Thesis for the degree of Doctor of Philosophy
in the Natural Sciences

Structural Insight Into the Bacterial Sialic Acid Catabolic Pathway

---

Rhawnie Caing-Carlsson

UNIVERSITY OF GOTHENBURG

Department of Chemistry and Molecular Biology

Gothenburg, 2018

Thesis for the degree of Doctor of Philosophy
in the Natural Sciences

Structural Insight Into the Bacterial Sialic Acid Catabolic Pathway

Rhawnie Caing-Carlsson

Cover: Structural alignment of FnNanK with hMNK in complex with N-acetylmannosamine and ADP

ISBN 978-91-7833-225-0 (Print)
ISBN 978-91-7833-226-7 (PDF)
Available online at http://handle.net/2077/57823

Department of Chemistry and Molecular Biology
Division of Biochemistry and Structural Biology
University of Gothenburg
SE-405 30, Göteborg, Sweden
Printed by BrandFactory AB
Göteborg, Sweden, 2018

To Timothy...

Abstract

A genetically diverse community of commensal and pathogenic bacteria thrive in the digestive system and urogenital tracts of animals. Many of these bacteria forage sialic acid from mucosal cell surfaces. Bacteria have evolved a system that utilizes host-derived sialic acid either as an alternative food source (catabolic pathway) or for molecular mimicry to evade the host's immune system (sialylation pathway). Their ability to utilize sialic acid confers a selective advantage by securing an ecological niche for colonization and persistence. Sialic acid catabolic and sialylation pathways are therefore potential targets for the development of novel antimicrobial therapies.

This thesis presents work aimed at determining X-ray structures of sialic acid catabolic enzymes and sialic acid transporters. An automated pipeline was developed to optimize the cloning, expression and purification of enzymes involved in the sialic acid catabolic and sialylation pathways. This led to the large scale production and purification of Nan kinase from *Fusobacterium nucleatum* (Fn)NanK, an enzyme that phosphorylates ManNAc to ManNAc-6-P in the catabolic pathway. The apo structure of FnNanK was determined at 2.2 Å resolution and displays motifs characteristic of the repressor open reading frame kinase (ROK) superfamily. Despite lacking a zinc-binding region previously implicated in stabilizing the enzyme's active site, FnNanK conserved all structural features required for enzymatic activity. A broad base strategy for the expression, solubilization and purification of a sialic acid TRAP transporter orthologues was pursued with the overriding goal of determining the crystal structure of a sialic acid TRAP transporter. Different constructs from four orthologues funneled down to the *Pasteurella multocida* TRAP transporter yielding crystals which diffracted

to 11 Å resolution. New crystallization strategies or other structural approaches may be necessary to propel this project to structure determination. Finally, the crystal structure of the sialic acid transporter SiaT from *Proteus mirabilis* was determined at 1.9 Å with bound substrate. SiaT adopts an outward-facing conformation that provides novel insight into the alternate access mechanism employed by transporters with inverted topology. The crystal structure also reveals a second sodium-binding site that aids substrate binding and stabilizes the outward-facing conformation.

Acknowledgements

The road to a PhD for me, was long and winding. Along the way, I have crossed paths with people whom I forged meaningful relationships with and who became part of my PhD journey.

Richard I would like to thank you for not turning a blind eye on the problem. As my examiner, you made sure that my extremely difficult and unfortunate situation with my former supervisor would not cost me my degree. Because of this you supported me when I was determined not to quit my PhD and you encouraged me when I was barely hanging in there. Thank you for being so generous of your time and for helping me with my writing. Thank you for a lot of things. I simply cannot thank you enough.

Gergely Thank you for taking me in, late into my PhD. You have been very supportive of me, finishing up. I enjoyed our talks on just about anything; science, academia, exotic fruits, travels, food and life. Thank you. **Rams** I appreciate that you let me work with the enzymes that led to two of my papers in this thesis. Thank you for the great hospitality when I was in India. I enjoyed meeting the people in your lab. It was a great experience. **Rosie** I sincerely wish you well. **Göran H.** Thanks for the support. I will exercise again, soon. **Dominique W.** Thank you for helping me get through the toughest time. I really appreciate it.

Weixiao and Parveen Career-driven and unstoppable. Science runs through your veins. Go for it guys and best of luck. **Elin D.** We are almost there. Im glad that we had each other when we were going through a difficult time. Somehow it made everything bearable. Good luck with the writing. **Mikael A.** It was great to have you around. **Lea** It was fun in Heidelberg.

Majo Thanks for being so caring. **Maja** and **Viktor**, it has been fun being in the same group with you guys.

People at Ram's lab. **JP**, **Lavanya**, **Thanuja**, **Nitish**, **Vinod**, **Archana** and everyone at Ram's lab. Thank you for those wonderful weekend trips.

Rajiv my officemate. Thanks for answering all my thesis questions. **Kristina** the lady in red, it is always interesting to strike a conversation with you. **Stephan**, **Florian**, the feminist. Thanks for the help. **Gisela**, you are a great listener. **Cissi**, shake it like a polaroid picture. **Swagatha** I am glad that you are in Sweden. Keep those chais and samosas coming. **Rebecca** the lab is buzzing again, now that you are back. **Petra B**, ninja! Soon it is your turn. Good luck! **Cecilia W** thanks for listening. **Per** grandpa's kitchen is therapy, I swear. **Giorgia**, **Daniel** who went gaga over Jose Gonzales?, **Rob B**, **Greger** keep the lab interesting. **Andreas** Good luck!

Valida Tack för att du är min vän. **Lars** Du är en ängel. **Anne** Thanks for taking care of the paperwork. **Bruno** you continue to smile at me even when I continue to misplace the office key. Thank you guys for keeping things working in the lab. **Tinna** the commander in the lab, thanks for keeping the lab in shape. **Örjan** Thanks for the courses.

Sebastian Friday Im in love, the Cure. **Leo**, **Elin C.** Everything will be fine, **Joachim**, **Linnea**, **Matthijs**, **Claire**, **Emil**, thanks for entertaining my last minute, panic thesis questions. Good luck. **Amke** and **Andrea**, welcome! The alumni, **Erik**, **David**, **Petra E**, **Mikael**, **Linda**, **Mike**, **Rob D.**, **Oskar**, **Ida**, **Anna**, **Jennie**, **Annette**, **Alex**, **Rajiv**, **Karin** and **Madde**, thanks for all the fun. **Johanna**, it is nice to have you here at bcbp. **Davide** and the rest of **Johanna's group**, Good luck!, **Björn**, your group is expanding. Good luck! **Emelie**, **Darius** and the rest of **Björns group**. **Amit**, it was fun with X-ray crystallography. **Nasha**, habibi thanks. **Heikki** thanks for screening. **Maria H**, I kept that envelope. **Stephan N**, thanks for being such a true friend. **Annette**, thank you for being there for me always.

Marie, **Marion** tack för allting. Thank you to all my friends and family. My **mom**, **dad** and **my brothers**, I love you guys. **Timothy**, you are the best thing that ever happened to me. I love you all the way to the moon and back. **Tommy**, love of my life, I know that it has been tough. Thanks for all the love, support and patience. *I love you*.

Publications

Paper I: Bairy, S., Gopalan, L. N., Setty, T. G., Srinivasachari, S., Manjunath, L., Kumar, J. P., Sai, G.R., Sucharita, B., Nayak, V., Ghosh, S., Sathyanarayanan, N., **Caing-Carlsson, R.**, Wahlgren, W.Y., Friemann, R., Ramaswamy, S. **(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(2), 420-428.

Paper II: **Caing-Carlsson, R.**, Goyal, P., Sharma, A., Ghosh, S., Setty, T. G., North, R. A., Friemann, R., Ramaswamy, S. **(2017)**. Crystal structure of N-acetylmannosamine kinase from Fusobacterium nucleatum. *Acta Crystallogr F Struct Biol Commun*, 73(Pt 6), 356-362.

Paper III: **Caing-Carlsson, R.**, Goyal, P., Wahlgren, W.Y., Dunevall, E., Ramaswamy, S, Friemann, R.**(2018)** Expression, purification and crystallization of a sialic acid tripartite ATP-independent periplasmic (TRAP) transporter. *Manuscript*

Paper IV: 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., Dobson, R.C.J., Abramson, J., Ramaswamy, S., Friemann, R. **(2018)**. Substrate-bound outward-open structure of a Na(+)-coupled sialic acid symporter reveals a new Na(+) site. *Nat Commun*, 9(1), 1753. doi:10.1038/s41467-018-04045-7

Contribution

Paper I: I did the initial bioinformatic work for the DNA sequences of the enzymes. I produced, purified and crystallized FnNanK and FnNagA.

Paper II: I produced, purified and crystallized the protein. I solved, refined and analyzed the x-ray structure. I performed the structural alignments and analysis. I took a major part in writing the manuscript and made some figures.

Paper III: I planned the experiment. I cloned the constructs and produced the proteins. I performed the detergent screens and purified the proteins. I set up the crystallization trials and screened the crystals at the synchrotron. I wrote the manuscript and made the figures.

Paper IV: I took part in the detergent screening. I produced and purified the protein.

List of Abbreviations

ABC	ATP binding cassette
ATP	Adenosine triphosphate
β-OG	n-octyl-β-D-glucoside
CMC	Critical micellar concentration
DDM	n-Dodecyl β-D-maltoside
DM	n-Decyl β-D-maltoside
EcNanK	Escherichi coli N-acetylmannosamine kinase
FSEC	Flourescent size exclusion chromatography
FnNanK	Fusobacterium nucleatun N-acetylmannosamine kinase
GFP	Green fluorescent protein
IMAC	Immobilized-Metal Affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LDAO	Lauryldimethylamine oxide
LmNanK	Listeria monocytogenes N acetymannosamine kinase
MFS	Major Facilitator Superfamily
MNG10	Decyl maltose neopentyl glycol
MNG12	Lauryl Maltose Neopentyl Glycol
ManNAc	N-acetylmannosamine
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PDB	Protein Data Bank
PEG	Polyethylene glycol
RFU	Relative flourescence unit
ROK	Repressor, open reading frame, kinase
SDS	Sodium dodecyl sulfate

SEC	Size exclusion chromatography
SGLT	Sodium galactose symporter
SRP	Signal recognition particle
SSS	Sodium solute symporter
TCA	The citric acid cycle
TMH	Transmembrane helices
TRAP	Tripartite ATP-independent Periplasmic
UDP-GlcNAc	Uridine diphosphate N-acetylglucosamine
hMNK	Bifunctional UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase

Contents

Chapter 1

Introduction

Billions and billions of microorganisms thrive in the Earth's biosphere [1]. Within this vast number, microbial diversity is indispensable for the bacteria's survival and proliferation. Bacteria constantly adapt and compete to secure ecological niches wherein through the process of natural selection, new phenotypes are acquired that confer fitness advantage.

Antibacterial therapy promotes such selection pressure on bacteria by counteracting its cytotoxic and cytostatic effects rendering the bacteria susceptible to the host immune response. Antibiotics typically work by inhibiting cell wall biosynthesis, disrupting nucleic acid and protein synthesis or other specific actions [2]. Resistance mechanisms against the actions of antibiotics are developing through a selection process.

The resistance mechanisms include diminished uptake of the antibiotics through low permeability of the outer membrane, efflux pumping to expel antibiotics out of the cytoplasm, impeding the binding of the antibiotics by modifying the target sites and expression of enzymes that inactivate antibiotics [3]. Antibiotic resistance arises when the ability of the drug to effectively inhibit bacterial growth is seriously compromised. Consequently, a higher concentration of antibiotics is needed to decrease the rate of bacterial replication. In addition, resistance can be acquired against more than one antibiotic drug simultaneously giving rise to multidrug resistant bacteria designated as 'superbugs' [4, 5].

Antibiotic resistance remains a global health threat that is gaining an

alarming momentum due to the misuse and overuse of the currently available antibiotics [6]. Infectious diseases continue to be one of the leading causes of death worldwide. The reemergence of the once contained diseases serve as a stark warning on what is to spiral out of control in an unprecedented global scale. Thus, the pressing antibiotic resistance problems need to be curbed on time. Since the late 1980s, no new antibiotics were discovered while one by one the presently available antibiotics become ineffective against to multidrug resistant bacteria, the antibiotic Ceftaroline being the latest one [2].

1.1 Antibiotics resistance spreads via horizontal gene transfer

The bacteria's ability to diversify in order to adapt and thrive in special niches and compete for the resources and acquire physiological traits that enhance virulence reflects the innovative nature of the bacterial genome propelled by lateral gene transfer. Through plasmids, integrons and transposable elements, fitness factors such as multidrug resistance are transferred [7].

There is a constant shuffling and reshufflings in the bacterial genome wherein genes that do not hold essential functions are segregated and deleted to be replaced by genes that endow cells with a fitness advantage [7]. DNA sequences encoding important protein factors such as resistance traits and new metabolic pathways are driven forward by natural selection. In bacteria it is a constant balancing act between genes acquired and genes lost. In this regard, lateral gene transfer contributes to the adaptability and speciation.

1.2 Virulence in host-microbe interaction

What is virulence? Virulence is the degree of damage inflicted on the host by any microbe capable of causing the disease. The definition of virulence is a two-way street wherein the invasiveness of the pathogen is largely dependent on the susceptibility of the immune response of the host under the attack [8].

Virulence mechanisms under host-microbe interaction largely constitutes of the use of microbial adhesins as a first line of contact to the host cell and the release of toxins to impair the host cells. It also includes any type of invasion of the innate and immune response and the upregulation of virulence-associated genes and secretion of virulence enhancing products [9]. Targeting any of the aforementioned virulence mechanisms can restrict or inhibit the pathogen's persistence and eventually prevent the disease.

