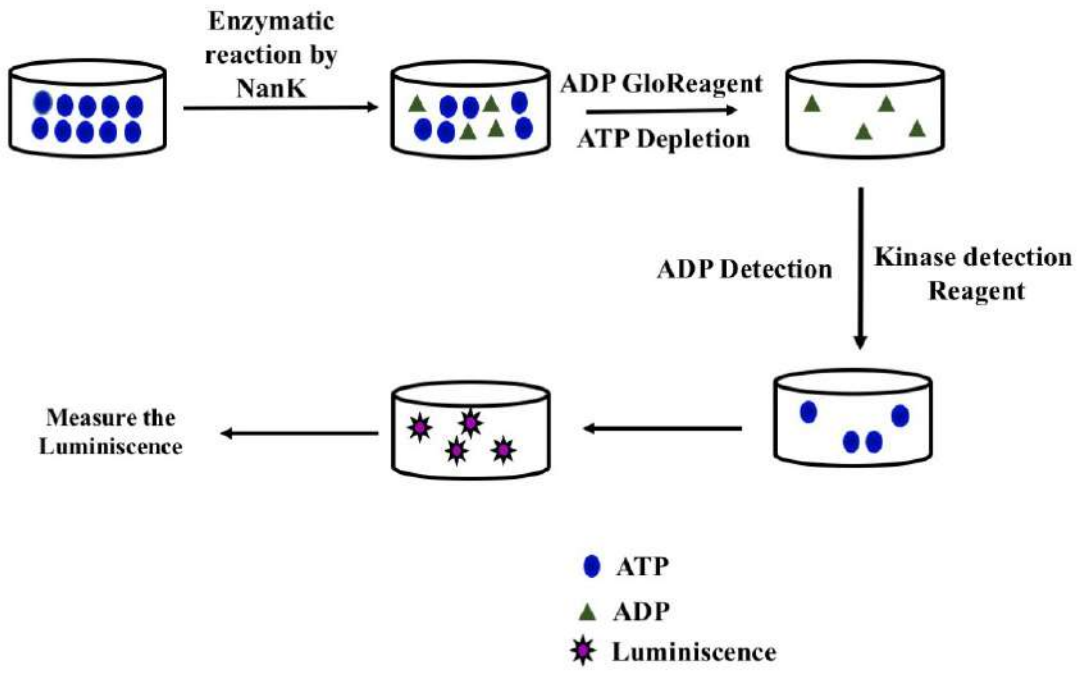
2.6 Biophysical studies

2.6.1 Isothermal calorimetry

Binding affinities and enthalpic and entropic measurements were carried out using Isothermal calorimetry.

2.6.2 Kinetic studies

Enzyme kinetic studies for NanK enzymes were carried using an ADP glo kinase assay kit from promega.



2.1 Schematic representation showing the ADP glo kinase assay protocol. After the kinase reaction, the remaining amount of ATP is depleted by the ADP glo reagent. Kinase detection reagent is added to convert the ADP generated during kinase reaction into ATP. The newly generated ATP will be measured based on the luciferase reaction using a plate reader. The luminescence measure corresponds to the kinase activity.

2.7 List of plasmids and primers:

The gene and primer sequences for *Hd*-SatA, incorporation pathway genes, catabolic pathway genes, and NanK genes are listed in sections I, II, and III.

Experiments related to cloning, expression, purification, and crystallization of catabolic and incorporation pathway enzymes are published in Journal - Microbial Biotechnology and it is attached as annexure 1.

Bairy S, Gopalan LN, **Setty TG,** Srinivasachari S, Manjunath L, Kumar JP, Guntupalli SR, Bose S, Nayak V, Ghosh S, Sathyanarayanan N, Caing-Carlsson R, Wahlgren WY, Friemann R, Ramaswamy S and Neerathilingam M. Automation aided optimization of cloning, expression and purification of enzymes of the bacterial sialic acid catabolic and sialylation pathways enzymes for structural data. ***Microb Biotechnol*** (2018) Mar; 11(2): 420–428.

Section I

Haemophilus ducreyi – SatA (*Hd*-SatA)

Hd-SatA in pET21a:

MASMTGGQQMGRGS FSSPSGSTIE AGIAYPISTG FDPLTSS GASSMAANLHIFE GLVDLHPVTR QPYLALAAKE PEQKDDLTYY ISLREGAMFH DGSPVTTEDV VYSFERVLDP AKASLFAQFI PFIASVTALD DNVVEFKLKY PFALFKERLT IIKIVPKHIV EAGQSAFDAK PIGSGPYKFV SATKDDRIVF EANTVYNGHY PAKVEKMTWF LLSDDAARVT AQESGRVQAI ESVPYLDAER LKRKNNVESV QSFGLLFLMF NCEKAPFDNP KVRQALHYAL DKQKLIDIVF LGNAKAATSY LQDTHPDYVK ASSQYDYDKA KAEKLLAEAG ITNLTFQLLA TDHAWVKECA PLILESWNAL SVVKVTLQHLQ SGALYSAHV DKGAYEVVIA PGDPSVFGND LDLLLSWWYR GDVWPKRRFR WANTAEYHEV QKLLDEAIKN PAGSKVAWQK AINIIAEQVP LYPIIHRKLP TAWNTKKLTD FQPLPTTGLS FLGVGRT LE HHHHHH

Section II

DNA and protein sequences of the sialic acid incorporation and catabolic pathway genes from *Haemophilus influenzae, Pasteurella multocida, Vibrio cholera, and Fusobacterium nucleatum* are listed separately, as given below.

Haemophilus influenzae : catabolic pathway

1. *Hi*-NanK : N-acetylmannosamine kinase

Accession number: WP_011271901

ATGCGTTGTTTAGCACTAGATATTGGTGGGACAAAAATTGCAGCGGCGATTGTAAAAAATGGCGA
AATTGAGCAACGTCAGCAAATTCATACACCACGTGAAAATGTCGTGGAGGGAATGCACCAAGCGT
TAGGCAAATTGCTTGCTGATTATGAAGGGCAGTTTGATTATGTAGCGGTAGCATCTACGGGGATTA
TCAATAATGGTATTTTAAGTGCATTAAATCCGAAAAACTTAGGTGGCTTGGCGGAATTTCCACTTA
AAGCGAGTATTGCAAAACATACGGATAAACCGATTGGTTTATTAAATGATGCACAAGCTGCGACT
TATGCGGAATATCAATTGCAAAATTTTGAACAGGTATCAAATTTTGTTTTTATTACCGTTTCAACAG
GAGTTGGCGGAGGGATTGTGCTGAACCAAATTTTGCAGACAGGTTCTCGTGGTATTGCAGGGCAT
ATTGGTCATACATTAGCTGATCCAAATGGTGCGATCTGTGGTTGTGGTCGTCGAGGTTGTGTCGAA
GCGATTGCTTCAGGGCGAGCAATTGAAGCAGTTTCTTCCCAATGGGAAGATCCGTGCGATCCAAA
AGAAGTGTTTGAGCGTTTTAGAAAAAATGATGAAAAAGCGACCGCACTTGTCGAGCGTTCTGCGA
AAGCCATTGCTAATTTAATTGCTGATTTAGTGATTAGTTTAGATATTCAAAAAATCGCGATTGGTG
GCAGTGTGGGATTAGCTGAGGGATATTTATCTTTGGTTGAAAAATATTTACAAGATTTTCCTTCGA
TATATTGTTGTGAAATTGAAACCGCTAAATTTGGACAAGATGCAGGTTTAATTGGCGCAGCTTATT
GGGTAAAAGATGTTTTACTGGATAAGCCTGAAGGAACAATTTATGGCTAA

MRCLALDIGGTKIAAAIVKNGEIEQRQQIHTPRENVVEGMHQALGKLLADYEGQFDYVAVASTGIINN
GILSALNPKNLGGLAEFPLKASIAKHTDKPIGLLNDAQAATYAEYQLQNFEQVSNFVFITVSTGVGGGI
VLNQILQTGSRGIAGHIGHTLADPNGAICGCGRRGCVEAIASGRAIEAVSSQWEDPCDPKEVFERFRKN
DEKATALVERSAKAIANLIADLVISLDIQKIAIGGSVGLAEGYLSLVEKYLQDFPSIYCCEIETAKFGQDA
GLI GAAYWVKDVL LDKPEGTIYG

2. *Hi*-NanE : N-acetylmannosamine-6-phosphate 2-epimerase

Accession number: WP_005660885

ATGTCTAAATTATCTTATCAAGAAGTACTTTCTCAAATTCAATATGGTCTCATTTCTTCTTGCCAAC
CTGTCGATGATGGCCCAATGGATAAACCTGAAATCGTTTCTGCTATGGCTCAGGCTTCAGTTATGG
GCGGCGCTTCTGGCTTACGTATTGAGGGTGTTGATAACCTTAAAGCAACCCGTCCTTTTGTGAATG
TTCCGATTATCGGAATTGTAAAACGTGATTTGCCTGATAGTCCCGTGCGTATTACGCCGTTTTTACA
AGATATTGAAGATCTAGCCAATGCAGGTGCGGATATTATTGCAGTAGATGGCACGAGCCGCCCAC
GTCCTGTGGATATTGAAAGTGCGGTGAAAAAAATCCACGAAATGGGCTGTTTGGCAATGGCAGAT
TGTTCTAATTTAGAAGAAGGCTTGTATTGTAAAGCGCTTGGTTTTGATATTGTCGGCAGTACGATG
TCGGGTTATACAGGTGGCGCAGTGCCAGAAGAACCAGATTATCAATTAGTGAAAGATTTAAAATC
AGCGGGTTGTTTTGTGATGGCTGAGGGGCGTTATAACACACCAGAATTGGCAAAAGTGGCAATTG
AAATTGGTGCAGATTGCGTGACTGTAGGTTCAGCCTTAACTCGTCTTGAACATATTGTGAGTTGGT
TTGCTAATTCAGTGAAATCTGCAAGATAG

MSKLSYQEVLSQIQYGLISSCQPVDDGPMDKPEIVSAMAQASVMGGASGLRIEGVDNLKATRPFVNVPI
IGIVKRDLPDSPVRITPFLQDIEDLANAGADIIAVDGTSRPRPVDIESAVKKIHEMGCLAMADCSNLEEGL

YCKALGFDIVGSTMSGYTGGAVPEEPDYQLVKDLKSAGCFVMAEGRYNTPELAKVAIEIGADCVTVG
SALTRLEHIVSWFANSVKSAR

3. *Hi*-NagA: N-acetylglucosamine-6-phosphate deacetylase

Accession number: WP_011271899

ATGAAATACGCATTAACCAACAGTGTGATCTACACTAAATACGAAATTCTACGAGATTATGCGGT
AGTAATTGATAGCGAAACTATTGAAGCAGTGATCCCCCAAGCTGAATTAGAAACAGGTATCAAAA
CCATTGATTTACAAGGTAATAATTTAACGGCTGGTTTTATTGATCTGCAATTAAATGGCTGTGGTG
GCGTGATGTTTAACGACCAAACTAGCGTTGAAACCTTGGAAATTATGCAGGAAACTAACTTAAAA
TCAGGTTGTACAAGTTTCCTTCCCACTTTTATTACCGCGCCTGATGAAAATATAAAAAGTGCGGTA
AAAATTATGCGTGAATATTTAAACAAGCATAAAAACCAAGCACTTGGTTTACATATTGAAGGGCC
TTATCTCAGCATTGAGAAAAAAGGCGTACACCGCCCCGAATATATTCGTGAAATTACGCCTGAAAT
GAAAGATTTTCTCTGTGAAAATGGCGATGTGATTACAAAAATAACCATTGCCGCAGAAAACCCAA
CCATTAACTATACGCCTGATTTTGTGAAAGCGGGCATTATTGTCTCTGTAGGTCATTCTAATGCCAC
TTATGAAGTGGCTAAAGCCGCATTCCACAAAGGCGCAACATTTGCAACCCACTTACATAATGCAAT
GTCGCCAATTAGTTCTGGACGTGAAATGGGCGTAGTGGGCGCAGTATTAGATTCTGATGTTTACAC
TGGCATTATCGTGGATGGTGTGCATATCAATTATGGCAATGTTCGCATTGATAAAAAAATCAAAGG
CGACAAACTTTGTATTGTGACTGACTCTATTGCAGCGGCAGGCGCACCGCCAGAACTAGAAAATTT
TACCTTTGTGGGAAAAACAATTTATATCAAAGATGGGCGTTGCTATGATGCTAATGGCACGATTGC
AGGGGCATCAATCACAATGATGGAATCCATTAAAAATGCGGTAGAATATGTTGAAATTCCTCTTGC
TGAAGCCATTAGAATGAGTAACCTTTATCCTGCACGAGCAATTGGTATAGATGATCGTTTAGGTTC
GGTAGAAAAAGGTAAGATAGCAAACTTAGCAGTGTTTACGCCAAATTACCAAGTAATCGGCACAG
TAGTAAACGGAAAGTGGAAAGAGAATTAATAG

MKYALTNSVIYTKYEILRDYAVVIDSETIEAVIPQAELETGIKTIDLQGNNLTAGFIDLQ LNGCGGVMFN
DQTSVETLEIMQETNLKSGCTSFLPTFITAPDENIKSAVKIMREYLNKHK NQALGLHIEG PYLSIEKKGV
HRPEYIREITPEMKDFLCENGDVITKITIAAENPTINYTP DFVKAGIIVS VGHSNATYEV AKAAFHKGAT
FATHLHNAMSPISSGREMGVVGAVLDSDVYTGIIVDGVHINYGNVRIDKKIKGDKLCIVTDSIAAAGAP
PELENFTFVGKTIYIKDGRCYDANGTIAGASITMMESIKNAVEYVEIPLAEAIRMSNLYPARAIGIDDRL
G SVEKGKIANLAVFTPNYQVI GTVVNGKWKE N

4. *Hi*-NagB: Glucosamine-6-phosphate deaminase

Accession number: WP_011271900

ATGCGTTTTATTCCATTACAAACAGAACAACAAGTCAGTTGTTGGGCTGCACAACATATCATCAAT
CGTATCAACGATTTTAAACCTACCGCCGAGCGTCCTTTTGTATTGGGGTTACCTACTGGTGGTACA
CCACTAAAAACTTATCAAGAACTGATTCGACTTTATCAAGCAGGCAAAGTGAGTTTCAAACACGT
AGTGACTTTCAATATGGATGAATATGTGGCTTTACCAGAGGAACACCCTGAAAGTTACCACAGCTT
TATGTATAACAATTTTTTCAATCATATCGATATTTTACCTGAAAATATCAATATTCTTAACGGCAAT
ACCGATGACCATAATGCAGAATGCCATCGTTATGAAGAAAAAATTAAATCCTACGGCAAAATTCA
TCTGTTTATGGGGGGGGTTGGCGTAGATGGACATATCGCATTTAATGAACCTGCTTCTTCATTAAG
TTCTCGCACCCGTATTAAAACGCTAACGCAAGATACATTAATTGCAAATTCACGCTTCTTTAATAA
CGATGTTACTCAAGTGCCGAAATATGCTTTAACCATTGGTGTCGGTACATTATTAGATGCAGAAGA
AGTGATGATTTTAGCAACAGGACATCAAAAAGCACTGGCAGTGCAAGCTGCCGTTGAAGGCAGTA
TTAACCATTTATGGACAGTCAGTGCATTACAAATGCACCGTCATTTCGTGCTTGTTTGTGATGAAG
CTGCTCAACAAGAGTTAAAAGTGAAAACCGTGAAATATTTTACTGAATTAGAAGGCTCGGTTGCT
GGTACGGATTATCAGGATAAATAG

MRFIPLQTEQQVSCWAAQHIINRINDFKPTAERPFVLGLPTGGTPLKTYQELIRLYQAGKVSFKHVVTFN
MDEYVALPEEHPESYHSFMYNNFFNHIDILPENINILNGNTDDHNAECHRYEEKIKSYGKIHLFMGGVG
VDGHIAFNEPASSLSSRTRIKTLTQDTLIANSRFFNNDVTQVPKYALTIGVGTLLDAEEVMILATGHQKA
LAVQAAVEGSINHLWTVSALQMHRHFVLVCD EAAQQELKVK TVKYFTELEG SVAGTDYQDK

Haemophilus influenzae : incorporation pathway

5. *Hi*-SiaB: Acylneuraminatecytidylyltransferase

Accession number: WP_114892122

ATGACAAGAATTGCAATTATTCCTGCGAGAGCGGGTTCTAAAGGAATTAAAGACAAAAATTTACA
GCTTGTGGGTGGTGTTTCTTTGGTGGGGCGTGCGATTTTGGCAGCGCAAGAATCGGGTATGTTTGA
TCAAATTGTGGTGACATCAGACGGTGAAAATATTTTAAAAGAAGCCACAAAATACGGCGCAAAAC
CAGTGGCACGCCCTGAAAGTTTGGCGCAAAGTGATACACGCACCATTGATGCAATTTTGCATTGCT
TAGAAACACTTAATATTTCACAAGGCACCGCTGCACTTTTGCAACCAACCAGCCCATTGCGTAATG
CTTTAGATATTCGTAATGCAATGGAAATTTTCCTTGGCGGCAAATATAAGTCAGTGGTTTCCGCTT
GCGAATGTGAGCATCATCCTTATAAATCCTTCACTTTAGAAGGTACCGAAGTTCAGCCTATTCACG
AATTAACGGATTTTGAAGCCCCTCGTCAAAAATTACCCAAATCTTATCGTGCAAATGGCGCAATTT
ATATTAATGACATCGAAAGTCTCTTTGAAGAAAAACGCTTTTTTATTGCGCCAATGCGTTTTTATTT
AATGCCAACTTATCGTTCCATTGATATTGACTCCACACTTGATTTGCAATTAGCCGAAAGTTTAATT
TCAAAAGAATTCTAA

MTRIAIIPARAGSKGIKDKNLQLVGGVSLVGRAILAAQESGMFDQIVVTSDGENILKEATKYGAKPVAR
PESLAQSDTRTIDAILHCLETLNISQGTAALLQPTSPLRNALDIRNAMEIFLGGKYKSVVSACECEHHPY
KSFTLEGTEVQPIHELTDFEA PRQKLPKSYR ANGAIYINDI ESLFEEKRFF IAPMRFYLMPTYRSIDIDST
LDLQLAESLI SKEF

6. *Hi*_LsgB:CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3 sialyltransferase

Accession number: WP_011272755

ATGAATCTCATTCTTTGTTGTACACCATTGCAAGTATTAATTGCGAGAAAAATAATTGAATTACAT
CCTAATGAACAATTCTTTGGCGTAATGTTTGGTAGGGTATGGGATAAAAAAAGAACCTTGTATGCG
AGTAAATTAGCGGAGGTTTGTAGTGATTCGATGAATATTGATACGGGAAAAGATCTAAAAGGTTT
CGATTTTCTCAAGTTAATGCGCCAGCTTAAGAATAAAATTACGCACAAAGGTTTTGATAAGGTTTT
TCTCGCCAATCTTAATTCATTATGGCTACAAACTTATCTTAGTCATGTTTCATTTAAAGAGCTTTAT
ACCTTTGATGATGGTTCAGATAATATTTTCCCACATCCTAATTTATTGAGAGAGCCTGACACATTTA
AATATAAACTGATTAAAGCTTTTATTGGCGATAAATATAGTGTGAATAAATTATTTAAAAAAATAA
AAAAACATTATACGGTTTATCCTAATTACAAAAATATTGTTTCTAATATTGAACCAATTTCTTTATG
GGATAATCAAATAGATTGCGAGATAGATGGTGAGGTCTCATTTTTTATTGGGCAACCATTGTTAAA
TACAAAAGAGGAAAATATCTCATTAATAAAAAAATTAAAAGAACAATTTTCTTTTGATTATTATTT
CCCTCATCCAGCAGAAGATTATCGAGTTGATGGGGTAAATTATGTTGAATCTGAATTAATTTTTGA
AGATTACGTATTTAAATATTTATCAAATAAAATAATAATTATTTATACGTTCTTCAGTTCTGTTGCA
TTTAATTTGTTAAGTCATCCTAATGTAGAAATTCGTTTTATTAGAACGAGTATTCCAAGATGGCAAT
TCTGTTACGATTCATTTCCAGACCTTGGATTAAAGATTTATAAGGAAATTTAA

MNLILCCTPLQVLIARKIIELHPNEQFFGVMFGRVWDKKRTLYASKLAEVCSDSMNIDTGKDLKGFDFL
KLMRQLKNKITHKGFDKVFLANLNSLWLQTYLSHVSFKELYTFDDGSDNIFPHPNLLREPDTFKYKLIK
AFIGDKYSVNKLFKKIKKHYTVYPNYKNIVSNIEPISLWDNQIDCEIDGEVSFFIGQPLLNTKEENISLIKK
LKEQFSFDYYFPHPAEDYRVDGVNYVESELIFEDYVFKYLSNKIIIIYTFFSSVAFNLLSHPNVEIRFIRTSI
PRWQFCY DSFPDLGLKIYKEI

7. *Hi*-SiaA: N-acetylneuraminic acid synthase-like protein

Accession number: WP_005693221

ATGAAATTTGTTTCTATAATTAGTAATTCTAGGATGTTATTTTTATTTTTGTTAATAACCGCCAAGT
CTGATAGAGAGAGTTGTTTGTATATATTTGACGAAAATATTCGAGGTGTTAATATAACTCCTACAT
TTGTTTCTACAAGAGCTAAGGGGTTATTTGATCTTTTTATAAAAAAACTTAGAAGTAGGGTAGCTT
TATCTTTCTTTTTACTAAAAAGAAAAATAAAGTTGGAAGAGACTATAGTTTATGGTGCAGATCATT

TGTCTCATTCATTGTTATTTCTGAAGAAGTGTTCTTTCTTTTTAATAGAAGATGGTACAGAAAATTA
CCACCAAAAAAGTTATAAGAGAAGTTGGAAAAATAAACTATTTTCAATACCAAAATTTGGAATGT
ATAAAAATGTAAAAAGGATCTATCTTACAAAGAGAGAGAATGTTCCTGATTGTATAAAAAGTAAA
GTAGAATACATAAATATCAAAGATCTTTGGTTTAAAAAGACAGAAGAGGAAAAATTAGAAATCTT
ATATCTATTGGGGATAGATATGAAAAAAATTCAGCTATTGATTGGCGAGCCTTTTATTTTATTTACT
CAACCTTTATCTGAAGATTATATTCTTACAGAAAATGAAAAAATTGAATTATATAAAAGCATTATT
GATAAGTATGATGCATCAAAATTGGTGATAAAGCCTCATCCAAGAGAGAAGACGGATTATTCTAG
AATTTTTCCAAATGTTAAGGTGTTTGATGAAACCTACCCTTCGGAAGTATTGGATATTTTGGAAGT
GAAATTTAGTAGAGTAATTACTCTTTTTTCTACAGCAGCTTTTTCTTACCCCAAAGAGAAAGTCGAT
TTTTATGGAACCAAAATACACCCTAAGTTATTAGCAAAGTTTGGAAATATTGAGTATGAATAG

MKFVSIISNSRMLFLFLLITAKSDRESCLYIFDENIRGVNITPTFVSTRAKGLFDLFIKKLRSRVALSFFLLK
RKIKLEETIVYGADHLSHSLLFLKKCSFFLIEDGTENYHQKSYKRSWKNKLFSIPKFGMYKNVKRIYLT
KRENVPDCIKSKVEYINIKDLWFKKTEEEKLEILYLLGIDMKKIQLLIGEPFILFTQPLSEDYILTENEKIEL
YKSIIDKYDASKLVIKPHPREKTDYSRIFPNVKVFDETYPSEVLDILEVKFSRVITLFSTAAFSYPKEKVDF
YGT KIHPKLLAKFGNIEYE

8. *Hi*-Lic3a: CMP-Neu5Ac--lipooligosaccharide alpha 2-3 Sialyltransferase

Accession number: WP_011271998

ATGTCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA
ATCAATCAATCAAAGTCTGTCATTATTGCAGGTAATGGAACAAGTTTAAAATCAATTGACTATAGT
TTATTACCTAAAGATTATGATGTTTTCCGTTGCAATCAATTTTATTTTGAAGATCATTATTTTCTTGG
CAAGAAAATAAAAAAGGTATTTTTTAATTGTTCTGTAATTTTTGAACAATACTATACGTTTATGCA
ATTAATTAAAAATAATGAATATAAATATGAATATGCTGATATTATCTTAGCATCTTTTCTGAATTTA
GGGGATTCAACATTAAAGAAAATCCAGCATTTAGAAAAATTACTGCCACAAATTGATCTTGGTCAT
TGCTATTTGAAAAAACTACGAGCTTTTAATGCTCATTTACAATATCACGAATTATATGAGAATAAG
AGGATTACATCAGGCGTTTATATGTGTGCAGTGGCGACTGCAATGGGTTATAAAGATCTTTATTTA
ACAGGTATTGATTTTTATCAAGAAAAAGGGAATCCTTACGCATTTCATCATCAAAAAGAAAATATT
ATTAAATTATTACCTTCTTTTTCACAAAATAAAAGTCAAAGCGATATCCATTCTATGGAATATGATT
TAAATGCACTTTATTTTTTACAAAAACATTATGGAGTAAATATTTATTGCATTTCGCCAGAAAGTCC
TCTATGTAATTATTTTCCTTTATCACCACTGAATAACCCAATTGCTTTTATTCCAGAAGAAAAGAAA
AATTACACACAAGATATTTTAATTCCGCCGAAGTTTGTATATAAAAAAATTGGTATATATTCCAAA
CCAAGAATTTACCAAAATCTGATTTTTCGGTTGTTCTGGGATATATTACGTTTACCTAATGATATAA
AACACGCCTTAAAATCAAGAAAATGGGATTAG

MSINQSINQS INQSINQSIN QSINQSKSVI IAGNGTSLKS IDYSLLPKDY DVFRCNQFYF EDHYFLGKKI
KKVFFNCSVIFEQYYTFMQLIKNNEYKYEYADIILASFLNLGDSTLKKIQHLEKLLPQID LGHCYLKKLR
AFNAHLQYHELYENKRITSGVYMCAVATAMGYKDLYLTGIDFYQEKGNPYAFHHQKENIIKLLPSFSQ
NKSQSDIHSMEYDLNALYFLQKHYGVNIYCISPESPLCNYFPLSPLNNPIAFIPEEKKNYTQDILIPPKFVY
KKIGIYSKPRIYQNLIFRLF WDILRLPNDI KHALKSRKWD

9. *Hi*-Lic3B: Alpha-2,3-sialyltransferase (CST-I)

Accession number: ABR14150

ATGAACGGTACAATATGCCCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAT
CAATCAATCAATCAATCAATCAATCAATCAATCAATCAAAGTCTGTCATTATTGCAGGTAATGGAA
CAAGTTTAAAATCAATTGACTATAGTTTATTACCTAAAGATTATGATGTTTTCCGTTGCAATCAATT
TTATTTTGAAGATCATTATTTTCTTGGTAAGAAAATAAAAAAGGTATTTTTTAATTATTCTGTAATT
TTTGAACAATACTATACGTTTATGCAATTAATTAAAAATAATGAATATGAATATGCTGATGTTATT
CTATCATCTTTTCTGAATTTAGGGGATTCAGAATTAAAGAAAATCCAGCGTTTAGAAAAATTACTA
CCACAAATCGATCTTGGTCATAGCTATTTAAAAAAACTACGAGCTTTTGATGCTCATTTACAATAT
CACGAACTATATGAGAATAAGAGGATTACATCAGGCGTTTATATGTGTGCAGTGGCAACTGCTAT
GGGTTATAAAGATCTTTATTTGACAGGCATTGATTTTTATCAAGAAAAAGGGAATCCTTACGCATT
TCATCATCAAAAAGAAAATATTATTAAATTACTACCTTCTTTTTCACAAAAGAAAAGTCAAAAT
GATATCCATTCTATGGAATATGATTTAAATGCACTTTATTTCTTACAAAAACATTATGGTGTAAATA
TTTACTGTATTTCGCCAGAAAGTCCTCTATGTAATTATTTTCCTTTATCACCACTGAATAACCCATT
TACTTTTATTCCCGAAGAAAAGAAAAATTACACACAAGATATTTTAATTCCGCCAAAGTCAATGTA

TAAAAAAATTGGTATATATTCCAAACCAAGAATTTACCAAAATCTGGTTTTTCGGTTGATCTGGGT
ATATTACGTTTACCTAATGATATAAAAAAAGCTTTGAAAGCAAAGAAAATGAGACTACGCAAATA
A

MNGTICPINQSINQSINQSINQSINQSINQSINQSKSVIIAGNGTSLKSIDYSLLPKDYDVFRCNQFYFEDH
YFLGKKIKKVFFNYSVIFEQYYTFMQLIKNNEYEYADVILSSFLNLGDSELKKIQRLEKLLPQIDLGHSY
LKKLRAFDAHLQYHELYENKRITSGVYMCAVATAMGYKDLYLTGIDFYQEKGNPYAFHHQKENIIKS
FSQKKSQNDIHSMEYDLNALYFLQKHYGVNIYCISPESPLCNYFPLSPLNNPFTFIPEEKKNYTQDILIPP
KSMYKKIGIYSKPRIYQ NLVFRLIWDILRLPNDIKKALKAKKMRLRK

Fusobacterium nucleatum: Catabolic pathway

10. *Fn*-NanA

Accession number: WP_011017181

ATGAAAGGGATATATTCAGCATTGATGGTTCCATACAATGAAGATGGAAGTATTAATGAAAAAGG
ATTAAGAGAAATTATAAGATATAACATTGATAAAATGAAAGTTGATGGATTATATGTAGGTGGAA
GTACTGGGGAAAACTTTATGATTTCTACTGAGGAAAAGAAAAGAGTTTTTGAAATTGCAATAGAT
GAGGCAAAAGATTCTGTAAATCTTATTGCTCAAGTAGGAAGTATAAACTTAAATGAGGCTGTGGA
ATTAGGAAAATATGTAACAAAATTAGGTTATAAATGTCTATCTGCTGTGACTCCTTTCTATTATAA
ATTTGATTTTTCTGAAATAAAAGACTACTATGAAACTATAGTTAGAGAAACAGGAAATTATATGAT
AATATATTCTATTCCATTTTTAACTGGAGTTAATATGTCTCTTTCTCAATTTGGTGAATTATTTGAA
AATGAAAAAATCATAGGAGTAAAATTCACTGCTGGTGATTTTTATCTTTTAGAAAGAGTTAGAAAA
GCTTTCCCAGATAAATTAATCTTTGCTGGTTTTGATGAAATGCTATTACCAGCTACTGTTCTTGGTG
TAGATGGTGCAATAGGAAGTACCTATAATATAAATGGAATAAGAGCTAAACAAATATTTGAATTA
GCAAAAAATTCTAAAATAGATGAAGCTTTAAAAATTCAACATACAACTAATGATTTGATAGAAGG
AATTTTATCCAATGGTCTATATCAAACAATAAAAGAAATATTAAAACTTGAAGGTGTTGATGCTGG
TTATTGTAGAAAACCTATGAAAAAAATCAGTCAAAAACAAATTGAATTTGCAAAAGAACTTCATA
AAAAATTTTTAAAAAATTAA

MKGIYSALMVPYNEDGSINEKGLREIIRYNIDKMKVDGLYVGGSTGENFMISTEEKKRVFEIAIDEAKD
SVNLIAQVGSINLNEAVELGKYVTKLGYKCLSAVTPFYYKFDFSEIKDYYETIVRETGNYMIIYSIPFLTG
VNMSLSQFGELFENEKIIGVKFTAGDFYLLERVRKAFPDKLIFAGFDEMLLPATVLGVDGAIGSTYNING
IRAKQIFELAKNSKIDEALKIQHTTNDLIEGILSNGLYQTIKEILKLEGVDAGYCRKPMKKISQKQIEFAK
ELHKKFLKN

11. *Fn*-NanK

Accession number: WP_011017180

ATGAATATTTTAGCAATAGATATCGGTGGAACAATGATAAAATATGGTTTAGTTTCTTTTGATGGA
AAAATTTTATCAACTGATAAAATAAAAACTGAAGCTAGTAAAGGTTTAAATAATATTTTAAATAA
AATAGATAATATTTTTAAAAGATATAAAGAGAATAATCCAGTTGGGATAGCTGTATCTGGGACAG
GTCAGATAAATGGTATGATAGGAAAAGTGATAGGGGGAAATCCTATTATACCAAACTGGATAGGC
ACAAATCTTGTTAAAATATTAGAAGAAAAATATAATTTACCCATTGTTTTAGAAAATGATGTGAAC
TGTGTTGCTCTTGGAGAAAAATGGGTAGGAGCAGGAAAAGATTTAAGTAATTTTATCTGTCTTACA
ATAGGAACAGGTATAGGTGGAGGAATTCTCTTAAATAATCAGCTTTTTAGAGGAGAAAATTTTGTA
GCAGGAGAATTCGGACATATATTAATAAAAAAAGGTGAATTTGAACAGTTTGCTTCTACAACAGC
ACTAATTAGACTAGTAAAAGAAAGAACTGGAAAAACATTAAATGGAAAAGAAATTTTTGATTTAG
AAAAAAAAGAAATACTGGAATATCAAGAAATTATTTCTGAGTGGATTGAAAATTTAACTGATGGA
CTTTCAAGTATAATTTATTGTTTTAATCCAGCAAATATAATACTAGGAGGAGGAGTTATAGAACAA
GGAGAACCTCTTATAAATAGAATAAAAAACAGTTTATTTAAAAAGATAGGTCCTCAATTTAAAGA
AAAGTTGAATATAACACAAGCAAAATTAGGGAATAATGCTGGAATGATAGGAGCAAGCTATTTAC
TTTTAGAGAAAATTAATAAAAGATAA

MNILAIDIGGTMIKYGLVSFDGKILSTDKI KTEASKGLNN ILNKIDNIFK RYKENNPVGI AVSGTGQING
MIGKVIGGNPIIPNWIGTNLVKILEEKYNLPIVLENDVNC VALGEKWVGA GKDLSNFICLTIGTGIGGGI
LLNNQLFRGE NFVAGEFGHI LIKKGEFEQF ASTTALIRLVKERTGKTLNG KEIFDLEKKE ILEYQEIISE
WIENLTDGLS SIIYCFNPAN IILGGGVIEQGEPLINRIKN SLFKKIGPQF KEKLNITQAK LGNNAGMIGA
SYLLLEKINK R