In this connection, bacterial sialobiology has a convincing role to play in developing new therapeutic interventions in preventing diseases and infections [10]. Many bacterial commensals and pathogens utilize sialic acid for persistence and to increase their pathogenicity. Sialic acid, which is abundant in the eukaryotic host, is acquired by both commensals and pathogens to be used as a food source (catabolic pathway) or as molecular mimicry (sialylation) as effective mechanism for colonization and persistence. Targeting these pathways can be instrumental in designing novel drugs for antimicrobial therapy.

1.3 What is sialic acid?

Sialic acid is the designation for a family of nine-carbon α-keto amino sugars. It is abundant in nature except in plants. It is mainly found within the deuterostome lineage under the domain Eukarya [11]. Scaffolding its structural diversification and modification is the core moiety of 2-keto-3-deoxy-5-acetamido-D-glycero-D-galacto-nonulosonic acid or N-acetylneuraminic

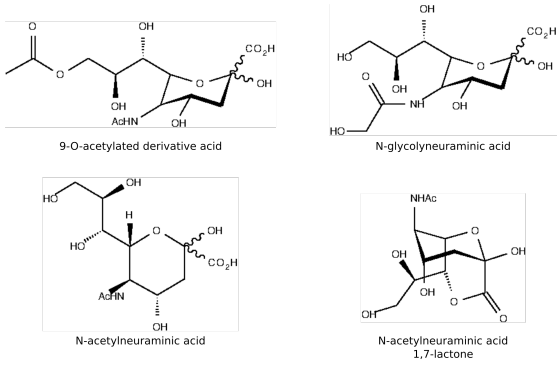

Figure 1.1: The most common types of sialic acids.

acid (Neu5Ac). (Figure 1.1). Serving as a precursor for any structural modification, Neu5Ac constitutes a 6-carbon carboxylic acid ring with an acetamido on C5 and a glycerol tail with hydroxyl groups C7, C8 and C9 [12]. Modifications in C5 position are prevalent and the most common ones are the substitution with a hydroxyl group forming N-glycolylneuraminic acid (Neu5Gc) [11].

Interestingly, Neu5Gc is ubiquitous in the deuterostome lineage but missing in humans. The distinct absence of Neu5Gc in humans is attributed to the deletion of the 92 bp exons in the hydroxylase gene rendering the enzyme that converts Neu5Ac to Neu5Gc inactive [13]. However trace amounts of Neu5Gc are detected in human tissues that are accounted to the dietary intake of mostly red meat [13, 14]. Further structural diversity extends to with the deaminated group of Neu5Ac and O-acetylation.

1.4 Sialic acid in vertebrates

Epithelial mucosal cell surfaces are studded with sialic acid serving as the terminal non-reducing sugar of the glycoproteins and glycolipids. Due to its location, sialic acid has an essential role in cell-cell interactions and self-recognition functions in numerous biological regulatory processes that is

crucial in maintaining homeostasis.

In humans, the most abundant source of sialic acid is in the form of polysialic acid (PSA) that is found in the brain. PSA is involved in cell signaling processes that are implicated in neural plasticity, growth and migration throughout the central nervous system (CNS) [15].

Additionally, glycoconjugates found in GD3 gangliosides carry the O-acetylated sialic acid. GD3 acts as a stress-induced apoptosis mediator by accumulating in the mitochondria and thereby compromising the integrity of the plasma membrane causing the release of apoptotic factors and caspases triggering Fas-mediated apoptosis [16, 17].

The role of sialic acid in self-recognition function is demonstrated in the down regulation of the alternative complement pathway through the binding of the Factor H to the glycerol tails of sialic acid. Studies reported that 9-O acetylation impedes the binding of Factor H thereby activating the innate immune response. Pathogenic bacteria that lack a coat of sialic acid and consequently are not recognized by Factor H, fail to evade the immune-surveillance and thus are subject to lysis via the complement system [18].

Another type sialic-acid binding protein found in the vertebrates is selectin. This family of lectins binds to sugar moieties such sialyl lewis x and sialyl lewis A on the cell surfaces. Selectins are expressed in leukocytes, activated platelets and endothelial cells. They are activated as the first line of defense of the immune system to recruite leukocytes to the inflammed region [16, 19, 20].

In addition, Siglecs is the largest family of mammalian sialic acid binding lectins. Interaction between the sialoglycoconjugates and CD33-related siglecs leads to inhibitory mechanism by enhancing the production of anti-inflammatory cytokines and diminishing the secretion of pro-inflammatory cytokines [21, 22].

1.5 *De novo* biosynthesis of sialic acid

The synthesis of Neu5Ac from UDP-GlcNAc involves five steps enzymatic reactions (Figure 1.2). In the cytoplasm UDP-GlcNAc from the hexosamine

pathway is epimerized by UDP-GlcNAc-2-epimerase yielding ManNAc and releasing UDP. Instantly, ManNAc is phosphorylated using ATP by ManNAc kinase producing ManNAc-6-phosphate. In humans, conversion of UDP-GlcNAc to ManNAc-6-phosphate is catalyzed by a bifunctional enzyme GNE [23]. Next, Neu5Ac9P synthase catalyzes the condensation of ManNAc-6P and phosphoenolpyruvate (PEP) yielding Neu5Ac9P. The 9-phosphate from Neu5Ac9P is then released by a phosphatase. The activation of Neu5Ac to CMP-Neu5Ac occurs in the nucleus where the cytidine triphosphate (CTP) and Neu5Ac are joined by CMP-Neu5Ac synthetase and the pyrophosphate is released by a phosphatase yielding CMP-Neu5Ac. Afterwards the nucleotide sugar is then transported into the Golgi body with the aid of CMP-Neu5Ac transporters. Once inside the Golgi apparatus, activated CMP-Neu5Ac is used as a substrate by sialyltransferases to proceed to glycosylation [23].

In vertebrates, Neu5Ac degradation is catalyzed by an aldolase, splitting Neu5Ac to ManNAc and PEP. ManNAc is then converted to GlcNAc by GlcNAc-2-epimerase and subsequently phosphorylated to GlcNAc-6P by a kinase whereby it enters the hexosamine pathway [16, 24].

Although rare, there are bacteria that are capable of synthesizing Neu5Ac de novo. The *E.coli* serotype K1 strain and some *Nesseria* strains have a biosynthetic pathway that utilizes homologues of UDP-GlcNAc-2-epimerase, Neu5Ac-9P-synthase and CMP-Neu5Ac-synthase. In prokaryotes, ManNAc is directly converted to Neu5Ac and immediately activated to CMP-Neu5Ac whereas in eukaryotes, the intermediate steps require certain phosphatases. The more rapid activation of Neu5Ac reflects the different fates of sialic acid between higher eukaryotes and bacteria. Cleaved sialic acid is either secreted or directed to lysosomes for recycling in eukaryotes while the microbes utilize sialic acid for molecular mimicry and as a nutritional source. In addition, bacteria have evolved dynamic modes of transport of the scavenged sialic acid [26].

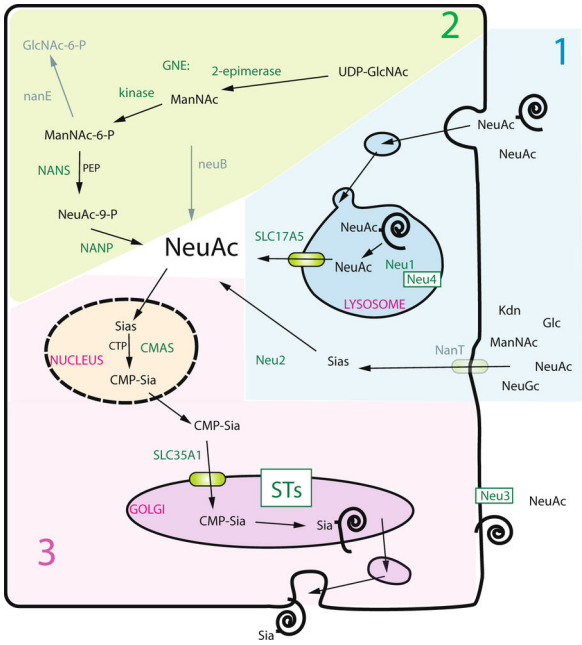

Figure 1.2: Sialic acid biosynthesis and metabolism in eukaryotes. The synthesis of Neu5AC from UDP-GlcNAc takes place in the cytoplasm (yellow background) while exogenous Neu5AC from the glycoconjugates (blue background) is acquired via the lysosomes. Neu5AC is activated in the nucleus and then transported to the Golgi body to proceed to glycosylation. (Reproduced from Petit, et al) [25].

1.6 Sialylation

Higher eukaryotes decorate their cell surfaces with sialic acid, a process that occurs in the Golgi apparatus (Figure 1.3). The physicochemical properties of an activated CMP-Neu5Ac impede its diffusion across the Golgi membrane. Hence a specific CMP-Neu5Ac transporter facilitates its entry into the cytoplasm. Once recognized by sialyltransferase as a substrate, it is incorporated onto the terminal region of the glycoproteins and glycolipids [27].

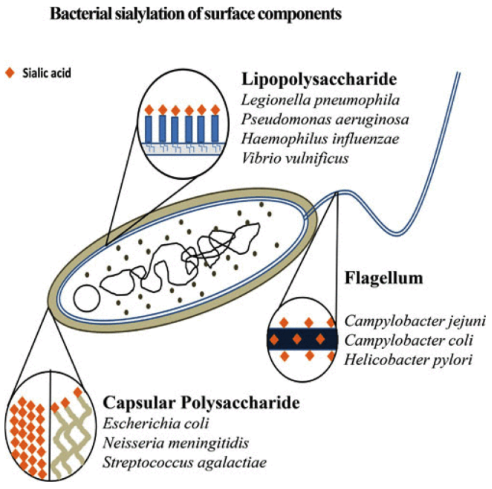

Figure 1.3: Bacterial sialylation in three types of surface structures; lipopolysaccharide, flagellar and capsular. (Reproduced from Haines-Menges et al [22]).

Since the outermost layer of mucosal surfaces in an animal host are saturated with sialic acid, it is not surprising that bacteria have evolved mechanisms to exploit this monosaccharide to their own advantage. Bacteria use adhesins to establish the first line of contact with the target cell. Sialidases are then produced by the bacteria to cleave sialic acid from the cell surface's glycoconjugates. Subsequently, the acquired sialic acid is transported into the cytoplasm to be used either for sialylation or foodstuff [14, 28].

Sialylation in bacteria occurs in three types of surface structures depending on the strain and species such as lipopolysacharride LPS in gram-negative bacteria, flagella and capsular polysaccharide [22]. Moreover, transport of Neu5Ac into the cytoplasm in *Tanerella forsythia* is reported to be important for biofilm formation and survival in epithelial cells [29].

Bacteria have adapted four main mechanisms for obtaining sialic acid: de novo biosynthesis, donor scavenging, trans-sialidase and precursor scavenging [30]. *De novo* biosynthesis of sialic acid is observed in *E.coli* K1 strain and other *Nesseria* strains wherein endogenous Neu5Ac is activated

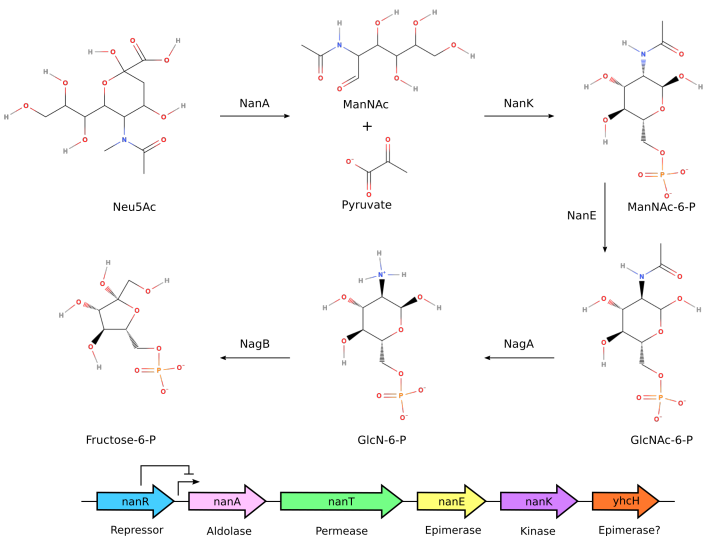

Figure 1.4: The sialic acid catabolic pathway. Degradation of Neu5Ac occurs in five steps yielding Fructose-6-P. The degradation is regulated by *nan-nag* cluster.(Adapted from Vimr [14])

by CMP-Neu5Ac synthetase and recognized by sialyltransferases as a substrate and then incorporated into the terminal end of the glycan moiety. Donor scavenging is observed in *N. gonorrhea* where its resistance to complement serum is restored after addition of exogenous CMP-Neu5Ac, using its own sialyltransferases to decorate cell surfaces. The trans-sialidase mechanism is demonstrated by *T. cruzi* with no inherent sialic acid biosynthesis and catabolic regulon, it hydrolyzes sialyl unit from the host's glycoconjugates to reconstruct its cell surface .

Moreover, *H. influenzae* lacks sialidases and cannot synthesis Neu5Ac *de novo*, however it carries a Neu5Ac degradation system utilizing it as carbon and nitrogen source. Apart from this, *H. influenzae* encodes its own CMP-Neu5Ac synthetase and sialyltransferase indicating its capability to incorporate Neu5Ac onto its cell surface. Scavenged Neu5Ac are either directed

to degradation or sialylation [27].