12. *Fn*-NanE : N-acetylmannosamine-6-phosphate 2-epimerase

Accession number: 5ZKN_A

ATGAATAAAATTTTAGAAAGTATAAGAGGAAAATTAATAGTTTCTTGTCAGGCTTTAGAAGATGA
ACCTCTACATAGTTCTTTTATAATGGGAAGAATGGCATATGCAGCTTATTCAGGTGGAGCTGCTGG
TATTCGTGCAAACACTGTTGAAGATATTAAAGAAATCAAAAAAAAATGTATCTCTTCCAATAATTGG
AATTATAAAAAAAGTTTATAACAATTCTGATGTATATATTACTCCAACAATAAAAGAAGTTGAAG
ATTTAATAAATGAAGGTGTACAAATAATAGCTATTGATGCAACAAAAAGAGAAAGACCTGATAGA
AAAGATTTAAAAAATTTTATAGCTGAAATTAAAGAAAAATACCCTAATCAACTTTTTATGGCAGAT
ATATCATCAGTTGATGAGGCTCTATATGCTGAAAAAATTGGATTTGATATAGTTGGAACAACTCTT
GTTGGTTATACTGATTATACAAAAAACTATAAAGCATTGGAAGAACTTAAAAAAGTTGTTAAGGTT
GTAAAAATTCCAGTGATAGCTGAGGGAAATATTGATACTCCTTTAAAAGCTAAAAAAGCTCTTGA
AATAGGTGCTTTTGCAGTTGTAGTTGGTGGAGCAATAACTAGACCTCAACAAATAACTAAAAAGTT
TGTAGATGAAATGAAGTAG

MNKILESIRG KLIVSCQALE DEPLHSSFIM GRMAYAAYSG GAAGIRANTV EDIKEIKKNV SLPIIGIIKK
VYNNSDVYITPTIKEVEDLINEGVQIIAIDATKRERPDRKDLKNFIAEIKEKYPNQLFMADISSVDEALYA
EKIGFDIVGTTLVGYTDYTKNYKALEELKKVVKVVKIPVIAEGNIDTPLKAKKALEIGAFAVVVGGAIT
RPQQITKKFV DEMK

13. *Fn*-NagA: N-acetylglucosamine-6-phosphate deacetylase [*Fusobacterium nucleatum*] Accession number: WP_008798351

ATGAAAAAAATATTATTAAAAAATGCAAAAATAGTTTTAGAAAATAAATTGATTAATGGT
TCTATTTTAATTTTTAAAAATAAAATAGAAAGAATTTTTACAGATAAATATAATTTATCT
GAATTTACTTTTGATGAAGTTATAGACTTAGAAGGTAAATATTTGGGACCTGCGTTTATT
GATATTCATATACATGGTGCTGATGGTGCTGATGCAATGGATAGTAGTGAAGAAGCTTTA
AGAAAAATTTCTAGTTATTTAGTTAAAGAAGGAACTGCAAATTTTTTAGCTACTACTTTA
ACAAGTTCAAAAGAAATTCTAAAAGAAATTTTAAAGGTTGTTGCTAATTTACAAAATAAA
GATATTGAAGGTGCAAATATTTTTGGAGTTCATATGGAAGGACCTTATTTTTCTATTGAA
TATAAAGGTGCTCAAAATGATAAATATATGTTACCTGCTAGTATAAAAGAGCTTGAAGAA
TATTTATCAGTCAAAGAAGGGATTATAAAATTATTTTCAATATCTCCTCATAATCAAGAA
AATTTAGAGGCTATTAAATTCTTATCTGATAGAGGAGTTATAGTTTCTGTTGGACATTCA
GCTGCAAGCTATGAAGATGTTATGAAAGCTATTGACCTTGGACTTTCTCATGCAACTCAT
ACTTATAATGGAATGAAAGGTTTTACTCATAGAGAACCGGGAGTTGTTGGAGCAGTATTT
AATTCAGATAATATTATGGCAGAAATTATCTTTGATAAATTTCATGTACACCCTGAGGCA
GTGAGAACTCTTATAAAAATAAAAGGTGTGGATAAAGTAATTTGTATTACAGATTCTATG
TCTGCAACAGGTTTAGCAGAAGGAAAATATAAATTAGGTGAGCTTGATGTAAATGTAAAG
GATGGACAAGCAAGGCTTGTTTCAAATAATGCATTAGCTGGCAGTGTACTTAGAATGGAT
GTAGCATTTAAAAATTTAATAGAATTAGGATATAGTATAACAGATGCTTTTAAAATGACT
TCAACTAATGCTGCAAAAGAATTCAAATTAAATACTGGAATTTTAAAAGAAGGAAAGGAT
GCTGATCTAGTTGTTTTAAATAAAGATTATAAAGTTTGTATGACTATGGTAAAGGGAAAA
ATCAAATTTAAAAGATAG

MKKILLKNAKIVLENKLINGSILIFKNKIERIFTDKYNLSEFTFDEVIDLEGKYLGPAFIDIHIHGADGADA
MDSSEEALRKISSYLVKEGTANFLATTLTSSKEILKEILKVVANLQNKDIEGANIFGVHMEGPYFSIEYK
GAQNDKYMLPASIKELEEYLSVKEGIIKLFSISPHNQENLEAIKFLSDRGVIVSVGHSAASYEDVMKAID
LGLSHATHTYNGMKGFTHREPGVVGAVFNSDNIMAEIIFDKFHVHPEAVRTLIKIKGVDKVICITDSMS
ATGLAEGKYKLGELDVNVKDGQARLVSNNALAGSVLRMDVAFKNLIELGYSITDAFKMTSTNAAKEF
KLNTGILKEGKDADLVVLNKDYKVCMTMVKGKIKFKR

14. *Fn*-NagB: Glucosamine-6-phosphate deaminase

Accession number: WP_008691945

ATGAGATTTATTGTAACTGATAATAAAAGAGCTGGTGATTGGGGAGCTGTTTATATTGCTAAGAAA
ATTAAAGAATTTAATCCCAGTCCAGAGAAAAAATTTGTATTAGGTTTACCAACTGGTAGTACACCA
CTTCAAATGTATAAAAGATTGATACAATTCAATAAAGAAGGAATTATTAGCTTTAAAAATGTGATT
ACATTCAATATGGATGAATATGTTGGGCTTCCTGAAACTCACCCACAAAGTTATCATTATTATATG
TATAATAATTTTTTTAATCATATTGATATAGATAAGGAAAATATCAATATTTTAAATGGTATGACT
AAAAAATATGAAGAAGAATGTAAAAGATATGAAGAAAAAATTTTAGAAGTTGGAGGAATAGATT
TATTTTTAGGTGGAGTTGGAGTTGATGGTCATATTGCATTTAATGAGCCTGGTTCATCTTTTAAATC
AAGAACAAGAGAAAAACAATTAACAGAAGATACAATCATTGCTAATTCAAGATTTTTTAATAATG
ACATTACAAAAGTTCCTCAATCTGCTTTAACTGTTGGAGTGGCTACAATTATGGATGCAAAAGAAG
TTTTAATAATGGTAGAAGGCAATAGCAAAGCAAGAGCCTTACATATGGGAATTGAAGAAGGAATA
AATCATATGTGGACCATATCAGCCTTACAATTACATGAAAAAGCTATTATTGTTGCTGATGAAGAT
GCTTGTGCAGAACTTAAAGTTGCAACCTACAAATATTATAAAGATATTGAAAAGAAAAATTATAA
TGTAGATAAGTTAATTGAAAACTTATATAAAAAATAA

MRFIVTDNKRAGDWGAVYIAKKIKEFNPSPEKKFVLGLPTGSTPLQMYKRLIQFNKEGIISFKNVITFNM
DEYVGLPETHPQSYHYYMYNNFFNHIDIDKENINILNGMTKKYEEECKRYEEKILEVGGIDLFLGGVGV
DGHIAFNEPGSSFKSRTREKQLTEDTIIANSRFFNNDITKVPQSALTVGVATIMDAKEVLIMVEGNSKAR
A LHMGIEEGIN HMWTISALQL HEKAIIVADE DACAELKVAT YKYYKDIEKK NYNVDKLIEN LYKK

Fusobacterium nucleatum : Incorporation pathway

15. *Fn*-SiaB : Acylneuraminate cytidylyltransferase [*Fusobacterium nucleatum*]

Accession number: WP_005889097

ATGTATAGAGAGAAAAAAATTCTTTGTGTCATTCCAGCACGTAAAGGAAGTAAAAGAATA
AAATGGAAGAATATTGTGCCTCTTGCTGGGAGTCCTATGTTGGAATATACAGTAAAATGT
GCTTTAAATTCTAAATATATTGATAGAGTTATTGTATCAACAGATAGTTATTATATAAAA
AAACTTGCAAAAAAAAATGGGAGCTGATACTCCATTTATAAGGCCTAAAAATTTAGCAACA
GATGATGCTAAAACAATAGATGTGTTGTTACATGCTGTAAAATATTGTGAAGAATTTGAA
AAAGAAAAATATGATTATTTGGTTTTATTACAAAACACAAGTCCTTTACGTAAAAGTTGG
CAAGTTGATGAGGCAATTGAAAAAATAGTATCTTCTACTCTAGATAGCTTAGTTTCTATT
TCAGAAGTAAGAGAACATCCAGTATTAATGAAAATACTTTCTAACAATAAAAAATTAATC
CCACTTTTAAATAACTTAAAAAAGAACAAATTTAGATCTATATATAGAATAAATGGAGCT
ATTTTTATAAATAAAATAGATAAAAATTTTAATAGTGATACTATTTTAACTAATAATCAA
CTCCCTTATATCATGAAAAGGGAAACTTCGATAGATATAGACACTATTGAAGATATAAAG
GTAGCAGAATATTATTTAGGAGTTGAAAAAATGAAAAAAAACCAGAAATATATTTTAAAG
GGAGAAGTATGGAGAAGTTAA

MYREKKILCVIPARKGSKRIKWKNIVPLAGSPMLEYTVKCALNSKYIDRVIVSTDSYYIKKLAKKMGA
DTPFIRPKNLATDDAKTIDVLLHAVKYCEEFEKEKYDYLVLLQNTSPLRKSWQVDEAIEKIVSSTLDSL
VSISEVREHPVLMKILSNNKKLIPLLNNLKKNKFRSIYRINGAIFINKIDKNFNSDTILTNNQLPYIMKRET
SIDIDTIEDIKVAEYYLGVEKMKKNQKYILKGEVWRS

16. *Fn*-Lic3a: lipooligosaccharide sialyltransferase [*Fusobacterium nucleatum* subsp. polymorphum ATCC 10953]

Accession number: WP_005897393

ATGAATTTATATATAATATATAACTATTGGGAGATATTATTATCTTTACTAGTTTTAGAA
AAAAATAAAGATGGAAATAATATTTTACTAATTGTAGAAAATGAAATTGAAGAAGAATTA
TTAAAGAGACTAGAAAAAAAAATATAAAGTATTAAGATTTAATATAAAACCAAATAGGTTT
TCGAAATTTTTAACATATTATTATAAAATTAATTATCAGTTTCCAAAAAAATTAGGGCGA

TTTTTATCAAGTATAGAAAGAATAATTTCGTTTTCGGACCAAGATGTAATCACTAGATAT
TTTATTAAAAATAAAAAATATATAGATTTATTTGAACATGGTGTAATAAATTATCAAACT
GAATTTGAAGGGATAGAACAAAAAATAAAGAGAATAATTTTTAAAATGGAAAAACCATAT
GGGAGAAATGAATATGTTAAAAATATTTTTCTTAGAGGTTCAGGAAAAATACCTGAAGAT
ATAAAAAATAAAGTAAAGATATTAGATTTAAAAGAACTATGGAGTCAATTAGATTTTAGT
AGTCAAAATAAAATATTAGATGTTTTTGGCTTGGATATAGAAAAATTAAAAATTATAGAA
ACTAAAAATACAATACTATTTACACAACCATTTTCAGAAGATAAAGTCATATCAGAAGAT
GAAAAGGTAGATTTGTATAAAAAAATTATAAAAAATTATGATAAAAGGAGTTTAGTAATA
AAAGCTCATCCAAGAGAAAAAACAGAGTATAATAAAATTTTTGAAGATATTATGATACTA
GAAAATAGTTTCCCAGCTGAATTGTTGTTATTAAGTAAATTTAAGTTTGAAAGAGTAGTC
ACAATATCTTCAACGGCAGTCTCAGTATTTTTTAATAAATCAGAGATAGATTTTTATGGT
TCAGAAGTTCATCCTAAAATTTTAAAATATTTTGGTAATTTAGATTTTTTTATGAAAAGA
AATAAGTTTATAGAAAAATAA

MNLYIIYNYWEILLSLLVLEKNKDGNNILLIVENEIEEELLKRLEKKYKVLRFNIKPNRFSKFLTYYYKIN
YQFPKKLGRFLSSIERIISFSDQDVITRYFIKNKKYIDLFEHGVINYQTEFEGIEQKIKRIIFKMEKPYGRNE
YVKNIFLRGSGKIPEDIKNKVKILDLKELWSQLDFSSQNKILDVFGLDIEKLKIIETKNTILFTQPFSEDKV
ISEDEKVDLYKKIIKNYDKRSLVIKAHPREKTEYNKIFEDIMILENSFPAELLLLSKFKFERVVTISSTAVS
VFFNKSEIDFYGSEVHPKILKYFGNLDFFMKRNKFIEK

Pasteurella multocida: catabolic pathway

17. *Pm*-NanK

Accession number: WP_005752223

ATGCGCTGTTTAGCATTAGATATTGGTGGAACAAAAATTGCCTCGGCTATTGTAACAGACGGCAAA
ATAGAACAACGCCAACAGATTGCAACGCCACAAGCTGATGCTGCCAATGCGATGCATGACACACT
GGCCAACATTCTTGCTTTGTATGCCGGACAATTCGATTACGTTGCAGTGGCTTCAACCGGCATCAT
TAATCATGGTGTGTTAACCGCCTTAAATCCAAAAAATTTAGGTGGTTTGGCGGAATTTCCCTTAAA
AGAAAGCATTGCGCGCCATACGGACAAACCGATTGGCTTACTAAATGATGTGCAAGCTGCCGCGT
GTGCGGAATATAAAGACGAAGATAAAAATGCGGTACAAAATTTTGTTTTTATCACCGTATCCACTG
GGGTGGGCGGTGGCATTATTTTAGAACGCCGCTTATTGACAGAACCTAACGGCGTTGCCGGGCAT
ATCGGACATACATTAGCAGACCCAAATGGTCCGGTTTGTGGTTGTGGGCGTGTAGGTTGTGTTGAA
GCCGTCGCGGCTGGACGCGCAATTGAAGCCGTTTCTAGCCAATGGAATCCACCTTGCACACCAAA
ACAAGCTTTCGAGTTATTTAGAAAAAATGATGAAAAAGCGACCGCACTTATCCAGCGTTCAGCTA
GCGCGATAGCCAACCTGATTGCAGATTTAGTGATTGGTTTAGATGTACAAAAAGTGGTAGTTGGTG
GTAGTGTTGGTTTAGCGGAAGGTTACCTCCCCTTAGTCAAACAGTACCTCAACACGATGCCACATT
TCTATCATTGTACCGTTGAACAAGCACGTCACGGACAAGATGCTGGGCTACTGGGTGCCGCTTGGT
GGGTAGCCGATTGCCTAAAACAAGGCGTTCATTTGAAATAG

MRCLALDIGGTKIASAIVTDGKIEQRQQIATPQADAANAMHDTLANILALYAGQFDYVAVASTGIINHG
VLTALNPKNLGGLAEFPLKESIARHTDKPIGLLNDVQAAACAEYKDEDKNAVQNFVFITVSTGVGGGII
LERRLLTEPNGVAGHIGHTLADPNGPVCGCGRVGCVEAVAAGRAIEAVSSQWNPPCTPKQAFELFRKN
DEKATALIQRSASAIANLIADLVIGLDVQKVVVGGSVGLAEGYLPLVKQYLNTMPHFYHCTVEQARHG
QDAGLL GAAWWVADCL KQGVHLK

18. *Pm*-NanE

Accession number: WP_005724602

ATGTCAAAACTATCACATCCGCAAGTTTTAGAACAAATTAAATATGGCTTAATCGCATCTTGCCAG
CCCGTAGATAATGGGCCAATGGACTCGCCAGAAATTGTTGCCGCCATGGCGCAAGCATCCGTTATC
GGTGGTGCTGCCGGGTTACGTATTGAAGGGATCGAAAATTTAAAAGCCACGCGTAACGTCGTCAA
TGTACCGATTATTGGTATCGTTAAACGCGATTTACCTGACAGTCCCGTTCGTATTAGCCCATTTTTA
CAAGACATTGAAGAGTTAGCCGCAGCAGGCGCCGATATTATTGCTTTTGATGGTACCGATCGCGTC
CGTCCAACTACACGTGAAGCCATTATTAAACGAATCAAAGAATTAGGTTGTTTAGCCATGGCGGAT

TGTTCTAATTTTGAAGAAGGGATGTATTGCCACAACTTAGGGGTAGAAATTATTGGTAGCACCATG
TCTGGTTATACTGGCGGTGAAATTCCAGCTGAACCGGATTATCAGTTAGTCAAAGATTTGAATGCT
GCTGGCTGCCGAGTGATGGCAGAAGGGCGCTATAACACCCCAGAATTGGCGAAAACCGCAATTGA
GATTGGTGCTTACTCTGTCACCGTAGGATCTGCGCTGACTCGCCTCGAACATATCGTAAGCTGGTT
TGCCGATGCCGTCAAATCGGCAAAATAA

MSKLSHPQVLEQIKYGLIASCQPVDNGPMDSPEIVAAMAQASVIGGAAGLRIEGIENLKATRNVVNVPII
GIVKRDLPDSPVRISPFLQD IEELAAAGAD IIAFDGTDRV RPTTREAIIK RIKELGCLAM ADCSNFEEGM
YCHNLGVEIIGSTMSGYTGGEIPAEPDYQLVKDLNAAGCRVMAEGRYNTPELAKTAIEIGAYSVTVGS
ALTRLEHIVSWFADAVKSAK

19. *Pm*-NagA

Accession number: WP_014326154

ATGTACGCGTTTGTTAATGCCGTAATTTACACTGCAAAAGACGTACTTTATGGTAAAGCGCTTGTT
GTCGATGGCGATAAAATTAGTGCAATTCTACCCGTAGAAGACGTGCCAGAAAATCTTCAGAAAAT
TGATTTACAAGGGAATAATTTAACCGCAGGTTTTATTGATCTTCAATTAAATGGCTGTGGTGGCGT
GATGTTTAATGAAGATATTAGCGTGAAGACGTTAGAGATCATGCAAGAAACTAATTTGAAATCCG
GCACGACAAGTTATTTACCAACCTTTATTACGTCACCTGATGAAGGAATGAAAGACGCAGTGAAG
GTGATGCGTGAGTATTTAACTCAGTATAAAAATCAAGCATTAGGTTTACACTTTGAAGGACCTTAT
TTGAGTGTTGAGAAAAAAGGAGTACATCGCGAAGAATATATTCGTGCGATAAGCCCAGAAATGAA
AACATTCTTGTGTGATAATGCAGATGTGATTACTAAAATCACGTTGGCAGCCGAAAATCCAACGGC
ACAGTATATTCCAGACTTTGTTGAAAAAGGGATTATCGTCTCGCTTGGACACTCGAATGCGACTTA
CGATGTCGCACAACAGGCTATTGAGAAAGGGGCGAGTTTTGCAACCCATTTACATAATGCAATGT
CACCAATCAGCTCTGGACGAGCGATGGGAGTAGTGGGCGCCGTATTAGATTCTGATATCTATACA
GGTATTATTGTGGATGGTTTACACGTTGATTATGGCAATATACGACTCGACAAGAAAGTAAAAGGT
GACAAATTGTGCATTGTGACAGATGCAACTGCTGCAGCCGGTGCAGACATTGACTCCTTTGTTTTC
GTCGGTAAAACCGTCTATGTGCGTGATGGTAAATGCTATGATAGCAATGGTACATTAGGTGGCGCT
GCAATCACGATGATTGAGTCAGTGAAAAATGCGGTACAGGAAGTTGGCATTCCATTGGATGAGAC
ATTGCGTATGTGTAATTACTATCCTGCTAAAGCAATTGGTGTGGACCATAAACTTGGCTCAATTGA
AGTGGGTAAAATCGCGAATTTAACAGCATTCACAAATGATTTTAACGTGTTGGGTACAGCCGTGA
ATGGTGAATGGAAAGCGAATTAA

MYAFVNAVIYTAKDVLYGKALVVDGDKISAILPVEDVPENLQKIDLQGNNLTAGFIDLQLNGCGGVM
FNEDISVKTLEIMQETNLKSGTTSYLPTFITSPDEGMKDAVKVMREYLTQYKNQALGLHFEGPYLSVEK
KGVHREEYIRAISPEMKTFLCDNADVITKITLAAENPTAQYIPDFVEKGIIVSLGHSNATYDVAQQAIEK
GASFATHLHNAMSPISSGRAMGVVGAVLDSDIYTGIIVDGLHVDYGNIRLDKKVKGDKLCIVTDATAA
AGADID SFVFVGKTVY VRDGKCYDSNGTLGGAAITM IESVKNAVQEVGIPLDETLR MCNYYPAKAI
GVDHKLGSIEVGKIANLTAF TNDFNVLGTA VNGEWKAN

20. *Pm*_NagB

Accession number: WP_005716895

ATGCGTTTAATTCCATTACACAATGTTGATCAAGTCGCAAAATGGTCTGCACGTTATATTGTTGATC
GTATTAATCAATTTCAACCAACAGAAGCGCGCCCATTCGTATTAGGTTTGCCTACGGGTGGTACTC
CGCTGAAAACCTATGAGGCGCTTATTGAGCTGTATAAAGCGGGTGAGGTGAGCTTTAAACATGTT
GTGACCTTCAATATGGATGAGTATGTTGGTTTACCAAAAGAACATCCTGAAAGCTACCATTCTTTC
ATGTATAAAAACTTTTTTGATCATGTTGATATTCAAGAGAAAAATATCAATATCTTAAATGGTAAT
ACCGAAGATCATGATGCCGAATGTCAGCGTTATGAAGAGAAAATCAAGTCTTATGGCAAAATTCA
TTTATTTATGGGCGGTGTTGGTGTAGATGGCCATATTGCTTTTAATGAACCGGCTTCATCCCTTTCC
TCACGTACGCGTATAAAAACGCTTACAGAAGATACGCTAATTGCCAACTCACGTTTCTTTGATAAC
GATGTGAATAAAGTGCCTAAATATGCACTCACCATTGGTGTGGGGACGTTACTTGATGCAGAAGA
AGTGATGATTTTAGTCACAGGTTATAATAAAGCTCAAGCTTTACAAGCCGCTGTAGAAGGTAGCAT
CAATCATTTATGGACCGTCACCGCATTACAAATGCATCGCCGAGCGATTATTGTGTGTGATGAACC
TGCGACGCAAGAGCTTAAAGTGAAAACAGTTAAATATTTTACTGAATTAGAAGCCTCTGCTATTCG
TAGCGTGAAATA

MRLIPLHNVDQVAKWSARYIVDRINQFQPTEARPFVLGLPTGGTPLKTYEALIELYKAGEVSFKHVVTF
NMDEYVGLPKEHPESYHSFMYKNFFDHVDIQEKNINILNGNTEDHDAECQRYEEKIKSYGKIHLFMGG
VGVDGHIAFNEPASSLSSRTRIKTLTEDTLIANSRFFDNDVNKVPKYALTIGVGTLLDAEEVMILVTGYN
KAQ ALQAAVEGSI NHLWTVTALQ MHRRAIIVCD EPATQELKVK TVKYFTELEA SAIRSVK

Pasteurella multocida: Neu5Ac incorporation pathway

21. *Pm*-SiaB

Accession number: WP_025248495

ATGACAAATATTGCGATCATTCCTGCACGAGCCGGCTCAAAAGGGATTCCAGATAAAAACCTGCA
GCCCGTGGGTGGGCATTCTCTGATTGGACGTGCTATTTTAGCGGCAAAAAATGCTGACGTTTTCGA
TATGATTGTCGTAACCTCTGATGGTGATAATATCTTACGTGAAGCAGAAAAATATGGGGCATTAGC
GCTGAAAAGACCCGCTGAGTTGGCCCAAGACAATTCAAGAACCATTGATGCTATTTTACATGCACT
AAAGAGCTTAAATATTCGCGAAGGCACTTGTACCCTATTACAACCGACTAGCCCATTACGTGATCA
TTTAGATATCAAAAATGCGATGGATATGTATGCCAATGGCGGAGTACACTCTGTGGTGTCGGCGTG
CGAGTGTGAACATCATCCCTATAAAGCTTTTGCGTTAAGCAAAGATCATGAAGTTTTGCCTGTACG
TGAAATTGCTGATTTTGAAGCCGCACGCCAAACTTTACCGAAAATGTATCGTGCAAATGGTGCCAT
TTATATTAATGACATTGCACAATTGCTAAAAGAAAAATATTTCTTCATTCCGCCTCTCAAATTCTAC
TTAATGCCAACTTATCGTTCTGTGGATATTGATGTGAAGCAGGATTTAGAACTCGCGGAAATTTTA
TCCAATAAATGA

MTNIAIIPARAGSKGIPDKNLQPVGGHSLIGRAILAAKNADVFDMIVVTSDGDNILREAEKYGALALKR
PAELAQDNSRTIDAILHALKSLNIREGTCTLLQPTSPLRDHLDIKNAMDMYANGGVHSVVSACECEHHP
YKAFALSKDHEVLPVREIADFEAARQTLPKMYRANGAIYINDIAQLLKEKYFFIPPLKFYLMPTYRSVDI
DV KQDLELAEIL SNK

22. *Pm*-LsgB

Accession number: WP_014325951

ATGAATTTGATTATTTGTTGTACACCGTTACAGGTGTTGATTGCAGAAAAAATTATCGCTAAATTTC
CGCATACGTCATTTTATGGTGTCATGCTTTCAACAGTCAGTAATAAAAAATTTGATTTTTATGCAAA
GCGGCTTGCGCAACAGTGCCAAGGTTTTTTTTCCATGGTGCAGCATAAGGATCGCTTCAATCTATT
AAAAGAAATTCTGTATTTAAAACGAACGTTTTCGGGTAAGCACTTTGATCAGGTTTTTGTGGCAAA
CATTAATGACTTACAAATTCAGTTTTTATTAAGTGCCATTGACTTTAATCTGTTAAATACCTTTGAT
GACGGCACAATTAATATTGTACCGAATAGTCTTTTTTACCAAGATGACCCTGCCACGTTGCAGCGA
AAACTGATTAATGTGCTGTTAGGTAATAAATACAGTATTCAATCATTACGCGCTTTATCCCATACA
CACTACACTATTTATAAAGGCTTCAAGAATATTATTGAACGGGTGGAGCCGATTGAATTGGTCGCA
GCAGATAACAGTGAAAAAGTCACTTCAGCGGTGATTAACGTATTGCTTGGGCAACCCGTTTTTGCT
GAAGATGAACGCAATATTGCCTTAGCGGAACGCGTGATCAAACAATTTAATATTCATTATTATTTG
CCTCATCCACGCGAAAAGTATCGTTTAGCCCAAGTCAATTACATTGATACGGAATTGATCTTTGAA
GATTATATTCTTCAGCAATGTCAAACCCACAAATACTGTGTTTATACATATTTTAGTAGCGCCATTA
TTAATATCATGAATAAAAGTGACAATATTGAAGTGGTAGCATTAAAAATTGACACAGAAAATCCC
GCCTATGATGCTTGTTATGATTTGTTTGATGAGCTAGGCGTTAACGTTATTGATATAAGAGAGTAA

MNLIICCTPLQVLIAEKIIAKFPHTSFYGVMLSTVSNKKFDFYAKRLAQQCQGFFSMVQHKDRFNLLKEI
LYLKRTFSGKHFDQVFVANINDLQIQFLLS AIDFNLLNTF DDGTINIVPN SLFYQDDPAT LQRKLINVLL
GNKYSIQSLRALSHTHYTIYKGFKNIIERV EPIELVAADN SEKVTSAVIN VLLGQPVFAE DERNIALAER
VIKQFNIHYYLPHPREKYRLAQVNYIDTEL IFEDYILQQC QTHKYCVYTY FSSAIINIMN KSDNIEVVAL
KIDTENPAYD ACYDLFDELGVNVIDIRE

23. Pm_Lic3A

Accession number: WP_015702737

ATGGATAAGTTCGCAGAACATGAAATACCGAAAGCCGTTATTGTTGCAGGGAATGGCGAGAGTTT
AAGTCAAATTGATTATAGGTTGTTACCGAAAAATTATGATGTGTTTCGTTGTAATCAATTTTATTTT
GAAGAACGCTATTTTTTAGGAAACAAGATAAAAGCGGTTTTCTTCACGCCAGGGGTCTTTCTTGAG
CAATATTATACACTTTATCATCTCAAGAGAAACAATGAGTATTTTGTTGATAATGTGATTCTCTCTT
CTTTTAATCATCCTACAGTAGATTTAGAAAAGAGTCAGAAAATACAAGCACTTTTTATTGATGTGA
TCAACGGATATGAAAAGCATTTATCTAAACTCACTGCTTTTGATGTTTATCTGCGCTATAAAGAAT
TATATGAGAATCAAAGAATTACATCTGGCGTATATATGTGTGCAGTTGCTATTGCGATGGGATATA
CAGATATTTACTTAACTGGTATAGATTTTTATCAAGCGAGCGAAGAAAACTACGCATTCGATAATA
AAAAGCCTAATATTATTAGGTTATTGCCTGATTTTCGGAAAGAAAAAACACTCTTTTCTTATCATA
GTAAAGATATTGATTTGGAAGCATTATCTTTTTTACAACAGCATTATCATGTTAATTTTTATTCGAT
TTCACCAATGAGCCCTTTGTCTAAACATTTTCCTATTCCAACTGTAGAGGATGATTGTGAAACAAC
TTTTGTTGCGCCACTAAAAGAAAATTACATTAATGATATATTGTTGCCTCCTCATTTTGTATATGAA
AAATTAGGGACCATCGTGTCTAAGAAATCACGTTTTCATTCTAACTTGATTGTCAGGTTGATTAGA
GACTTATTGAAATTACCGAGTGCACTTAAACACTATTTAAAAGAAAAATAG

MDKFAEHEIPKAVIVAGNGESLSQIDYRLLPKNYDVFRCNQFYFEERYFLGNKIKAVFFTPGVFLEQYY
TLYHLKRNNEYFVDNVILSSFNHPTVDLEKSQKIQALFIDVINGYEKHLSKLTAFDVYLRYKELYENQRI
TSGVYMCAVAIAMGYTDIYLTGIDFYQASEENYAFDNKKPNIIRLLPDFRKEKTLFSYHSKDIDLEALSF
LQQHYHVNFYSISPMSPLSKHFPIPTVEDDCETTFVAPLKENYINDILLPP HFVYEKLGTI VSKKSRFHSN
LIVRLIRDLL KLPSALKHYLKEK

Vibrio cholera: catabolic pathway

24. *Vc*-NanA

Accession number: WP_000736950

ATGAAAAAACTAACAGGTTTGATTGCTGCTCCACATACGCCTTTCACAAAGGATAACAAG
GTTAATTTTGCGGCGATTGACCAAATTGCGGAGCTTCTCATTGAGCAAGGAGTGAAAGGT
GCTTATGTCTGCGGAACGACAGGTGAGGGGATCCATTGCTCTGTCGAAGAAAGAAAAGCC
ATTGCAGAACGTTGGGCGAAAGCCGTTGATGGGAAGTTAGATGTTATTTTACATACAGGC
GCTTTAAGTATTGTTGATACCATCAACTTGACAGAGCACGCTGAAACACTGGATATCTTT
GCTACCTCCGCTATTGGCCCATGCTTTTTCAAACCGGGAAGTGTTGATGATTTAGTTGAG
TATTGTGCTCAAGTTGCTGCTGCAGCTCCTTCAAAAGGTTTTTATTACTACCACTCAGGA
ATGTCGGGTGTGAATCTAGATTTAGAGCAGTTCTTGATTAAAGGCGAACAGCGCATTCCT
AATCTTTATGGCGCAAAATTCAATAATGCTGATTTGTATGAATACCAGCGTTGCGTGCGA
GTATCGAATCGCAAATTTGATATTCCATTTGGTGTGGATGAGTTTTTACCTGCTGGTTTA
GCCATTGGTGCGGTAGGCGCAGTTGGTAGCACTTATAACTATGCTGCACCATTGTATTTG
AAAATCATTGAAGCATTTAACCATGGCAAACATGATGAAGTCGCAGCGTTGATGGATAAA
GTGATTGCGATTATTCGAGTATTGGTCGAATACGGCGGAGTTGCGGCAGGTAAAGTTGCC
ATGCAGTTACATGGTATTGATGCTGGGGATCCCCGCTTACCTATTCGTTCATTGAATGAC
AAGCAGAAAGCGGATGTGCTTGCGAAGATGCGTGATGCTGGGTTCCTGTCGATCTAA

MKKLTGLIAAPHTPFTKDNKVNFAAIDQIAELLIEQGVKGAYVCGTTGEGIHCSVEERKAIAERWAKA
VDGKLDVILHTGALSIVDTINLTEHAETLDIFATSAIGPCFFKPGSVDDLVEYCAQVAAAAPSKGFYYY
HSGMSGVNLDLEQFLIKGEQRIPNLYGAKFNNADLYEYQRCVRVSNRKFDIPFGVDEFLPAGLAIGAV
GAVGSTYNYAAPLYLKIIEAFNHGKHDEVAALMDKVIAIIRVLVEYGGVAAGKVAMQLHGIDAGDPR
LPIRSLND KQKADVLAKMRDAGFLSI

25. *Vc*-NanK N-acetylmannosamine kinase [Vibrio cholerae O1 biovar El Tor str. N16961]

Accession number: WP_001259414

ATGCGCACACTAGCCATCGATATCGGCGGAACAAAAATAGCATTAGCGATTGTAGAAGAA
GGTACGATCATACAGCGTTACCAGATAGCAACGCCTGTGGTGCAGGACGTCACCAAGTTT
GTTCAAGCGATTCTCGAAAAGGTCACTGAATGGCTCCCTTCTATCGATTATGTAGGCGTA
TCAACAACAGGTTATGTAACGCCAGAAGGCATCACATCGATAAATCCAGAGACTCTTAAT
TTCCCTGTACCTTTTCCATTAGCTCAAACGCTAGAACAACTCACTAATAAACCAGTCTCC
ATTCTTAATGATGCACAGGCCGCGGCATGGTTTGAATTTGTGCAGTTAAAAAATCCTAGC
TTAAATATGGCTTTTATTACGGTTTCAACAGGGGTTGGTGGCGGAATCATCATTGATGGC
AAGCTTCACAAAGGAAACAGCGGTTTAGCGGGACACATAGGTCACATGTCGGTGGCCATT
GAAGGTCCATTATGCGGATGTGGTCAGCGAGGTTGTGTTGAGTCCATGGCTTCAGGTAAC
GCGATTCAAAAAGAGAGCGAAGCCACTTTTACCGAGACAATGAGCAATGTTGAGTTATTT
AAGCAAGCAGCGTTCAACCCGAAAGCAGAAGCGATCATCAATCGCAGTGTGCAAGCAGTA
GCCACATTATGCTGTAATCTTAAAGCATGCTTAGATCTGGATATTATTGTTTTGGGTGGT
GGAATTGGTCTCGCAGAGGGCTACCTTGAACGCCTCAACAAAGCTATTCAATCTAGGCCA
TCCGTATTTCACATCCCTGTCACTCCTGCTCATGGCGATTATGATGCTTGCCTACTCGGA
GCAGCATTTCAGTTTAAGGAGTAA

MRTLAIDIGGTKIALAIVEEGTIIQRYQIATPVVQDVTKFVQAILEKVTEWLPSIDYVGVSTTGYVTPEGI
TSINPETLNFPVPFPLAQTLEQLTNKPVSILNDAQAAAWFEFVQLKNPSLNMAFITVSTGVGGGIIIDGKL
HKGNSGLAGHIGHMSVAIEGPLCGCGQRGCVESMASGNAIQKESEATFTETMSNVELFKQAAFNPKAE
AIINRSVQAVATLCCNLKACLDLDIIVLGGGIGLAEGYLERLNKAIQSRPSVFHIPVTPAHGDYDACLLG
AAFQFKE

26. *Vc*-NanE: N-acetylmannosamine-6-phosphate 2-epimerase [Vibrio cholerae]

Accession number: 5ZJB_A

ATGAGAAAGAATTTTTTGAATATCGAAGAGTTAAAGCGATTTTTAAACGGCCAAACTGTC
GTTTCCATACAACCTGTAACAGGCAGCCCACTCGATAAAACAGATTTCATTGTGGCGATG
GCGATCGCCGTCGAACAAGCGGGTGCAAAAGCGCTGCGCATTGAAGGTGTCAGTAACGTG
GCTGCCGTATCCGCCGCAGTAACAATTCCGATTATAGGTATCGTTAAGCGTGACTTGCCT
GATAGCCCAGTTCGGATCACCCCTTTTGTTTCAGATGTGGATGGCTTAGCGAATGCTGGC
GCAACGGTCATTGCTTTTGATGCCACCAATAGAACACGCCCTGAAAGTCGAGAACGTATT
GCGCAAGCAATCAAAAACACGGGATGTTTTGCGATGGCCGATTGCTCCACATTCGAGGAT
GGATTATGGGCTAATTCACAAGGTGTTGAAATTGTTGGATCAACGCTGTCAGGTTATGTC
GGAGATATTGAGCCCACGGTTCCAGATTTCCAACTTGTGAAAGCGTTTTCTGAAGCCGGC
TTTTTTACCATGGCGGAAGGTCGATACAATACGCCTGAGTTAGCAGCGAAAGCTATTGAG
TCTGGTGCGGTTGCCGTAACGGTTGGTTCTGCATTGACTCGATTAGAAGTGGTTACTCAA
TGGTTCAATAACGCAACCCAAGCTGCGGGAGAAAGGAAATGCGCACACTAG

MRKNFLNIEELKRFLNGQTVVSIQPVTGSPLDKTDFIVAMAIAVEQAGAKALRIEGVSNVAAVSAAVTI
PIIGIVKRDLPDSPVRITPFVSDVDGLANAGATVIAFDATNRTRPESRERIAQAIKNTGCFAMADCSTFED
GLWANSQGVEIVGSTLSGYVGDIEPTVPDFQLVKAFSEAGFFTMAEGRYNTPELAAKAIESGAVAVTV
GS ALTRLEVVTQWFNNATQAAGERKCAH

27. *Vc*-NagB : glucosamine-6-phosphate deaminase [Vibrio cholerae O1 biovar El Tor str. N16961]
Accession number: WP_001237050

ATGAGACTTATCCCACTGAAAGCGGCAGCACAAGTTGGTAAATGGGCGGCTGCACACATT
GTTAAGCGTATCAACGAGTTTCAACCAACAGCTGAACGTCCTTTTGTATTGGGACTCCCT
ACTGGTGGCACACCACTCGCGACTTATAAAGCGCTGATTGAAATGCATAAAGCGGGTGAA
GTGAGCTTCAAACACGTCGTCACTTTCAACATGGATGAGTATGTTGGCCTAGCCGCAGAC

CACCCTGAGTCTTACCGCTCATTCATGTACAACAACTTCTTCAATCATATTGATATCCAA
GAAGAAAATATCAATCTGCTGAACGGTAATACCGATGATCACGAAGCAGAATGCAAGCGC
TACGAAGACAAAATCAAATCTTACGGTAAGATCAACCTGTTCATGGGTGGTGTAGGTAAC
GATGGCCACATCGCCTTTAATGAACCAGCGTCTTCTCTCTCTTCACGCACTCGTATCAAA
ACCCTGACGGAAGATACTCGTATCGCTAACTCACGTTTCTTTGATGGTGATATCAATCAA
GTGCCTAAGTATGCACTGACTATCGGTGTCGGCACTCTGCTGGATGCACAAGAGATCATG
ATTCTAGTCACTGGCCACAACAAGGCACTGGCACTACAAGCAGCGGTTGAAGGGAGTGTA
AACCACCTATGGACTGTCTCTGCACTGCAACTACACCCTAAAGCGGTGATCGTGTGTGAC
GAGCCATCAACCCAAGAGCTGAAAGTGAAAACCGTTAAATATTTCACTGAGTTAGAAGCC
AAAAACATCGTAGGCTTCTAA

MRLIPLKAAAQVGKWAAAHIVKRINEFQPTAERPFVLGLPTGGTPLATYKALIEMHKAGEVSFKHVVT
FNMDEYVGLAADHPESYRSFMYNNFFNHIDIQEENINLLNGNTDDHEAECKRYEDKIKSYGKINLFMG
GVGNDGHIAFNEPASSLSSRTRIKTLTEDTRIANSRFFDGDINQVPKYALTIGVGTLLDAQEIMILVTGHN
KALALQAAVEGSVNHLWTVSALQLHPKAVIVCDEPSTQELKVKTVKYFTELEAKNIVGF

Vibrio cholera: Neu5Ac incorporation pathway

28. ***Vc*-SiaB :** N-acylneuraminate cytidylyltransferase: **hypothetical protein [Vibrio cholerae]**
Accession number: WP_000064388

ATGTCGAATGAATATGTTGCTCTTATAACTGCCAGAGGCGGCTCTAAGGGACTTCTACGT
AAAAACGTACTTCCTTTACACGGAATACCGTTGATTGGATGGACAATAAAGGCTGCACAA
GGTTGTTCTTACATAAGTAAAGTTTTTGTTTCAACAGATGATTATGAAATAGCCAAAATT
AGTGAAGGTTTGGGTGCTCTTGTGATTAATAGGCCTGAGGAGTTGGCTACGGATACAGCG
AGTAGCATAGATGTGATATTACATGCTATTTCTTGGCTCGAACAAAAAGAAGTCCAAAAA
TATGAAGGTATGATTTTACTGCAACCGACCTCTCCACTTCGTACTTCACATCACATTAAA
GAAGCAATTGAGTTGTATGAAAAAACGGCTGCCAAATTTGTAATTAGTGTATTTGAACCA
ACACACACACCAATTAAATCATATTTAGAAAATGACGACGGTACGATATCTGGCTTGTAC
AGTAACGAAGCTCCATATCAAAGGCGGCAAGATCTTCCAAGAGCATATCAGCCAAATGGG
GCAATCTATGCATTTTCTATTGATGAATTCAAATTAAATAACCATTTTCCTCGAAATAAG
GTATTCCCTTATGTGATGTCTGAAGTTGAATCTGCAGATATTGACACGCTAGAAGATCTG
AGAAAAGTAGAAGAACAACTGAAAATTAAGGAAATTAATAAATGA

MSNEYVALITARGGSKGLLRKNVLPLHGIPLIGWTIKAAQGCSYISKVFVSTDDYEIAKISEGLGALVIN
RPEELATDTASSIDVILHAISWLEQKEVQKYEGMILLQPTSPLRTSHHIKEAIELYEKTAAKFVISVFEPT
HTPIKSYLENDDGTISGLYSNEAPYQRRQDLPRAYQPNGAIYAFSIDEFKLNNHFPRNKVFPYVMSEVE
SADIDTLEDLRKVEEQLKIKEINK

Section III

The DNA, protein, and primer sequences of NanK from *Haemophilus influenzae, Pasteurella multocida, Vibrio cholera, and Fusobacterium nucleatum* are listed below:

Mutant sequences of *Pm*-NanK and *Hi*-NanK:

1. *Pm*-NanK (D115L)

Forward primer: 5'-GCG GAA TAT AAA CTC GAA GAT AAA AAT GCG GTA C-3'
Reverse primer: 5'-CCG CAT TTT TAT CTT CGA GTT TAT ATT CCG CAC AC-3'

Recombinant protein sequence:

MRCLALDIGG TKIASAIVTD GKIEQRQQIA TPQADAANAM HDTLANILAL YAGQFDYVAV
ASTGIINHGV LTALNPKNLG GLAEFPLKES IARHTDKPIG LLNDVQAAAC AEYKLEDKNA
VQNFVFITVS TGVGGGIILE RRLLTEPNGV AGHIGHTLAD PNGPVCGCGR VGCVEAVAAG
RAIEAVSSQW NPPCTPKQAF ELFRKNDEKA TALIQRSASA IANLIADLVI GLDVQKVVVG
GSVGLAEGYL PLVKQYLNTM PHFYHCTVEQ ARHGQDAGLL GAAWWVADCL KQGVHLK

2. *Pm*-NanK (D115L, H273F)

Forward primer: 5'-GAA CAA GCA CGT TTC GGA CAA GAT GCT GGG-3'
Reverse primer: 5'-CCC AGC ATC TTG TCC GAA ACG TGC TTG TTC-3'

Recombinant protein sequence:

MRCLALDIGG TKIASAIVTD GKIEQRQQIA TPQADAANAM HDTLANILAL YAGQFDYVAV
ASTGIINHGV LTALNPKNLG GLAEFPLKES IARHTDKPIG LLNDVQAAAC AEYKLEDKNA
VQNFVFITVS TGVGGGIILE RRLLTEPNGV AGHIGHTLAD PNGPVCGCGR VGCVEAVAAG
RAIEAVSSQW NPPCTPKQAF ELFRKNDEKA TALIQRSASA IANLIADLVI GLDVQKVVVG
GSVGLAEGYL PLVKQYLNTM PHFYHCTVEQ ARFGQDAGLL GAAWWVADCL KQGVHLK

3. *Hi*-NanK (T131V)

Forward primer: 5'-TTT ATT ACC GTT TCA GTA GGA GTT GGC GGA GGG-3'
Reverse primer: 5'-CCC TCC GCC AAC TCC TAC TGA AAC GGT AAT AAA-3'

MRCLALDIGG TKIAAAIVKN GEIEQRQQIH TPRENVVEGM HQALGKLLAD YEGQFDYVAV
ASTGIINNGI LSALNPKNLG GLAEFPLKAS IAKHTDKPIG LLNDAQAATY AEYQLQNFEQ
VSNFVFITVS VGVGGGIVLN QILQTGSRGI AGHIGHTLAD PNGAICGCGR RGCVEAIASG
RAIEAVSSQW EDPCDPKEVF ERFRKNDEKA TALVERSAKA IANLIADLVI SLDIQKIAIG
GSVGLAEGYL SLVEKYLQDF PSIYCCEIET AKFGQDAGLI GAAYWVKDVL LDKPEGTIYG

Chapter 3

Structural and functional characterization of SatA binding to sialic acids in ABC transport system from Haemophilus ducreyi

JBC ARTICLE

Molecular characterization of the interaction of sialic acid with the periplasmic binding protein from *Haemophilus ducreyi*

Received for publication, August 2, 2018, and in revised form, October 11, 2018 Published, Papers in Press, October 12, 2018, DOI 10.1074/jbc.RA118.005151

Thanuja Gangi Setty[‡§1], Jonathan C. Mowers[¶], Aaron G. Hobbs[¶], Shubha P. Maiya[‡], Sanaa Syed[‡], Robert S. Munson, Jr.[‖], Michael A. Apicella[**], and Ramaswamy Subramanian[‡¶2]

From the [‡]Institute for Stem Cell Biology and Regenerative Medicine, GKVK Post, Bangalore 560065, India, the Departments of [¶]Biochemistry and [**]Microbiology and Immunology, Carver College of Medicine, University of Iowa, Iowa City, Iowa 52242, the [‖]Center for Microbial Interface Biology, Ohio State University, Columbus, Ohio 43210, and the [§]University of Trans-Disciplinary Health Sciences and Technology (TDU), Bengaluru, Karnataka 560064, India

Edited by Chris Whitfield

3.1 Abstract

The primary role of bacterial periplasmic binding proteins is sequestration of essential metabolites present at a low concentration in the periplasm and making them available for active transporters that transfer these ligands into the bacterial cell. The periplasmic binding proteins (SiaPs) from the tripartite ATP-independent periplasmic (TRAP) transport system that transports mammalian host–derived sialic acids have been well studied from different pathogenic bacteria, including *Haemophilus influenzae, Fusobacterium nucleatum, Pasteurella multocida,* and *Vibrio cholerae*. SiaPs bind the sialic acid N-acetylneuraminic acid (Neu5Ac) with nanomolar affinity by forming electrostatic and hydrogen-bonding interactions. Here, we report the crystal structure of a periplasmic binding protein (SatA) of the ATP-binding cassette (ABC) transport system from the pathogenic bacterium *Haemophilus ducreyi*. The structure of *Hd*-SatA in the native form and sialic acid bound forms (with Neu5Ac and N-glycolylneuraminic acid (Neu5Gc)), determined to 2.2, 1.5, and 2.5 Å resolutions, respectively, revealed a ligand-binding site that is very different from those of the SiaPs of the TRAP transport system. A structural comparison along with thermodynamic studies suggested that similar affinities are achieved in the two classes of proteins through distinct mechanisms, one enthalpically driven and the other entropically driven. In summary, our structural and thermodynamic characterization of *Hd*-SatA reveals that it binds sialic acids with nanomolar affinity and that this binding is an entropically driven process. This information is important for future structure-based drug design against this pathogen and related bacteria.

3.2 Introduction

Protein–protein interactions and protein–ligand interactions play a key role in all living organisms, ranging from prokaryotes to eukaryotes (1, 2). These interactions are highly specific and regulate several downstream signalling pathways and cellular metabolic processes (3, 4). Understanding the molecular details of protein–ligand interactions is a key step in the structure assisted drug discovery process. Electrostatic interactions, hydrogen-bonding interactions, van der Waals interactions, and hydrophobic interactions contribute to the binding affinity of a ligand to a protein. Thermodynamically, both enthalpy (ΔH) and entropy (ΔS) contribute to the binding affinity. The difference in enthalpy and entropy between the bound and unbound forms dictates the spontaneity of the binding reaction. The enthalpy of the system is mainly derived from noncovalent interactions, such as hydrogen bonding, ionic interactions, and van der Waals contacts between polar side chains of the amino acids and ligand. In addition, hydrophobic interactions between ligands and proteins are responsible for the displacement of water, which contributes to either positive or negative net entropy and generates randomness or order within the system. In biological systems, enthalpy–entropy compensation of protein–ligand interactions dictates how spontaneous the reactions will be (5, 6).

Sialic acids are nine-carbon containing sugars, and the carboxylate group at the C1 position imparts the acidic nature to the molecule. In vertebrates and in associated symbiotic and pathogenic bacteria, these sugars are present as the outermost sugars on glycoproteins and glycolipids of the outermost cell surface (7, 8). Based on structural differences, sialic acids exist in ~50 different isoforms, and the two most common sugars include N-acetylneuraminic acid (Neu5Ac) and N-glycolylneuraminic acid (Neu5Gc). Chemically, Neu5Gc differs from Neu5Ac by a single hydroxyl group at the C5 position (9–11). Most pathogenic bacteria that reside in the mucus rich regions lack the biosynthetic pathways to make Neu5Ac, so they scavenge and transport host-derived sialic acids for their long-term survival. Sialic acids are transported across the bacterial periplasmic membrane by tripartite ATP-independent periplasmic (TRAP), ABC, or major facilitator superfamily (MFS) transport systems. After transport of sialic acid into the cytoplasm, it is incorporated as the outermost sugar on the lipooligosaccharide (LOS) by CMP-sialic acid synthetase (SiaB) and sialyltrans- ferases (SiaA, Lic3A, Lic3B, or LsgB). Pathogenic bacteria incorporate sialic acid during lipid glycosylation to evade the host immune system by "molecular mimicry" (12). Previously, it was reported that mutating either SiaB or the sialyltransferase gene Lic3A, which encodes for α.2,3-sialyltransferase, exhibited less resistance to the killing effects of human serum (13, 14). In

addition to lipid glycosylation, part of the internalized sialic acid is often converted into fructose 6-phosphate by a pathway of enzymes that include NanA, NanK, NanE, NagA, and NagB that enables the bacteria to use them as an energy source (15).

Gram-negative bacteria, such as *Haemophilus influenzae, Pasteurella multocida, Fusobacterium nucleatum,* and *Vibrio cholera*, utilize the TRAP transport system for sialic acid transport (16, 17). The sialic acid binding proteins (SiaPs) of the TRAP system have been structurally and functionally characterized (18–20). These proteins contain two domains, and sialic acid binds between these two domains in a conserved binding pocket. In SiaPs, ligand binding generates conformational changes in the hinge region, which cause the movement of two domains to close on the substrate. This mechanism is called either the Venus flytrap mechanism or Pac-Man model (21, 22). Thermodynamic studies have shown that the binding of sialic acids to SiaPs is an enthalpically favored process.

However, *Haemophilus ducreyi*, another Gram-negative bacteria, has been reported to use the ABC transport system for sialic acid transport, utilizing ATP hydrolysis to import sugars across the periplasmic membrane (23, 24). *H. ducreyi* is a pathogenic Gram-negative coccobacillus belonging to the pasteurellaceae family. It causes a sexually transmitted genital disease called chancroid (25). This disease is a major public health concern in developing and low socio-economic level countries due to an increased risk of co-transmission with HIV (26, 27). The pathogenicity of *H. ducreyi* in humans is caused by several virulence factors, including the hemolytic toxin (28, 29) and cytolethal-distending toxin (30–32). In addition, the LOS of *H. ducreyi* acts as a putative virulence factor that aids in bacterial adherence to human skin fibroblasts and mucosal epithelial cells (33, 34). LOS terminates with N-acetyllactosamine and sialyl N-acetyllactosamine, which is heavily sialylated (35).

In *H. ducreyi*, the ABC transport system that transports sialic acid is structurally encoded by the four-gene operon SatABCD. In this operon, SatA encodes a periplasmic binding protein that binds to sialic acid in the periplasm. SatBCD encodes the ABC- type integral membrane transporter that transports the delivered sugar using the energy derived from ATP hydrolysis (36). Similar to the maltose binding protein system, it is expected that SatA bound to Neu5Ac/Neu5Gc anchors onto the transmembrane domain of the ABC transporter. This signals ATP binding, the sialic acid is then released from SatA, and the sugar is transported from the periplasm to the cytoplasm by SatBCD (37–39).

ABC transporters import and efflux a variety of different substrates across the lipid bilayer, including ions, peptides, proteins, and sugars. Although several periplasmic binding proteins of ABC transporters that transport a wide range of other substrates have been described

previously, the structural and molecular mechanism of sialic acid binding to SatA in ABC transporters is not known. The sequence similarity between SiaP from TRAP transporters and SatA from ABC transporters is ~20 –25%; thus, it is important to structurally and functionally characterize SatA from ABC transporters.

Here, we report the structures of the sialic acid binding protein of H. ducreyi, which uses the ABC transporter system (*Hd*-SatA), in unliganded and ligand-bound forms. In addition, we have thermodynamically characterized Hd-SatA binding to its ligands Neu5Ac and Neu5Gc. Site-directed mutagenesis of residues that bind to the sugars followed by measurement of binding affinities shed light on the contribution of these residues to the binding phenomenon.

3.3 Results

3.3.1 *Hd*-SatA unliganded, Neu5Ac and Neu5Gc-bound structures

Hd-SatA was crystallized in three different forms – unliganded, Neu5Ac and Neu5Gc bound. *Hd*-SatA has low sequence identity (20-25%) with other periplasmic sugar binding proteins, and there are no homologous structures available in the protein data bank. We first determined the crystal structure of selenomethionine derivatized *Hd*-SatA bound to Neu5Gc to 2.4 Å resolution by MAD (Multiple-wavelength Anomalous Dispersion) method (Fig. 1). These crystals belong to the space group C121, and the model was refined to an R-factor of 0.17 and free R-factor of 0.25. The crystal structure shows two molecules in the asymmetric unit and the electron density clearly shows the presence of Neu5Gc in the binding pocket. The crystals of the *Hd*-SatA:Neu5Ac complex diffracted to a higher resolution of 1.49 Å, and these crystals belong to the $P2_12_12_1$ space group. This crystal structure was determined by molecular replacement using the *Hd*-SatA–Neu5Gc structure as the search model. It was refined to an R-factor of 0.17 and a free R-factor of 0.197. In the *Hd*-SatA–Neu5Ac structure, the electron density map clearly shows the presence of Neu5Ac in the binding pocket (Figure 3.2 inset). The crystals of the *Hd*-SatA unliganded form diffracted to 2.1 Å, and the structure was determined by molecular replacement using the *Hd*-SatA–Neu5Ac structure as the search model. These crystals belong to the space group P1211 and were refined to an R-factor of 0.20 and a free R-factor of 0.26. The crystal structures of unliganded *Hd*-SatA and *Hd*-SatA–Neu5Ac forms have one molecule in the asymmetric unit.

The details of the crystallography data and refinement statistics are presented in Table 1. The coordinates and structure factors for *Hd*-SatA unliganded, *Hd*-SatA–Neu5Ac and *Hd*-SatA–Neu5Gc have been deposited in the Protein Data Bank (PDB) with accession numbers 5ZA4, 5Z99, and 5YYB, respectively.

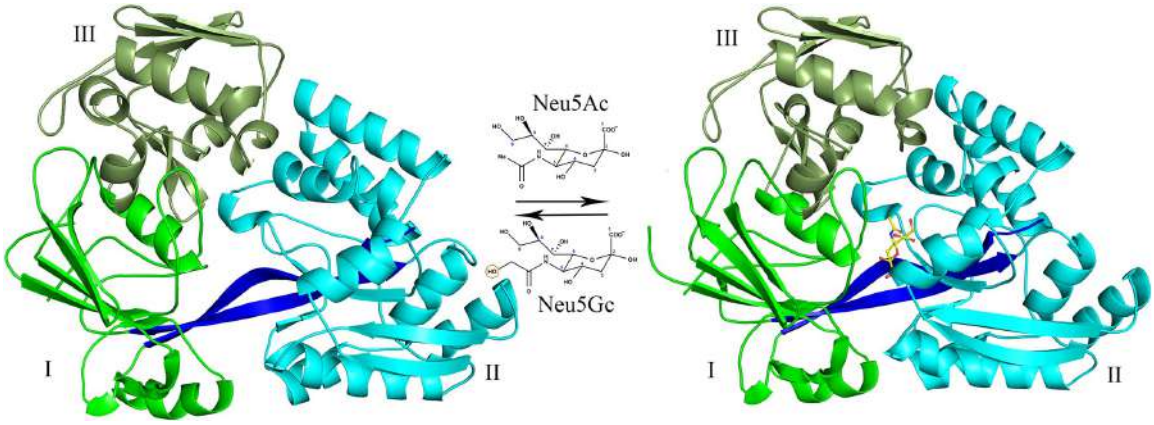


Figure. 3. 1. Structure of *Hd*-SatA in complex with Neu5Ac. A cartoon representation shows the crystal structures of *Hd*-SatA in unliganded form (left) and ligand-bound conformations (right). Neu5Ac is represented by a ball-and-stick model. Structural representations of Neu5Ac and Neu5Gc are also shown. The domains are marked as I, II, and III. The program PyMOL (Schrödinger) (54) was used to create the figure.

	Unliganded	With Neu5Ac	With Neu5Gc*
Wavelength	1.03	0.9793	0.9546
Resolution Range(A0)	29.43-2.19 (2.27-2.19)	19.52 - 1.49 (1.52 - 1.49)	28.99 - 2.48 (2.59 - 2.48)
Space group	P 1 21 1	P 2_1 2_1 2_1	C 1 2 1
Unit cell			
a(Å)	43.9	73.4	134.4
b(Å)	79.3	89.6	70.1
c(Å)	68.4	90.8	113.1
α, β, γ (degrees)	90, 91.2, 90	90, 90, 90	90, 114.2, 90
Total reflections	90,314 (7117)	604,746 (23207)	124,762 (11816)
Unique reflections	24,030 (1961)	93,977 (4237)	33,076 (3289)
Multiplicity	3.8 (3.6)	6.4 (5.5)	3.8 (3.6)
Completeness(%)	99.5 (94.9)	96.8 (88.6)	97.2 (85.6)
Mean I/Sigma(I)	7.00 (3.00)	19 (4.9)	7.2 (2.3)
Wilson B-factor	31.7	11.2	36.6
R-merge	0.178 (1.175)	0.058 (0.538)	0.116 (0.549)

R-meas	0.209 (1.382)	0.063 (0.591)	0.135(0.641)
R-pim	0.108 (0.72)	0.024 (0.237)	0.069 (0.329)
CC1/2	0.985 (0.37)	0.999 (0.807)	0.993 (0.502)
Reflections used in refinement	24,004 (2281)	93,931 (9021)	33,071 (2840)
Reflections used for R-free	1191 (102)	4864 (459)	1647 (118)
Rwork	0.20 (0.26)	0.17 (0.26)	0.17 (0.26)
Rfree	0.26 (0.29)	0.19 (0.28)	0.25 (0.38)
Number of non-hydrogen atoms	3974	4555	7869
macromolecules	3847	3972	7667
ligands	-	21	44
Solvent	127	562	202
Protein residues	488	485	964
RMS(bonds)	0.008	0.005	0.01
RMS(angles)	0.96	0.8	1.02
Ramachandran favoured(%)	94.86	96.27	93.96
Ramachandran allowed (%)	4.94	3.73	5.73
Ramachandran outliers(%)	0	0	0
Average B-factor	28.91	18.23	40.86
Macromoleucules	28.83	16.75	40.29
Ligands	-	9.97	32.53
Solvent	27.26	29.01	38.44
PDB Code	5ZA4	5Z99	5YYB

Statistics for the highest-resolution shell are shown in parentheses.

*This data set is also the the data set from the peak wavelength of Se-Met protein used to determine the structure.

Table 3.1 Crystallography data and refinement statistics

The overall structure contains three α/β domains, highlighted in Figure 1. The structure contains an N-terminal domain that is larger than the C-terminal domain due to the presence of an extended region in the N-terminus and can be divided into two domains, I and III. The N-terminal domain contains amino acids from the N-terminus and C-terminus. Domain I contains residues 1-38, 160-239 and 461-487 (from the C-terminus), which form 6 β strands that are sandwiched between 3 α helices. Domain III is composed of residues 39-159, which form 6 β strands and 2 α helices. Domain II contains the C-terminus residues 244-456, which form 6 β strands enclosed by 7 α helices. The N and C terminal domains are connected by two β strands, which are composed of residues 240-243 and 457-460, and these two strands may serve as hinge regions. The crystal structure clearly shows that the ligand is buried deep inside the binding pocket formed between domains I and II, and there are no apparent interactions of the ligand with domain III. Upon ligand binding, both domains I and II come together to trap the ligand, akin to the Venus fly trap mechanism, which is similar to the pattern that is observed in other sugar binding proteins.

The Dali server search results showed that many structures are structurally similar to *Hd*:SatA but they have less than 25% sequence identity. Some of the structures with the highest matching Z scores are dipeptide binding protein DppA from *Escherichia coli* (with Z score of 40.9) (PDB ID: 1DPP) (40), periplasmic oligopeptide binding proteins AppA from *Bacillus subtilis* (PDB ID: 1XOC) (41), antibiotic agrocinopine A transport protein in *Agrobacterium tumefaciens* (PDB ID: 4ZE9) (42)and periplasmic lipoprotein HbpA in *Haemophilus parasuis* (PDB ID: 3M8U) (43).

3.3.2 The sialic acid binding pocket of *Hd*-SatA

The binding of Neu5Ac in the binding pocket of *Hd*-SatA is depicted in Figure 3.2. There are several hydrogen-bonding interactions observed between different amino acids in the binding pocket and the Neu5Ac molecule. In SiaPs of TRAP transporters, the major interactions are salt bridges between an Arg and the carboxylate group (C1 position) of sialic acid (18, 19). In *Hd*-SatA, there are no amino acids in the binding pocket that make salt bridges with the ligand. There is only one Arg in the binding pocket, at position 408, and it forms a hydrogen bond with the hydroxyl group at the C9 position. His^{333} and Ser^{362} are two polar residues closest to the carboxylate group at C1, and they form hydrogen bonds with the carboxylate group. Ser^{18} and Ser^{362} form a hydrogen bond with the hydroxyl group at C8, and Ser^{27} forms a hydrogen bond with hydroxyl groups at positions C8 and C9. Pro^{16} has a water-

mediated hydrogen- bonding interaction with O2. Tyr^{15}, Thr^{477}, and Asp^{383} also display water-mediated hydrogen-bonding interactions with the hydroxyl group at position C4. The binding pocket of *Hd*-SatA–Neu5Gc is similar to *Hd*-SatA–Neu5Ac, and there are no structural changes between these two sugar-bound forms.

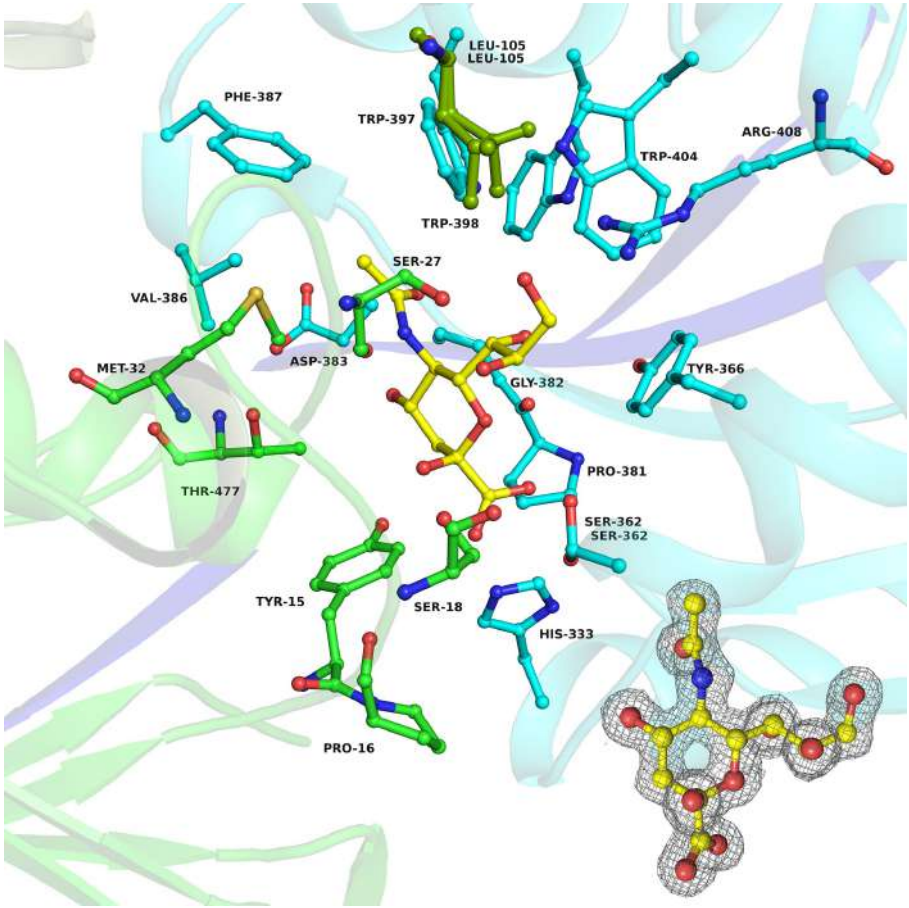


Figure. 3.2. Binding pocket of *Hd*-SatA with Neu5Ac. A ball-and-stick diagram shows Neu5Ac and its interacting residues in Hd-SatA. Inset, 2Fo - Fc map at 1 **σ** level showing the electron density map of Neu5Ac in the binding pocket of *Hd*-SatA. The program PyMOL (Schrödinger) (54) was used to create the figure.

In summary, SiaPs from TRAP transporters have more charged residues in their binding pocket, and their interactions with sialic acids are mostly mediated by hydrogen bonds, ionic interactions, and salt bridges. However, *Hd*-SatA has few charged residues and more polar residues in the binding pocket, and its interaction with Neu5Ac is mainly mediated by hydrogen bonds and hydrophobic interactions.

3.3.3 Hinge region analysis

The superposition of open and closed conformations of *Hd*-SatA resulted in a root mean square deviation of 3.496 for 3641 Cα atoms. The relative movement of domains in the open and closed conformations was analyzed using DYNDOM. DYNDOM classified the structure as a two-domain protein, where domain I is considered a fixed domain and is composed of residues 1–240 and 460 – 487. Domain II is considered a mobile domain and is composed of residues 244 – 456. Domain II moves at a rotation of 31° relative to domain 1 and contributes a small component of translation by 0.5 Å. The two hinge regions, which are responsible for the domain movement, are formed by residues 241–243 and 457– 459 (Figure 3.3). Comparative analyses of phi and psi, angles of unliganded *Hd*-SatA and *Hd*-SatA–Neu5Ac bound structures show that there are large changes in the amino acids in the two hinge regions. Main chain torsion angles between Phe^{241} and Tyr^{242}, between His^{456} and Arg^{457}, and between Arg^{457} and Lys^{458} are primarily responsible for the hinge movement. In both cases, the largest differences are in the phi angle. Further structural analysis reveals that there are significant conformational changes of amino acids in the binding pocket upon ligand binding. In the open conformation, Arg^{457} from hinge two is oriented toward the binding pocket, and His^{333} and Tyr^{15} from domains I and II are far away from the binding pocket. Upon ligand binding, the Arg^{457} side chain moves away from the binding pocket to form hydrogen bonds with Ser^{222} and Tyr^{225}, and amino acids His^{333} and Tyr^{15} move closer to the ligand to form hydrogen bonds with the Neu5Ac. Thus, there are a series of conformational changes upon ligand binding that lead to the opening and closing of the two domains of the protein (Figure. 3.3).