1.7 Sialic acid catabolism

The ability to use sialic acid as an alternative food source is instrumental for many types of bacteria's colonization and persistence. Studies report a direct correlation between pathogen's enhanced virulence and sialic acid catabolism [22, 28].

Once bound sialic acid is cleaved by the sialidases from the host mucosal cell surface, the amino sugar is transported by permeases into the cytoplasm. Afterwards, N-acetylneuraminate lyase (NanA) cleaves Neu5Ac into ManNAc and pyruvate, the latter then enters the TCA pathway. ManNAc proceeds to be phosphorylated by N-acetylmannosamine kinase (NanK) yielding ManNAc-6P which is then epimerized to GlcNAc-6P by N-acetylmannosamine-6-phosphate-2-epimerase (NanE). The *nag* operon is induced by GlcNAc-6P, which encodes the N-acetylglucosamine-6P deactylase (NagA) and subsequently the glucosamine-6P-deaminase (NagB) that breaks down GlcNAc-6P to fructose 6-P and ammonia. (Figure 1.4)[14].

1.8 Bacterial sialometabolism in mucosal cell surfaces

Starting from the mouth, the oral pathogen that causes inflamed gums *T. forsythia* is widely implicated in biofilm formation. Its genome encodes for sialidases, sialic acid transporter and sialic acid catabolism. Mutation studies in its sialidases indicated terminal growth in media with sialic acid as a sole carbon source and a hampered ability to adhere on cell surfaces [31, 32].

Passing through the respiratory tract where the pathogen non-encapsulated *H. influenzae* could establish a niche, causes bronchitis and otitis media. Although primarily a sterile environment, the respiratory tract's mucosal linings, which are rich in sialic acid are a good habitat for *H. influenzae*. The

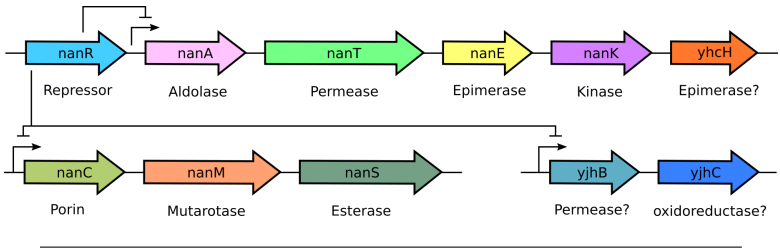

Figure 1.5: The sialic acid regulon

scavenged sialic acid ends up either as surface decoration or food source. Studies where in the first committed degradation enzyme NanA is deleted increased sialylation and consequently serum resistance was observed [22, 33, 34].

The mucosal layer of the gastrointestinal tract functions as a protective barricade for infections. Diverse microbes have inhabited the spanning length of the gastrointestinal tract as their niches, whose rich diversity is comparable to the stratified species variation of a rainforest [10, 35]. The most dominant Gram-negative facultative commensal in the human gut is *E.coli* K12, its ability to use sialic acid as an alternative food source gives *E.coli* K12 a competitive edge, thus securing the GIT as its niche [14]. Another known gut pathogen *S. enterica*, enhances its virulence by utilizing Neu5Ac for adhesion and colonization [36].

Worth mentioning is the colonization of the gut by the *Vibrio sp.* Encoding the sialoregulon in their genome, *V. cholera* shows a growth advantage during the early stages of infections upon utilization of sialic acid [26]. Further, NanA lyase mutations in *V. vulnificus* reported diminished virulence and susceptibility to cytotoxic response. NanA lyase experiments suggest that NanA, whose transcription is induced by the presence of sialic acid, is necessary for growth in minimal medium supplemented with sialic acid as the sole carbon source [22, 37].

Thriving in the urogenital tract is *G. vaginalis*, the causative agent for bacterial vaginosis which uses endogenous sialidases to acquire sialic acid

which is then degraded to carbon and nitrogen source. Furthermore, increased virulence is implicated as the sialidases attacks the mucosal protective barrier by foraging sialic acid into depletion [38].

Taken together, the ability to utilize sialic acid by the commensal and pathogenic bacteria as an alternative food source confers a competitive advantage that leads to securing niches that span the entire mucosal layer in a vertebrate.

1.9 The canonical nan operon in *E. coli* K12

The nan operon is a sialic acid regulating gene cluster consisting of nanATEK-yhcH and a nanR repressor located upstream. NanR repressed the transcription of nanA by binding onto the operator, thereby blocking the RNA polymerase that transcribes NanA. More recently, a corregulated sialic acid operon nanCMS yjhBC was identified (Figure 1.5) [14, 39]. In this nan operon, NanC encodes for an outermembrane porin, NanM for a sialate mutarotase and NanS for a sialate O-acetyl esterase. The crystal structure of 12 stranded beta-barrel NanC belongs to a family of diffusion channels with high selectivity for acidic oligonucleotide [40, 41].

Generally sialic acid is released as an α-anomer from the host glycoconjugates but imported by the transporters into the cytoplasm in its β-anomeric form [14, 42]. Spontaneous rotation from the α-anomeric form to the β-anomeric occurs at a slow pace but with the aid of NanM the rapid conversion of α-anomeric sialic acid to β-anomeric structure boosts the scavenging mechanism of the commensals and pathogens alike, conferring a competitive edge.

Further, NanS converts O-acetylated sialic acid to Neu5Ac and the yhcH gene is possibly an epimerase as its crystal structure indicates an epimerase activity [43]. Similar to nanT, gene yjHB possibly transports a distinct type of sialic acid while gene yjhc is a putative oxidoreductase.

As proposed, the sialocatabolic pathway starts with the transport of free flowing sialic acids into the periplasm by outer membrane proteins OmpF, OmpC and NanC through passive diffusion. In the periplasm, sialic acids

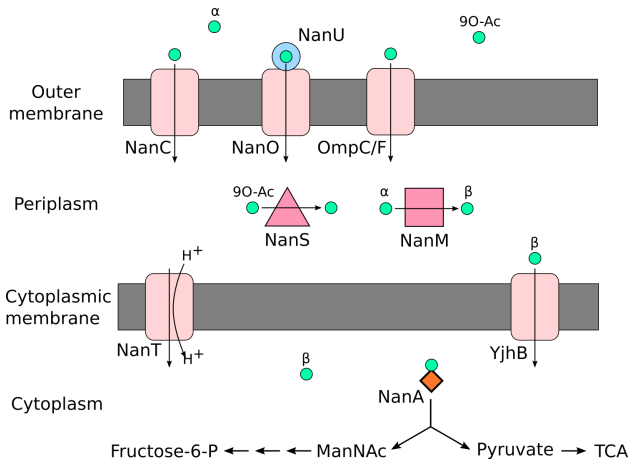

Figure 1.6: The proposed mechanism of sialic acid catabolic pathway. The uptake of sialic acid into the periplsm by nan porins, followed by the convertion of sialic acid to recognizable substrates for the sialic acid transporter. NanA cleaved Neu5Ac to ManNAc and pyruvate.

are converted by NanS and NanM into recognizable substrates for NanT and possibly also for Yjhb, to be transported into the cytoplasm. Once inside the cytoplasm Neu5Ac is cleaved by NanA to ManNAc and pyruvate. The resulting ManNAc is finally degraded into carbon, nitrogen and energy (Figure 1.6) [14].

1.10 Sialic Acid Transporters

Bacteria have devised diverse ways of adaptation for growth and survival in response to a dynamically changing environment where the resources are limited and the competition is fierce. Many commensals and pathogenic bacteria occupy niches and persist in mucin-rich environments such as gastrointestinal and respiratory tracts of humans and other members of the deuterostome lineage. In higher eukaryotes, the cell surface glycans have sialic acid as its terminal sugar. Bacteria exploit the externally acquired

sialic acid by using it as molecular mimicry to bypass the immunosurveillance of the host or as an alternative source for carbon, nitrogen and energy [14, 22, 34, 44].

The uptake of the externally acquired sialic acid requires a functional sialic acid transporter to move the sugar across the membrane into the cytoplasm. Throughout the bacterial kingdom, there are various ways of transporting sialic acid into the cytoplasm. To date, there are four known families of sialic acid transporters representing both primary and secondary transporters [45, 46]. The family of ATP-binding cassette (ABC) transporters is the primary transporter in which that the hydrolysis ATP drives the transport of the sugar into the cytoplasm [47]. Representing the secondary transporters are the tripartite-ATP-independent periplasmic (TRAP) transporters, major facilitator superfamily (MFS) and the sodium solute symporters (SSS) [46]. All secondary transporters harness the energy from the electrochemical gradient to drive on transport across the membrane.

In order for bacteria to exploit sialic acid, an active sialic acid transporter is necessary to move the sugar into the cytoplasm where it can end up in two different pathways: the sialylation pathway or the catabolic pathway [26, 28]. Mutation studies that lead to loss of function of the sialic acid transporters demonstrate the capacity to utilize sialic acid is linked to its ability to transport the sugar into the cytoplasm. Disruption in the uptake function of nanT sialic acid transporter in *E. coli* results in the bacteria's inability to grow on sialic acid as the sole carbon source [39]. Mutation and the subsequent disruption of the sialic acid TRAP transporter in *H. influenza*, diminished sialylation that results in bacteria's high sensitivity of bacteria to human serum [48]. In a separate growth assay, the sialic acid transporter gene from *Salmonella enterica* serovar *Typhimurium* was able to restore growth on sialic acid as a sole source of carbon in ΔnanT deleted *E. coli* strain [49].

Humans are capable of *de novo* biosynthesis of sialic acid and also have corresponding sialic acid transporters that have a low homology with the

bacterial sialic acid transporters. Development of novel inhibitors that inactivate the sialic acid transporters in bacteria and thereby reduce its pathogenicity, is a promising target for drug development. In this section, the different types of sialic acid transporters are introduced providing a background to Paper 3 and Paper 4 (Chapters 4 and 5).

1.10.1 Outer membrane protein

In Gram-negative bacteria there is an additional outer membrane layer that controls the traffic of molecules into and out of the periplasmic space. There are outer membrane porins that are used for the efficient uptake of sialic acid into the periplasm. The porin NanC [40, 41] encoded from the sialic acid operon nanCMS facilitates the movement of sialic acid across the outermembrane. Once inside, the α-anomeric sialic acid is converted to β-anomeric sialic acid by a mutarotase NanM from the same operon, prior to the sugar's uptake into the cytoplasm [50].

More recently, the outermembrane nanOU system was identified as a sialic acid transport system in *T. forsythia* and *B. fragilis*. The nanOU system consists of two subunits; the NanO, a Ton-B dependent porin and a NanU, a sialic acid binding protein that captures sialic acid with high affinity then delivers it onto the porin to be translocated into the periplasm (Figure 1.7) [51].

1.10.2 ABC transporters

The sialic acid ABC transporter is encoded by the operon satABCD wherein the SatA is a subsrate binding protein for high affinity capture of sialic acid in the periplasm. The SatB subunit is an integral membrane protein that translocates the substrate into the cytoplasm while SatD subunit is the nucleotide binding domain. In addition, SatC subunit is a fusion between a permease and a nucleotide binding domain (Figure 1.7) [52].

Sialic acid ABC transporters are present in a wide range of bacteria such as the genus *Streptococcos*, *Corynebacterium*, *Actinobacillus* and *Haemophilus*

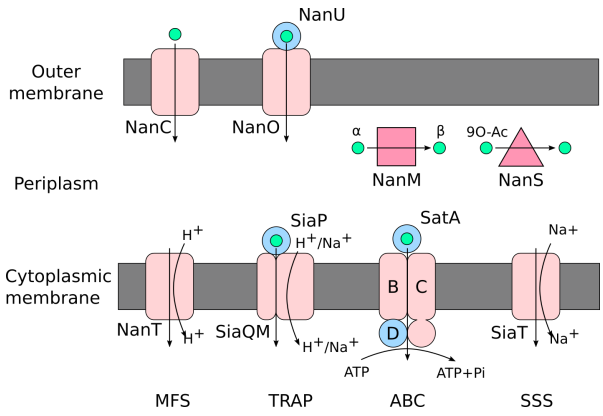

Figure 1.7: Sialic acid transporters. The NanC and NanOU porins in the outer membrane import sialic acid into the cytoplasm. Enzymes NanM and NanS convert sialic acid to recognizable substrate prior to transport into the cytoplasm. Different types of sialic acid transporters, MFS, TRAP, ABC and SSS have various mode of substrate transport. (Adapted from North et. al [45])

[53, 54]. This was first characterized in *H. ducreyi* as it was demonstrated to be essential for its sialylation [52].

1.10.3 TRAP transporters

The TRAP transporter was first characterized in the C4-dicarboxylate transporter in *Rhodobacter capsulatus* [55]. As a secondary transporter, the transport of the substrate across the membrane is driven by an electrochemical gradient. Moreover, TRAP transporters are prevalent in bacteria and archae but are absent in eukaryotes. It transports a wide range of solutes such as amino acids, C4-dicarboxylate, ectoine, gluconate, aromatic substrate and sialic acid [56, 57].