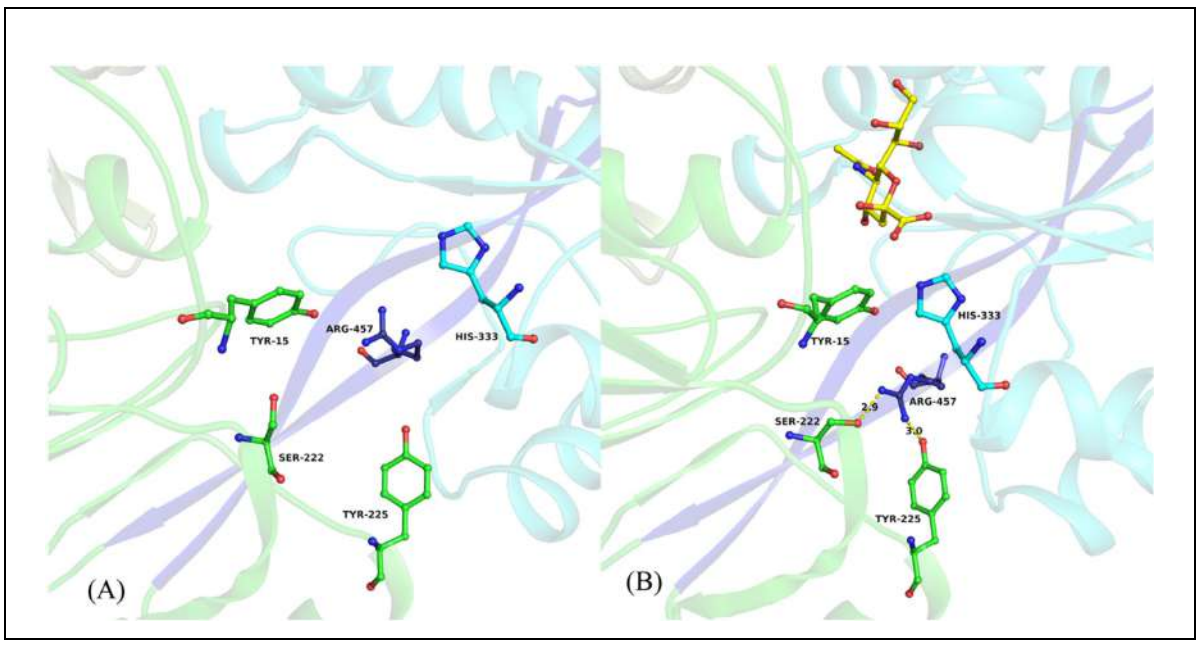


Figure. 3.3. Close up view of hinge region and amino acids from the binding pocket of *Hd*-SatA. Neu5Ac and its interacting residues in the binding pocket of *Hd*-SatA are shown in ball and stick model. Close up view clearly shows the conformational changes in the amino acids (A) before and (B) after Neu5Ac binding. Neu5Ac is shown in Figure B. The two β strands that form the hinge are shown in dark blue colour. Figures were made using PyMOL (Schrödinger) (54).

3.3.4 Thermodynamics of Neu5Ac and Neu5Gc binding to *Hd*-SatA and its mutants

Isothermal titration calorimetry (ITC) was used to measure the thermodynamics of the interactions between *Hd*-SatA and its ligands Neu5Ac and Neu5Gc. Fitting the data using single site binding mode analysis indicates that *Hd*-SatA binds to Neu5Ac and Neu5Gc in a 1:1 stoichiometric ratio with nano molar binding affinity. Further binding studies showed that the entropy contribution has a larger effect on the overall free energy of the system compared with enthalpy (K_d =133 +/- 22 nM, ΔH= -2.21 +/- 0.023 kcal/mol and -TΔS= -7.172 kcal/mol). These results are similar to *Hd*-SatA binding to Neu5Gc (K_d =277 +/- 58 nM, ΔH= -2.986 +/- 0.056 kcal/mol and -TΔS= -5.96 kcal/mol) (Figure. 3.4). Therefore, the interaction between *Hd*-SatA and sialic acid is an entropically driven process.

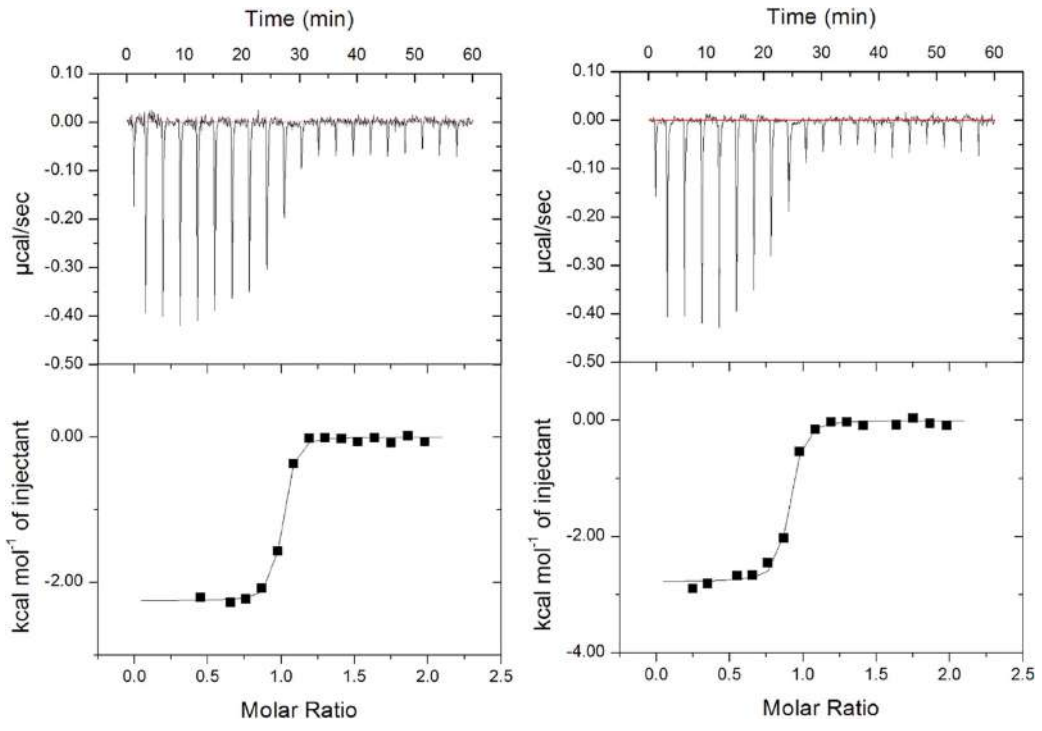


Figure. 3.4. Thermodynamic studies of *Hd*-SatA. Isothermal calorimetric titration shows the binding of *Hd*-SatA to Neu5Ac (A) and Neu5Gc (B)

Based on the determined structures and thermodynamic properties of SiaPs from TRAP transporters, we anticipated that the enthalpic contribution to the free energy of binding in *Hd*-SatA can be increased by substituting a few polar residues to charged residues in the binding pocket. To test this, we made total of 10 substitutions in the binding pocket of *Hd*-SatA- S27N, S27K, S362K, S362R, S362A, H333K, H333R, H333A, R408A and R408K and their thermodynamic properties were probed using ITC. Ser27, which forms hydrogen bonds with the hydroxyl groups at positions C8 and C9, was modified to Lys and Asn. ITC binding studies showed that S27K had no effect on the interaction, but S27N bound endothermically to Neu5Ac (Figure 3.5) and exothermically to Neu5Gc.

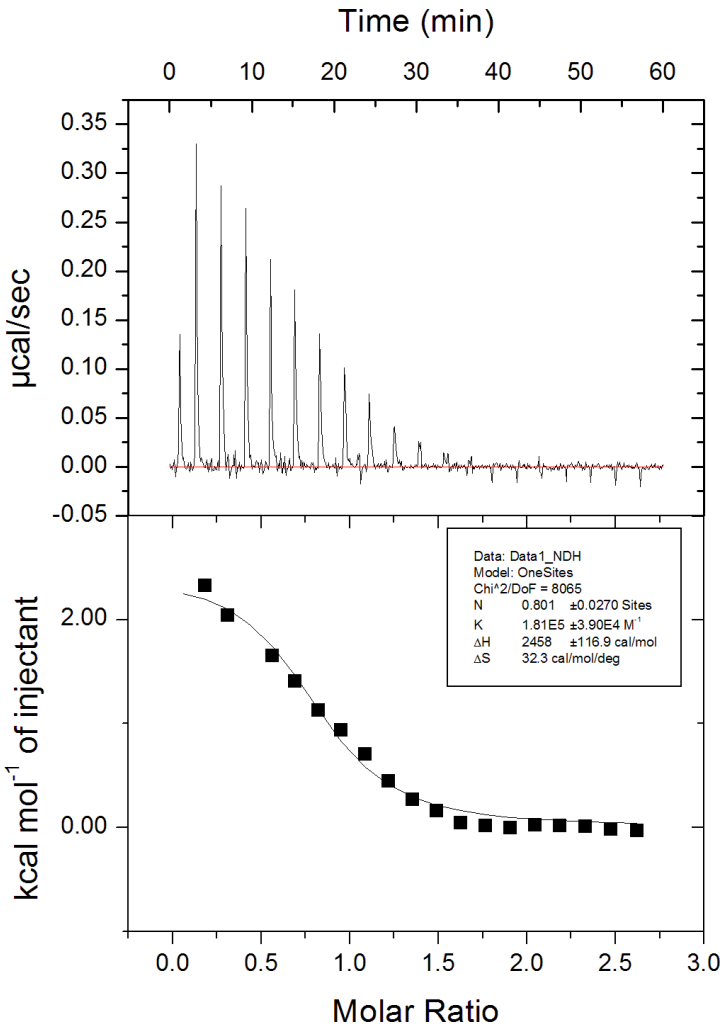


Figure 3.5. Isothermal calorimetric titration showing the binding of *Hd*-SatA S27N to Neu5Ac.

Ser362 and His333, which form hydrogen bonds with the C1 carboxylate of Neu5Ac, were mutated to S362K, S362R, S362A, H333K, H333R, and H333A, respectively. Ser362 and His333 were individually mutated to Arg and Lys to investigate whether an ATP transporter can be made to have similar binding properties as periplasmic binding proteins in TRAP transporters via salt bridge formation (18). However, these substitutions had binding properties similar to the WT protein, although there was no measurable binding affinity with the H333K substitution. The R408A and R408K substitutions did not show any measurable binding to Neu5Ac. The binding affinities, including the enthalpic and entropic contributions of sialic acids binding to *Hd*-SatA WT and its mutants, are presented in Table. 3.2. The thermodynamic studies show that there were few minor changes in the thermodynamic properties of the protein due to site-directed mutagenesis.

S.No.	***Hd-SatA Protein***	***Ligand***	***K_d***	***ΔH (kcal/mol)***	***-TΔS (kcal/mol)***
1.	Wt	Neu5Ac	133 +/- 22 nM	-2.21 +/- 0.023	-7.172
		Neu5Gc	277 +/- 58 nM	-2.986 +/- 0.056	-5.96
2.	S27N	Neu5Ac	5.52 +/- 1.18μM	2.458+/- 0.116	-9.625
		Neu5Gc	57.8 +/- 19.3 μM	-0.635 +/- 0.11	-5.155
3.	S27K	Neu5Ac	24 +/- 3.04 μM	-2.725 +/- 0.126	-3.576
4.	S362R	Neu5Ac	4.83 +/- 0.44 μM	-1.107 +/- 0.017	-6.138
5.	S362A	Neu5Ac	418 +/- 179 nM	-0.47 +/- 0.0133	-8.224
6.	S362K	Neu5Ac	5.37 +/- 0.22 μM	-1.63 +/- 0.03	-5.542
7.	H333R	Neu5Ac	115 +/- 52 nM	-2.850 +/- 0.055	-6.615
8.	H333A	Neu5Ac	31 +/- 10 nM	-5.724 +/- 0.052	-4.529
9.	H333K	Neu5Ac	No binding		
10.	R408A	Neu5Ac	No binding		
11.	R408K	Neu5Ac	No binding		

Table. 3.2. Binding affinities and thermodynamic parameters of native protein and proteins with site specific amino acid substitutions.

3.3.5 Database search to identify protein binding to the same ligand with different binding modes

In a PDB database search, we found two proteins, a maltose transporter (PDB code 3PUY) and a bacterial maltosyltrans- ferase (PDB code 4U33), that bound to the same ligand

(maltose) with different modes of binding. Previously, it was reported that maltose binding to an MBP-maltose transporter is an entropically driven process (44). The MBP-maltose transporter has a large hydrophobic binding pocket; hence, the displacement of water molecules might be the reason for the entropically favored binding process. However, unlike the data presented here, no experimental evidence is present to establish whether the thermodynamics of binding is different in the maltose-binding protein of bacterial maltosyltransferase.

3.4 Discussion

Pathogenic bacteria have evolved different mechanisms to evade the host immune system, including coating their outer membranes with host nine carbon sugars, known as sialic acids, that are scavenged from the hosts using TRAP/ABC/MFS transporter systems. Pathogenic bacteria display these sugars as the outermost moiety to mimic eukaryotic outer membrane composition, which aids in eluding the host immune response. Although several studies have been reported in the last few decades related to sialic acid scavenging proteins and its incorporation, structural information at the atomic level on sialic acid binding modes as well as the amino acid environment of the different transport systems are just beginning to be revealed. Here, we have structurally and functionally characterized the sialic acid binding protein from H. ducreyi that belongs to the ABC transport system. The results from structural and thermodynamic studies support the conclusion that *Hd*-SatA binds to sialic acids with nanomolar affinity, and this process is entropically driven.

Hd-SatA is a three-domain protein, where sialic acid binds between domains I and II. Domain III is a small part of the N-terminal domain, which does not interact with the ligand. Initially, Fukami-Kobayashi et.al. classified substrate binding proteins based on their ß sheet topology as Class I and Class II proteins (45). After this initial classification, many more proteins were structurally and functionally characterized with a wide range of ligands. Nearly a decade later, Ronnie et al. further classified these substrate binding proteins into a total of six groups (Cluster A – Cluster F) based on their structural similarity (46). In this classification, SiaPs from TRAP transporters were categorized in Cluster E. All other clusters have proteins that belong to ABC transport systems. SatA from *H. ducreyi* can be categorized into Cluster C because it has an extra N-terminal domain. Cluster C has other proteins that belong to Class II ABC transporters, where they bind wide to range of ligands, including di and oligopeptides, nickel and arginine. Based on previous structural reports, it was hypothesized that proteins like AppA from *Bacillus subtilis* and OppA from *Lactococcus lactis* have an extra domain to

accommodate large ligands, such as oligopeptides (41, 47); however, in the case of *Hd*-SatA, the reason and necessity for the extra domain is still unknown. SiaPs from *Haemophilus influenzae, Pasteurella multocida, Vibrio cholera* and *Fusobacterium nucleatum*, which utilize the TRAP transport system, have been structurally and functionally characterized (18–20). The common features of sialic acid binding proteins from ABC and TRAP transport systems are: i) The ligand binds between two globular domains and ii) the amino acids in the hinge regions are responsible for opening and closing the two domains upon ligand binding. But the salient features of SiaPs from TRAP transporters are i) two arginine residues (Arg^{127} and Arg^{147}: numbers correspond to *H. influenzae* SiaP) in the binding pocketform salt bridges with the C1 carboxylate group of the sialic acid, and ii) the amino acids in the hinge regions are responsible for opening and closing the two domains upon ligand binding. But the salient features of SiaPs from TRAP transporters are i) two arginine residues (Arg^{127} and Arg^{147}: numbers correspond to H. influenzae SiaP) in the binding pocket form salt bridges with the C1 carboxylate group of the sialic acid; and ii) Glu and His from the hinge region form and break a series of hydrogen bonds upon ligand binding (18). These features show that the binding pocket has charged residues and can form salt bridges and hydrogen bonds with the negatively charged ligand. In case of Hd-SatA, there are only two charged residues in the binding pocket, Arg408 and Asp383. The other residues in the binding pocket are either polar or hydrophobic (Tyr^{15}, Ser^{27}, Gly^{28}, Ala^{29}, Leu^{105}, Pro^{381},His^{333}, Asp^{383} and Trp^{397}). This is an important difference between SiaP and SatA, where one ligand (sialic acid) is binding to two different proteins (SiaP and SatA) in two different amino acid environments. Results from thermodynamic studies confirmed that the binding of sialic acid in *Hd*-SatA is a more entropically favored process, whereas binding is a more enthalpically favored process in the SiaP of TRAP transport system (Table 3.3). The structures provide a molecular explanation for the observed thermodynamic properties.

S.No.	**Sialic acid binding protein**	*Ligand*	K_d	*ΔH (kcal/mol)*	*-TΔS (kcal/mol)*
ABC Transporter					
1.	***Haemophilus ducreyi***	Neu5Ac	133 +/- 22 nM	-2.21 +/- 0.023	-7.172
		Neu5Gc	277 +/- 58 nM	-2.986 +/- 0.056	-5.96
TRAP Transporters					
*2.	***Haemophilus influenzae***	Neu5Ac	28 nM	-15.300	5
		Neu5Gc	83 nM	-12.970	3.308
*3.	***Fusobacterium nucleatum***	Neu5Ac	45.5 nM	-9.165	-0.849
		Neu5Gc	45.7 nM	-12.074	2.059
*4.	***Pasturella multocida***	Neu5Ac	19.7 nM	-6.56	-3.948
		Neu5Gc	30.7 nM	-7.469	-2.777
*5.	***Vibrio cholrae***	Neu5Ac	306 nM	-7.715	-0.508
		Neu5Gc	1090 nM	-5.2	-1.127

*Thermodaynamic data for TRAP transporters is taken from Gangi Setty et al. (2014) to compare with ABC transporters sialic acid binding proteins.

Table 3. 3: Comparison of thermodynamic characters of ABC and TRAP transporters. Table showing binding affinities of ABC and TRAP transporter sialic acid binding proteins to Neu5Ac and Neu5Gc, and their enthalpic and entropic contributions.

We calculated the binding pocket volumes in the sialic acid binding proteins of the TRAP and ABC transport systems using the CASTp (Computed Atlas of Surface Topography of proteins) server (48). The binding pocket volumes of *Hi*-SiaP in the open and closed conformations are 834.5 Å3and 191.9 Å3. This shows that upon ligand binding, there is a modest change in the volume of 643 Å3. In *Hd*-SatA, the binding pocket volumes of the open and closed conformations are 1490.4 Å3 and 242 Å3. Thus, upon ligand binding, there is a large volume change of 1200Å3 (Figure. 3.6).

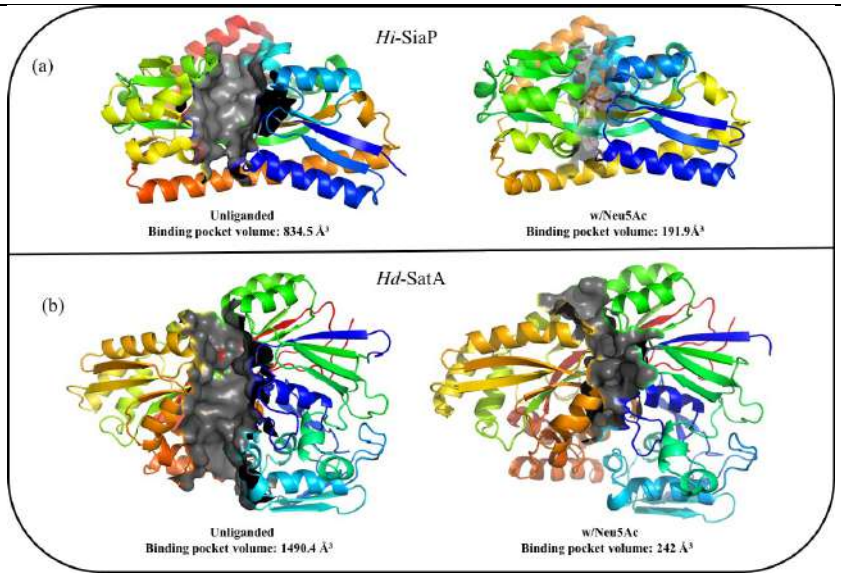


Figure 3.6. Binding pocket cavity analysis of Sialic Acid Binding proteins from TRAP and ABC transporters. Cartoon representations showing the binding pockets of open and closed conformations of (a) *Hi*-SiaP (PDB- ID's: Closed form – 3B50, Open form – 2CEY) (b) *Hd*-SatA (PDB- ID's: Open Form – 5ZA4, Closed form – 5Z99) in presence and absence of its ligand Neu5Ac. Cavity analysis is done using CastP server.

This large volume change upon ligand binding may displace a lot of water molecules, which may account for the increased entropic contribution in ABC transporters. This analysis shows that the binding mechanism of a sialic acid binding proteins is determined by the amino acid composition in the binding pocket and the cavity size in the protein.

These, preliminary structural and protein homology studies suggest the binding of the sialic acid by heterologous peptides is the result of convergent evolution; the sialic acid binding proteins of the ABC transport system and TRAP transport system evolved independently to accomplish similar sialic acid binding functions. However, this hypothesis requires further testing.

This interesting result persuaded us to search the PDB database to explore if there are other proteins in the protein database that bind to the same ligand with a different motif of the binding pocket. Our Database search shows that the maltose transporter from *E. coli* and bacterial maltosyl transferase from *Mycobacterium tuberculosis*, both bind maltose but utilize two distinct binding pockets. The binding pocket of the Maltose transporter (PDB ID: 3PUY) is mainly composed of polar and hydrophobic residues, such as Tyr, Phe, Gly, Leu, Asn, Phe and Ser. This binding pocket amino acid composition is similar to the binding pocket of *Hd*-

SatA. It was previously reported that binding of maltose to the MBP-maltose transporter is an entropically driven process. This supports the concept that, in both the MBP-maltose transporter and *Hd*-SatA the hydrophobic binding pocket is responsible for entropic binding to their respective ligands. In the bacterial maltosyl transferase, the binding pocket has many charged residues, including Arg, Lys, Asp and Glu, which is similar to TRAP transporter SiaPs. Based on the previous structural and thermodynamic studies of SiaPs from the TRAP system, we hypothesize that binding of Maltose to bacterial Maltosyl transferase might be an enthalpically driven process (Figure. 3.7).

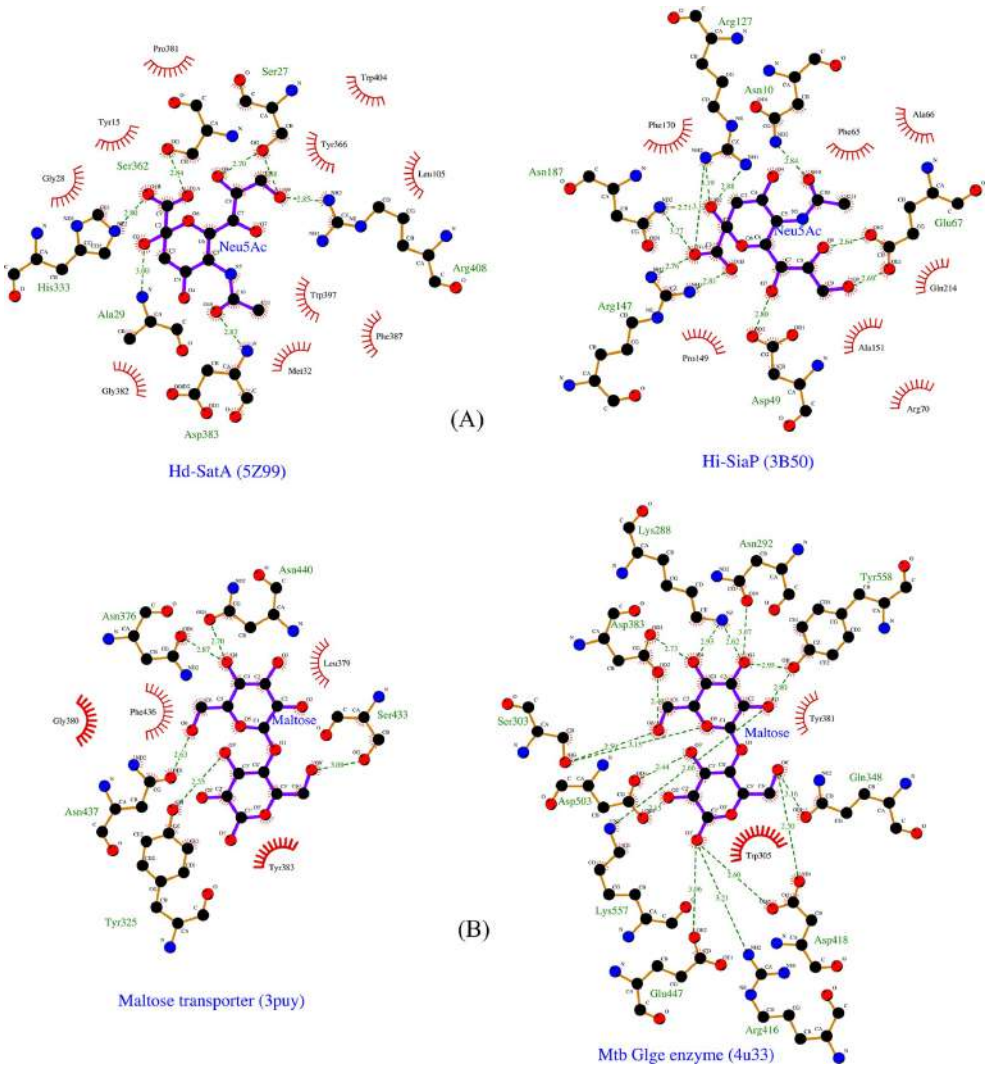


Figure. 3.7 Ligplot diagrams showing the ligands and its interacting residues different proteins. A. Ligplot representation of the binding pocket of Hd-SatA and H. influenzae SiaP in the presence of their ligand Neu5Ac. B. Ligplot representation of maltose tranporter and bacterial maltosyltransferase in the presence of their ligand maltose. Ligands are

represented in purple color. Carbon, nitrogen, and oxygen are represented in black, blue, and red, repectively. The residues with red spikes that are oriented toward the ligand represent hydrophobic interactions.

In summary, pathogenic bacteria have evolved to bind sialic acids with similar binding affinities, using different binding topologies. The topological and chemical properties of the binding pocket leads to different energetic process that govern binding. This study of SatA of *H. ducreyi* from the ABC transport system and analysis of previously reported SiaPs from the TRAP transport system together provides important and divergent structural details about sialic acid binding proteins in two different transport systems. This information is important for future structure-based drug design against these pathogenic bacteria.

3.5 Experimental procedures

3.5.1 Cloning and mutagenesis of *Hd*-SatA WT

The DNA encoding Neu5Ac-binding protein (SatA) from *H. ducreyi* in PET101D-TOPO was amplified by PCR and sub- cloned into the BamHI and XhoI sites of a pET21a (Novagen) vector as a fusion with a C-terminal His tag (NCBI accession number WP_010945473). The gene sequence was verified by DNA sequencing and designated Hd1669 (FSS-End) pET21a. Site specific substitutions S362R, S362A, S362K, S27K, S27N, H333R, H333K, H333A, R408A, and R408K were generated in *Hd*-SatA using site-directed mutagenesis with corresponding mutagenic primers. These substitutions were confirmed by sequencing.

3.5.2 Protein expression and purification

The protein expression of the recombinant plasmids was carried out in *E. coli* BL21(DE3) star cells. The cells were transformed with the plasmid and grown overnight at 37 °C in lysogeny broth medium containing 100 μg/ml ampicillin. The cells were reinoculated with 1% overnight culture in fresh lysogeny broth medium containing ampicillin. Then these cells were cultured at 37 °C until the absorbance reached 0.6, and the cultures were induced with 100 μM isopropyl β-D-thiogalactopyranoside. The cells were incubated at 25 °C for 4 h. Later, the cells were harvested at 13,000 rpm for 30 min. For protein purification, each 1-liter pellet was resuspended in 25 ml of resuspension buffer containing 20 mM HEPES, 150 mM NaCl, 5 mM imidazole (pH 8.0), which was further supplemented with a protease inhibitor mixture tablet lacking EDTA (Roche Applied Science). The resuspended pellet was treated

with lysozyme and DNase and incubated on ice for 30 min following lysis using an Emulsiflex C3 from Avestin at 15,000 psi. The lysate was centrifuged at 13,000 rpm for 30 min and further purified on a Talon affinity column (Bio-Rad) using Profinia (Bio-Rad). The column was first equilibrated with resuspension buffer, and the lysate was loaded onto the column. Following a wash with 10 column volumes of wash buffer (20 mM HEPES, 500 mM NaCl, 5 mM imidazole, pH 8.0), the protein was eluted with four column volumes of elution buffer containing 20 mM HEPES, 150 mM NaCl, and 500 mM imidazole, pH 8.0. The eluted protein was dialyzed overnight in 20 mM HEPES and 10 mM NaCl (pH 8.0). The protein was further purified by anion exchange chromatography (column Hi Trap Q FF, GE Healthcare). The column was first equilibrated with Buffer A (20 mM HEPES, 10 mM NaCl, pH 8.0) and loaded with the dialyzed protein. The protein was eluted with a gradient of Buffer A to the same buffer containing 1M NaCl. The protein eluted at a salt concentration of 150 mM. The protein fractions were pooled and purified using size-exclusion chromatography (Sepharose 200 column, GE Healthcare) using 20 mM Tris-Cl, 100 mM NaCl (pH 8.0). The samples corresponding to the protein peak were pooled, and concentration was measured using the Bradford assay (Bio- Rad). To express selenomethionine-derivatized protein of *Hd*- SatA, a protocol similar to that of Doublie´ et al. (49) was used. Further, the protein was purified using the same method as the native protein.

3.5.3 Crystallization, structure determination, and refinement

Hd-SatA bound with Neu5Gc—To obtain crystals of the protein ligand complex, selenomethionine-derivatized *Hd*-SatA protein at a concentration of 50 mg/ml was mixed with Neu5Gc at a molar ratio of 1:10. The hanging-drop vapor diffusion method was used for crystallization trials. Crystals were obtained in a buffer containing 20 mM MES (pH 6.2), 60 mM NaCl with 30% PEG 1500 as a precipitant. The *Hd*-SatA– Neu5Gc crystals were mounted in loops, and X-ray data were collected at three different wavelengths (Molecular Biology Consortium, Advanced Light Source, using a NOIR1 detector). For phasing, data were collected from selenomethionine incorporated crystals of Neu5Gc bound *Hd*-SatA at selenium peak wavelength (0.9778 Å). Further selenomethionine substituted protein structure refinement was carried out using the automated SOLVE-RESOLVE pipeline (50). Further refinements were done using Phenix, and model building was carried out using COOT (51). The coordinates and structure factors for the *Hd*-SatA–Neu5Gc complex have been deposited in the PDB with accession number 5YYB.

Hd-SatA bound with Neu5Ac—*Hd*-SatA protein at a concentration of 50 mg/ml was mixed with Neu5Ac at a molar ratio of 1:10. The crystals were obtained in buffer containing 20 mM MES, pH 6.2, 60 mM NaCl with 30% PEG 1500 as a precipitant by the hanging-drop diffusion

method. The *Hd*-SatA–Neu5Ac complex crystals were mounted in loops, and X-ray data were collected at the IMCA-CAT beamline using the MAR-CCD detector at the Advanced Photon Source (Argonne, IL). The data were indexed, integrated, and scaled using HKL2000, and further molecular replacement was carried out using the *Hd*-SatA–Neu5Gc structure as a search model. Manual model building was performed using COOT, and further refinements were performed using Phenix. The coordinates and structure factors for the *Hd*-SatA–Neu5Ac complex have been deposited in the PDB with accession number 5Z99.

Hd-SatA without ligand—The crystals of *Hd*-SatA were grown by the hanging-drop vapor diffusion method by mixing 1 µl of protein at 40 mg/ml concentration with the 1 µl of the reservoir. Crystals were obtained from the reservoir solution containing 0.2M sodium iodide and 20% (w/v) PEG 3350. These crystals were mounted in loops, and X-ray data were collected from the IMCA-CAT beamline using the MAR-CCD detector at the Advanced Photon Source (Argonne, IL). Molecular replacement was performed using Neu5Ac-bound *Hd*-SatA as a search model. Manual model building was performed using COOT, and further refinements were performed using Phenix. The coordinates and structure factors for *Hd*-SatA have been deposited in the PDB with accession number 5ZA4.

3.5.4 Thermodynamic studies of *Hd*-SatA

Thermodynamic studies were performed using isothermal calorimetry (MicroCal ITC microcalorimeter, Malvern, GE Healthcare). All titrations were carried out at 25 °C, and the protein samples were in a buffer containing 20 mM HEPES and 10 mM NaCl (pH 8.0). Protein concentrations were measured using the Bradford assay (Bio-Rad). The concentrations of *Hd*-SatA WT and proteins with site specific amino acid substitutions were varied in different titrations. *Hd*-SatA WT ITC titrations with Neu5Ac and Neu5Gc was carried out from three independent preparations and technical triplicates for each preparation. All ITC data from proteins with mutations in *Hd*-SatA were technical replicates. Each experiment consisted of 20 injections, with the first injection 1 µl and the rest of the injections 2 µl with 180 s between the titrations. The nonspecific heats of dilution liberated during the titration of the sugars and protein were calculated by averaging the heat liberated during the last 3–5 injections after saturation. The raw heat released during each injection was calculated by deleting these heats of dilution.

Global fit analysis from technical replicates was done for *Hd*-SatA WT using SEDPHAT (52). The analysis was repeated with three different protein preparations, and all global fit analysis led to similar values. All of the mutant protein data were analyzed using the ORIGIN ITC software package (Microcal) using a single-site binding model, and errors are fit

to curves. In the ORIGIN software, nonlinear least squares analysis was used to calculate the values of stoichiometry, affinity, change in enthalpy (ΔH), and change in entropy (ΔS).