The sialic acid TRAP transporter is composed of three subunits. Like the ABC transporter, it has a substrate-binding domain SiaP that captures sialic acid with high affinity in the periplasm and delivers it onto the intergral membrane domains. Structural studies on the SiaP subunit indicate a

highly conserved salt bridge formation between an Arg and the carboxylate group of the substrate (Figure 1.7) [58, 59]. In *H. influenzae, P. multocida* and *F. nucleatum*, these two integral membrane components are fused forming a single subunit while in *V. cholera* the integral membrane components are two separate subunits making it a true tripartite system. Topology prediction for the transmembrane domains indicates 17 transmembrane helices (TMHs). The small SiaQ domain is made up of 4 TMHs while the large SiaM domain is composed of 12 TMHs. The two domains are fused with an extra helix. Interestingly, the SiaM domain is the translocator and belongs to the IT superfamily of proteins while the SiaQ domain has an unclear function [60].

Mutations on the substrate binding SiaP or in the transmembrane domains result in the loss of sialic acid uptake, which leads to decreased sialylation and subsequently high sensitivity to human serum [48, 56]. In chinchilla model of otitis media, disruption of the TRAP transporter in *H. influenzae* decreases its virulence [61].

1.10.4 Major Facilitator Superfamily

In 1985, the first sialic acid transporter NanT was experimentally characterized in *E. coli* [39]. The sialic acid transporter was identified by isolating *E. coli* mutants with knock out function of Neu5Ac uptake and consequently by mapping the mutation to the locus designated as nanT. The sialic acid NanT is a member of the MFS superfamily. A unique feature of NanT is that it is comprised of 14 TMHs instead of the more common 12 TMHs in the MFS family. The proposed translocation mechanism that is employed within this family, is the "rocker switch" mechanism [62, 63] in which the inward and outward facing conformations alternate and the substrate binding site can only be reached in a successive manner. In *Tanerella forsythia*, NanT is important for the biofilm formation and enhanced virulence (Figure 1.7) [29] .

1.10.5 Sodium Solute Symporter

More recently, the newest addition to the types of sialic acid transporter is the sodium solute symporter (SSS) family, first characterized in *S. enterica* serovar *Typhimurium* [49]. SiaT is utilized by a broad spectrum of both Gram-negative and gram- positive bacteria such as *S. aureus* [64], *C. difficile* [65], *V. fischeri* [26], *L. sakei* [66] and *P. profundum* [26]. As a sodium symporter, it transports sodium ions together with the substrate driven by the sodium gradient across the membrane (Figure 1.7) [49].

There is one known crystal structure of an SSS transporter, the *V. parahaemolyticus* sodium galactose symporter (SGLT) [67]. It is an inward-facing structure and the core domain structure of vSGLT is comprised of two inverted repeats wherein each repeat is composed of 5 TMHs. Interestingly, although sharing a low sequence homology, this fold is also observed in other sodium symporters such as the sodium leucine symporter (LeuT) [68] and the sodium benzyl-hydantoin symporter (Mhp1) [69].

The aim of my studies was to provide structural insight on the enzymes involved in the catabolic pathway and the corresponding sialic acid transporters that are essential for the import of sialic acid into the cytoplasm. The bulk of this thesis describes the production and crystallization strategies aimed for structure determination of the enzymes and transporters in the sialic acid catabolic pathways. Further, crystal structures of a Nan kinase in the catabolic pathway and a sialic acid transporter from the soldium solute symporter (SSS) family are presented.

1.11 Scope of this thesis

This thesis presents the construct screening, expression, purification, crystallization and structure determination of enzymes and trasnporters involved in the sialic acid catabolic pathway. Chapter 2 introduces the X-ray crystallogaphy methodology. In Chapter 3, **Paper 1** and **Paper 2** will be discussed. In Chapter 4, the entire pipeline of membrane protein crystallization from detergent screening to crystallization is taken into account, summarizing

Paper 3. Chapter 5 discusses the crystal structure of the first sialic acid transporter, described in detail in **Paper 4**. Finally, Chapter 6 discusses the conclusion and future perspective.

Chapter 2

X-ray Crystallography

X-ray crystallography remains the most widely-used technique in structural biology. Since the determination of myoglobin structure in 1957 [70] crystal structures continue to advance our understanding of the biological processes of life. X-ray snapshots of these molecules of life in atomic detail have facilitated the development of medicine through structure-based drug design. The increasing number of protein structures being determined has established X-ray crystallography as a reliable and accessible technique.

This chapter presents the methodology in X-ray crystallography, a technique used to obtain the three-dimensional molecular structure from a crystal.

2.1 What is in a crystal?

When molecules in solution come together and self- assemble into a highly ordered and periodic manner, a *crystal* is formed. An *asymmetric unit* is the smallest building block of a *unit cell*. For chiral macromolecules like proteins, there are 65 combinations of rotational and translational symmetry that can be applied to the asymmetric unit for reconstructing the entire unit cell. The number is limited to 65 out of 230 *space groups* because of the chirality of the proteins. Some combinations of symmetry operations such as mirror images and inversions are not possible since they do not generate an identical copy of the molecule but instead change the molecule's handedness.

2.2 No crystal, no x-ray crystallography

For a crystal grower, the path to a single and well-defined crystal is never trivial. Moreover, a crystal grower has to devise a systematic approach or else the time required to test all parameters is not feasible. Knowledge of the intrinsic properties of the desired protein, a good grip of the basic principle of crystallization and recognizing the caveats of the available statistics to be tested are important in crystallization. At some point during crystallization trials, a crystal grower has to make a decision whether to continue pursuing yet another crystallization condition or to abandon the construct and take up a new one.

Crystal formation is possible because well-ordered protein molecules self-assemble periodically. This formation entails *conformational homogeneity* so that units can stack side by side in highest order possible. The intermolecular forces that hold these protein molecules are specific and weak, owing to the proteins *flexibility* and are thereby easily broken. These intermolecular interactions are made up of low binding energies such as hydrogen bonds, van der Waal contacts, salt bridges and hydrophobic interactions. In between this weak and specific network of intermolecular interactions are large gaps of *solvent channels* that serve as the medium for small molecules to migrate and bind to active sites and heavy atoms that reach and bind the molecules during soaking.

2.3 The crystallization diagram explained

The purpose of a crystallization experiment is to reduce the protein solubility in the solution to the point that phase separation occurs. First, a protein-rich phase is formed that is supersaturated then a nucleation event triggers the molecules to self-assemble into a well-ordered crystal lattice.

The crystallization diagram is separated into three regions; called the *stable, metastable* and *unstable*. Within the stable region, the solution is in a single phase and the protein molecules are spread out and hardly making contact with each other. The protein molecules are surrounded by the

solvent (water and precipitant). As the protein solubility decreases, the solubility line is crossed and the phase transitions into the metastable region. Within this region, the solution becomes supersaturated and a protein-rich phase is formed. Here, collisions and contacts between protein molecules are more frequent. However, the metastable region is subdivided into two zones; *heterogeneous nucleation zone* and *homogeneous nucleation zone*. In the heterogenous nucleation zone, nuclei sites are difficult to form but can grow on externally introduced nuclei. Few protein molecules gather momentarily together and then eventually dispersed. Whereas in the homogenous nucleation zone, the solution is supersaturated enough to spontaneously induce nucleation sites. After nuclei form, the molecules move towards the nucleation zone for slower crystal growth that yields better crystals[71].

More often than not, microcrystal showers appear in the drop. This happens when supersaturation is nearly approaching the unstable phase. Here, multiple nuclei sites are formed and upon induction of many nuclei, the supply of protein molecules becomes limited and growth becomes stunted. These microcrystals could be seeded into fresh drops (Figure 2.1).

2.4 Growing crystals by vapor-diffusion

There are various crystallization techniques that can be employed to grow crystals. The one used in this thesis is the vapor-diffusion technique, both the hanging-drop and sitting drop versions. The hanging drop method is used for manual setups while the sitting drop method is more suited and optimized for robotic setups. The hanging drop method typically uses a 24-well Linbro plate or another plate with similar caliber, equal amounts (typical volume of 1 μl) of protein solution and reservoir solution are mixed on a siliconized cover slide and the cover slide is flipped over to seal the greased-rim well containing the mother liquor. In this closed system, the water diffuses from the protein-reservoir hanging drop into the reservoir, since the reservoir has double the concentration compared to the drop. As

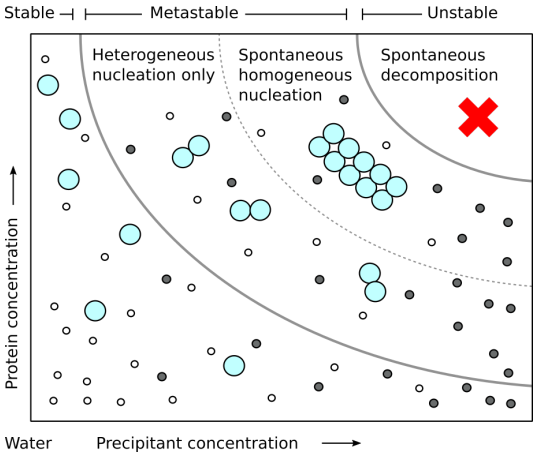

Figure 2.1: The crystallization diagram is divided into three regions; the stable, metastable and the unstable. The heterogenous nucleation zone and homogenous zones nucleation zone subdivided the metastable region.The open small circles are water molecules while the closed small circles are precipitant molecules.

the water vapor diffuses out of the drop, the protein and precipitant concentration increases in the drop and phase separation takes place. Because of supersaturation, nucleation is induced and crystals may form [71, 72].

The sitting drop method applies the same principle in a closed system but the drop is placed in a shelf slightly elevated from the mother reservoir. This method is optimized for automation using a 96-well plate (Figure 2.2).

In Paper 2 and Paper 3, sparse matrix screening kits for soluble and membrane proteins were initially used to sample the crystallization conditions that are known to drive crystal formation. Grid screens were set up for both proteins wherein the concentrations of salt, polyethylene glycol PEG and their pH were varied. In addition, additive screens were used for Paper 3 to improve the diffraction quality of the crystals.

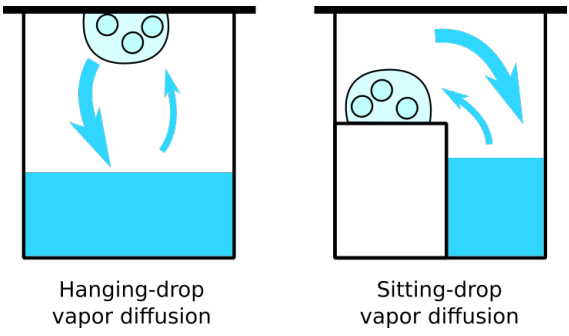

Figure 2.2: Two methods of vapor diffusion: The hanging drop method (left) and the sitting drop method (right). The closed system wherein water diffuses from the drop into the reservoir

2.5 Cryocooling and cryoprotection

Once a crystal is formed, the next step is to harvest the crystal from the mother liquor. This is done by scooping or fishing the crystal with a nylon loop and immediately plunging it into the liquid nitrogen to be flash cooled. *Cryocooling* the crystals reduces the severity of the radiation damage caused from the exposure of the ionizing X-ray radiation. However, cooling down the crystals to cryogenic temperatures also often causes the mother liquor surrounding the crystal to form crystalline ice. The presence of ice around and inside the crystal greatly diminishes the diffraction quality of the crystal. To prevent this from happening, cryoprotecting the crystal has become vital during harvesting. To cryoprotect the crystal, it is carefully dipped into the cryoprotectant immediately before flash cooling.

There are a variety of cryoprotectants that can be used, the most common ones are low molecular glycerol, PEG, ethylene glycol and sucrose which could be mixed in a buffer. In Paper 2 and Paper 3, 10 % of glycerol is mixed with the mother liquor as cryoprotectant and crystals was carefully dipped before flash cooling.

2.6 Diffraction of X-rays by a crystal

X-rays are high energy electromagnetic radiation. As they pass through a crystal, they scatter on the electrons of the atoms. If the resulting scattered waves superimpose in phase, the waves get amplified producing constructive interference. The scattering from a single molecule is very weak and very hard to detect. However in a crystal lattice, because of the periodic arrangement of the molecules, the total scattering intensities are amplified due to constructive interference. Bragg's law describes the conditions necessary for the emanating waves to interfere constructively and give maximum diffraction. Bragg's law predicts the reflection of X-rays on the set of lattice planes hkl, as a function of the diffraction angle θ and the lattice spacing d and X-ray wavelength λ if

$$2dsin\theta = n\lambda \tag{2.1}$$

where n is an integer.

2.7 The diffraction geometry

The basic idea of a diffraction experiment is that there is a source of X-rays, a rotating device for the crystal (goniostat) and an X-ray detector placed behind the crystal to record the intensities of the incoming reflections. How does this setup correlate to the diffraction condition? In 1921, Paul Ewald applied the Bragg's law (diffraction condition) into the diffraction geometry in a diffraction experiment as a way of sampling the reciprocal space.

The scattering vector S_{hkl} from the indices h,k,l of crystal planes in real space corresponds to the reciprocal space vector d^*_{hkl} that are both perpendicular to the reflecting planes. On an *Ewald construction*, given with a sphere radius of $1/\lambda$, the relationship of the reciprocal lattice vector to the reflection condition is described by the equation $d^*_{hkl} = \frac{1}{d_{hkl}} = \frac{2sin\theta}{n\lambda}$. Therefore, $d^*_{hkl} = S = \frac{2sin\theta}{\lambda}$, diffraction occurs if the S vector is collinear and has equal radius with the reciprocal space vector d^*_{hkl}. This is what happens

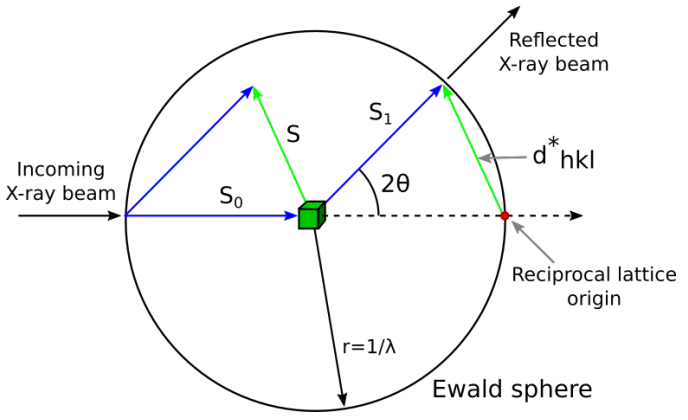

Figure 2.3: The Ewald construction applies the Bragg's law into the diffraction geometry in a diffraction experiment as a way of sampling the reciprocal space.(Adapted from Rupp [71])

when the reciprocal lattice point h,k,l lies on the Ewald sphere (Figure 2.3). Only a fraction of reciprocal lattice points lie on the Ewald sphere. At any given crystal orientation, by incrementally rotating the crystal, this brings other lattice points onto the Ewald sphere, thereby sampling more lattice points and yielding more diffraction spots.

2.8 Data collection

A space group has to be identified so that each recorded reflection is labelled with (at least) one Miller index (*indexing step*). Since the lattice symmetry is often unknown, the pattern of their diffraction vectors is used to rationalize the potential candidates (*autoindexing*). By integrating the indexed spots, the total intensity of the reflection is calculated. Indexing and integrating can be done with programs like MOSFLM [73] and XDS. In data reduction, program SCALA [74] or Aimless [75], scales the intensities of equivalent reflections and merges partial observations to full and finally merges symmetry equivalent observations together.

2.9 The structure factor and B-factor

During a diffraction experiment, an X-ray detector measures the photon count at a given position. The intensity of the signal or the number of photons in each reflection is determined by the molecules in the real space crystal lattice. The measured intensity of each reflection with Miller indices h,k,l is directly correlated to the magnitude of the structure factor amplitude. Hence, "the structure factor is the summation of the individual scattering contributions of each and every atom" [71]. It also a complex vector with amplitude and phase information. However, the phase information is lost during the detection process, giving rise to the phase problem in X-ray crystallography. The phase information is very crucial in the reconstruction of the electron density by Fourier Transform of the x-ray scattering amplitudes.

As it happens, the mean positions of the atoms are disrupted due to thermal vibration and other type of displacements. Displacement of the atoms attenuates the scattering contribution of the atoms, resulting in poor diffraction intensity. Higher diffraction angles are more affected by atomic displacements and crystals with high disorder limit the crystallographic resolution (d_{max}). The extent to which the atoms are displaced also signifies the degree of disorder in the crystal lattice. Attenuated scattering in this scenario is measured by the *isotropic displacement parameter* or the *B-factor*. Usually the B-factor is inversely correlated to the occupancy of the atom. If the B-factor is high, this could mean that the diminished scattering contribution is due to less than full occupancy of the atom.

2.10 The electron density reconstruction

A Fourier transform converts a list complex structure factors (reciprocal space) to the electron density (real space) of the scattering molecule. In the reconstruction of the electron density, there are two Fourier coefficients that are essential; the structure factor amplitude and the phases.

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}^{-h}\sum_{k}^{-k}\sum_{l}^{-l} F_{hkl} \cdot exp[-2\pi i(hx + ky + lz - \alpha_{hkl})] \tag{2.2}$$

The structure factor amplitude is proportional to the square root of the measured intensity that is obtained during the diffraction experiment while the phases come from external sources such as experimental phasing or from a structurally similar protein.

2.11 The Patterson function

Unlike the electron density map that needs both the structure factor amplitudes and the phases to be reconstructed, the Patterson map can be directly computed solely from the observed intensities. The map describes the interatomic distance vector *uvw* between atoms. The value of the Patterson function during its application in acquiring the missing phase information, cannot be overstated. The Patterson map is useful in determining the marker atom in both isomorphous and anomalous difference data and finding out the orientation of the search model in a unit cell for molecular replacement.

$$P(u,v,w) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l} F_h^2 cos2\pi(hu + kv + lw) \tag{2.3}$$

2.12 The phase problem

In *experimental phasing*, acquisition of the phases relies on the difference intensity between data sets of the heavy or anomalous marker atoms. Single isomorphous replacement (SIR) and multiple isomorphous replacement (MIR) methods require derivative crystals soaked with heavy atoms. The intensity difference between the derivative crystal data and native crystal data is compared. In anomalous diffraction methods from crystals containing anomalous scatters, single-wavelength anomalous diffraction (SAD) and

multiple-wavelength anomalous diffraction (MAD), uses anomalous difference and dispersive difference data measured at different wavelengths respectively. Both difference methods require isomorphism when multiple crystals are used.

If there is a known structure that is similar to the sought-after protein structure, molecular replacement is the most common phasing method that is used. The method may work with as low as $\sim$ 30% sequence homology. Initially, the orientation of the search model is determined by applying three-dimension cross rotation search using a Patterson function. This is followed by probing the actual location of the search model in the unit cell through a translational search. Programs commonly used for molecular replacement are PHASER [76] and Molrep [77].

In reconstructing the electron density, it is the phase of the structure factor that holds most of the structural information. In obtaining the phases through *molecular replacement*, *phase bias* is an inherent problem, because the starting electron density map will be highly biased towards the search model.

2.13 Model building and refinement

Model building and refinement consists of alternating rounds of building the structural model by fitting it into the electron density in real space and by fixing geometric error in restrained reciprocal space refinement. The process is iterative: one goes back to the model building and guided by improved electron density maps and then refine again.The idea is as this process is cycled several times, the maps progressively improve and the final molecular structure describe the experimental data well and the spatial distribution of atoms make chemical and physical sense. COOT [78] is the computer graphic that used is in model building while refinement is often carried out by REFMAC [79].

Difference maps are used to aid in interpreting the electron density and minimizing the model bias. Difference maps are generated from different combinations of the Fourier coefficients. To generate the difference map,

$(F_o - F_c) \cdot exp(i\phi_c)$, the Fourier coefficients used are the observed structure factors from the measured intensity F_o and the calculated F_c structure factors from the model, the phases ϕ_c is also calculated from the model. The negative density region in the $(F_o - F_c) \cdot exp(i\phi_c)$ map indicate where the density should not be present while the positive density regions indicate where there is a missing electron density. The $(2F_o - F_c) \cdot exp(i\phi_c)$ map is more suitable for the initial model building as the positive density signal is strengthened.

To assess the overall fit between the observed data and the model, the linear residual or the R-value is calculated. This uses the F_{obs} which is the structure factor amplitude from the measured intensity and the F_{calc} , the calculated structure factor amplitude from the model:

$$R = \frac{\sum_h |F_{obs} - F_{calc}|}{\sum_h F_{obs}} \tag{2.4}$$

During the course of the structural refinement, the R-value decreases as the fit between the observed structure factor amplitudes (the diffraction experiment) and computed structure factor amplitudes (the model) is improved.

2.14 Cross validation

As the parameters are adjusted during the refinement, the model improves and the R-values decreases. However, overparameterization is a real risk in reciprocal space refinement and the R-value will not indicate a problem. Specifically, adding unnecessary parameters to the reciprocal space refinement, results in a dropping R-value. To act as a restraint for R-value, Axel Brunger introduced cross validation in the form of 'R-free' [80].

The idea is by setting aside $\sim$ 5% of the reflections from the experimental data and excluding them from the refinement, a comparison can be judged between how well the model fits with the excluded 'test data set' which is 'R-free' and how well the model fits with the working data set,

which is R-work. How 'R-free' and R-work follow each other is monitored during successive model rebuilding and refinement steps. If parameters are unnecessarily added that results in no physical improvement on the model, 'R-free' will stop improving and may even increase whereas the R-work value will still continue to drop. Too wide a gap between R-free and R-work is an indication for overfitting and the accuracy of the model is compromised, whereas too close 'R-free' and R-work is an indicator that more parameters can still be introduced to improve the model.

While 'R-free' can serve as a cross validation in reciprocal space [80], omit maps can provide a cross validation in real space. As the name suggests, *omit maps* are generated by omitting the structural model around questionable region from the density calculation and scrutinizing if the problematic region is still present in the omit map.

Structural validation is the next step when refinement reaches convergence. The program MolProbity [81] generates an error report of the structural model. After checking the geometry of the backbone torsion angles ϕ and ψ and clearing the structure for Ramachandran outliers, the polished structure is now ready for biological interpretation. Depositing the newly determined structure to the *Protein Data Bank* (PDB) also provides an additional validation report.

Chapter 3

Fn N-acetylmannosamine kinase

In Paper I, we describe a complete automated method of optimizing the cloning, expression and purification of eighteen enzymes [82]. These enzymes are orthologues from the *nan-nag* gene cluster regulating the sialic acid catabolic and surface sialylation pathways from the pathogens *Vibrio cholera*, *Pasteurella multocida*, *Fusobacterium nucleatum* and *Haemophilus influenzae*. These studies aim to develop a less time-consuming, efficient and cost-effective method of screening wide-ranging multiple constructs for production of functional recombinant proteins to be used in crystallographic studies and functional assays. Results for the expression and solubility tests conducted for FnNanK from this paper are applied for the crystallization and structure determination of FnNanK described in Paper 2.

3.1 Automated cloning

We implemented an automated cloning strategy for preparing the constructs for optimization in expression and purification of the sialic acid catabolic and sialylation pathways enzymes. For efficient and quick dispensing of liquids, a Liquid Handling System was used during the entire automation process. The desired gene flanked by the 15bp attB sites to be cloned into the Gateway vector pET300/NT-DEST was PCR amplified using a thermocycler in 96-well plates. The Gateway cloning technology [83] allows the

efficient transfer of DNA sequences into multiple vectors without the subcloning steps involving restriction enzymes. After generating the entry vector comprising the gene of interest and a cloning vector engineered with lambda phage recombination sites, it is convenient to transfer the gene of interest into multiple vectors for expression or functional studies. The constructs were transformed into competent Dh5α cells in a 96-well plate and after incubation the reaction mixture was plated onto a 6-well plate with LB Agar. Distinct colonies were isolated for plasmid amplification and sequencing (Figure 3.1).

3.2 Expression and media optimization

After confirmation by sequencing, all 18 constructs were transformed into three different *E.coli* expression strains namely BL21(DE3), BL21(DE3)PLysS and Rosetta 2. Subsequently, colonies that grew after the transformation were selected by an automated colony picker and inoculated into two different culture media, 2x YT and ZYM media for seed culture. Overnight seed cultures were transferred into 24-well plates with 2x YT or ZYM media for expression screen. The plate containing ZYM autoinduction media [84] yielded optimal expression of the constructs whereas the 2x YT media plate has very low to no expression. This is largely because of the varying time for each construct to reach the optimal optical density. Generalized IPTG induction may have affected the expression quality of each protein.

Afterwards, the cell pellets were normalized and lysed using Bacterial Protein Extraction Reagent (BPER) and transferred into a 96-well plate. The expression and solubility of each construct was analyzed by running the total lysate and the soluble fractions on Caliper Gx (Figure 3.1). Of the 18 constructs, 11 yielded soluble protein expression. The *E. coli* BL21(DE3) strain is the most successful for 9 of the 11 constructs while 4 and 3 constructs gave soluble expression for BL21 (DE3) PLysS and Rosetta (DE3) respectively.

3.3 Purification and scaling up

Viable constructs for overexpression were purified using Ni-NTA-based 96-well filter plates and load, flow through and elutes fractions were collected separately and analyzed by running them on Caliper GXII system. Deriving from the automated optimization screening, the large scale production and purification of constructs FnNanA, FnNanE, FnNanK, HiNanK, followed by crystallization yielded crystals. The automated pipeline of cloning, expression and purification described here is efficient and cost-effective. Production of proteins suited for structural studies usually starts with screening of a broad base of constructs. Along the way, bottlenecks are encountered and these problems are circumvented in a trial and error process. Streamlining the screening for the optimal conditions for expression, solubility and purification of several constructs in 96-well plates, facilitates the identification of viable constructs. Further automation optimizes conditions that therefore yield successful purification with minimal manual intervention. Unlike previous automated pipelines [85, 86], Paper 1 describes an integrated pipeline with absolute minimal manual intervention in cloning, expression and purification.

The successful large scale purification of FnNanK was based on the automated small scale screening for optimal expression and purification of 18 sialic acid catabolic and sialylation enzymes from Paper 1. The crystallization and the ensuing structure determination of FnNanK was facilitated by the automated process.

3.4 Fn N-acetylmannosamine kinase crystal structure

Fusobacterium nucleatum (Fn) N-acetylmannosamine kinase (NanK) catalyzes the transfer of a phosphoryl group from ATP to C6 of ManNAc yielding ManNAc-6-P in a degradation pathway that ultimately breaks down sialic acid to fructose-6-P and thereby entering glycolysis (Figure 3.2).