3.5.5 Database search to identify protein that bind to the same ligand with different binding modes

To identify proteins where the same ligand bound to completely different amino acids, the complete database of 121,700 PDB structures was downloaded. Of these structures, 98,011 had no ligand bound to them, so these structures were eliminated. From the remaining structures, ligands were extracted, and a directory was created for each ligand. Approximately 23,700 structures were sorted into groups based on their ligand specificity. As we were interested in finding two proteins with dissimilar binding sites, only the ligands to which more than one protein bound were retained. Finally, this whole process resulted in the sorting of 6015 ligands. The majority of current methods use binding pocket geometry rather than the physicochemical properties or the microenvironment of the binding site; therefore, they may report the similar binding sites as dissimilar ligand-binding sites. For example, a replacement from Arg to Lys may be considered as a dissimilarity. Therefore, we chose the pocket feature algorithm, which considers the similarity of the binding site as a function of shared microenvironments rather than the geometric orientations of the amino acids in the binding site. "Microenvironment" refers to residues around a given center within a radius value of 7.5 Å. This feature calculates the microenvironment in protein A and B; further, the Tanimoto coefficient was calculated based on similar properties (53). A score of 7 or greater signifies similar binding pockets, whereas scores between 0 and 1 signify dissimilar binding sites. We selected a score of 0 to identify completely dissimilar sites. All ligands with a molecular mass less than 110 Da were deleted, including the proteins that bound to these molecules. Among these, ligands bound to sites other than the active site were also ignored. The resulting structures were then visually examined.

3.6 Accession numbers

Structure factors and coordinates of *Hd*-SatA unliganded, *Hd*-SatA–Neu5Ac, and *Hd*-SatA–Neu5Gc structures have been deposited in Protein Data Bank with accession codes 5ZA4, 5Z99, and 5YYB, respectively.

3.7 Author contributions

T. G. S., J. C. M., G. A. H., S. P. M., and R. S. designed the research and performed all experiments. M. A. A. and R. S. M. conceived the project and helped design experiments and interpret the data. T. G. S., S. P. M., and J. C. M. prepared constructs and did expression studies of the protein. Selenomethionine protein purifications, crystallization, and data collection were carried out by G. A. H. Data processing was done by T. G. S. and S. R. ITC experi- ments were done by T. G. S. and S. P. M. Modeling studies were done by S. S. Results were discussed with all of the authors. T. G. S. and R. S. wrote the manuscript with contributions from coauthors.

3.8 Acknowledgements

We thank ESRF Access Program of RCB (supported by the Grant No. BT/INF/22/SP22660/2017 of the Department of Biotechnology). We thank the Advance Photon Source (IMCA-CAT beamline 17-ID) and Advanced light source (Beamline 4.2.2, MBC) for providing the beam time for this project. This research was supported by DBT-Indo Swedish Grant (BT/IN/SWEDEN/41/SR/2013); DBT-B-life grant (BT/PR5081/INF/156/2012); NCBS X-ray facility grant (BT/PR12422/MED /31/287/214) and a grant from the DST for SR. TG would like to acknowledge senior research fellowship support from CSIR.

8.9 Abbreviations

Hd, Haemophilus ducreyi; SatA, Sialic acid transport A gene; SiaP, Sialic acid binding Protein; ABC transporter, ATP-binding cassette transporter; TRAP transporter, Tripartite ATP-independent periplasmic transporter; Neu5Ac, N-Acetyl neuraminic acid; Neu5Gc, N-glycolyl neuraminic acid; ITC, Isothermal calorimetry;

3.10 References

1. de Las Rivas, J., and Fontanillo, C. (2010) Protein-protein interactions essentials: Key concepts to building and analyzing interactome networks. PLoS Comput. Biol. 6, 1–8
2. Kuzmanov, U., and Emili, A. (2013) Protein-protein interaction networks: probing disease mechanisms using model systems. Genome Med 5, 37
3. Pawson, T., and Nash, P. (2000) Protein − protein interactions define specificity in signal transduction. Genes and Develpoment 14, 1027–1047
4. Sevimoglu, T., and Arga, K. Y. (2014) The role of protein interaction networks in systems biomedicine. Comput. Struct. Biotechnol. J. 11, 22–27
5. Saini, R. D. (2017) Thermodynamics of protein-ligand interactions and their analysis. J. proteins proteimics 8, 205–217
6. Du, X., Li, Y., Xia, Y., Ai, S., Liang, J., Sang, P., and Ji, X. (2016) Insights into Protein – Ligand Interactions : Mechanisms , Models , and Methods. Int. J. Mol. Sci. 17, 1–34
7. Schauer, R., Srinivasan, G. V., Wipfler, D., Kniep, B., and Schwartz-albiez, R. (2011) in The molecular immunology of complex carbohydrates-3 pp. 525–548
8. Vimr, E. R., Kalivoda, K. A., Deszo, E. L., and Steenbergen, S. M. (2004) Diversity of Microbial Sialic Acid Metabolism. Microbiol. Mol. Biol. Rev. 68, 132–153
9. Hayakawa, T., Satta, Y., Gagneux, P., Varki, A., and Takahata, N. (2001) Alu-mediated inactivation of the human CMP- N-acetylneuraminic acid hydroxylase gene. PNAS 98, 11399–11404
10. Varki, A. (2010) Colloquium paper: uniquely human evolution of sialic acid genetics and biology. Proc. Natl. Acad. Sci. U. S. A. 107 Suppl, 8939–8946
11. Varki, A., and Schauer, R. (2009) Chapter 14 Sialic Acids. , 1–17
12. Varki, A. (2008) Sialic acids in human health and disease. Trends Mol. Med. 14, 351–360
13. Hood, D. W., Cox, A. D., Gilbert, M., Makepeace, K., Walsh, S., Deadman, M. E., Cody, A., Martin, A., Mansson, M., Schweda, E. K. H., Brisson, J.-R. R., Richards, J. C., Moxon, E. R., and Wakarchuk, W. W. (2001) Identification of a lipopolysaccharide alpha-2,3-sialyltransferase from Haemophilus influenzae. Mol Microbiol 39, 341–350
14. Hood, D. W., Makepeace, K., Deadman, M. E., Rest, R. F., Thibault, P., Martin, A., Richards, J. C., and Moxon, E. R. (1999) Sialic acid in the lipopolysaccharide of Haemophilus influenzae: strain distribution, influence on serum resistance and structural characterization. Mol. Microbiol. 33, 679–692

15. Apicella, M. A. (2012) Nontypeable Haemophilus influenzae : the role of N -acetyl-5-neuraminic acid in biology. Front. Cell. Infect. Microbiol. 2, 1–7

16. Severi, E., Hood, D. W., and Thomas, G. H. (2015) Sialic acid utilization by bacterial pathogens. Microbiology 153, 2817–2822

17. Almagro-moreno, S., and Boyd, E. F. (2009) Sialic Acid Catabolism Confers a Competitive Advantage to Pathogenic Vibrio cholerae in the Mouse Intestine. Infect. Immun. 77, 3807–3816

18. Setty, T. G., Cho, C., Govindappa, S., Apicella, M. A., and Ramaswamy, S. (2014) Bacterial periplasmic sialic acid-binding proteins exhibit a conserved binding site. Acta Crystallogr. Sect. D Biol. Crystallogr. 70, 1801–1811

19. Johnston, J. W., Coussens, N. P., Allen, S., Houtman, J. C. D., Turner, K. H., Zaleski, A., Ramaswamy, S., Gibson, B. W., and Apicella, M. A. (2008) Characterization of the N-acetyl-5-neuraminic acid-binding site of the extracytoplasmic solute receptor (SiaP) of nontypeable Haemophilus influenzae strain 2019. J. Biol. Chem. 283, 855–865

20. Mu, A., Severi, E., Mulligan, C., Watts, A. G., Kelly, D. J., Wilson, K. S., Wilkinson, A. J., and Thomas, G. H. (2006) Conservation of Structure and Mechanism in Primary and Secondary Transporters Exemplified by SiaP , a Sialic Acid Binding Virulence Factor from Haemophilus influenzae *. J. Biol. Chem. 281, 22212–22222

21. Felder, C. B., Graul, R. C., Lee, A. Y., Merkle, H., Sadee, W., and Chemistry, P. (1999) The Venus Flytrap of Periplasmic Binding Proteins : An Ancient Protein Module Present in Multiple Drug Receptors. AAPS pharSci 1

22. Oh, B., Pandits, J., Kang, C., Nikaidoii, K., Gokcen, S., Amesll, G. F., and Kim, S. (1993) Three-dimensional Structures of the Periplasmic Lysine/Arginine/ Ornithine-binding Protein with and without. J. Biol. Chem. 268, 11348–11355

23. Thomas, G. H. (2016) Sialic acid acquisition in bacteria – one substrate , many transporters. Biochem. Soc. Trans. 44, 760–765

24. Haines-menges, B. L., Whitaker, W. B., Lubin, J. B., and Boyd, E. F. (2014) Host Sialic Acids : A Delicacy for the Pathogen with Discerning Taste. Microbiol. Spectr., 1–17

25. Spinola, S. M., Bauer, M. E., and Munson, R. S. (2002) Immunopathogenesis of Haemophilus ducreyi infection (Chancroid). Society 70, 1667–1676

26. Cohen, M. S. (1998) Sexually transmitted diseases enhance HIV transmission: no longer a hypothesis. Lancet 351, S5--S7

27. Inamadar, A. C., and Palit, A. (2002) Chancroid: an update. Indian J. Dermatol. Venereol. Leprol. 68, 5–9

28. Alfa, M. J., DeGagne, P., and Totten, P. A. (1996) Haemophilus ducreyi hemolysin acts as a contact cytotoxin and damages human foreskin fibroblasts in cell culture. Infect. Immun. 64, 2349–2352

29. Wood, G. E., Dutro, S. M., and Totten, P. A. (1999) Target cell range of Haemophilus ducreyi hemolysin and its involvement in invasion of human epithelial cells. Infect. Immun. 67, 3740–3749

30. Cortes-Bratti, X., Chaves-Olarte, E., Lagergård, T., and Thelestam, M. (1999) The cytolethal distending toxin from the chancroid bacterium Haemophilus ducreyi induces cell-cycle arrest in the G2 phase. J. Clin. Invest. 103, 107–115

31. Ceelen, L. M., Decostere, A., Ducatelle, R., and Haesebrouck, F. (2006) Cytolethal distending toxin generates cell death by inducing a bottleneck in the cell cycle. Microbiol. Res. 161, 109–120

32. Cope, L. D., Lumbley, S., Latimer, J. L., Klesney-Tait, J., Stevens, M. K., Johnson, L. S., Purven, M., Munson, R. S., Lagergard, T., Radolf, J. D., and Hansen, E. J. (1997) A diffusible cytotoxin of Haemophilus ducreyi. Proc. Natl. Acad. Sci. U. S. A. 94, 4056–4061

33. Alfa, M. J., and DeGagne, P. (1997) Attachment of Haemophilus ducreyi to human foreskin fibroblasts involves LOS and fibronectin. Microb. Pathog. 22, 39–46

34. Gibson, B. W., Campagnari, A. A., Melaugh, W., Phillips, N. J., Apicella, M. A., Grass, S., Wang, J., Palmer, K. L., and Munson, R. S. (1997) Characterization of a transposon Tn916-generated mutant of Haemophilus ducreyi 35000 defective in lipooligosaccharide biosynthesis. J. Bacteriol. 179, 5062–5071

35. Melaugh, W., Campagnari, A. A., and Gibson, B. W. (1996) The lipooligosaccharides of Haemophilus ducreyi are highly sialylated. J. Bacteriol. 178, 564–570

36. Post, D. M. B., Mungur, R., Gibson, B. W., and Munson, R. S. (2005) Identification of a novel sialic acid transporter in Haemophilus ducreyi. Infect. Immun. 73, 6727–6735

37. Hollenstein, K., Frei, D. C., and Locher, K. P. (2007) Structure of an ABC transporter in complex with its binding protein. Nature 446, 213–216

38. Oldham, M. L., Davidson, A. L., and Chen, J. (2008) curr opin struct biol. curr opin struct biol 18, 726–733

39. Davidson, A. L., Dassa, E., Orelle, C., and Chen, J. (2008) Structure, function, and evolution of bacterial ATP-binding cassette systems. Microbiol. Mol. Biol. Rev. 72, 317–364

40. Dunten, P., and Mowbray, S. L. (1995) Crystal structure of the dipeptide binding protein from Escherichia coli involved in active transport and chemotaxis. Protein Sci. 4, 2327–2334

41. Levdikov, V. M., Blagova, E. V, Brannigan, J. A., Wright, L., Vagin, A. A., and Wilkinson, A. J. (2005) The structure of the oligopeptide-binding protein, AppA, from Bacillus subtilis in complex with a nonapeptide. J. Mol. Biol. 345, 879–892

42. Sahili, A. El, Li, S., Lang, J., Virus, C., Planamente, S., Ahmar, M., Guimaraes, B. G., Aumont-nicaise, M., Vigouroux, A., Reader, J., Queneau, Y., and Faure, D. (2015) A Pyranose-2-Phosphate Motif Is Responsible for Both Antibiotic Import and Quorum- Sensing Regulation in Agrobacterium tumefaciens. PLOS Pathog., 1–24

43. Vergauwen, B., Elegheert, J., Dansercoer, A., Devreese, B., and Savvides, S. N. (2010) Glutathione import in Haemophilus influenzae Rd is primed by the periplasmic heme-binding protein HbpA. PNAS 107, 13270–13275

44. Thomson, J., Liu, Y., and Sturtevant, J. M. (1998) A thermodynamic study of the binding of linear and cyclic oligosaccharides to the maltodextrin-binding protein of Escherichia coli. Biophys. Chem., 101–108

45. Fukami-kobayashi, K., Tateno, Y., and Nishikawa, K. (1999) Domain Dislocation : a Change of Core Structure in Periplasmic Binding Proteins in their Evolutionary History. J. Mol. Biol. 286, 279–290

46. Berntsson, R. P., Smits, S. H. J., Schmitt, L., Slotboom, D., and Poolman, B. (2010) A structural classification of substrate-binding proteins. FEBS Lett. 584, 2606–2617

47. Maqbool, A., Horler, R. S. P., Muller, A., Wilkinson, A. J., Wilson, K. S., and Thomas, G. H. (2015) The substrate-binding protein in bacterial ABC transporters: dissecting roles in the evolution of substrate specificity. Biochem. Soc. Trans. 43, 1011–1017

48. Dundas, J., Ouyang, Z., Tseng, J., and Binkowski, A. (2006) CASTp : computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acids Res 34, 116–118

49. M, S. D. T., Kapp, U., Aberg, A., Brown, K., Strub, K., and Cusack, S. (1996) Crystallization and preliminary X-ray analysis of the 9 kDa protein of the mouse signal recognition particle and the selenomethionyl-SRP9. FEBS Lett. 384, 219–221

50. Terwilliger, T. (2004) SOLVE and RESOLVE : automated structure solution , density modification , and model building research papers. J. synchrotron Rad 11, 49–52

51. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. Acta Crystallogr. D. Biol. Crystallogr. 66, 486–501

52. Houtman, J. C. D., Brown, P. H., Bowden, B., Yamaguchi, H., Appella, E., Samelson, L. E., and Schuck, P. (2007) Studying multisite binary and ternary protein interactions by

global analysis of isothermal titration calorimetry data in SEDPHAT: Application to adaptor protein complexes in cell signaling. Protein Sci. 16, 30–42

53. Liu, T., and Altman, R. B. (2011) Using Multiple Microenvironments to Find Similar Ligand-Binding Sites : Application to Kinase Inhibitor Binding. PLoS Comput. Biol. 7

54. The PyMol Molecular Graphics System, Version 2.1, Schrödinger, LLC.

Chapter 4

Structural and functional characterization of N-acetyl mannosamine kinases from pathogenic bacteria

To Zinc or not to Zinc –

Structure and function of *N*-acetylmannosamine kinases from pathogenic bacteria.

Thanuja Gangi Setty[1,2], Arunabha Sarkar[3], David Coombes[4], Renwick C. J. Dobson[4], and Ramaswamy Subramanian[1,5*].

[1]Institute for Stem Cell Science and Regenerative Medicine, GKVK Post, Bangalore, KA 560065, India; [2]The University of Trans-Disciplinary Health Sciences & Technology (TDU), Bangalore, KA 560064 India; [3]National Centre for Biological Sciences – TIFR, Bangalore, 560065, India; [4]Biomolecular Interaction Centre and School of Biological Sciences, University of Canterbury, New Zealand; [5]Department of Biological Sciences, Bindley Bioscience Center, Purdue University, West Lafayette, IN 47907, USA.

*To whom correspondence should be addressed: S. Ramaswamy: Institute for Stem Cell Science and Regenerative Medicine, GKVK Post, Bangalore, 560065. India; ramas@instem.res.in; Tel. +91 80-67176002

Running title: Insights into Zn-binding motif of N-acetylmannosamine kinases

Keywords: Sialic acid; *N*-acetyl mannosamine*; Haemophilus influenzae; Fusobacterium nuleatum; Vibrio cholera; Pasteurella multocida;* Structure; Enzyme kinetics; ITC.

The abbreviations used are: Neu5Ac, N-acetylneuraminic acid; Neu5Gc, N-glycolylneuraminic acid; NanK, N-acetylmannosamine kinase; LOS, Lipooligosaccharide; LPS, Lipopolysaccharide; ITC, Isothermal calorimetry; AEC, 3-amino-9-ethylcarbazole.

4.1 ABSTRACT

Pathogenic bacteria can often catabolize sialic acids as their sole nutrient source of carbon and nitrogen. *N*-acetyl mannosamine kinase (NanK) in the sialic acid catabolic pathway phosphorylates N-acetylmannosamine, and it belongs to ROK superfamily of enzymes. The Zn-binding motif is one of the conserved signature motifs in this family of enzymes and is known to be important for their structural and functional properties. Previous structural studies have shown that the Zn-binding motif is absent from some including the *Fusobacterium nucleatum* NanK(*Fn*-NanK), but the effect of its absence on the enzyme activity is not known. Using kinetic and thermodynamic studies, we have studied the functional properties of *Fn*-NanK and compared it with other Zn-binding NanKs from *Haemophilus influenzae (Hi), Pasteurella multocida (Pm),* and *Vibrio Cholerae (Vc)*. There are no major differences in the kinetic properties between these NanKs. Thermodynamic studies show that, NanK enzymes follow either ordered (*Pm*-NanK and *Vc*-NanK) or random binding (*Hi*-NanK and *Fn*-NanK) pattern to their ligands ManNAc and AMPPNP. We also report the crystallographic structures of *Pm*-NanK:unliganded, *Pm*-NanK holo (AMPPNP), *Pm*-NanK:ManNAc, *Hi*-NanK:ManNAc and *Hi*-NanK:ManNAc-6P:ADP. Our studies suggest that even in the absence of Zn-binding motif, functional properties of *Fn*-NanK are not significantly altered. These findings indicate that low sequence homology and presence/absence of Zn have no major impact on the overall three dimensional structure or the molecular mechanisms of these enzymes. Interestingly Thr 131 that binds to both the sugar and the nucleotide in the active sites does not affect binding, but is critical for catalysis.

4.2 Introduction

Zinc is an essential trace metal and acts as a cofactor in several proteins and enzymes for their biological functions. It is a nontoxic metal and it has coordination flexibility with different macromolecules. Due to its physical and chemical properties, it is utilized by several enzymes in different microorganisms, plants and animals for carrying out diverse biological functions (1, 2). In eukaryotic organisms, Zn is usually associated with two histidines and two cysteines, known as Cys_2His_2 zinc finger domain, and often involved in protein-nucleic acid and protein-protein interactions. In prokaryotes, this domain was first identified in a transcriptional regulator (Ros) from *Agrobacterium tumefaciens* (3). In addition to its role as a cofactor, Zn is required for catalytic, structural, and functional properties of many proteins and enzymes. Zinc is also known to play an important role in folding and oligomerization of proteins (4–7).

Sialic acids are a family of cyclic nine-carbon amino sugar acids and decorated as the outer most moiety of lipooligosaccharide /lipopolysaccharide (LOS/LPS) on the cell membranes of organisms ranging from prokaryotes to eukaryotes (8–10). Majority of the pathogenic bacteria, such as *Haemophilus influenzae (Hi), Pasteurella multocida (Pm), Vibrio cholerae (Vc),* and *Fusobacterium nucleatum (Fn)* that reside in the host mucous rich environments like gut and lungs scavenge sialic acids from the host and incorporate it as the outermost sugar on their LOS/LPS to evade the host immune system and is known as 'molecular mimicry' (11–14). Furthermore, bacteria use these sugars as carbon and nitrogen source. During this process, part of the scavenged sugar is converted to fructose 6-phosphate by enzymes of the Nan-Nag gene cluster (NanA, NanK, NanE – NagA, NagB) in the sialic acid catabolic pathway (Figure 4.1). Interestingly, this gene cluster has been shown to be present in most of the pathogenic and commensal bacteria (15). Earlier reports have also suggested that the Nan-Nag gene cluster of *V. cholera* is helpful in initial bacterial colonization. In addition, it was shown that the bacteria utilizing sialic acids as their sole carbon source have increased *N*-acetyl mannosamine kinase (NanK) expression (12). These studies underline the importance of sialic acids for the survival of pathogens in the host.

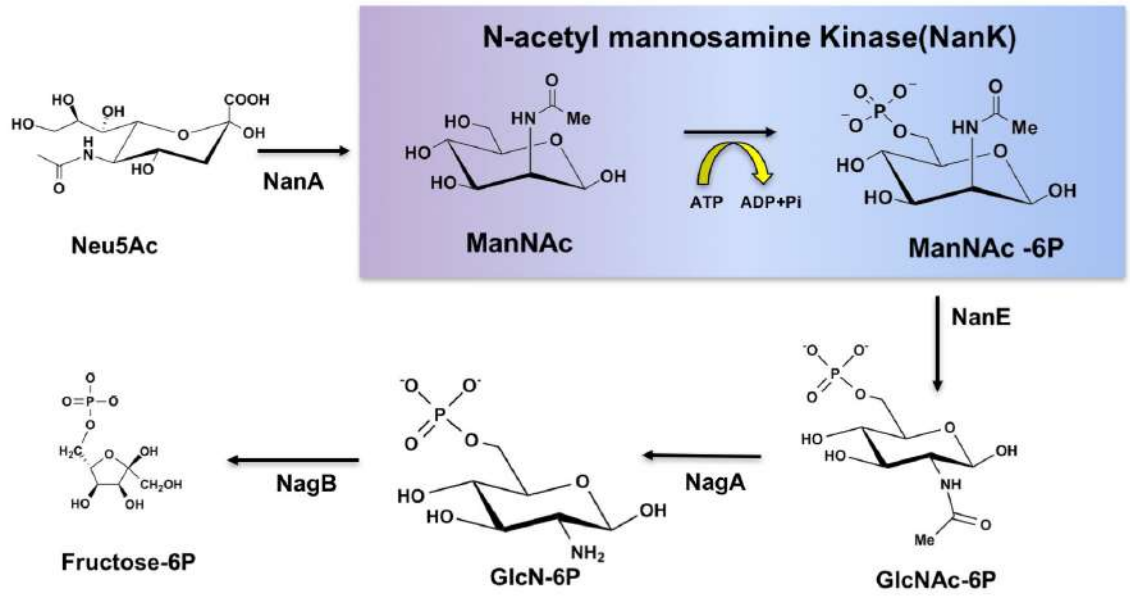


Figure 4.1. Schematic representation of sialic acid catabolic pathway in the Gram negative bacteria: The enzymatic reaction of phosphorylation of ManNAc to ManNAc-6P by the enzyme NanK (*N*-acetyl mannosamine kinase) in presence of ATP is highlighted in the box. The chemical structures are drawn using ChemDraw. SiaP, Sialic acid binding protein; Neu5Ac, *N*-Acetyl neuraminic acid; NanA, *N*-acetylneuraminic acid lyase; ManNAc, *N*-acetyl mannosamine, ManNAc-

6P, *N*-acetyl mannosamine 6-phosphate; NanE, *N*-acetylmannosamine-6-P epimerase; NagA, *N*-acetylgucosamine-6-phosphate deacetylase; NagB, Glucosamine-6-phosphate deaminase

ROK (bacterial **R**epressors, uncharacterized **O**pen reading frames, and sugar **K**inases) family of enzymes code for bacterial transcriptional repressors, sugar kinases and large number of uncharacterized proteins. The members of this family are functionally very diverse and are found in prokaryotes as well as eukaryotes (16, 17). *N*-acetyl mannosamine kinase (NanK) belongs to the ROK family and it is the second enzyme in the sialic acid catabolic pathway. NanK phosphorylates *N*-acetyl mannosamine (ManNAc) into *N*-acetyl mannosamine 6-phosphate (ManNAc) in the presence of ATP, an important step in the sugar metabolism (Figure 4.1) (18, 19). In general, during phosphorylation, the binding of ligand and nucleotide to the kinase induces conformational changes in the protein to facilitate the transfer of **γ**-phosphate of ATP to the substrate in presence of a divalent metal ion Mg^{2+} (20–22).

Interestingly, NanK enzymes from both bacteria and eukaryotes are involved in the phosphorylation of ManNAc. However, the human form (hMNK) has been shown to be involved in the biosynthetic pathway of Neu5Ac (8, 23). The structural and functional aspects of hMNK from GNE/MNK complex (22, 24) and the crystal structures of bacterial *Fn*-NanK and *Ec*-NanK have been published (22, 25) (EcNanK: Unpublished structure from New York Structural genomics consortium-PDB ID 2AA4). All these ROK family NanKs possess conserved signature motifs: i) DXDGT, located in the ATP binding site at the N-terminus of the protein; ii) EXGH motif; and iii) a catalytic aspartate residue. Surprisingly, excluding *Fn*-NanK, all other NanKs, from human, *E. coli*, *H. influenzae* (*Hi*), *P. multocida* (*Pm*), and *V. cholera* (*Vc*) contain a Zn-binding CXCGXXGC motif (Figure 4.2) (19). In contrast to traditional His_2Cys_2 Zn finger domain, the coordination of Zn in these NanK enzymes is His_1Cys_3 association as Zn coordinates with one histidine and three cysteines.

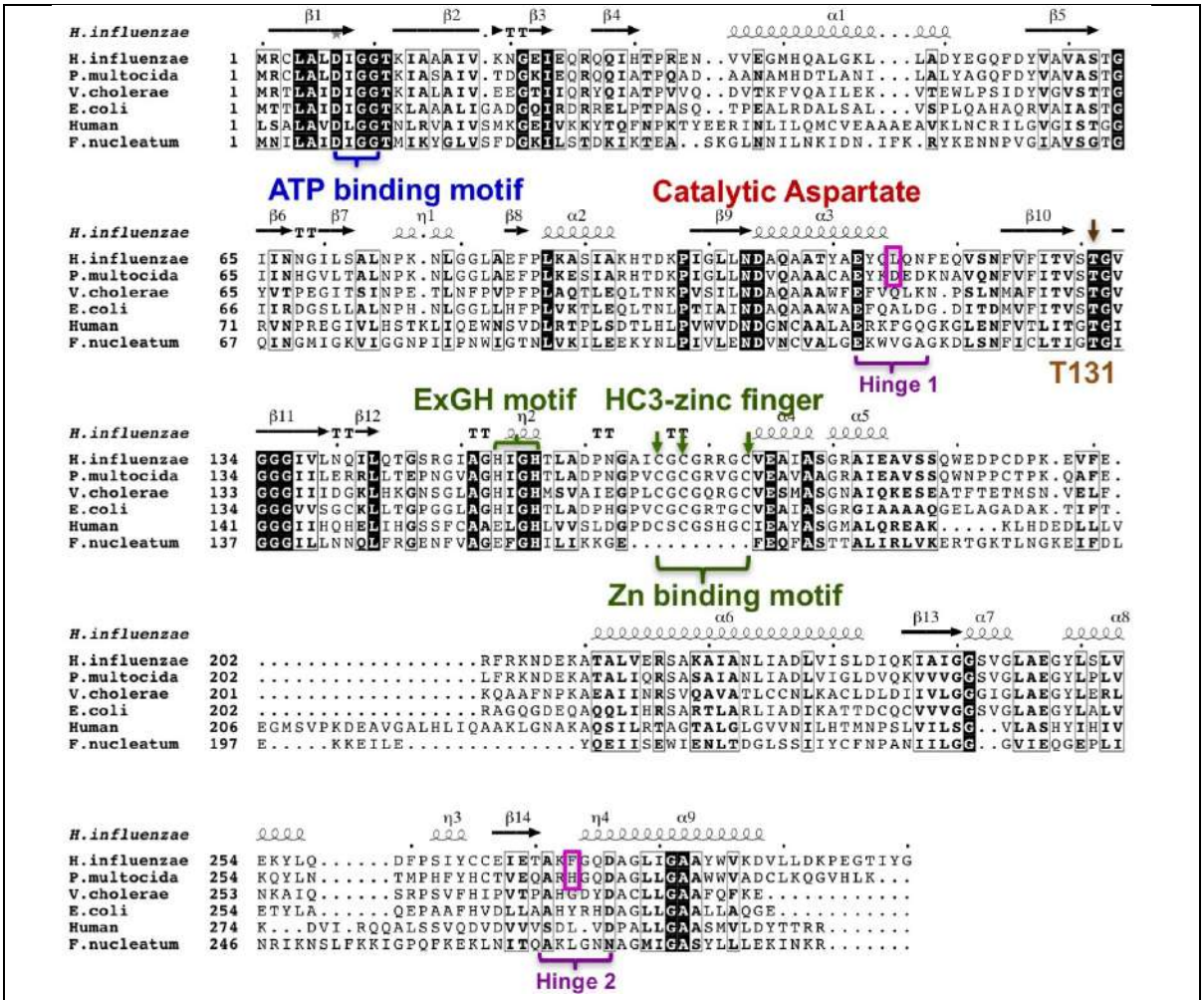


Figure 4.2. Multiple sequence alignment of NanK enzymes from pathogenic bacteria and human. Sequence alignment of NanK enzymes from *H. influenzae, P. multocida, V. cholera, F. nuleatum, E. coli* and *human*. Conserved ROK enzyme signature motifs, hinge regions, and conserved T131 are highlighted in the alignment figure. The alignment also highlights the absence of Zn-binding motif in *F. nucleatuem.* Sequence alignment was performed using clustal W and the figure was generated by ESPript (39)

The sequence similarity between *Fn*-NanK to the other NanKs (*Hi*-NanK, *Pm*-NanK and *Vc*-NanK) is only 20-25%, thus it is very important to understand the functional differences between these NanK enzymes in addition to the role of Zn-binding motif.

Towards this, we have grouped *Hi*-NanK, *Pm*-NanK, and *Vc*-NanK into Zn-binding NanKs and *Fn*-NanK as non-Zn-binding NanK. To understand the role of the Zn-binding domain, we report here the kinetic and thermodynamic properties of NanK enzymes from these two groups. Further, to illustrate the functional difference between the NanK enzymes, we also report the structures of *Pm*-NanK in unliganded form, *Pm*-NanK holo (in complex with

AMPPNP), *Pm*-NanK:ManNAc, *Hi*-NanK:ManNAc, and *Hi*-NanK:ManNAc-6P:ADP. Based on our structural analysis, Thr131 seems to be one of the important residues during the **γ**-phosphate transfer from ATP. Further, we have also cloned *Hi*-NanK T131V mutant, purified and characterized the functionality of this enzyme by thermodynamic and kinetic studies. In addition, based on the structural analysis, *Pm*-NanK Asp115 and His273 residues are presumed to be important for its sequential binding properties of the enzyme to its substrates. To test this hypothesis, D115L, H273F, and D115L H273F substitutions were generated in *Pm*-NanK, purified and their thermodynamic parameters are determined. In summary, this systematic and comprehensive biophysical characterization of these two groups of enzymes elucidates the significance of unique amino acids and the plausible role of Zn-binding motif during the phosphorylation event.

4.3 RESULTS

4.3.1 NanK enzymes: Low sequence similarity but have highly conserved signature motifs

Multiple sequence alignment of NanK amino acid sequences from human and different pathogenic bacteria was carried out using ClustalW (26). The analysis reveals that the sequence identity between different NanK enzymes is only 20-25%. Similarly, the sequence identity of *Fn*-NanK with other NanK enzymes from pathogenic bacteria is 25%. Among these sequences, *Hi*-NanK and *Pm*-NanK shows higher sequence identity of approximately 70% (Table 4.1). Though the sequence similarity between the NanK enzymes appears to be moderate, these kinases show high homology at the conserved signature motifs of the ROK family such as DXGXT motif and aspartate catalytic residue. Further, excluding *Fn*-NanK, all the other bacterial NanKs also shows the presence of a conserved Zn-binding motif (Figure 2)(19, 22, 25).

	F. nucleatum	*Human*	*V. cholerae*	*E. coli*	*H. influenzae*	*P. multocida*
F. nucleatum	**100.00**	**27.72**	**23.83**	**24.37**	**25.44**	**26.15**
Human	**27.72**	**100.00**	**28.87**	**25.87**	**23.79**	**24.83**
V. cholerae	**23.83**	**28.87**	**100.00**	**40.77**	**41.81**	**41.11**
E. coli	**24.37**	**25.87**	**40.77**	**100.00**	**49.65**	**50.35**
H. influenzae	**25.44**	**23.79**	**41.81**	**49.65**	**100.00**	**70.37**
P. multocida	**26.15**	**24.83**	**41.11**	**50.35**	**70.37**	**100.00**

Table 4.1. Table showing the percentage sequence similarity between NanKs from different pathogenic bacteria and human.

4.3.2 Zn-binding and non Zn-binding NanK enzymes show similar kinetic properties

NanK enzymes catalyze the phosphorylation of ManNAc. To understand the significance of Zn-binding motif in these enzymes, we have characterized the steady state kinetic properties of *Hi*-NanK, *Pm*-NanK and *Vc*-NanK (Zn-binding NanK enzymes), and *Fn*-NanK (non Zn-binding NanK) to its substrate ManNAc. The apparent K_m, V_{max} and K_{cat} for different NanK enzymes with ManNAc are presented in Table 4. 2. The enzyme activity assays show that *Hi*-NanK, *Pm*-NanK, and *Vc*-NanK have K_m ~ 65 µM, whereas *Fn*-NanK has Km ~ 450 µM (Figure 4. 3). The kinetic study shows that, though the *Fn*-NanK lacks Zn-binding motif, its kinetic properties are very similar to other NanK enzymes.

	K_m (µM)	**V_{max} (µmoles $min^{-1}mg^{-1}$)**	**K_{cat} (min^{-1})**	**K_{cat}/K_m ($\mu M^{-1}min^{-1}$)**
H. influenzae	66 ± 4.7	13.3±2	457	7
P. multocida	63.3 ± 4.3	9.7 ± 1	305	5
V. cholerae	71 ± 8.1	6.7 ± 2	211	3
F. nucleatum	715 ± 55	10.4 ± 0.3	330	0.5

Table 4.2. The table showing the comparative kinetic parameters of Zn-binding and non-Zn-binding NanKs from pathogenic bacteria.