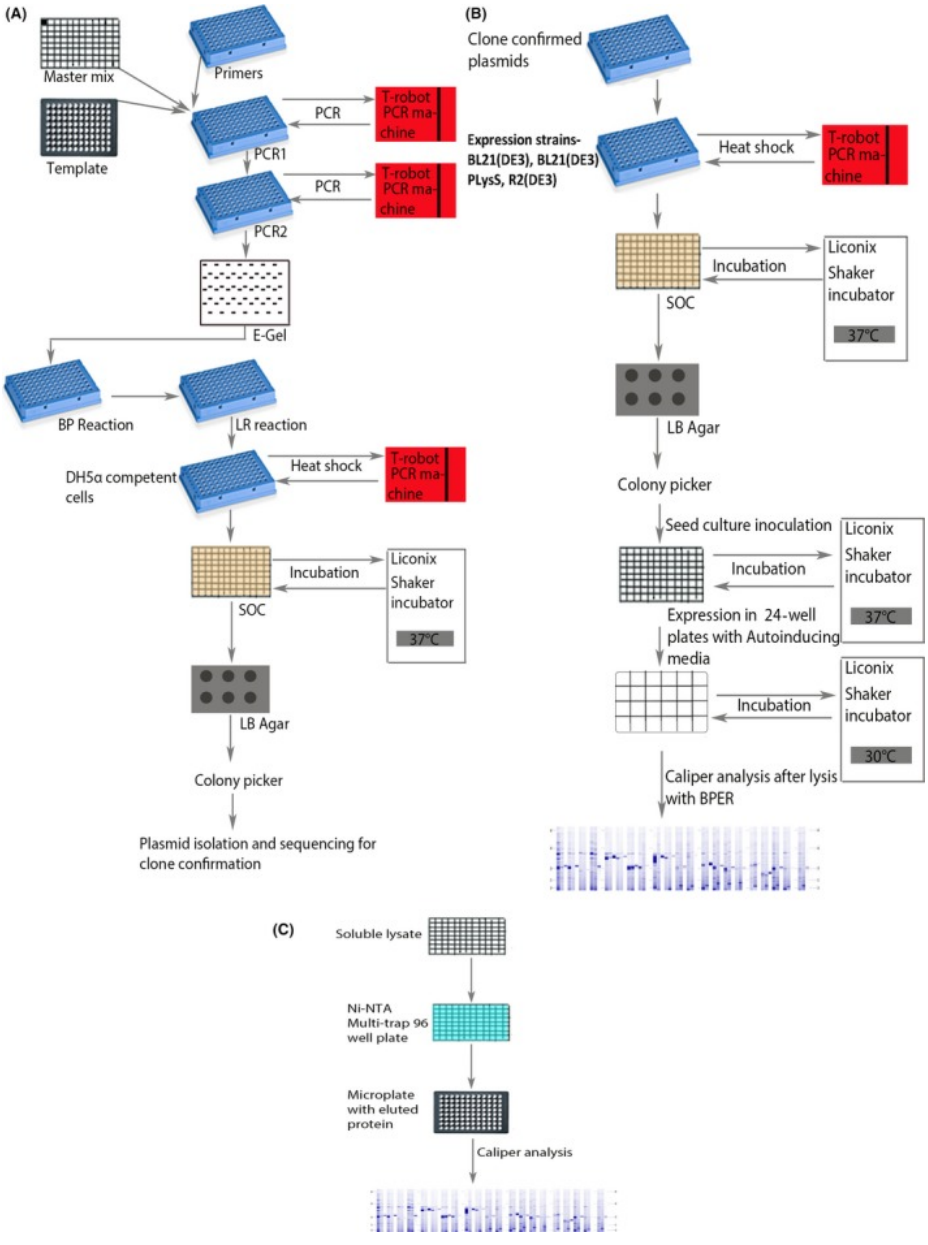

Figure 3.1: Flow charts for automated cloning (A), expression(B) and purification (C) on the liquid handling system.The entire pipeline is automated with the most minimal manual intervention (Reproduced from Bairy et al [82].)

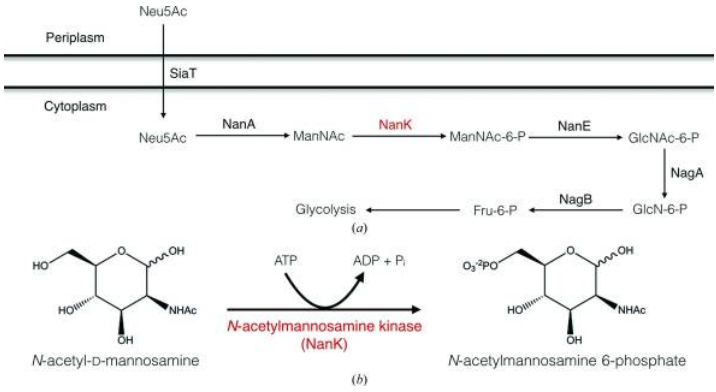

Figure 3.2: The second step in bacterial sialic acid catabolic pathway; N-acetylmannosamine kinase phosphorylates ManNAc to ManNAc-6-P. (Figure reproduced with permission of the International Union of Crystallography. [87])

In Paper 2, the apo crystal structure of FnNanK was determined at 2.23 Å in resolution and structurally aligned with other bacterial Nan kinases available in Protein Data Bank PDB [87]. FnNanK structural alignment with the human N-acetylmannosamine kinase domain of the bifunctional UDP-N-acetylglucosamine 2-epimerase / N-acetylmannosamine kinase (hMNK) [88], the only crystal structure of nan kinase bound with both ManNAc and ATP and well-characterized, provided insight into the putative active site of FnNanK.

3.5 The ROK superfamily

FnNanK belongs to the repressor, open reading frame kinase (ROK)[89] superfamily which largely consists of transcriptional repressors, sugar kinases and other unknown gene clusters. Members of the ROK family share signature features such as: (i) nucleotide-binding site with a DxGxT consensus sequence on the N-terminal region; (ii) a highly conserved aspartate within the active site loop; (iii) an ExGH motif that interact with the sugar

substrate; (iv) and a zinc-binding motif with the consensus sequence Cx-CGxxGCx(E/D) [90]. Zinc-binding is believed to have an integral role for upholding the structural stability in the active site.

3.6 Protein production, purification and crystallization

FnNanK was overexpressed and purified using immobilized metal affinity and size exclusion chromatography. The homogeneity of the collected fractions were analyzed through SDS PAGE. The collected fractions were concentrated to 14 mg/ml which was then used for the ensuing crystallization trials. Within one week, rod-shaped crystals formed under reservoir conditions of 0.2 M Lithium sulfate monohydrate, 0.1 M Tris-HCl pH 8.5 and 30% (w/v) PEG 4000 through sitting drop vapor diffusion method. The crystal diffracted to 2.2Å and the structure was solved by molecular replacement.

3.7 FnNanK overall crystal structure

FnNanK has a homodimeric structure resembling a butterfly [87] (Figure 3.3). Each V-shaped monomer contains 11 alpha helices and 11 beta-sheets with alpha/beta domains and an inner cavity. The domains are held together by hinge loops. Each protomer consists of an N-terminal domain and a C-terminal dimerization domain. The N-terminal domain begins from the N-terminus until residue 118 then proceeded from residue 272 ending towards the C-teminus. The C-dimerization domain is from residue 126-268 (Figure 3.4). Within the cavity of the V-shaped monomer, situated between the N-terminal and C-terminal dimerization interfaces, is the putative binding site of the substrate ManNAc (Figure 3.3).

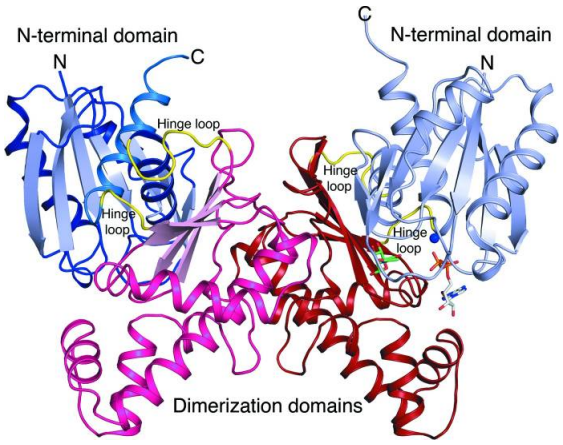

Figure 3.3: The homodimeric structure of FnNanK determined at 2.2 Å in resolution.(Reproduced with permission of the International Union of Crystallography. [87])

3.8 The putative binding site

Superimposition of the FnNanK structure with the ligand-bound human NanK (hMNK; 2yhy PDB) shows high similarity in the overall fold with a root mean square (r.m.s.) deviation of 2.5 Å, alignment length of 280 Cα atoms and 23% sequence identity. The highly conserved residues in human NanK (Asn516, Asp517, Arg477, Glu566, His569 and Glu588) that are required for the coordination of ManNAc are superimposable with corresponding residues (Asn106, Asp107, Gln67, Glu156, His159 and Glu168) in FnNanK.

Further structural alignment of the residues in the putative binding site region with two other bacterial Nan kinases, *E.coli* NanK (EcNanK; 2aa4) and *L. monocytogenes* NanK (LmNanK; 4htl) shows that they are superimposable with the r.s.m.d of 2.6 Å and 1.9 Å, respectively. EcNanK residues (Asn104, Asp105, Ile66, His153, His156 and Glu 175) and LmNanK residues (Asn102, Asp103, Tyr64, Glu152, Tyr155 and Asn172) which are involved in the substrate binding are superimposable to the putative binding site region

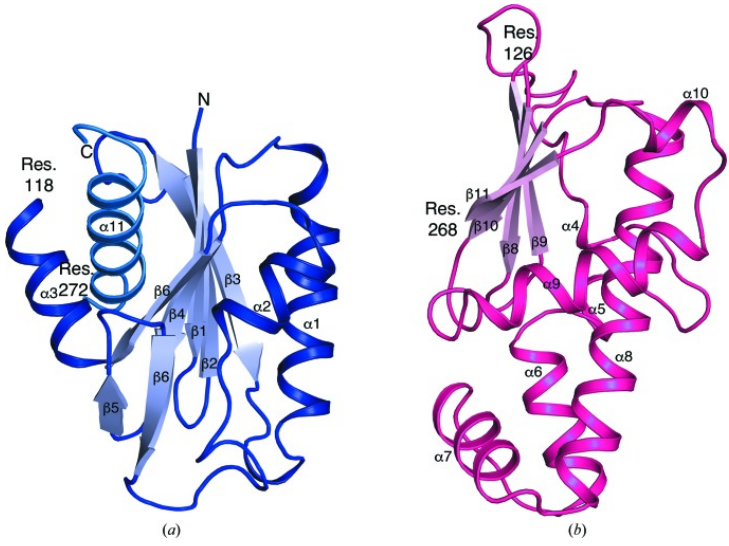

Figure 3.4: Each monomeric structure of FnNanK consists of N-terminal domain (blue) and C- terminal dimerization domain (pink).(Reproduced with permission of the International Union of Crystallography. [87])

of FnNanK (Figure 3.5).

3.9 Lack of the zinc-binding site

FnNanK lacks the zinc-binding motif that is considered to be a shared feature of the members of the ROK superfamily. The consensus sequence Cx-CGxxGCx denoting the zinc-binding region is absent in FnNank, while the site is eliminated by substitution in LmNanK . His159 in FnNanK is highly conserved corresponding to residues His156 in EcNanK and His569 in human NanK. In hMNK structure [88], this highly conserved histidine residue is coordinated to both the zinc ion and the substrate ManNAc. Both the corresponding strictly conserved histidines in hMNK and EcNank are notably shifted compared to the uncoordinated histidine His159 in FnNanK.

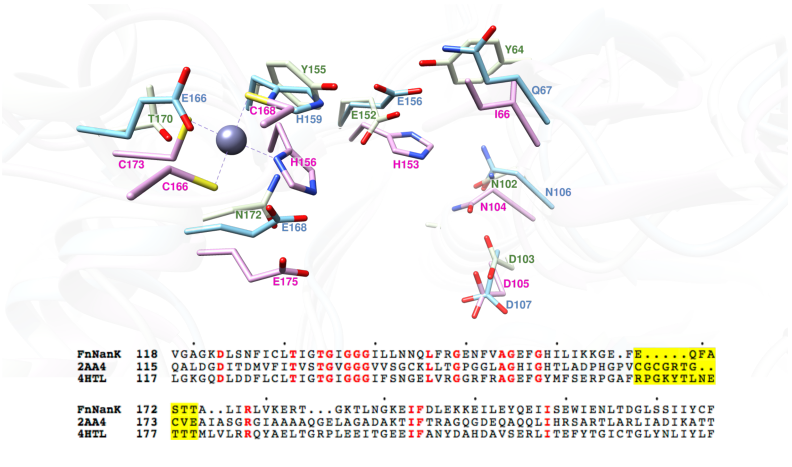

Figure 3.5: The lack of zinc-binding site. Structural alignment of FnNank with *E. coli* NanK and *L.monocytogenes* NanK.(Reproduced with permission of the International Union of Crystallography. [87])

Instead of three cysteines that supposedly coordinate with the zinc ion, a glutamate residue is present in FnNanK (Figure 3.5).

More than being a signature feature of the ROK superfamily, the zinc-binding motif is implicated in upholding the structural integrity of the active site. Studies in cysteine deletion mutagenesis in the cysteine-rich region in the zinc-binding site exhibit inactivation of the *B. sublitis* glucokinase [91]. As observed in the hMNK structure, the zinc ion coordinates with three cysteines and the strictly conserved histidine residue His569, which in turn coordinates with the substrate in the active site, linking the zinc-motif region to the ManNAc binding site. Mutating two cysteines associated in the zinc-binding in *E.coli* Mlc repressor, compromised its repressor function [92]. Despite of the absence of the zinc-binding region in FnNanK, all its major structural features are still conserved. To gain insight into the conformational change that takes place upon binding of a substrate, an FnNanK structure bound ManNAc and ATP analogues is required.

3.10 Summary

Automation of the process of optimizing cloning, expression and purification ofthe 18 enzymes associated to the sialic acid degradation and surface sialylation paved the way for finding optimal conditions for the subsequent large scale production, purification and crystallization of these enzymes. One of them was FnNanK.

The FnNanK crystal structure shares structural motifs characteristic of the ROK family. One of these motifs is the presence of a zinc binding site which is known to be important for structural architecture in homologous enzymes. This motif however, is notably absent in FnNanK. Even though FnNanK lacks the zinc-binding region, its main structural features are well preserved.

Chapter 4

Sialic Acid TRAP Transporters

Embedded within the lipidic bilayer of the cellular membrane, membrane proteins have an integral role in numerous biological processes. They serve as receptors, channels and transporters. Their crucial functions in many cellular processes make them a major drug discovery target. Structural information provides valuable insight to important mechanisms guiding drug development.

In 1985, the photosynthetic reaction center of *Rhodopseudomonas viridis* was the first membrane protein structure to be determined by x-ray crystallography [93]. Since then, 827 unique membrane protein structures have been solved [94]. Nevertheless, the first membrane protein structure determined for heterologous expression, the potassium channel KcsA [95] was elucidated more than a decade after *R. viridis* photosynthetic reaction center. However the growth of membrane protein structure determination was not exponential. Membrane protein structure determination lags behind soluble proteins [96].

The onset of heterologous expression systems used for high yielding overexpression of recombinant protein; development of new detergents for improved protein solubility and crystallizability; and recent technological advancements in synchrotron and beamlines, are all factors that greatly contributed to the steady growth of membrane protein structure entries in Protein Data Bank [97–99].

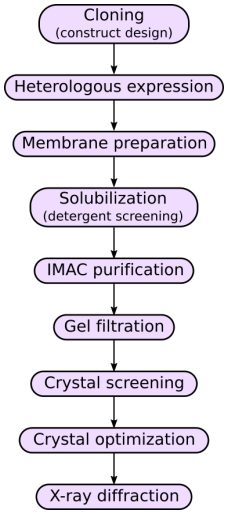

Figure 4.1: Workflow of the production and purification of membrane protein aimed for structural studies.

Despite this progress, the path to membrane protein structure determination is paved with obstacles and bottlenecks. Membrane protein production and crystallization remain an empirical process (Figure 4.1). In this chapter, the methodology of membrane protein production, solubilization, purification and crystallization is introduced. In the same vein, the entire pipeline of sialic acid TRAP transporters crystallizability in Paper 3 will be discussed.

4.1 Construct design and expression system

Once the decision to use *Escherichia coli* as the host organism is taken, the entire pipeline of membrane protein production to purification can be outlined. Of the bacterial expression systems, *E. coli* is often the preferred host for heterologous membrane protein overexpression. Its advantages are short generation time, tolerance of high cell density, it is inexpensive and suitable for transformation by exogenous DNA and the wild-type *E. coli*

strains are modified into a variety of expression strains that are designed for different expression system [100, 101].

In Paper 3, we used the araBAD and T7 expression systems, which are inducible by arabinose and IPTG respectively. Four orthologues of the sialic acid TRAP transporter were cloned into pBad and pWarf vectors for regulated production of the desired recombinant protein. It is established that heterologous recombinant membrane protein expression is considered toxic for the bacterial host cells [102, 103]. Unchecked expression of the recombinant membrane protein saturates the cellular capacity of the host and overloads the Sec translocase pathway. The accumulation of protein triggers cell stress response that impede cell growth and even cause cell death [104].

Control of expression is the key in recombinant membrane protein production. Quantity does not necessarily entail good quality, stable and properly folded membrane proteins. For this purpose, a low copy vector such as pBad [105] and pWarf [106] are suitable for membrane protein expression. Their tightly regulated and finely tunable systems are designed to circumvent the problems of misfolding and aggregation resulting in the formation of inclusion bodies.

As described in Paper 3, the araBAD expression system yielded better results for PmTRAP compared to the T7 expression system. Although both systems can be regulated, the mechanisms employed are distinct. The T7 expression system [107] is known to suffer from leaky expression, because the lacUV5 promoter that transcribes T7 RNA polymerase is not completely shut off by LacI [108, 109]. Thus, there is a steady pool of T7 RNA polymerase being transcribed and this subsequently leads to basal expression of the recombinant protein that is toxic to the host cells.

To reduce leaky expression of the T7 expression system, we used the Lemo21(DE3) strain [110] that counteracts the low level expression of T7 RNA polymerase with T7 lysozyme, a known T7 RNA polymerase inhibitor. Expression of the T7 Lys is controlled by the rhaBAD promoter, which is inducible by L-rhamnose. In practice, the amount of L-rhamnose is optimized in order to increase the yield of the recombinant protein during test expression. For Pm TRAP the yield was still poor, despite this optimization

procedure (Figure 4.2a).

Compared to the T7 expression system, the araBAD system is a positively controlled expression system. Adjacent to araBAD operon, AraC functions both as the activator/repressor on the araBAD promoter. In the absence of an arabinose inducer, araC binds to two sites on the operon forming a loop that hinders RNA polymerase from binding to the promoter. Upon binding to the inducer, the loop is relaxed as both araC and CAP function as activators by promoting the binding of RNA polymerase to the promoter.

Temperature affects protein overexpression. Lowering the temperature slows down the rate of protein production thereby giving the Sec translocase machinery ample time to fold the recombinant membrane protein properly and thereby prevent the saturation of the signal recognition particle (SRP) pathway. The optimal temperature for membrane protein production is reported to be 20-30 ^{o}C [111]. In Paper 3, we lowered the incubation temperature from 30 ^{o}C to 21 ^{o}C after induction for an 18 h culture for better yield.

4.2 Detergent and solubilization

Like lipids, detergents are amphiphatic and possess a hydrophilic headgroup and a hydrophobic tail. Detergents are used to extract membrane proteins from their natural lipidic environment and shield the hydrophobic regions of membrane proteins. The protein becomes soluble in the aqueous media, and in optimal cases, with its native conformation intact [112]. If the hydrophobic transmembrane regions are not shielded this may lead to aggregation of membrane proteins.

At low concentrations, detergents are present as monomers. When the detergent concentration reaches a critical point, the detergent monomers form spherical micelle structures as the hydrophobic tails associate in the core of the micelle and the hydrophilic head groups form hydrogen bonds with water. The minimal detergent concentration required for detergent monomers to form micelles is known as the *critical micellar concentration*

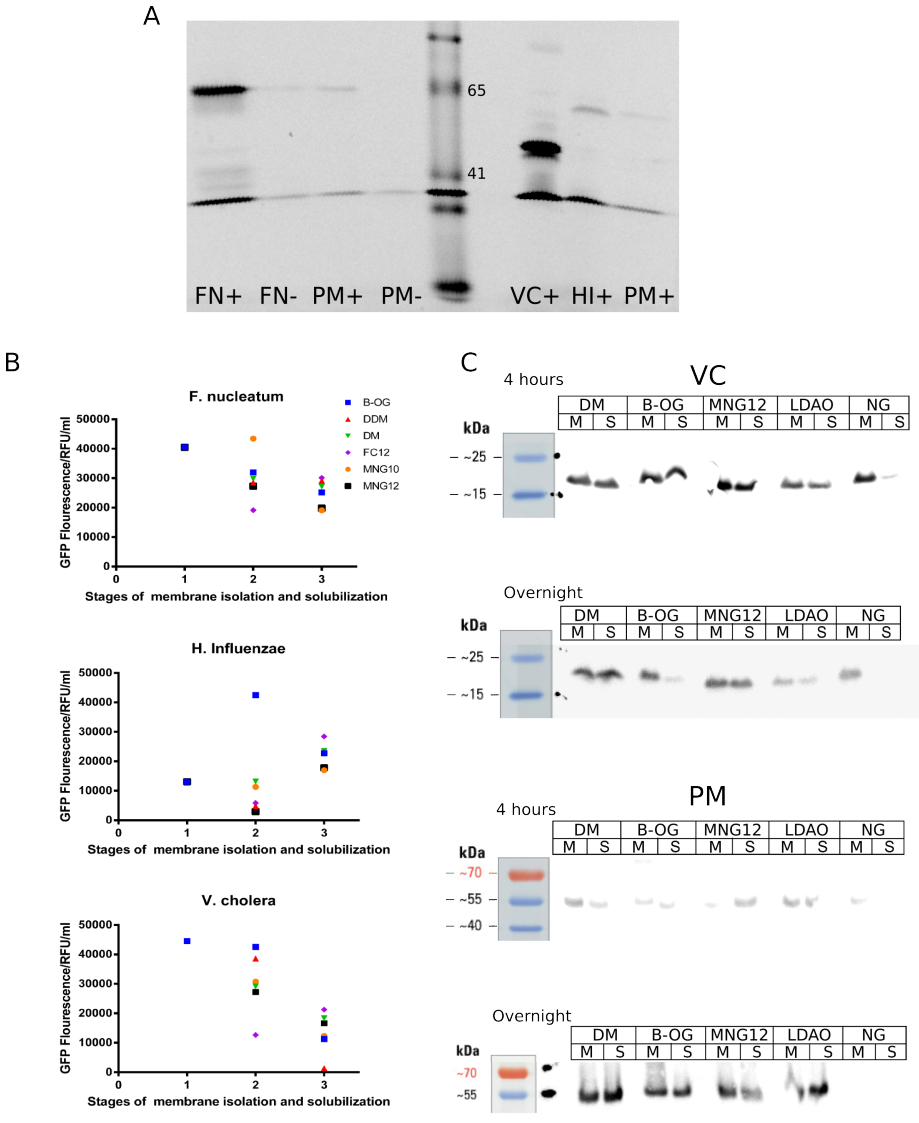

Figure 4.2: TRAP transporters expression and detergent screening. A. pWarf(+) constructs in-gel flourescence to assess levels of expression. The feature at 41 kDa is an artefact. B. Whole cell relative flourescence unit (RFU) for pWarf(+)constructs for Fn, Hi, and Vc detergent screen. C. Vc and PM pBad constructs manual detergent screen after 4 h and overnight treatment of different detergents.

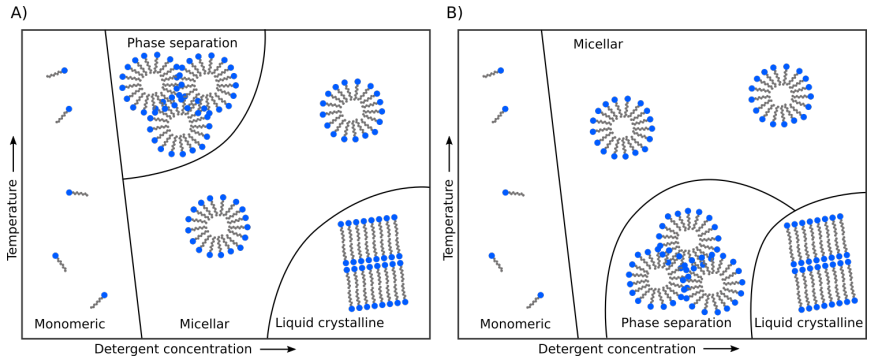

Figure 4.3: Detergent phase diagrams. Phase separation from two types of detergents. A. Phase separation of nonionic detergents has lower consulate boundary (line separating micelle solution and phase separation). B. Phase separation in zwitterionic detergents and glycosidic detergents [103]. (Adapted from Linke [112].)

(CMC) and the minimum temperature above which a micelle can form is called the *Kraft temperature*. Another important aspect of detergents is the size of the micelle, which is determined by the average number of detergent monomers forming the micelle. Detergents undergo phase separation induced at a certain temperature (cloud point) wherein the solution separates into detergent-poor and detergent-rich phase [112–114]. In detergent rich phase, immiscible non-micellar aggregates assemble which can be problematic during crystallization. The phase behavior of the detergent is largely dependent on the physicochemical properties of the detergent together with the ionic strength and the pH conditions of the buffer (Figure 4.3).

The classes of detergents are determined according to their head group charge. *Ionic detergents* are harsh detergents carrying either anionic or cationic group charge. This class of detergent disrupts not only the lipid-lipid and lipid-protein bonds but also the protein-protein bond. *Non-ionic detergents* have uncharged head groups making them a mild and gentle detergent. Not surprisingly, this class of detergent is routinely used for membrane protein extraction. The sugar based maltoside and glucoside are the most

prevalent ones. The *zwitterionic detergents* carry the combined properties of ionic and non-ionic detergents. They are harsher than non-ionic detergent, but not as aggressive as the ionic detergents. Zwitterionic detergents like Fos-choline are capable of disrupting the lipid bilayer, while retaining the proteins native conformation [115, 116].

Solubilization of membrane proteins starts at lower concentration of the detergents, when they penetrate into the membrane bilayer and disrupt it. As the detergent concentration increases, the lipid bilayer disintegrates and the hydrophobic tails of the detergent enwrap the hydrophobic transmembrane region of the embedded protein forming the protein-detergent-lipid complex. This process stabilize membrane proteins in an aqueous solution (Figure 4.4). Harsher detergent treatment leads to *delipidation* of the protein-detergent-lipid complex and removal of the essential lipids can destabilize the structure of membrane proteins.

If after extensive ultracentrifugation of the detergent incubated membrane, the membrane protein remains in the supernatant, it indicates that the solubilization was successful. In Paper 3, multiple detergent solubilization protocols were tested on TRAP transporter orthologues. These orthologues were fused to green fluorescent protein (GFP)[117–119] and produced in T7 expression systems. The fused GFP protein allows one to keep track of the expressed protein during the solubilization process. For the araBAD expression system, the orthologues were tagged with 6x His in the N-terminal region. A conventional detergent screen was applied and the membrane protein purified.