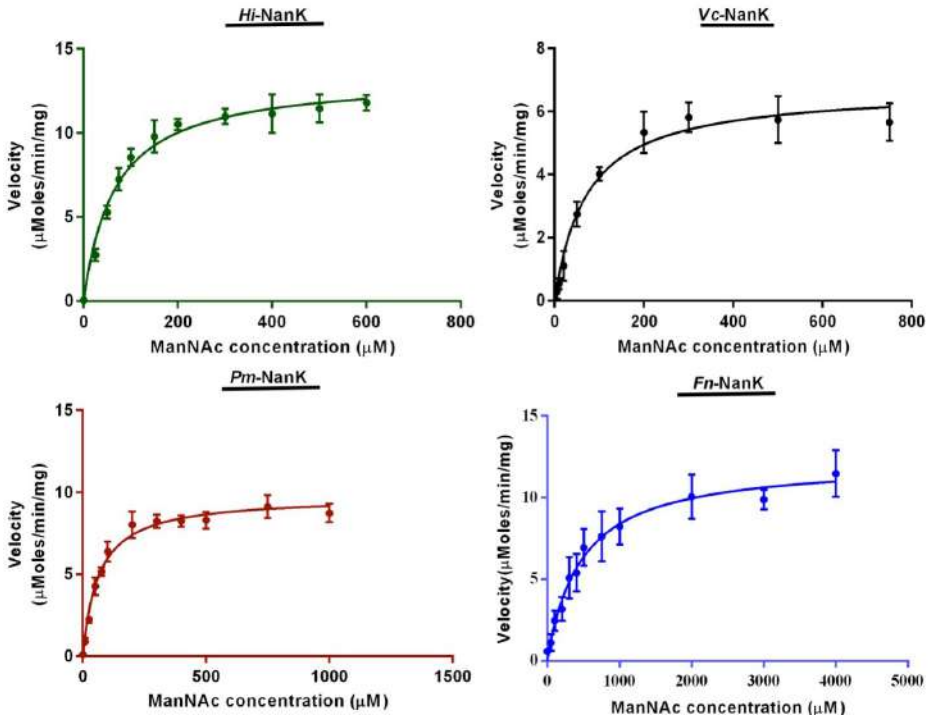
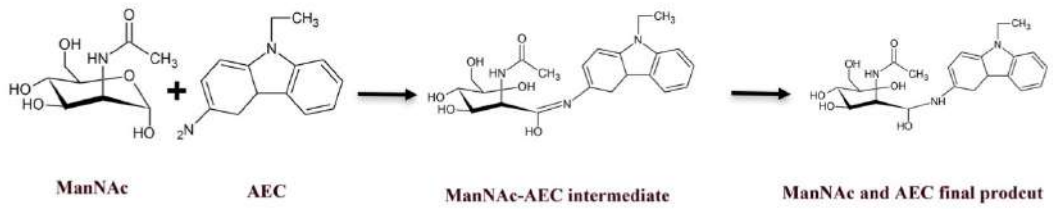


Figure 4.3. Enzyme kinetic studies of NanKs. Enzyme kinetics of NanK for the conversion of ManNAc to ManNAc-6P were determined by using various concentrations of ManNAc and the luminiscence was measured using plate reader. Data was plotted by Michelis-Menten equation and V_{max} and K_m were calculated. The final K_m was calculated from experimental triplicates and technical duplicate values.

4.3.3 Reductive amination of NanK enzyme products with AEC and subsequent MS confirms phosphorylation

In addition to kinetic analysis, characterization of the product of the Zn-binding (*Pm*-NanK) and non-Zn-binding NanK (*Fn*-NanK) enzymes were performed by Mass-spectrometry after AEC (3-amino-9-ethylcarbazole) derivatization, to confirm phosphorylation activity. In this procedure, the substrate and products of the NanK enzyme were first reduced by AEC in presence of sodium cyanoborohydrate to ease their quantification by mass spectrometry (27). Subsequent MS revealed the daughter ion signature that can only emanate from the expected product of the NanK enzyme. These studies confirm that both Zn and Non-Zn binding NanKs can catalyze the phosphorylation reaction and generate the product *N*-acetyl mannosamine 6-phosphate. (Figure 4.4A, 4.4B)

(A)



(B)

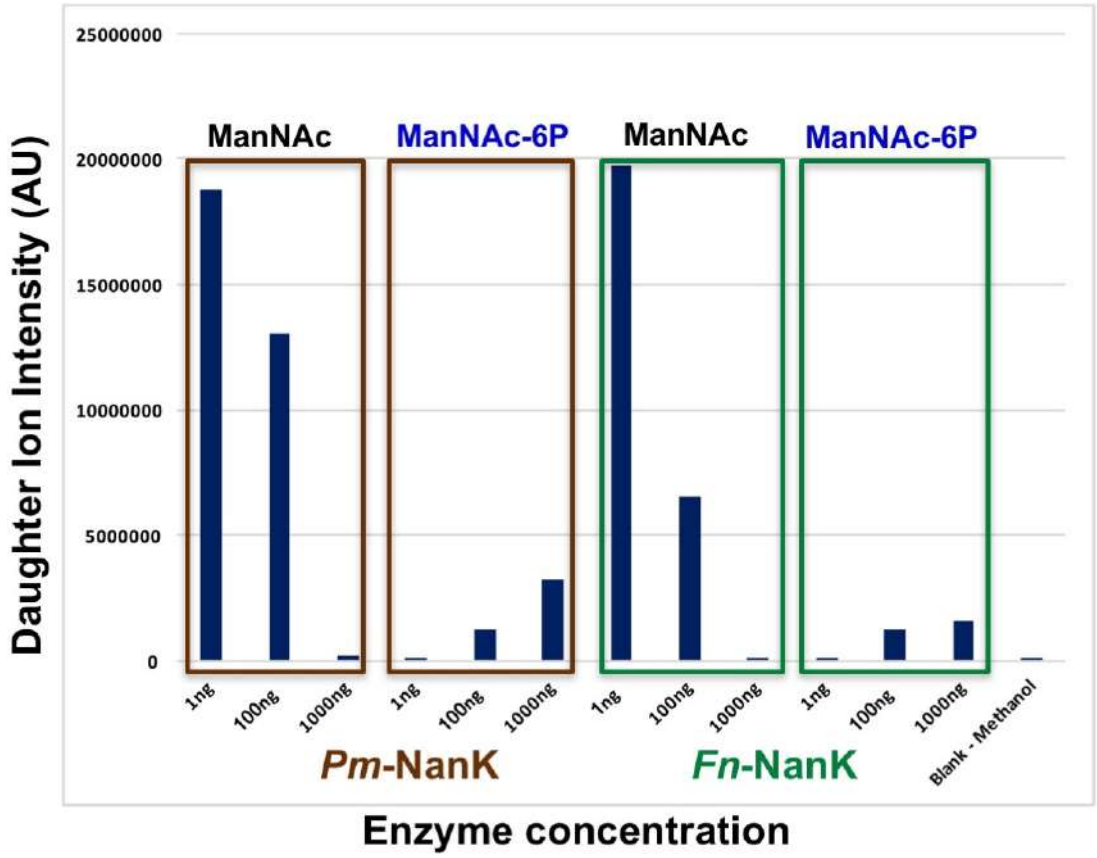


Figure 4.4. Mass spectrometry analysis of enzyme reaction of *Pm*-NanK and *Fn*-NanK.

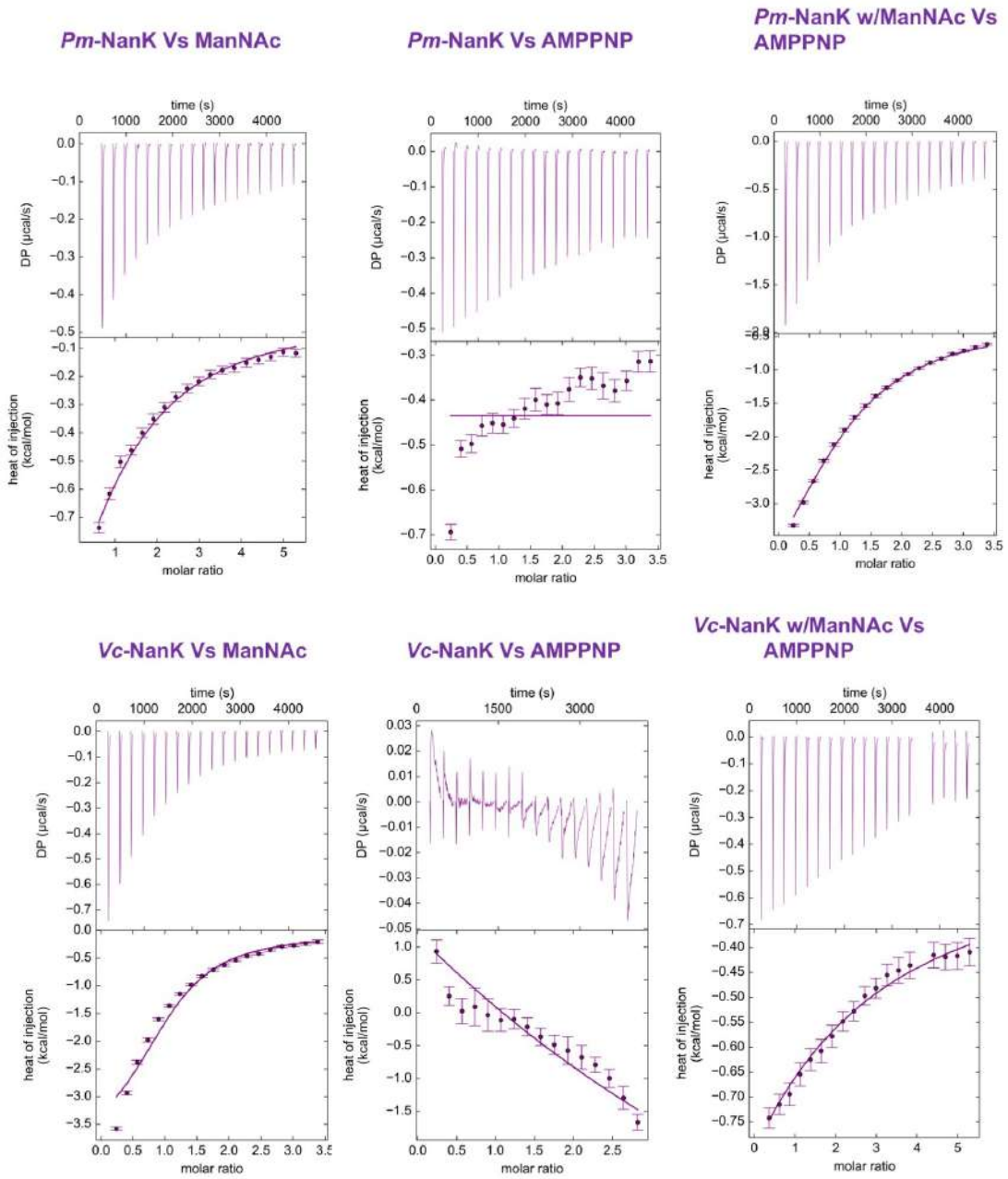
(A) Chemical reaction showing the AEC derivatization of substrate ManNAc.

(B) Bar graph representing the amount of the substrate and product formed after the enzymatic reaction by *Pm*-NanK and *Fn*-NanK.

4.3.4 Thermodynamic characterization of ManNAc and AMPPNP binding to NanK eznymes reveals two different patterns of binding

The kinetic and mass spectrometric results did not show any major differences between Zn and non-Zn-binding NanK enzymes. We then used isothermal titration calorimetry (ITC), to investigate the differences in the thermodynamic properties between these two groups of NanK enzymes binding to substrates ManNAc and AMPPNP. Figure 4 represents the ITC thermograms, where the top panel illustrates the raw heat liberated during protein ligand titrations and the bottom panel demonstrates the integrated enthalpy changes for these titrations. Single site binding mode analysis of the data shows that NanK enzymes bind to ManNAc and AMPPNP in a 1:1 stoichiometric ratio. To test whether the binding of NanKs to ManNAc and AMPPNP is an ordered or random, we have carried out combination of titrations. The ITC results show that *Pm*-NanK and *Vc*-NanK binds to ManNAc and AMPPNP with micro molar affinity, but they bind to AMPPNP only in the presence of ManNAc (Figure 4. 5A). These results suggest that prior to AMPPNP binding, *Pm*-NanK: ManNAc and *Vc*-NanK: ManNAc complex formation is necessary. However, our ITC results show that binding of *Hi*-NanK and *Fn*-NanK to ManNAc and AMPPNP are independent with approximately equi-micromolar affinities (Figure 4. 5B). In summary, these results demonstrate that *Pm*-NanK and *Vc*-NanK follow ordered binding, whereas *Hi*-NanK and *Fn*-NanK follow random binding to the substrates. The binding affinities, enthalpic and entropic contributions of different NanK enzymes with their substrates are presented in the Table 4. 3.

(A)


Pm-NanK Vs ManNAc
Pm-NanK Vs AMPPNP
Pm-NanK w/ManNAc Vs AMPPNP
Vc-NanK Vs ManNAc
Vc-NanK Vs AMPPNP
Vc-NanK w/ManNAc Vs AMPPNP
time (s)
DP (µcal/s)
heat of injection (kcal/mol)
molar ratio

(B)

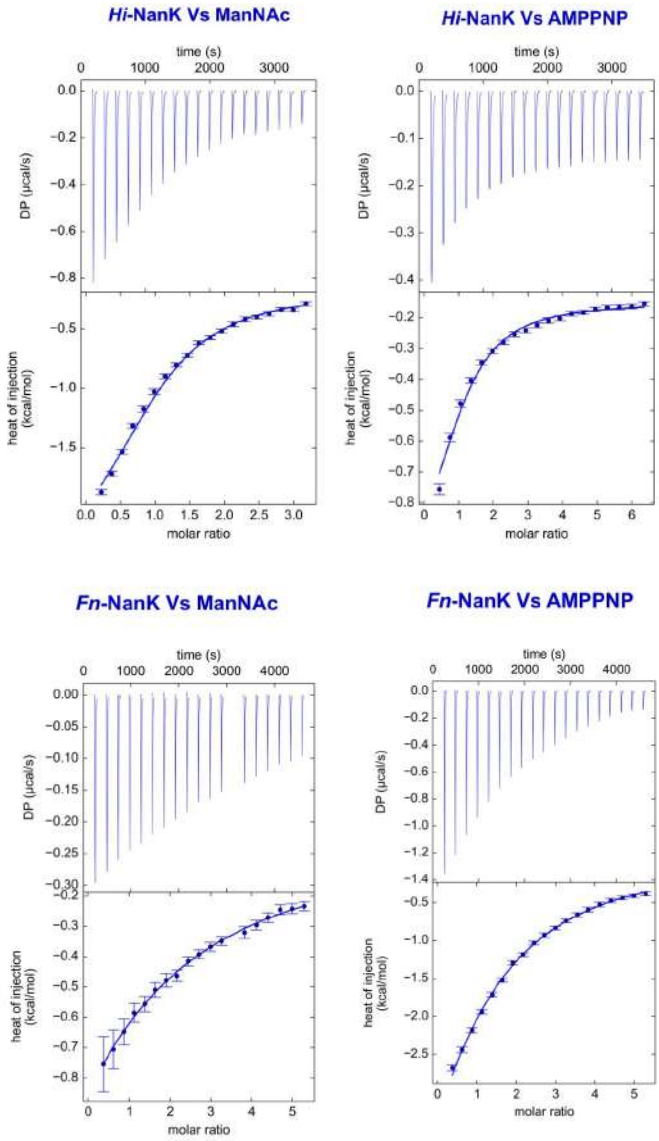


Figure 4.5. ITC studies of *Pm, Vc, Hi* and *Fn* - NanKs binding to its substrates - ManNAc and AMPPNP. In each chromatograph, top panel shows the raw data and the bottom panel shows the integrated curves. (A) Micro calorimetric titration of *Pm* and *Vc* -NanK with ManNAc and AMPPNP. Titration results show that *Pm* and *Vc* -NanK binds to AMPNP only in the presence of ManNAc. (B) Micro calorimetric titration of *Hi* and *Fn*-NanK with ManNAc and AMPPNP. Titration results show that *Hi* and *Fn*-NanK binds to both the substrates independently with micro molar affinities.

	NanK	(in presence of)	Ligand	K_d (μM)	ΔH (Kcal/mol)	TΔS (Kcal/mol)	ΔG (Kcal/mol)
			Sequential binding				
1	*P. multocida*	-	ManNAc	88 ± 0.0528	-1.408 ± 0.00042	4.12	-5.52
		AMPPNP	ManNAc	99 ± 0.014	-6.88 ± 0.00047	-1.419	-5.4
		-	AMPPNP	No measurable binding			
		ManNAc	AMPPNP	146 ± 0.138	-5.466 ± 0.0031	-0.235	-5.16
2	*V.cholera*	-	ManNAc	19.7 ± 0.0394	-3.961 ± 0.003	2.458	-6.4
		AMPPNP	ManNAc	31.3 ± 0.0496	-9.84 ± 0.0043	-3.695	-6.1
		-	AMPPNP	No measurable binding			
		ManNAc	AMPPNP	1200 ± 3	-3.874 ± 0.0099	0.107	-4
			Random binding				
3	*H. influenzae*	-	ManNAc	76 ± 0.026	-2.435 ± 0.0004	3.179	-5.61
		AMPPNP	ManNAc	73 ± 0.037	1.37 ± 0.0006	3.706	-2.4
		-	AMPPNP	53.5 ± 0.017	-0.969 ± 0.00014	4.858	-5.9
		ManNAc	AMPPNP	71 ± 0.0228	-0.391 ± 0.00004	5.266	-5.7
4	*F. nucleatum*	-	ManNAc	503 ± 3.92	-4.722 ± 0.0374	-0.223	-4.9
		AMPPNP	ManNAc	48.5 ± 0.01018	-4.194 ± 0.00041	1.691	-5.9
		-	AMPPNP	227 ± 0.154	-10.551 ± 0.0045	-5.58	-4.9
		ManNAc	AMPPNP	30 ± 0.0081	-8.238 ± 0.0009	-2.076	-6.1

Table 4.3. Binding affinities and thermodynamic parameters of NanKs binding to its ligands ManNAc and AMPPNP (ATP analogue). Isothermal calorimetry (ITC) was used to measure the binding affinities, enthalpy, and entropy for the binding of NanKs with its ligands. The data were analyzed using Origin analysis software.

4.3.5 Crystallographic details of Pm-NanK and Hi-NanK

In order to understand the structural basis of differential binding properties of NanK enzymes, here we determined and reported crystal structures of *Pm*-NanK and *Hi*-NanK. Further, we compared this structural information with previously reported NanK structures.

Pm-NanK in unliganded, holo (AMPPNP bound form) and with ManNAc-bound forms were crystallized. *Pm*-NanK:ManNAc crystal structure was refined to a resolution of 2.7 Å and it has one molecule in the asymmetric unit with positive electron density in the binding pocket demonstrating the presence of ManNAc (full occupancy). Crystal structures *Pm*-

NanK:Apo and *Pm*-NanK:holo (0.76 occupancy for AMPPNP) were refined to a resolution of 1.9 Å and they reveal the presence of two molecules in the asymmetric unit (Figure 4. 6).

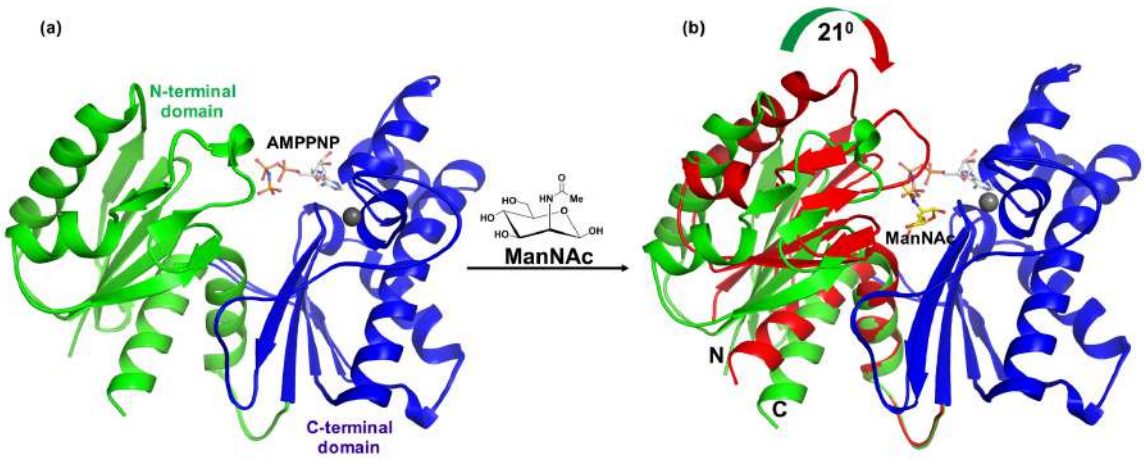


Figure 4.6. Structures of *Pm*-NanK in open and closed conformations. (a) Cartoon representation showing the superposition of *Pm*-NanK-unliganded and *Pm*-NanK:AMPPNP conformations (N-terminal domain colored in green and C-terminal domain is colored in blue). AMPPNP is represented in ball and stick model. This superposition clearly shows that upon binding of the nucleotide, there is no conformational change in the enzyme. (b) Cartoon representation of superposition of C-terminal domains from *Pm*-NanK:AMPNP and Pm-NanK:ManNAc (N-terminal domain colored red and C-terminal domain is colored in blue) structures. ManNAc is represented in ball and stick form in yellow color. Superposition clearly shows that upon ManNAc binding there is rotation of the N- terminal domain by about 21^0 towards C-terminal domain. The figures are generated using PyMol (Schrödinger) (40)

Similarly, two different forms of *Hi*-NanK, one bound to ManNAc and the other bound to ADP:ManNAc-6P were crystallized. HiNanK:ADP:ManNAc-6P crystal structure was determined to a resolution of 2.65 Å and these crystals were obtained by co-crystallization, wherein *Hi*-NanK was incubated with ATP, $MgCl_2$ and ManNAc at room temperature before setting up of crystallization trays. Interestingly, the final crystal structure shows the presence of positive electron density for both products ManNAc-6P and ADP in the binding pocket. These results illustrate the transfer of γ-phosphate from the ATP to ManNAc due to the enzyme catalytic activity (Figure 4. 7). Likewise, *Hi*-NanK:ManNAc crystals were also obtained by co-crystallization, wherein *Hi*-NanK was incubated with sodium orthovandate, $MgCl_2$, and

ManNAc. However, the crystal structure shows the presence of only ManNAc (full occupancy) and not sodium orthovandate and/or ADP in the binding pocket. The crystal structure was refined with data extending to 2.27Å resolution.

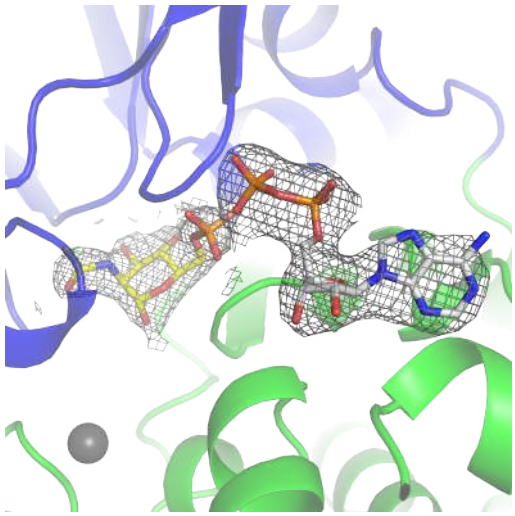


Figure 4.7. Crystal structure showing the binding pocket of *Hi*-NanK. (a) *Hi*-NanK bound to its products ManNAc-6P and ADP displays that the enzyme is catalytically active. 2Fo-Fc electron density maps for the ManNAc-6P and ADP contoured at 2**σ**, represented by a grey mesh. The figure is generated by using the program Pymol (40).

Although, some of the NanK crystal structures have one molecule in the asymmetric unit, the gel filtration studies have confirmed that these kinases exist as biological dimers and it is generated by crystallographic symmetry.

The crystallographic details and the refinement statistics of *Pm*-NanK and *Hi*-NanK are presented in Table 4. 4.

	Pm-NanK	*Pm*-NanK: AMPPNP	*Pm*-NanK: ManNAc	*Hi*-NanK: ManNAc	*Hi*-NanK: ManNAc-6P:ADP
Wavelength	0.97856 A	0.97857	0.97857	0.86500	0.97857
Resolution range	45.58 - 1.90 (1.92 - 1.90)	45.69- 2.00 (2.05 - 2.00)	48.06 - 2.9 (3.08 - 2.9)	46.01-2.22 (2.29 - 2.22)	46.30-2.64 (2.77 - 2.64)
Space group	P 3_2 2 1	P 3_2 2 1	C 2 2 2_1	I 4_1 2 2	I 4_1 2 2
Data collection statistics					
Unit cell a, b, c (Å) **α**, **β**, **γ** (degrees)	126.3, 126.3, 82.4 90, 90, 120	126.4, 126.4, 82.9 90, 90, 120	103.8, 173.7, 48.1 90, 90, 90	91.1, 91.1, 184.0 90, 90, 90	92.6, 92.6, 183.3 90, 90, 90
Total reflections	579087 (29614)	514396 (36298)	62338 (9925)	69837 (4746)	71480(9401)
Unique reflections	59936 (3836)	51733 (3739)	9983 (1573)	19197 (1661)	12021(1528)
Multiplicity	9.7 (7.7)	9.9 (9.7)	6.2 (6.3)	3.6 (2.9)	5.9(2.6)
Completeness (%)	99.9 (99.5)	99.9 (98.5)	99.5 (99.8)	98.1 (93.8)	99.6(98.0)
Mean I/sigma(I)	17.2 (1.8)	15.0 (2.3)	13.4 (3.7)	11.3 (2.1)	11.9(1.9)
Wilson B-factor	27.14	35.27	56.20	55.67	57.23
R_{merge}	0.08 (1.2)	0.07 (1.2)	0.09 (0.45)	0.07 (0.43)	0.09(0.95)
R_{meas}	0.09 (1.3)	0.08 (1.2)	0.09 (0.49)	0.08 (0.52)	0.104(1.04)
R_{pim}	0.03 (0.56)	0.03 (0.39)	0.04 (0.20)	0.04 (0.28)	0.04(0.41)
CC1/2	0.9 (0.8)	0.9 (0.9)	0.9 (0.9)	0.9 (0.3)	0.9(0.8)
Refinement Statistics					
Reflections used in refinement	59897 (5917)	51682 (5110)	9975 (977)	17929 (1706)	11928(1175)
Reflections used for R-free	2997 (321)	2624 (230)	471 (31)	909 (96)	572(57)
R-work	0.196 (0.347)	0.192 (0.298)	0.194 (0.345)	0.217 (0.332)	0.214(0.323)
R-free	0.230 (0.394)	0.229 (0.342)	0.243 (0.528)	0.266 (0.362)	0.276(0.435)
Number of non-hydrogen atoms	4730	4730	2180	2053	2234
macromolecules	4374	4325	2151	1982	2175
ligands	14	64	22	16	47
solvent	342	341	7	55	12
Protein residues	584	584	291	269	290
RMS (bonds)	0.006	0.007	0.009	0.008	0.008
RMS (angles)	0.74	0.83	1.06	0.93	0.96
Ramachandran favored (%)	98.28	98.10	93.77	95.85	93.75
Ramachandran allowed (%)	1.7	1.9	5.88	4.2	6.3
Ramachandran outliers (%)	0.00	0.00	0.35	0.00	0.00
Rotamer outliers (%)	0.00	0.68	0.00	0.00	0.00
Clashscore	2.7	3.3	8.7	5.9	7.0
Average B-factor	38.6	43.7	63.4	44.6	64.8
macromolecules	38.2	43.2	63.5	44.7	64.7
ligands	63.9	54.1	57.1	32.5	69.7
solvent	42.6	47.7	49.5	42.4	57.8

Table 4.4. Crystallographic data collection and refinement statistics of NanKs

4.3.6 Structural analysis of Pm-NanK and Hi-NanK shows that ManNAc binding induces conformational change in these enzymes

Similar to the previously reported bacterial and hMNK structures (22, 24, 25), the structures reported here from *H. influenzae* and *P. multocida* contain N- and C- terminal domains, composed of two α/β domains and connected by hinge regions. The monomeric structure is 'V' shaped. The C-terminal domain of NanK possess the dimerization domain and the resulting dimeric structure is shaped like a butterfly. N-terminal domain of *Hi*-NanK consists of residues 1 to 117 and residues 270 to 291 of C-terminus. These are arranged into 5 long and 3 short β-strands, sandwiched between 4 α-helices. The C-terminal domain contains residues from 118 to 269 and arranged in 4 α-helices and 4 β-strands, sandwiched between N- and C-terminal domains, respectively. *Pm*-NanK and *Hi*-NanK display similar structural folds and the α-helices are numbered from α1 to α11 and β-sheets are numbered from β1 to β12 (Figure 4.8). The crystal structures of *Pm*-NanK unliganded and *Pm*-NanK:AMPPNP bound form shows open conformations, whereas *Pm*-NanK:ManNAc and *Hi*-NanK:ManNAc structures exist in a closed conformation. These results demonstrate that upon ManNAc binding, N-terminal domain closes over ManNAc to trap the substrate, resulting in the closed conformation of NanK (Figure 4.6). The RMS deviation after superposition of *Pm*-NanK:AMPPNP and *Pm*-NanK:ManNAc structures is 2.6Å for 2024 Cα atoms. The RMS deviation after superposition of *Pm*-NanK:AMPPNP and HiNanK:ManNAc-6P structures is 2.8 Å for 1898 C α atoms.

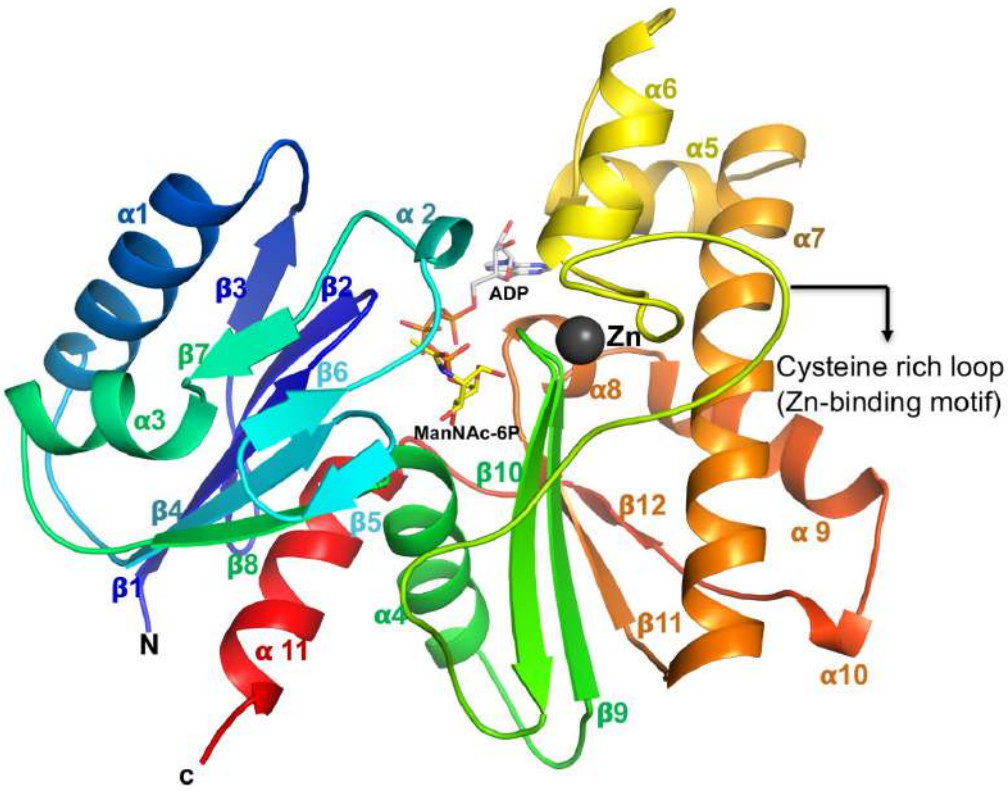


Figure 4.8. Cartoon representation of *Hi*-NanK in complex with its products ManNAc-6P and ADP. The structure of *Hi*-NanK is in cartoon style and it is colored in rainbow format from blue at N-terminus to red at C-terminus. Secondary structural elements are labelled as ***α*** and **β**. MaNAc-6P and ADP are shown in ball and stick models in yellow and gray colors, respectively. Zn is colored as a gray color sphere and the Zn-binding motif is labelled in the figure. The figures are generated using PyMol (Schrödinger) (40)

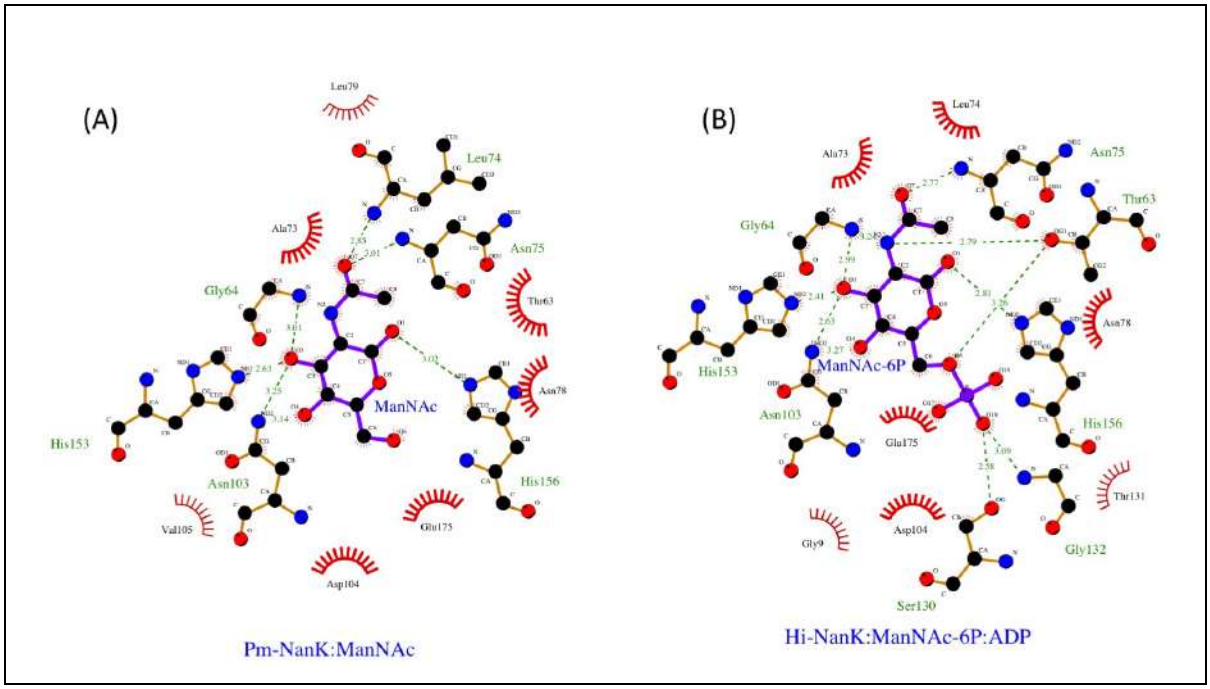
4.3.7 The ManNAc binding site

The binding pocket for ManNAc is buried deep inside the cleft region between N and C-terminal domains. Electron density difference map, F_O-F_C and $2F_O$-F_C shows the presence of ManNAc in the binding pocket and it appears in a chair conformation in the complex structures. Binding sites in both *Pm*-NanK and *Hi*-NanK are very well conserved and majority of the residues contributing to substrate binding are from the N-terminal domain. Only His153, His156, and Glu175 from the C-terminal domain play a role in substrate binding. Both His153 and His156 are located on the long loop that connects β10 and α5 helix, whereas Glu175 is located on the α5 helix. Crystal structure shows that the hydroxyl group at C1 position forms hydrogen bonds with His156 and Glu175. His153 and Gly64 form hydrogen bonds with the hydroxyl group at C3 position. Similarly, Asn103 and Asp104 form hydrogen bonds with the

hydroxyl group at C4 position, while Leu74 and Asn75 form hydrogen bonds with the hydroxyl group at C7 position. Phosphorylation occurs at the hydroxyl group at C6 position, which interacts with the catalytic Asp104 (Figure 4.9A, B).

4.3.8 The nucleotide binding domain in NanKs

The holo form of *Pm*-NanK adopts an open conformation and the majority of the interactions are with the C-terminal domain. The (F_O-F_C) and ($2F_O$-F_C) maps show the presence of AMPPNP in the binding pocket and it is located towards the C-terminal domain. In *Pm*-NanK holo structure the $\boldsymbol{\gamma}$-phosphate of AMPPNP shows hydrogen bonding interaction with Ser130. Additionally, adenine and ribose moiety of AMPPNP are sandwiched between residues Gly180, Pro196, Phe200, Ala246, Ser242 and Val243 from C-terminal domain. In the closed conformation of *Hi*-NanK:ManNAc-6P:ADP structure, pyrophosphate makes extensive interactions with residues from both the N- and C-terminal domains. In this structure, apart from the residues of C-terminal domain, the main chain of Lys12 forms hydrogen bond with **β**-phosphate of ADP (Figure 4.9C, D). Based on the distances, in the open conformation of *Pm*-NanK:holo, Thr131 forms weak hydrogen bond with $\boldsymbol{\gamma}$ phosphate of AMPPNP, whereas in a closed conformation of Hi-NanK:ManNAc-6P:ADP, it forms a strong hydrogen bond with β phosphate of ADP.



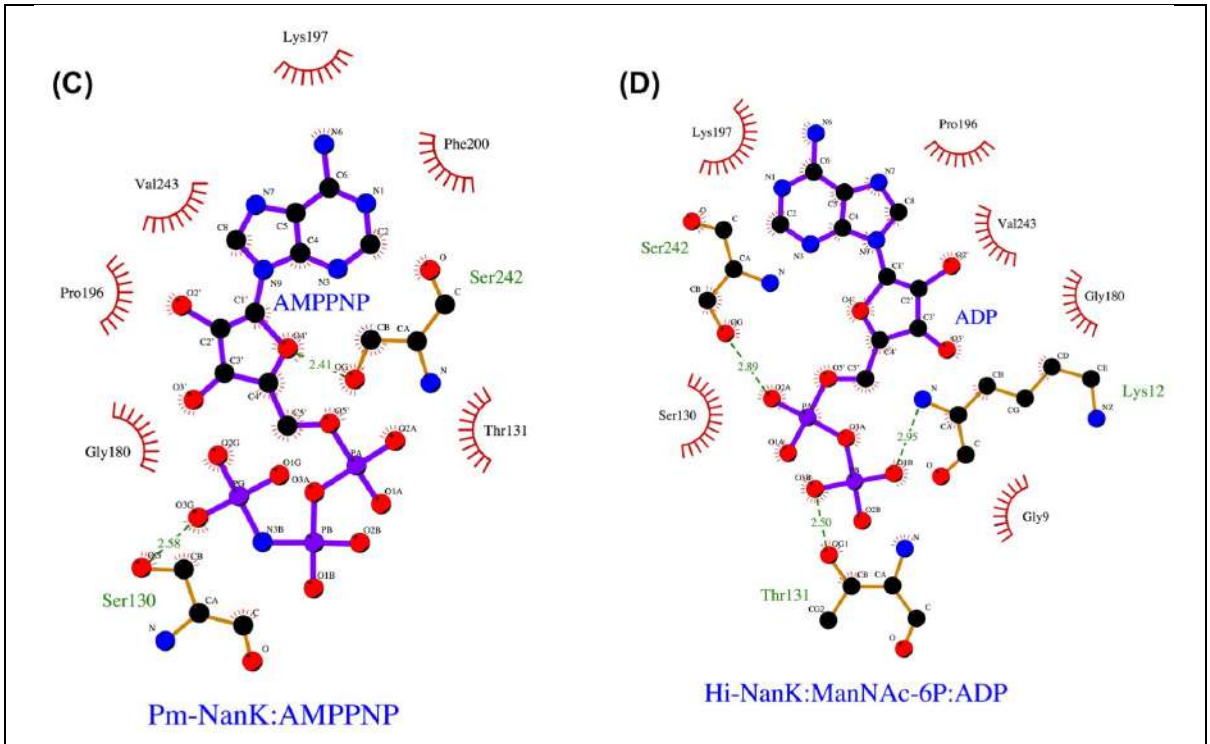


Figure 4.9. LigPlot representation showing the differences in amino acid interactions upon ligand (substrates and products) binding in *Hi*-NanK and *Pm*-NanK. (A) ManNAc binding pocket from Pm-NanK:ManNAc, (B) ManNAc-6P binding pocket from Hi-NanK:ManNAc-6P:ADP, (C) AMPPNP binding pocket from Pm-NanK:holo, and (D) ADP binding pocket from Hi-NanK:ManNAc-6P:ADP.

4.3.9 Domain rotation analysis using Dyndom

We have analyzed the conformational changes of the open and closed forms of *Pm*-NanK:AMPPNP and *Pm*-NanK-ManNAC using Dyndom (28). Dyndom conformational analysis considers N-terminal domain containing residues from 1-63 and 277-end as moving domain, whereas the C-terminal domain containing amino acids 64-276 as fixed domain. The analysis shows that, upon ManNAc binding, there are global changes in the structure of the enzyme, where the N-terminal domain rotates by 22.8^0 towards the C-terminal domain. This is a significantly larger movement compared to hMNK, where only a 12^0 rotation has been reported (22). The analysis also shows residues 64-66, 70-83, 104-108 and 270-278 as hinge regions, while residues 104-108 and 270-278 are located between the two domains. Further, Dyndom conformational analysis of *Pm*-NanK:AMPPNP and *Hi*-NanK:ManNAc-6P structures suggests similar hinge regions.

4.3.10 Thr131 is a catalytically important residue

The structural analysis of *Pm*-NanK*:holo* and *Hi*-NanKs:ManNAc-6P:ADP shows hydrogen bonding interactions between Thr131 and **γ**- and **β**-phosphates of AMPPNP and ADP and it is located in the loop between **β**9 and **β**10. Sequence and structural alignment of NanK enzymes also shows that this residue is conserved among the different NanK enzymes (Figure 4.2). Based on these results, we hypothesized that Thr131 plays a catalytic role during the phosphorylation event. To test this hypothesis, we cloned, over-expressed and purified *Hi*-NanK T131V mutant and studied its thermodynamic and kinetic properties in comparison to the wild-type (WT) NanK. Surprisingly, the *Hi*-NanK T131V mutant does not show any change in its thermodynamic properties and the binding affinities of *Hi*-NanK T131V to ManNAc and AMPPNP are very similar to *Hi*-NanK WT (Figure 4.10). However, the kinetic studies illustrate that the *Hi*-NanK mutant enzyme appears to have lost its enzymatic activity and is similar to a kinase dead mutant. Overall, these results suggest that Thr131 mutation did not affect the overall fold of the enzyme or its ability to bind substrates, but it plays an important role during substrate phosphorylation by positioning/transferring the **γ**-phosphate from ATP.

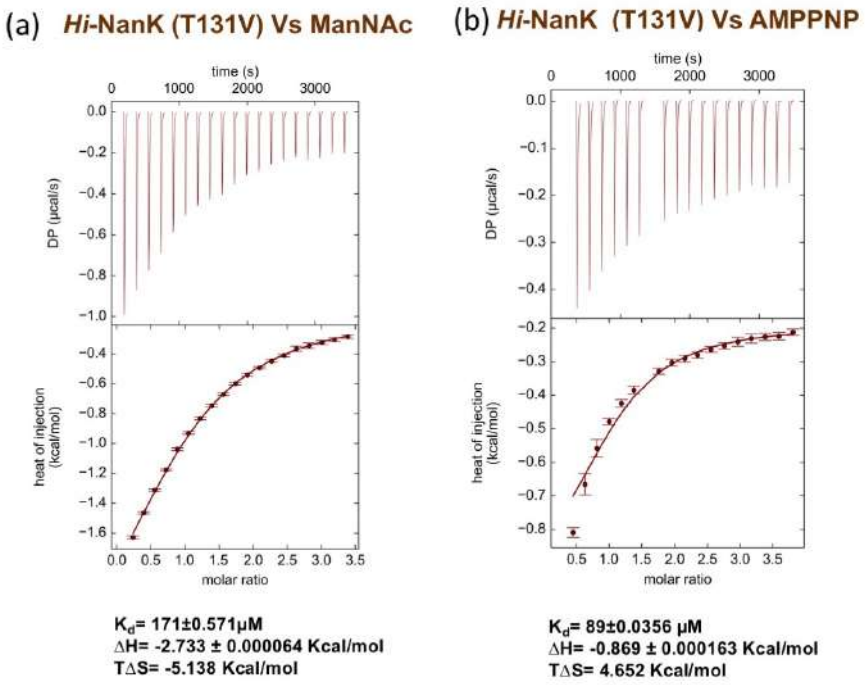


Figure 4.10. Thermodynamic studies on *Hi*-NanK (T131V) NanK. In the chromatograms, top panel shows the raw data and the bottom panel shows the integrated curves. Micro-calorimetric titrations of the *Hi*-NanK T131V mutant to its substrates (a) ManNAc and (b) AMPPNP show that the mutations did not change its binding properties.

4.3.11 Mutational analysis of residues in the hinge region

We analyzed the structures of *Pm*-NanK and *Hi*-NanK in order to locate residues that could be involved in regulating ordered or random binding patterns of the enzymes. Since the residues in the binding pockets are very well conserved, we hypothesized that residues in the hinge regions might be playing an important role in the differential binding properties of these two enzymes. To test this hypothesis, we cloned, over-expressed and purified the mutants *Pm*-NanK D115L (mutation in the α-helix between N- and C- terminal domains), *Pm*-NanK H273F (mutation in the hinge region, located in the loop connecting last β-12 strand to last α-11 helix), and the double mutant *Pm*-NanK D115L/H273F (marked in magenta box – Figure 2, in the hinge region). Further, these mutants were tested by ITC to check if the random binding pattern of *Hi*-NanK can be achieved in *Pm*-NanK mutants. Surprisingly, no changes were observed in the binding properties of *Pm*-NanK mutants.

4.3.12 Zn helps in positioning of the substrate in the binding pocket

Although, Zn salts were not added during the protein purification or crystallization experiments, except in *Pm*-NanK holo, the other refined structures of *Pm*-NanK and *Hi*-NanK shows density for Zn ion, which is positioned closer to the ManNAc binding pocket. We did try modeling other ions like Mg^{2+}, but they either did not satisfy the coordination or the electron density during refinement. Similar to the previously reported NanK structures ((22, 24), PDB ID: 2AA4), tetrahedral coordination of Zn is satisfied by 3 cysteines (Cys166, Cys168, and Cys173) and one His156 from C-terminal region (numbering corresponds to *Hi*-NanK), that are positioned approximately at a distance of 2.3 Å. These amino acids are located in the long loop that connects β10 and α5 helix and in the α5 helix of the C-terminal domain. In all these structures, the B factors for Zn^{2+} are similar to B factors of its interacting cysteines and histidine, which corresponds to full occupancy for Zn^{2+} in all the structures. In addition to its interactions with Zn in the Zn-binding motif, His156 also forms hydrogen bond with the hydroxyl group at C1 position of ManNAc. These structural details clearly illustrate that Zn plays an important role in proper positioning of His156, which further helps in the positioning of ManNAc in the active site (Figure 4.11). This analysis also shows the significance of Zn in the structural integrity of the NanK enzymes.

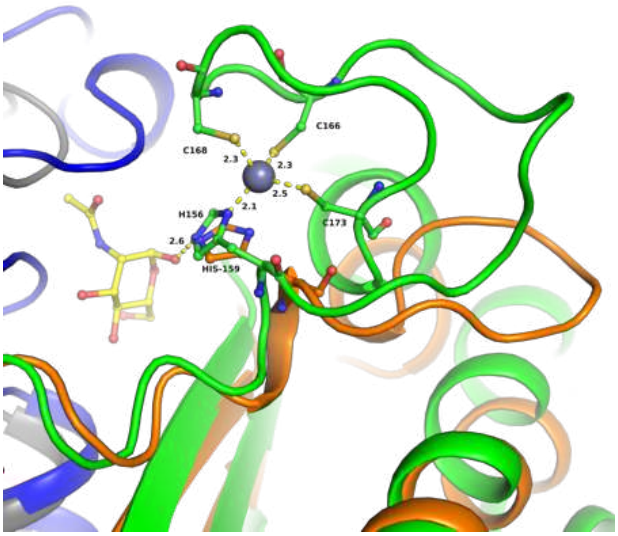


Figure 4.11. Superposition of *Hi*-NanK and *Fn*-NanK structures clearly shows the absence of Zn-binding motif in *Fn*-NanK. In *Hi*-NanK (green color), metal ion Zn is coordinated by 3 cysteines and one His156. His156 in *Hi*-NanK can form bond simultaneously with both ManNAc and Zn. In contrast, *Fn*-NanK shows the presence of only His159 and the absence of Zn-binding motif.

4.4 DISCUSSION

N-acetyl mannosamine kinases are involved in the anabolic and catabolic pathways of sialic acids in the human and bacteria, respectively. These are the members of ROK superfamily of enzymes, and most of these kinases have a conserved Zn-binding motif, where Zn is coordinated by one histidine and three cysteines. The Zn-binding motif containing zinc ions are known to play an important role in structural integrity and catalytic activity of enzymes. Previous structural studies on hMNK and *E. coli* NanK (PDB ID: 2AA4) enzymes demonstrate the presence of the conserved motifs specific to ROK family enzymes in their structures (22). Surprisingly, NanK from *F. nucleatum* only has His153 and not the three cysteines, resulting in the absence of the Zn-binding motif and zinc ion. Previous functional studies on ROK family of enzymes - transcriptional repressor Mlc from *E. coli* and GlcK from *Bacillus subtilis* has revealed that the mutations in cysteine residues of the Zn-binding motif lead to either lack or reduced enzyme activity (29, 30). Amino acid insertions, mutations or deletions may lead to new structural and functional properties of the proteins and they are very significant in protein evolution (31, 32). Nevertheless, the previous structural studies on *Fn*-NanK illustrated that in

spite of absence of the Zn-binding motif, the three dimensional structure is not affected (25). However, the effect on the functional properties of *Fn*-NanK due to the absence of divalent metal Zn and Zn-binding motif are unknown.

To understand the significance of Zn-binding motif, we have carried out functional studies on *Hi*-NanK, *Pm*-NanK and *Vc*-NanK (Zn-binding NanKs), and *Fn*-NanK (non-Zn-binding NanK) binding to its substrates ManNAc and AMPPNP and then compared their properties. Interestingly, the K_m of *Fn*-NanK is only 5-6 times higher compared to the other NanKs, while the K_{cat}, V_{max}, and catalytic efficiency of all the NanK enzymes are similar to each other. Further, mass spectrometry results show that even in the absence of the Zn-binding motif, *Fn*-NanK can phosphorylate the substrates to form products similar to *Pm*-NanK. These functional studies indicate that the phosphorylation function of *Fn*-NanK is not hampered by the absence of the Zn-binding motif.

On the other hand, our thermodynamic studies by ITC show that, *Pm*-NanK and *Vc*-NanK follow an ordered binding pattern, whereas *Hi*-NanK and *Fn*-NanK follow a random binding pattern to their substrates ManNAc and AMPPNP. Interestingly, sequence alignment shows that *Hi*-NanK and *Pm*-NanK are 75% identical but they show difference in their substrate binding characteristics. In contrast, although *Fn*-NanK lacks the Zn-binding motif and shares only 25% sequence similarity to *Hi*-NanK, they both follow similar binding patterns. These results suggest that Zn-binding motif is not a crucial factor for the order of binding.

To further understand the role of Zn- binding motif and to explore the structural reasons responsible for differential binding properties, we carried out structural studies on NanKs from *H. inflenzae* and *P. multocida*. Our structural studies reveal that even though there is less sequence similarity between different NanK enzymes, they have several common features: (a) the overall structural folds are very similar, (b) they are two domain proteins and undergo conformational change only in the presence of ManNAc, (c) ManNAc makes extensive interactions with the amino acids in the N-terminal domain and it is deeply buried inside the binding pocket between N- and C- terminal domains, and (d) AMPPNP shows extensive interactions with amino acids in the C-terminal domain of NanK.

Superimposing the putative ManNAc binding pocket structures of *Hi/Pm*-NanK, hMNK, and *Fn*-NanK shows very well conserved nucleotide and ManNAc binding pocket with minor differences. Since the C-terminal domain is the dimerization domain, we fixed C-terminal domain and superposed *Pm*-NanK:AMPPNP (open conformation) with the *Hi*-NanK-ManNAc-6p (closed conformation) structure. Superposition analysis shows that, upon ManNAc binding, the loop between β1 and β2 sheets containing ATP binding motif DIGG

makes a huge movement and moves closer to the AMPPNP, which further facilitates the phosphorylation event. A representative example is Gly9 (from DIGG motif of ATP binding loop) which moves about 5.8 Å in the ManNAc bound conformation compared to the open conformation - this brings it closer to the ADP. Additionally, in the open conformation of *Pm*-NanK:holo, Lys12 (present in the ATP binding loop) is approximately 10 Å away from the β-phosphate of AMPPNP and in the closed conformation of *Hi*-NanK, Lys12 gets positioned 2.9 Å away from the β-phosphate of ADP, which helps hydrogen bond formation (Figure 4.12). Previous structural studies of hMNK have shown that Mg^{2+} ion coordinates axially with both β-phosphate of ADP and Asp413 in the active center and is crucial for ADP binding (22). Unexpectedly, density for Mg^{2+} is not observed in any of the structures reported here. Notably, the overall distance between Asp7 and β-phosphate-oxygen of ADP from *Hi*-NanK:ManNAc-6P:ADP is similar to the distance between Asp413 and β-phosphate-oxygen of ADP from hMNK. This analysis further demonstrates that Mg^{2+} may not be necessary for binding of the substrates and is required for correct orientation of ATP and ManNAc for the phosphorylation event.

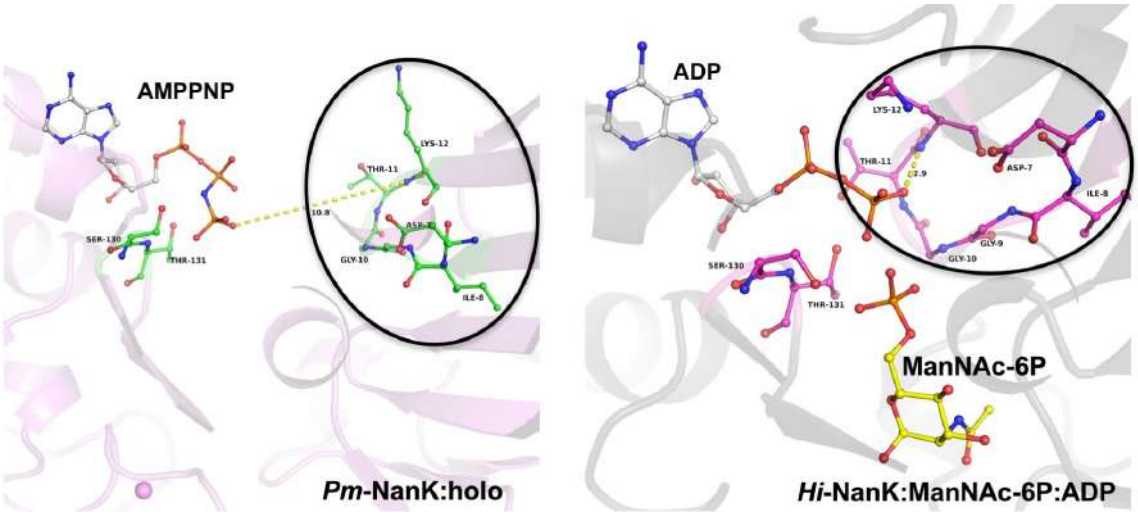


Figure 4.12. Close-up view of the amino acids interacting with the nucleotides in *Pm*-NanK (open conformation) and *Hi*-NanK (closed conformation). The residues interacting with AMPPNP and ADP are shown in ball and stick model from the open (green color) and closed (magenta color) conformations of *Pm*-NanK:holo and *Hi*-NanK:ManNAc6P-ADP complex, respectively. The close-up view shows that in the ManNAc bound closed conformation, there is a significant movement of the ATP binding motif in order to interact with ATP, which further facilitates the phosphorylation. The figures are generated using PyMol (Schrödinger) (40).

A closer look at the structure shows that Thr131 forms hydrogen bonds with **β**-phosphate of the nucleotide in open and closed conformations of NanK. *Hi*-NanK:ManNAc-6P:ADP structure illustrates that **β**-phosphate is positioned in between Thr11 (from ATP binding motif loop) and Thr131. This is similar to hMNK structure wherein **β**-phosphate gets positioned in between Thr417 and Thr544 residues. Our functional studies of *Hi*-NanK T131V suggests that it is an important residue for enzyme catalysis. Summary of all the above interactions are shown in Figure 4.13.

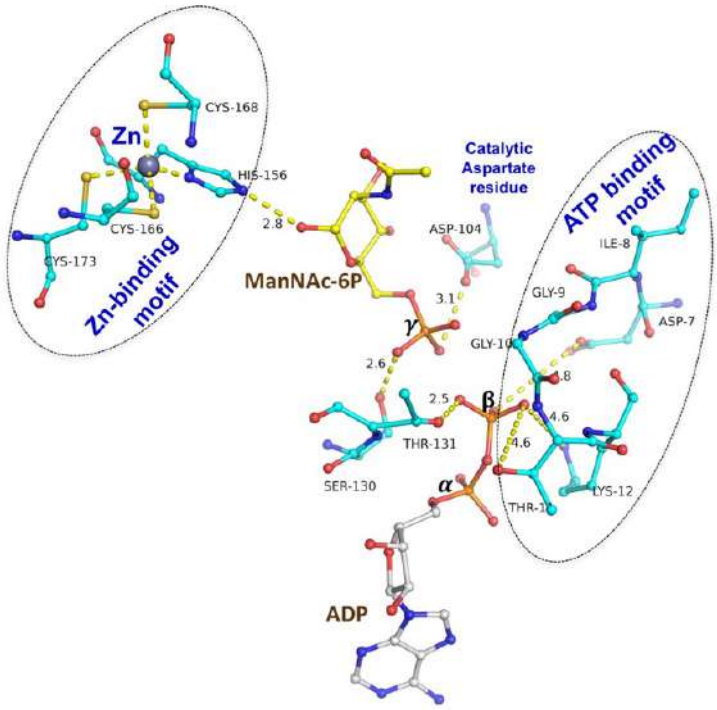


Figure 4.13. Cartoon representation of *Hi*-NanK: ManNAc-6P:ADP showing the interactions of the residues in the binding pocket with ManNAc-6P and ADP.

Structural superposition of *Fn*-NanK, *Hi*-NanK and *Pm*-NanK structures shows that, in all the three structures, histidine from ExGH motif is holding the substrate in position at the active site for enzyme activity. Additionally, in case of Zn-binding NanKs the position of His156 (numbering corresponds to *Hi*-NanK) is maintained by Zn, which is further coordinated by Zn-binding motif containing 3 cysteines. But, the closure examination of the structural details of the *Fn*-NanK shows that, though it lacks the Zn, Zn-binding motif/cysteine rich loop, the position of ManNAc is maintained by His159. Further, the position of shorter loop holding the His159 is maintained by a combination of hydrophobic and hydrogen bonding interactions

of Phe167 which is present on the helix following this short loop. The side chains of Phe167 is held tightly in the hydrophobic pocket formed by Ile136, Ile160, Ala171 and Gly218. Further, main chain amino group of Phe167 makes hydrogen bonds with main chain carbonyl residues of Gly158 and Ile160.

Based on these functional and structural characterization studies of Zn-binding and non Zn-binding NanK enyzmes here we hypothesize that, His159 in *Fn*-NanK alone might be sufficient to orient ManNAc in the active site to preserve the structural and functional properties of *Fn*-NanK as a kinase.

Additionally, *Pm*-NanK hinge region mutants did not show any change in the binding properties to its substrates, so the structural reasons for these differential binding properties of the NanK enzymes has to be addressed in future.

In summary, our work shows that in spite of low sequence similarity among NanK enzymes, different pathogenic bacteria have evolved to carry out similar function of phosphorylation by conserving most of their signature motifs. The presence or absence of Zn does not seem to make a difference for the overall structure or the activity. NanK structural and functional studies from four uncharacterized pathogenic bacteria and their comparison with the previously reported NanK enzymes, paves the way for elucidating the molecular level details of these kinases in the sialic acid catabolic pathway.

4.5 Experimental procedures

4.5.1 Cloning and expression of Hi-NanK, Pm-NanK, Fn-NanK and Vc-NanK

The genes corresponding to NanK from *Fusobacterium nucleatum (Fn*-NanK: NCBI reference sequence no: WP_011017180.1), *Haemophilus influenza (Hi*-NanK; NCBI reference sequence no: WP_011271901.1)*, Pasteurella multocida (Pm*-NanK; NCBI reference sequence no: WP_005752223.1), *and Vibrio cholera (Vc*-NanK: NCBI reference sequence no: WP_001259414.1) were synthesized from GeneArt Gene Synthesis (Thermo Fisher Scientific). These genes were cloned into pET300/NT-DEST using GATEWAY cloning technology (33). Further, these plasmids were transformed into BL21(DE3)* cells for protein expression. *Hi*-NanK T131V, *Pm*-NanK D115L, *Pm*-NanK H273F, and *Pm*-NanK D115L H273F substitutions were generated by site directed mutagenesis and these substitutions were further confirmed by DNA sequencing. The cells were grown at 37 ^{0}C to an absorbance of 0.5-0.6 at 600 nm, induced with 0.5 mM IPTG and further grown at 20^{0}C temperature overnight. Following day, cells were centrifuged at 13000 rpm for 30 min and the pellet suspended in

lysis buffer (50 mM phosphate buffer pH 7.4, 300 mM NaCl, 10 mM imidazole and protease inhibitor cocktail without EDTA (Roche)).

4.5.2 Purification of Hi-NanK, Pm-NanK, and Vc-NanK

The cells were lysed using EmulsiFlex-C3 and the lysate was centrifuged at 13000 rpm for 30min. The supernatant was loaded on to an equilibrated Ni-NTA column for affinity purification. Post binding, the beads were washed with 5 column volumes (CVs) of lysis buffer and then with 10 CVs of wash buffer 1 (50 mM phosphate buffer pH 7.4, 500 mM NaCl, 20 mM imidazole), followed by 10 CVs of wash buffer 2 (50 mM phosphate buffer pH 7.4, 500 mM NaCl, 50 mM imidazole). Subsequently, the protein was eluted in buffer containing 50 mM phosphate buffer 7.4, 300 mM NaCl, and 250 mM imidazole. Purity of the purified protein was confirmed by SDS-PAGE gel electrophoresis. The protein fractions were pooled and purified using size exclusion chromatography (superdex 200, GE Healthcare). The column was pre-equilibrated with a buffer containing 20 mM Tris-Cl pH 7.5, 50 mM NaCl, 5% glycerol, and 5 mM DTT. The purity of these samples was validated using SDS-PAGE gel electrophoresis. The purification of *Fn*-NanK was carried out using the protocol as described previously (25). Finally, the fractions were pooled, concentrated and used for thermodynamic and kinetic experimentation as well as structural studies.

4.5.3 Kinetic enzymatic assays

Kinase activities for different NanKs were carried out using ADP-Glo kinase assay kit (Promega). The reaction was performed in a buffer containing 40 mM Tris-Cl pH 7.5, 20 mM $MgCl_2$, and 0.1 mg/ml BSA. The kinase assay with ManNAc was performed using 2 ng of enzyme, 100 μM ATP and varying concentrations of ManNAc. The reactions were incubated at room temperature for 15 min. Reagent 1 was added to 5 μl of reaction mixture to deplete the unutilized ATP and the reaction was further continued for 40 min. Next, reagent 2 was added to the reaction mixture that converts ADP (liberated during phosphorylation reaction) to ATP, which further gets converted into light by luciferase. The luminescence was measured post 1 h incubation using a plate reader (Tecan). The data was analyzed and the apparent K_m values for ManNAc was calculated by nonlinear fitting of the data into Michaelis-Menten equation. Each point represents the average of three independent samples and the curves were plotted using GraphPad Prism (version 7.0b). These kinetic enzymatic assays were performed in duplicates with technical triplicates.

4.5.4 Analysis of ManNAc and ManNAc-6P by Mass spectrometry

To measure the amount of the substrate and product formed after the NanK enzymatic activity, three different concentrations (1, 100 and 1000 ng) of *Pm*-NanK and *Fn*-NanK enzymes were used for each reaction and the protocol as described previously was followed (27). Briefly, *Fn*-NanK or *Pm*-NanK at a final concentration of 1, 100, and 1000 ng were added to 50 µl of the reaction mixture containing 50 mM Tris-HCl, pH 8.0, 5 mM ManNAc, 10 mM ATP, and 10 mM $MgCl_2$. These reactions were carried out at 310K for 20 min and stopped by adding TCA to a 10% final concentration, followed by placing the reaction mixture on ice for 10 min. These reactions were centrifuged at 13000 rpm for 10 min, the supernatants were collected and subjected to mass spectrometry analysis. The 10 µl supernatant was diluted with 90 µl MS grade water in a low protein binding microcentrifuge tube. Subsequently, freshly made 100 µl of 25 mM 3-amino-9-ethylcarbazole and 50 µl of 50 mM NaCNBH3 were added to the supernatant. The reaction mixtures were further incubated at 70^0C for 60 min and then kept on ice for 1 min. Next, 300 µl of dichloromethane:hexane (2:1) mixture was added to the above 300 µl of MS grade water. These reaction mixtures were vortexed and centrifuged at 10000 rpm for 5 min. Approximately, 300 µl of the upper aqueous phase was transferred to a fresh low protein binding micro centrifuge tube without disturbing the lower phase. Finally, the sample was further injected into TSQ Vantage-Agilent 1290 UHPLC(LC-MS) machine for MRMs of parent and daughter ions of both reactant and product at a constant injection rate using a Hamilton syringe with flowrate of 5 µl/min in positive ion mode. Daughter ion of 3-amino-9-ethylcarbazole characteristic (AEC) derivatization moiety of 210 Da from parent ion was used for quantification.

4.5.5 Isothermal calorimetry (ITC)

The binding affinities between NanK and its substrates, *N*-Acetyl mannosamine and AMP-PNP (ATP analogue) were measured using a MicroCal ITC system (Malvaren). The reactions were carried out in a buffer containing 20 mM Tris-HCl pH 8.0, 300 mM NaCl, 5% glycerol, and 1 mM DTT. The protein concentration was determined by measuring the absorbance at 280 nm. The protein concentration in the reaction cell was varied from 100 µM to 200 µM. Correspondingly, the syringe concentration used in the experiments was 10-20 times higher than the cell concentration. In Zn-binding and non Zn-binding NanKs, two different series of thermodynamic reactions were carried out to measure the binding affinities of NanKs to their substrates.

The first set of titrations was carried out between different NanKs and ManNAc in the presence and absence of AMPPNP. The second set of titrations were carried out between different NanKs and AMPPNP in the presence and absence of ManNAc. Concentrations of NanK, AMP-PNP and ManNAc were varied in the titrations and the experiments were performed in duplicates with technical triplicates. The non-specific heat released by the dilution of protein, nucleotide and substrate was calculated by averaging the heat liberated during the last 3–5 injections, post saturation. The values were further subtracted from the raw heat released during each injection, which eliminates the heat of dilution. The calorimetric data sets were further analyzed using Origin ITC analysis software (MicroCal., USA) for single site binding model. Nonlinear least square analysis was used to calculate the stoichiometry, enthalpy (ΔH), and binding affinity.

4.5.6 Protein crystallization and data collection

Pm-NanK

Post purification, the protein was concentrated to 20 mg/ml. Crystallization trays were set up by Mosquito (Nano drop liquid handling machine, TTP labtech) using hanging drop vapor diffusion method.

Pm-NanK unliganded: The crystals were obtained in a buffer containing 0.1 M Tacsimate pH 5.0, 20% v/v isopropanol, 20% w/v PEG 4000 with an additive containing 2 M NaCl.

Pm-NanK holo (with AMPPNP): The crystals were obtained in a buffer containing 0.1 M Tacsimate pH 5.0, 20% v/v isopropanol, 20% w/v PEG 4000 with an additive containing 100% v/v ethylene glycol. These crystals were further soaked for 7 days in the same buffer containing AMPPNP and ManNAc (however, the obtained crystal structure contains only AMPPNP).

Pm-NanK bound with ManNAc: The protein and ManNAc were mixed at 1:10 ratio and incubated on ice for 1 hour. The crystals were obtained in a buffer containing 0.2 M lithium citrate tribasic tetrahydrate and 20% w/v PEG 3350. *Pm*-NanK unliganded, *Pm*-NanK holo and *Pm*-NanK:ManNAc crystals were mounted in loops with glycerol as cryoprotectant and the x-ray diffraction data was collected at 100 K at the Proxima 1 beamline, SOLEIL synchrotron, France.

Hi-NanK

The crystallization trays were setup by Mosquito (Nano drop liquid handling machine, TTP labtech) using hanging drop vapor diffusion method.

Hi-NanK bound with ManNAc:

The crystal trays were setup at a protein concentration of 20 mg/ml. Further, the protein was incubated with ADP, $MgCl_2$, sodium ortho vandate, and ManNAc at 1:5:5:5 molar ratios at RT for 30 min. The crystals were obtained in a buffer containing 0.2 M sodium malonate pH 7.0 and 20% w/v PEG 3350.

Hi-NanK bound with products ManNAc-6P and ADP:

The crystals trays were setup at protein concentration of 21 mg/ml. The protein was further incubated with ATP, $MgCl_2$, and ManNAc at 1:10:10:20 molar ratio. The crystals were obtained in the buffer containing 0.2 M ammonium phosphate monobasic, 0.1 M Tris-Cl pH 8.5, and 50% v/v MPD.

Hi-NanK:ADP:ManNAc-6P crystals were mounted in loops and the x-ray diffraction data was collected at 100 K using the Proxima 1 beamline, SOLEIL synchrotron, France. *Hi*-NanK:ManNAc complex crystals were mounted in loops and x-ray diffraction data was collected at the ID29 beamline, ESRF, Grenoble.

4.5.7 Refinement of NanK protein structures

All structures were processed by using XDS (34) and scaled using aimless in the CCP4 program suite (35). The structure of *Pm*-NanK:ManNAc was determined by molecular replacement using *E. coli*-NanK structure (PDB-ID 2AA4) as the starting model. Further *Pm*-NanK:ManNAc structure was used as the search model for *Pm*-NanK apo and holo structures. The structure of *Hi*-NanK:ADP:ManNAc-6P complex were also determined by molecular replacement using the structure of *Pm*-NanK:ManNAc complex as starting model. Further, Hi-NanK:ADP:ManNAc-6P complex structure was used as a search model for the molecular replacement of *Hi*-NanK:ManNAc structure. Finally, manual model building was carried out using COOT (36) and further refinements and processing were performed using Phenix (37, 38). Atomic coordinates and structure factors for the 5 structures were deposited in Protein Data Bank (PDB) and their IDs are as follows: *Pm*-NanK:apo - 6JDH, *Pm*-NanK:holo – 6JDO, *Pm*-NanK:ManNAc - 6JDA, *Hi*-NanK:ManNAc-6P:ADP - 6JDB, and *Hi*-NanK:ManNAc - 6JDC. Crystallographic data for data collection and refinement are presented in Table 4.

4.6 ACCESSION NUMBERS

Structure factors and coordinates of Pm-NanK unliganded, Pm-NanK holo, Pm-NanK:ManNAc, Hi-NanK:ManNAc-6P:ADP, and Hi-NanK:ManNAc structures have been deposited in Protein Data Bank with accession codes 6JDH, 6JDO, 6JDA, 6JDB and 6JDC, respectively.

4.7 ACKNOWLEDGEMENTS

We thank the ESRF Access Program of RCB (supported by Department of Biotechnology Grant BT/INF/22/SP22660/2017). We thank SOLEIL synchrotron (Proxima 1 beamline) and ESRF (ID29 beamline) for providing the beam time for data collection. This research was supported by DBT-Indo Swedish Grant (BT/IN/SWEDEN/41/SR/2013); DBT-B-life grant (BT/PR5081/INF/156/2012); NCBS x-ray facility grant (BT/PR12422/MED /31/287/214) and a grant from the DST for SR. We thank Nitish Sathyanarayana and Dr. Swagatha Ghosh and Dr. Rhawnie Caing-Carlsson for their contributions towards the initial stages of the work; Dr. Chetan Arya for the help with the chemdraw figures of sialic acid pathway; Dr. Sanchari Banerjee for critical reading of the manuscript. T.G would like to acknowledge senior research fellowship support from Council of Scientific and Industrial Research (CSIR).

4.8 AUTHOR CONTRIBUTIONS

T.G., and S.R. designed the research; T.G. carried out most of the experiments; A.S. carried out the mass spectrometric studies; T.G and S.R. wrote the manuscript with the contribution from coauthors. D. C. and R.C.J. D contributed to analysing kinetic data, critical reading of the manuscript and discussion of the results.

4.9 References

1. Vallee, B. L., and Auld, D. S. (1990) Zinc coordination, Function, and Structure of Zinc Enzymes and Other Proteins. *Biochemistry* **29**
2. Bert L. Valle and Kenneth Falchuk, H. (2019) The Biochemical Basis of Zinc Physiology. *Physiol Rev* **73**, 79–118
3. Katayama, A., Tsujii, A., Wada, A., Nishino, T., and Ishihama, A. (2002) Systematic search for zinc-binding proteins in Escherichia coli. *Eur J Biochem* **2413**, 2403–2413
4. Malgieri, G., Palmieri, M., Russo, L., Fattorusso, R., Pedone, P. V, and Isernia, C. (2015) The prokaryotic zinc-finger : structure , function and comparison with the eukaryotic counterpart. *FEBS J.* **282**, 4480–4496
5. Baglivo, I., Russo, L., Esposito, S., Malgieri, G., Renda, M., Salluzzo, A., Di, B., Isernia, C., Fattorusso, R., and Pedone, P. V (2009) The structural role of the zinc ion can be dispensable in prokaryotic zinc-finger domains. *PNAS* **106**, 6933–6938
6. Coleman, J. E. (1992) ZINC PROTEINS : Enzymes , Storage Replication Proteins. *Annu Rev Biochem*, 897–946
7. Hambidge, M. (2000) Zinc and Health: Current status and future directions. *J. Nutr.* **130**, 1437S–1446S
8. Tanner, M. E. (2005) The enzymes of sialic acid biosynthesis. *Bioorg. Chem.* **33**, 216–228
9. Varki, A., and Gagneux, P. (2013) Multifarious roles of sialic acids in immunity. *Ann N Y Acd sci* **1253**, 16–36
10. Li, Y., and Chen, X. (2012) Sialic acid metabolism and sialyltransferases : natural functions and applications. , 887–905
11. Foxwell, A. R., Kyd, J. M., and Cripps, A. W. (1998) Nontypeable Haemophilus influenzae : Pathogenesis and Prevention. *Microbiol. Mol. Biol. Rev.* **62**, 294–308
12. Almagro-moreno, S., and Boyd, E. F. (2009) Sialic Acid Catabolism Confers a Competitive Advantage to Pathogenic Vibrio cholerae in the Mouse Intestine. *Infect. Immun.* **77**, 3807–3816
13. Severi, E., Hood, D. W., and Thomas, G. H. (2015) Sialic acid utilization by bacterial pathogens. *Microbiology* **153**, 2817–2822
14. Vimr, E. R., Kalivoda, K. A., Deszo, E. L., and Steenbergen, S. M. (2004) Diversity of Microbial Sialic Acid Metabolism. *Microbiol. Mol. Biol. Rev.* **68**, 132–153
15. Almagro-Moreno, S., and Boyd, E. F. (2009) Insights into the evolution of sialic acid catabolism among bacteria. *BMC Evol. Biol.* **9**, 118
16. Nocek, B., Stein, A. J., Jedrzejczak, R., Cuff, M. E., Li, H., Volkart, L., and Joachimiak, A. (2011) Structural studies of ROK fructokinase YdhR from Bacillus subtilis: Insights into substrate binding and fructose specificity. *J. Mol. Biol.* **406**, 325–342
17. Titgemeyer, F., Reizer, J., Reizer, A., and Saier, M. H. (1994) Evolutionary relationships between sugar kinases and transcriptional repressors in bacteria. *Microbiology* **140**, 2349–2354
18. Larion, M., Moore, L. B., Thompson, S. M., and Miller, B. G. Divergent Evolution of Function in the ROK Sugar Kinase Superfamily: Role of Enzyme Loops in Substrate Specificity.
19. Conejo, M. S., Thompson, S. M., and Miller, B. G. (2010) Evolutionary bases of carbohydrate recognition and substrate discrimination in the ROK protein family. *J. Mol. Evol.* **70**, 545–556

20. Matte, A., Tari, L. W., and Delbaere, L. T. J. (1998) How do kinases transfer phosphoryl groups ? *Structure* **6**, 413–419
21. Hunter, T. (1995) Protein kinases and phosphatases: The Yin and Yang of protein phosphorylation and signaling. *Cell* **80**, 225–236
22. Martinez, J., Nguyen, L. D., Hinderlich, S., Zimmer, R., Tauberger, E., Reutter, W., Saenger, W., Fan, H., and Moniot, S. (2012) Crystal structures of N-acetylmannosamine kinase provide insights into enzyme activity and inhibition. *J. Biol. Chem.* **287**, 13656–13665
23. Reinke, S. O., Lehmer, G., and Reutter, W. (2009) Regulation and pathophysiological implications of UDP-GlcNAc 2-epimerase / ManNAc kinase (GNE) as the key enzyme of sialic acid biosynthesis. *Biol chem* **390**, 591–599
24. Tong, Y., Tempel, W., Nedyalkova, L., MacKenzie, F., and Park, H. W. (2009) Crystal structure of the N-acetylmannosamine kinase domain of GNE. *PLoS One* **4**, 1–9
25. Caing-carlsson, R., Goyal, P., Sharma, A., Ghosh, S., Setty, T. G., North, R. A., Friemann, R., and Ramaswamy, S. (2017) Crystal structure of N -acetylmannosamine kinase from Fusobacterium nucleatum research communications. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **F73**, 356–362
26. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W : improving the sensitivity of progressive multiple sequence alignment through sequence weighting , position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680
27. Han, J., Tschernutter, V., Yang, J., Eckle, T., and Borchers, C. H. (2013) Analysis of selected sugars and sugar phosphates in mouse heart tissue by reductive amination and liquid chromatography-electrospray ionization mass spectrometry. *Anal. Chem.* **85**, 5965–5973
28. Hayward, S., and Berendsen, H. J. C. (1998) Systematic Analysis of Domain Motions in Proteins From Conformational Change : New Results on Citrate Synthase and T4 Lysozyme. *Proteins Struct. Funct. Genet.* **154**, 144–154
29. Gerber, K., Seitz, S., Welte, W., Diederichs, K., and Boos, W. (2005) The Crystal Structure of Mlc , a Global Regulator of Sugar Metabolism in Escherichia coli *. *J. Biol. Chem.* **280**, 29073–29079
30. Mesak, L. R., Mesak, F. M., and Dahl, M. K. (2004) Bacillus subtilis GlcK activity requires cysteines within a motif that discriminates microbial glucokinases into two lineages. *BMC Microbiol.* **4**, 1–10
31. Wolf, Y., Madej, T., Babenko, V., Shoemaker, B., and Panchenko, A. R. (2007) Long-term trends in evolution of indels in protein sequences. *BMC Evol. Biol.* **10**, 1–10
32. Zhang, Z., Huang, J., Wang, Z., Wang, L., and Gao, P. (2011) Impact of indels on the flanking regions in structural domains. *Mol. Biol. Evol.* **28**, 291–301
33. Bairy, S., Gopalan, L. N., Setty, T. G., Srinivasachari, S., Manjunath, L., Kumar, J. P., Guntupalli, S. R., Bose, S., Nayak, V., Ghosh, S., Sathyanarayanan, N., Caing-Carlsson, R., Wahlgren, W. Y., Friemann, R., Ramaswamy, S., and Neerathilingam, M. (2018) Automation aided optimization of cloning, expression and purification of enzymes of the bacterial sialic acid catabolic and sialylation pathways enzymes for structural studies. *Microb. Biotechnol.* **11**, 420–428
34. Kabsch, W. (2010) XDS. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 125–132
35. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D. Biol. Crystallogr.* **67**, 235–242

36. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486–501
37. Adams, P. D., Pavel, V., Chen, V. B., Ian, W., Echols, N., Moriarty, N. W., Read, R. J., Richardson, D. C., Jane, S., and Thomas, C. (2010) PHENIX : a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **D66**, 213–221
38. Adams, P. D., Ralf, W., Grosse-Kunstleve, Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) PHENIX : building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* **D58**, 1948–1954
39. Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, 320–324
40. DeLano, W. L. (2018) The PyMol Molecular Graphics System, version 2.1, Schrödinger, LLC, New York

Chapter 5

Future directions

Sialic acids are simple nine-carbon sugar acids that are decorated as the outermost sugars of host glycan molecules. To resist the host innate immune response, many pathogenic bacteria have evolved to adorn their cell surfaces with sialic acids in a process termed 'molecular mimicry'. In addition, these sugar acids are good food sources to bacteria as they provide carbon, nitrogen, and energy. A majority of the pathogenic bacteria live in mucus-rich regions of the host and share their niche with symbiotic bacteria. Sialidases secreted by commensal bacteria will help in cleaving the host sialic acids, which will then be taken up by pathogenic bacteria. This method allows the pathogen to colonize and is a good example of niche adaptation. Recent studies have shown that gut microbiota encounter a fierce competition for sialic acid Neu5Ac. Consequently, pathogenic bacteria have evolved with multiple routes to acquire these sugars from the host.

Pathogenic bacteria like *H. influenzae, F. nucleatum, P. multocida,* and *V. cholera* uses the TRAP transport system (SiaPQM) for the transport of sialic acids. The periplasmic sialic acid binding proteins (SiaP) of the TRAP transport system from various pathogenic bacteria have been well characterized (1)(2). Mainly, these proteins have a high affinity for substrate binding and enforce unidirectionality in substrate transfer. Unlike the other pathogenic bacteria, *H. ducreyi* uses the ABC transport system (SatABCD) for the transport of sialic acids. This thesis details the structural and functional characterization of the previously uncharacterized SatA from ABC transport system. Our structural studies on SatA show that sialic acid binds between two domains even though the protein contain three domains, similar to SiaP of TRAP transport system. In contrast to SiaP, our thermodynamic studies show that the sialic acid binding to SatA is an entropically driven process and protein possess more hydrophobic residues in the binding pocket (3).

Although the biophysical properties of SiaP and SatA are very well studied, the structural and functional properties of their membrane components SiaQM and SatBCD are not yet described/illustrated.

When compared to ABC transporters, TRAP transporters are ATP-independent and cotransport two Na+ ions along with the substrate. Metagenomic and metaproteomic studies show that TRAP transporters are more abundantly expressed in bacteria that live under oligotrophic and deep sea conditions, where the Na+ (high) and nutrient (low) availabilities are in reciprocal ratio (4). TRAP transporters contain three subunits, the periplasmic substrate binding protein and the small and large transmembrane domains (TMDs). Based on previous

biochemical studies, the following five-step model has been hypothesized for sialic acid transport through the TRAP transport system (Figure. 5.1) (5).

1. The binding of the ligand induces a conformational change in the SiaP, so that it can close upon the substrate.
2. The ligand-bound SiaP binds to/interacts with the TMDs (called the resting stage).
3. After the docking of SiaP onto TMDs, two Na+ ions are translocated and this leads to the outside open conformation of TMDs.
4. The outside open conformation of TMDs will allow the release of the substrate from SiaP into the translocation channel.
5. Subsequently, TMDs change their conformation from the outside-open to outside-closed, which enables release of the substrate and Na+ ions into the bacterial cytoplasm.

Interestingly, previous studies also reported that the binding of the unliganded SiaP to TMDs will possibly export the substrates from the cytoplasm (6). Future studies on the structural and functional characterization of these membrane proteins will provide a better understanding of their export and import mechanisms.

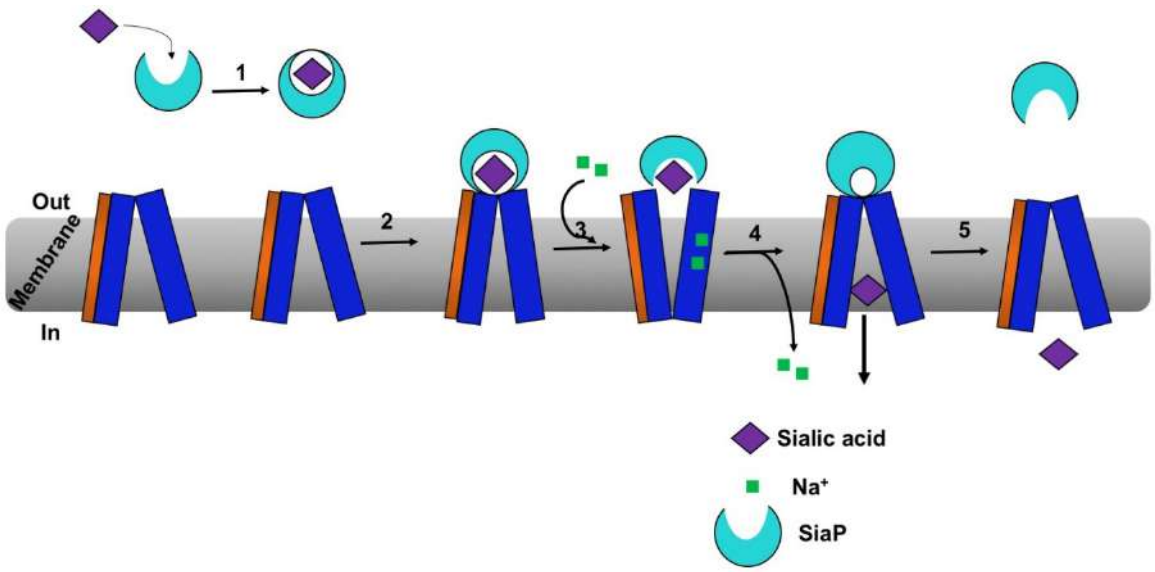


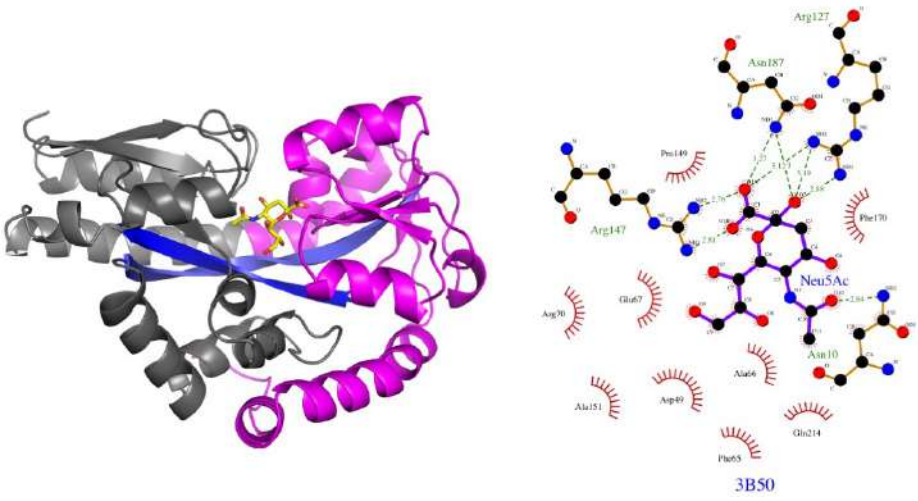
Figure. 5.1. Hypothetical model on sialic acid transport through TRAP transport system. (1) Binding of ligand to SiaP, (2) Docking of ligand bound SiaP on SiaQM, (3) Translocation of Na+ ions, (4) Release of substrate into the translocation channel, and (5) Release of substrate into the cytoplasm. Figure is adapted from Ref. (5)

Similarly, though several structures of the ABC transport system that transport a wide range of substrates have been previously reported, the structure of sialic acid binding protein

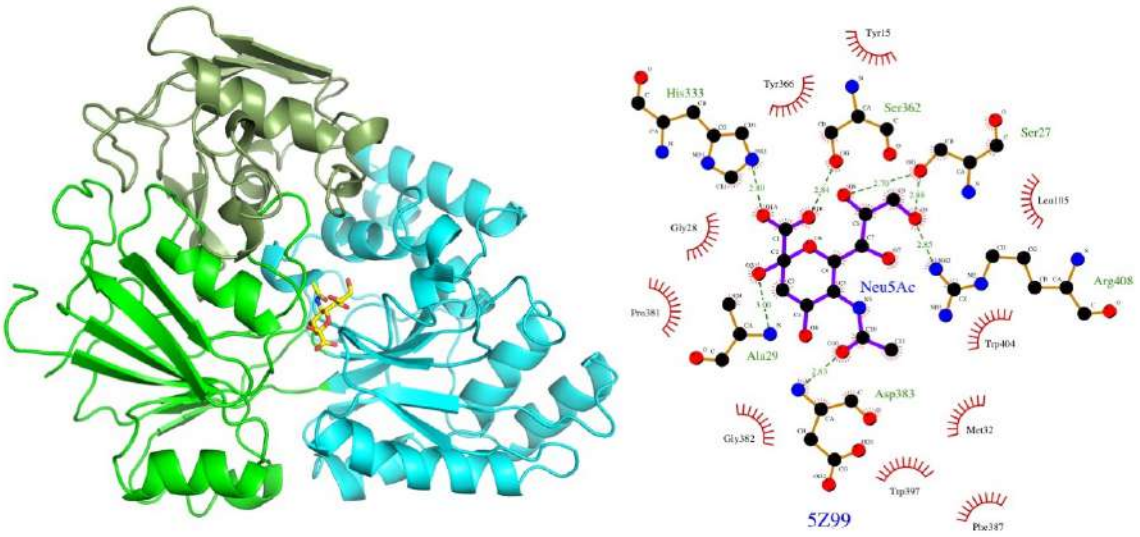
(SatA) along with that of SatBCD has not yet been solved. In summary, studies that focus on the structural and functional characterization of the membrane components (SiaQM and SatBCD) will aid in the identification of differences between these two transport systems with regard to ligand binding and transport mechanisms.

Post internalization of Neu5Ac, part of the sialic acid is incorporated as the outermost sugar, while the remaining sialic acid goes into the catabolic pathway. SiaB and NanA are the first enzymes in these two pathways and they convert Neu5Ac into CMP-Neu5AC and ManNAc, respectively. There are several biophysical studies on SiaB and NanA from different organisms (7–10). The structural information indicates that the binding pockets of SiaB and NanA are different from that of SiaP and SatA(Figure. 5.2). However, it will be useful to find a single drug molecule that can target these different sialic acid binding proteins (SiaP, SatA, SiaB, and NanA) and may act as an effective drug against the different pathogenic bacteria.

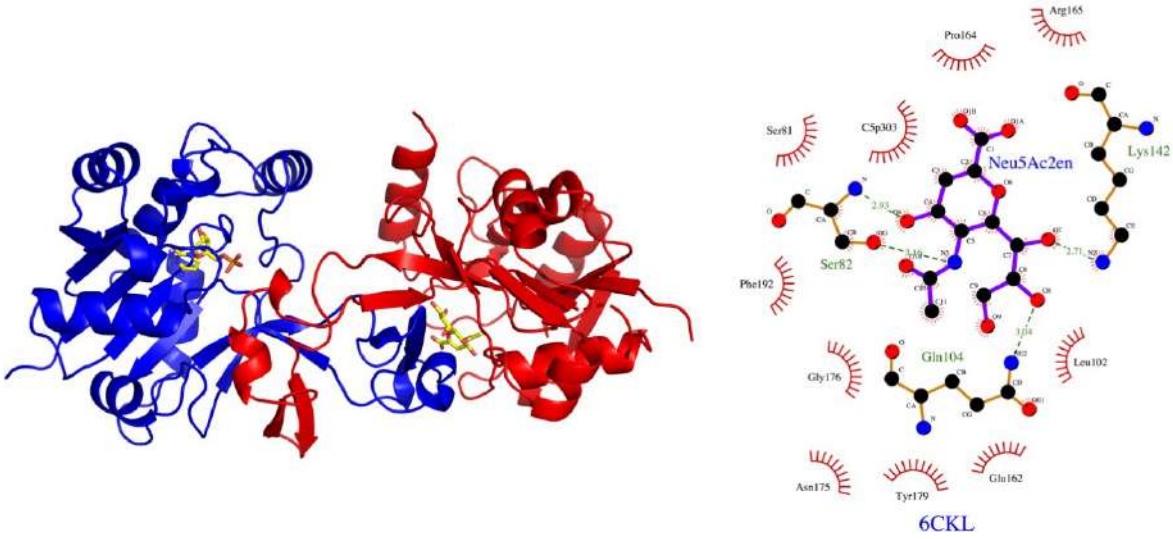
(a) **SiaP – *Haemophilus influenzae***



(b) SatA - *Haemophilus ducreyi*



(c) SiaB - *Neisseria meningitides*



(d) **NanA - *Pasteurella multocida***

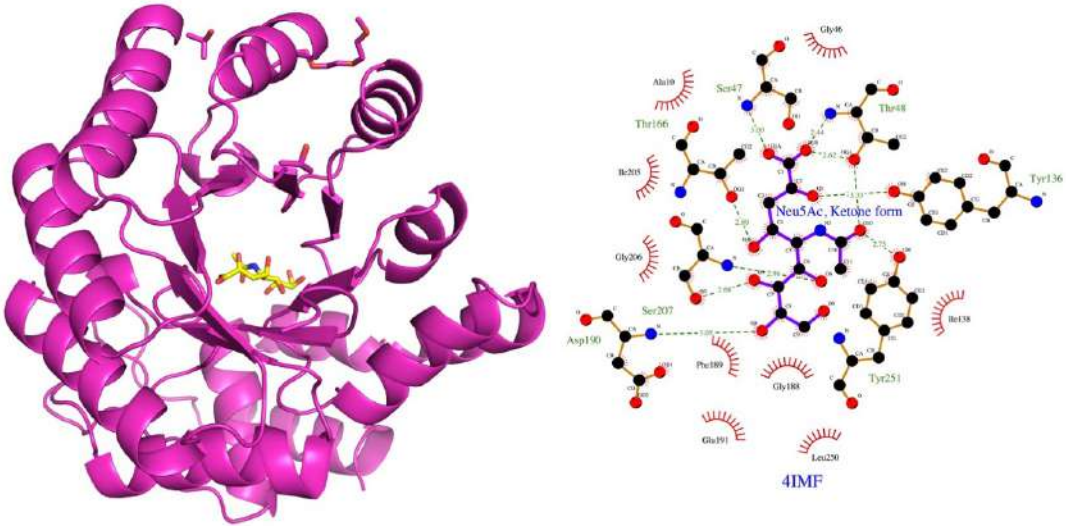


Figure. 5.2. Cartoon representation showing sialic acid binding proteins - SiaP, SatA, SiaB, and NanA. Ligplot representations showing the differences in their binding pocket environments.

In the catabolic pathway, Nan-Nag cluster of genes help to catabolize the sugars into carbon, nitrogen, and energy sources. In some organisms, during sequential multi-enzyme processing of substrates, the proteins co-localize into a multi-enzyme complex (supramolecular complexes) to avoid accumulation of intermediates and to decrease the interference with other cellular components. In contrast, in other organisms, the enzymes are channeled through a substrate tunnel to prevent the loss of intermediates and to improve catalytic performance. In case of bacteria, though they lack cellular compartments, based on necessity, the proteins can be assembled to form thin protein shells that typically encapsulate sequential pathway-related enzymes called bacterial micro compartments (BMCs) (11).

N-acetylmannosamine kinase, a second enzyme in the catabolic pathway, phosphorylates ManNAc into ManNAc-6P. Interestingly, NanK in humans is involved in the Neu5Ac synthetic pathway and forms a complex with epimerase, UDP-GlcNAc 2-epimerase/ManNAc kinase (GNE). Previous studies demonstrated that a knockout of this gene results in the loss of sialic acid production in the cells (12). Additionally, UDP-GlcNAc 2-epimerase inactivation also leads to early embryonic lethality in mice, indicating that sialylation by this bifunctional enzyme is important during development (13). In general, bifunctional enzymes contain two active centers in two domains that are connected by a linker. These enzymes are involved in a variety of metabolic pathways and catalyze two consecutive/non-consecutive reactions or two opposing reactions. Further, these enzymes will help in increasing the success rate of the reaction by preventing loss of intermediates into the bulk. Previously, it was hypothesized that there was a possibility for substrate channeling in human GNE, as it is a rate-limiting step in sialic acid biosynthesis. In addition, it was proposed that the product of the epimerase reaction was channeled as a substrate for the kinase reaction (14). Nevertheless, at present the complex structure of this enzyme is not available. In contrast to human GNE, in the catabolic pathway of bacteria, the kinase-mediated phosphorylation occurs before the epimerase reaction. Likewise, the product ManNAc-6P, formed from the NanK enzymatic reaction, serves as a substrate for NanE, which gets epimerized into GlcNAc-6P. Based on the human GNE domain assembly, we speculate that there is formation of a bifunctional enzyme in bacteria, wherein NanK and NanE form a complex (Figure 5.3).

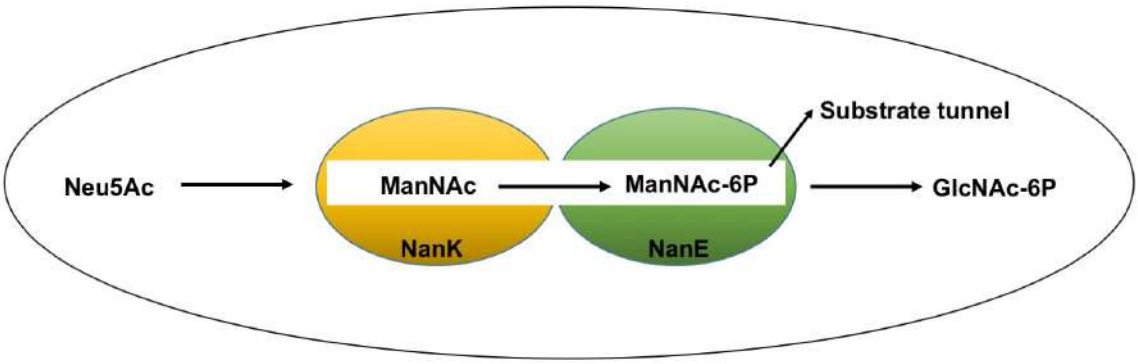


Figure. 5.3. Hypothetical model showing the substrate channelling between NanK and NanE enzyme complex in pathogenic bacteria.

It will be very interesting to study the formation of this complex and its bifunctional enzyme activity. This hypothesis can be tested using size exclusion chromatography and kinetic assay experiments on purified proteins. These studies may provide evidence for substrate channeling in pathogenic bacteria.

In summary, our structural and functional studies on sialic acid binding protein, SatA of the ABC transport system, show that the sialic acid binding pocket is completely different compared to the previously characterized SiaP of the TRAP transport system. In addition, the biophysical characterization of N-acetyl mannosamine kinase from different pathogenic bacteria provides the molecular basis for enzyme function, with special insights on the Zn-binding motif.

The overall goal of the work on these proteins is to identify possible therapeutic targets that will reduce the virulence of these bacteria. The current study reveals several interesting scientific insights into the molecular mechanisms of binding and catalysis of sialic acids. I believe that these structure-activity studies will contribute to the development of effective intervention strategies for diseases caused by this class of Gram-negative bacteria.

5.1 References:

1. Setty, T. G., Cho, C., Govindappa, S., Apicella, M. A., and Ramaswamy, S. (2014) Bacterial periplasmic sialic acid-binding proteins exhibit a conserved binding site. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **70**, 1801–1811
2. Johnston, J. W., Coussens, N. P., Allen, S., Houtman, J. C. D., Turner, K. H., Zaleski, A., Ramaswamy, S., Gibson, B. W., and Apicella, M. A. (2008) Characterization of the N-acetyl-5-neuraminic acid-binding site of the extracytoplasmic solute receptor (SiaP) of nontypeable Haemophilus influenzae strain 2019. *J. Biol. Chem.* **283**, 855–865
3. Setty, T. G., Mowers, J. C., Hobbs, A. G., Maiya, S. P., Syed, S., Munson, R. S., Apicella, M. A., and Subramanian, R. (2018) Molecular characterization of the interaction of sialic acid with the periplasmic binding protein from Haemophilus ducreyi. *J. Biol. Chem.* **293**, 20073–20084
4. Bergauer, K., Fernandez-Guerra, A., Garcia, J. A. L., Sprenger, R. R., Stepanauskas, R., Pachiadaki, M. G., Jensen, O. N., and Herndl, G. J. (2018) Organic matter processing by microbial communities throughout the Atlantic water column as revealed by metaproteomics. *Proc. Natl. Acad. Sci.* **115**, E400–E408
5. Mulligan, C., Fischer, M., and Thomas, G. H. (2011) Tripartite ATP-independent periplasmic (TRAP) transporters in bacteria and archaea. *FEMS Microbiol. Rev.* **35**, 68–86
6. Mulligan, C., Geertsma, E. R., Severi, E., Kelly, D. J., Poolman, B., and Thomas, G. H. (2009) The substrate-binding protein imposes directionality on an electrochemical sodium gradient-driven TRAP transporter. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 1778–1783
7. Huynh, N., Aye, A., Li, Y., Yu, H., Cao, H., Tiwari, vinod kumar, Shin, D., Chen, X., and Fisher, A. J. (2013) Structural basis for substrate specificity and mechanism of N-acetyl-D-neuraminic acid lyase from Pasteurella multocida#. *Biochemistry* **52**
8. Krapp, S., Münster-Kühnel, A. K., Kaiser, J. T., Huber, R., Tiralongo, J., Gerardy-Schahn, R., and Jacob, U. (2003) The crystal structure of murine CMP-5-N-acetylneuraminic acid synthetase. *J. Mol. Biol.* **334**, 625–637
9. Kumar, J. P., Rao, H., Nayak, V., and Ramaswamy, S. (2018) Crystal structures and kinetics of N -acetylneuraminate lyase from Fusobacterium nucleatum . *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **74**, 725–732
10. Bose, S., Purkait, D., Joseph, D., Nayak, V., and Subramanian, R. (2019) Structural and

functional characterization of CMP- *N* -acetylneuraminate synthetase from *Vibrio cholerae*. *Acta Crystallogr. Sect. D Struct. Biol.* **75**, 564–577 [online] http://scripts.iucr.org/cgi-bin/paper?S2059798319006831.

11. Pröschel, M., Detsch, R., Boccaccini, A. R., and Sonnewald, U. (2015) Engineering of Metabolic Pathways by Artificial Enzyme Channels. *Front. Bioeng. Biotechnol.* **3**, 1–13
12. Keppler, O. T., Hinderlich, S., Langner, J., Schwartz-Albiez, R., Reutter, W., and Pawlita, M. (1999) UDP-GlcNAc 2-Epimerase: A regulator of cell surface sialylation. *Science (80-.).* **284**, 1372–1376
13. Schwarzkopf, M., Knobeloch, K.-P., Rohde, E., Hinderlich, S., Wiechens, N., Lucka, L., Horak, I., Reutter, W., and Horstkorte, R. (2002) Sialylation is essential for early development in mice. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 5267–70
14. Hinderlich, S., Weidemann, W., Yardeni, T., Horstkorte, R., and Huizing, M. (2015) UDP-GlcNAc 2-epimerase/ManNAc kinase (GNE): A master regulator of sialic acid synthesis. *Top. Curr. Chem.* **366**, 97–138