The expression levels of Vc, Hi and Fn TRAP transporters expressed in pWarf system were quantified by measuring relative fluorescence units (RFU) of the whole cell, the supernatant and pelleted fractions. The analysis was repeated with six different detergents. The amount of solubilized Fn TRAP was similar for all detergents, except MNG-10. Hi TRAP was solubilized best using β-OG, while Vc TRAP was soluble in all tested detergent with the exception of Fos-Choline (Figure 4.2b).

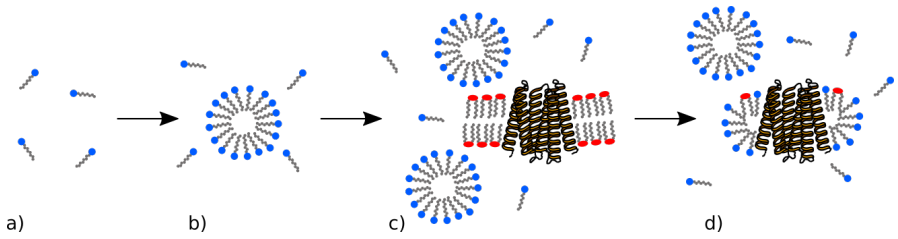

Figure 4.4: Membrane protein solubilization. A. Detergent monomer. B. Detergent in micellar formation. C. At low concentration the detergent disrupts the membrane bilayer. D. With increasing detergent concentration the detergents wraps around the transmembrane hydrophobic region forming the protein-detergent-lipid complex.

Another way to characterize the solubility of the membrane protein incubated in different detergents is through Western Blot analysis. The intensity of the bands corresponds to the amount of membrane protein solubilized after detergent treatment. Pm and Vc TRAP transporters produced via pBad vector were detergent-treated with the membrane fractions (M) before centrifugation and the supernatant fraction (S) after centrifugation compared through Western Blot analysis. For Vc TRAP, 2 h of solubilization shows bands from the supernatant fraction with membranes treated with DM, β-OG, MNG and LDAO. On the other hand after overnight solubilization, the protein was only stable in DM and MNG detergents. Furthermore, the 2 h treatment with DM and β-OG made a small fraction of Pm TRAP soluble, and solubilzation with MNG was more successful. Longer incubation improved the solubilization yield with detergents DM, MNG and β-OG and LDAO showed also showing more distinct bands (Figure 4.2c).

A novel class of detergent maltose neopentyl glycol (MNG) [120] used in Paper 3 proved to be an ideal choice of detergent for solubilization and purification. Its head and tail are fused to a central quarternary carbon derived from neopentyl glycol. Two maltose units are connected to this central atom and form a hydrophilic head. The other side of the central carbon is occupied by two n-decyl chains forming a hydrophobic tail. This detergent has the advantage of having a lower CMC value compared to

other conventional alkyl-glycoside like DDM.

4.3 Chromatography and Purification

The chromatographic method in protein purification aims to separate the desired protein from contaminants. Immobilized metal affinity chromatography (IMAC) is the first purification step used to separate the recombinant protein after the solubilization process. The main idea is to introduce a unique chemical feature such as poly-histidine tag in the desired protein through recombinant DNA techniques. The most widely used is the poly-histidine tag comprising a string of six to ten histidine residues. Usually the histidine tag is located at its N- or C- terminus making it readily accessible for transition metal ions (Ni^{2+}, Co^{2+}, Cu^{2+}, Fe^{2+} and Zn^{2+}) immobilized by a chelating ligand (nitrilotriacetic [NTA], iminodiacetic [IDA]) in a matrix. In the TRAP transporters the six consecutive histidine amino acid residues are inserted before the N-terminus of the membrane protein. The Ni-NTA IMAC method was applied in this thesis and it proved to be highly efficient and a fast purification step.

Size exclusion chromatography (SEC), as it name suggests, separates protein molecules by their size. Large molecules are eluted first as they cannot occupy the porous volume of the stationary phase while the small molecules pass through porous bead volume delaying its elution time. Large aggregates of protein molecules travel in the most mobile void volume and they are eluted first. SEC can also be used to evaluate the homogeneity of the purified protein. Conformational homogeneity of the desired protein is an important requirement in crystallization and it is indicated with a sharp and narrow Gaussian peak in SEC. Multimodal or broad peaks usually indicate oligomers or the presence of more than one population. Thus, SEC is an important tool for evaluating the homogeneity of the desired protein before proceeding with crystallization.

The detection of protein can occur with UV absorption at 280 nm wavelength. Alternatively, if the protein is tagged with a fluorescent protein tag the emission of the tag is detected at 512 nm. This technique is flourescence

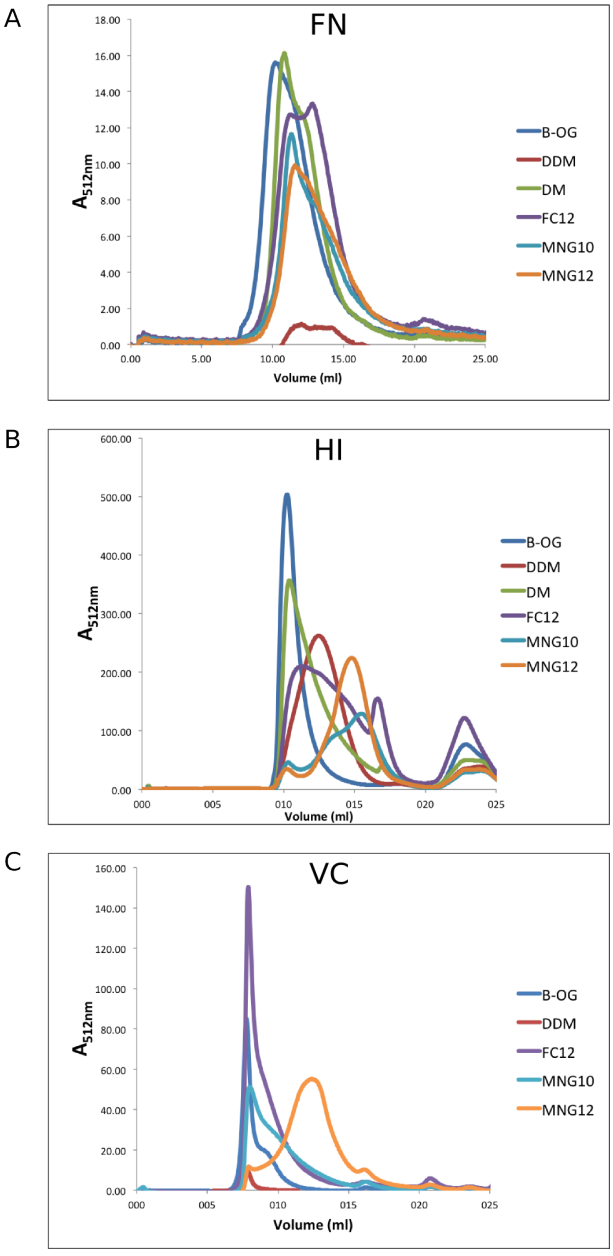

Figure 4.5: Flourescent size exclusion chromatography (FSEC) profiles of TRAP transporters. A. FSEC profile from the pWarf(+)Fn construct. B. FSEC profile from the pWarf(+) Hi construct. C. FSEC profile from the pWarf(+) VC construct .

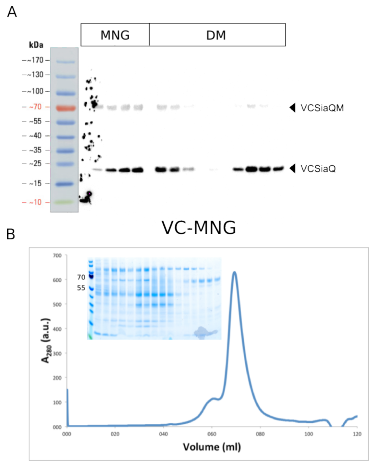

Figure 4.6: Results from the purification of Vc TRAP in the pBad system. A. Western Blot after IMAC purification and using DDM and MNG-12 as detergents. B. SEC elution profile and SDS PAGE using MNG-12 detergent.

size exclusion chromatography (FSEC)[121]. Although fluorescent proteins tend not to affect the stability of the tagged protein, adverse effects cannot be excluded. Initially, FSEC was used in detergent screening of TRAP transporters (Figure 4.5). With the exception of MNG12-treated membranes, all detergent treatment of Vc-TRAP protein resulted in aggregation. This conclusion was drawn from the large amount of protein observed in the void volume. For Fn TRAP, the broadness of the elution peak is an indication of dimeric or oligomeric states of the protein. Sharp Hi TRAP elution profiles indicate monodispersity in the cases where the membranes were treated with detergents DDM and MNG12 (Figure 4.5).

Fn TRAP in pWarf construct was purified with MNG12 detergent. However, the cleavage of the GFP fusion at the HRV3C site was unsuccessful after testing different concentrations of the protease (result not shown). Most likely, the HRV3C site was inaccessible for the protease because it is buried inside the Fn TRAP fold. The purification of the pWarf constructs therefore

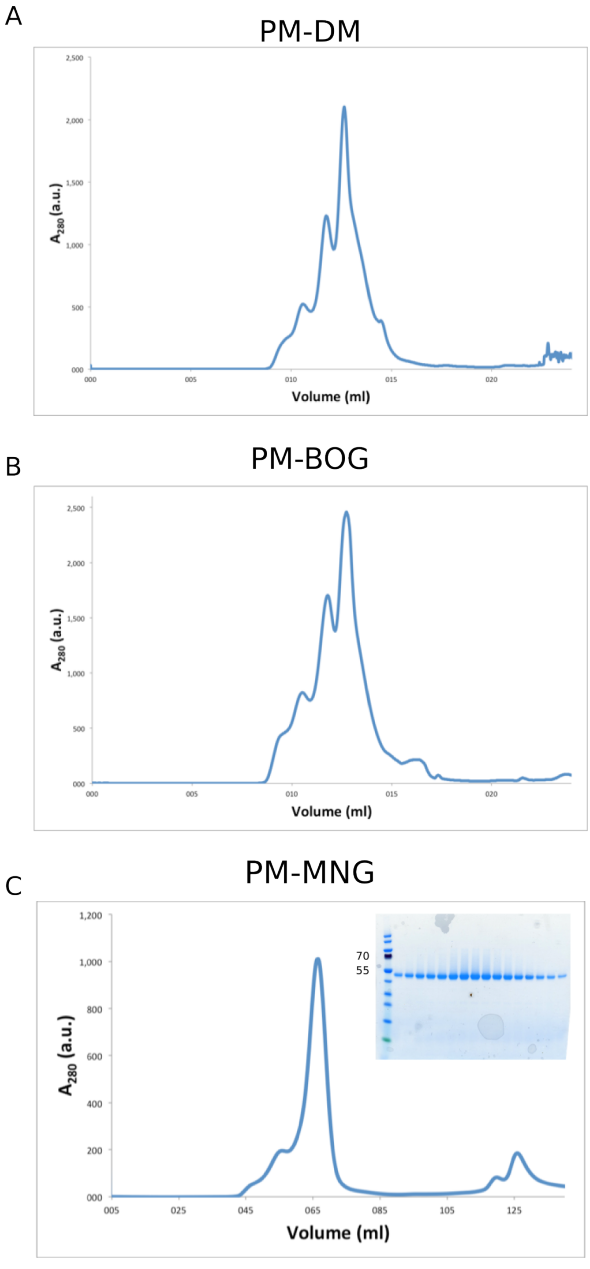

Figure 4.7: Results from the purification of Pm TRAP in pBad system. A. SEC elution profile when using the detergent DM. B. SEC elution profile when using the β-OG detergent. C. SEC elution profile and SDS PAGE using the MNG-12 detergent.

stopped at this point since Vc and Pm pBad constructs were more successful during purification. Moreover, Pm in pBad had promising results in the sparse matrix crystallization screen. I did not have the time nor the resources to pursue both pBad and pWarf systems at the same time. For this reason, I opted to invest my time and energy on pursuing the Vc and Pm pBad constructs.

Unlike the other screened TRAP transporters orthologues, Vc TRAP's Q and M transmembrane domains were not fused and are transcribed by two separate open reading frames (ORF). Western Blot analysis after IMAC shows that more distinct bands are observed for the Q domain as the histidine tag is located upstream the domain. Weak bands are discernible for the QM domain indicate that intact QM domain population is purified but the M domain could be solubilized during SDS treatment (Figure 4.6a).

SEC elution profile for Vc TRAP complex MNG12 detergent indicate a stable profile with a small shoulder (Figure 4.6b). For Pm TRAP, SEC elution profile complexed with MNG 12 detergent shows monodispersity compared to Pm TRAP in DM and β-OG detergent (Figure 4.7a and b). Pm TRAP-MNG12 complex was used for crystallization (Figure 4.7c).

4.4 Membrane protein crystallizability

Crystallizability is an important quality when evaluating membrane protein overexpression, solubilization and purification. In Paper 3, we described that Pm TRAP transporters which crystallizes in five different conditions. The crystals diffracted to 15-11 Å resolution which is a promising indication (Figure 4.8). For near atomic resolution structure determination more ordered crystals are needed. To increase the crystalline order, optimization of crystallization conditions is needed which is highly empirical and time consuming. The presence of the detergent is an intrinsic hinder for ordering of membrane proteins in the crystalline state. Therefore it is very important to find the most suitable detergent for this purpose. The detergents that form a broad collar around the membranes proteins hydrophobic

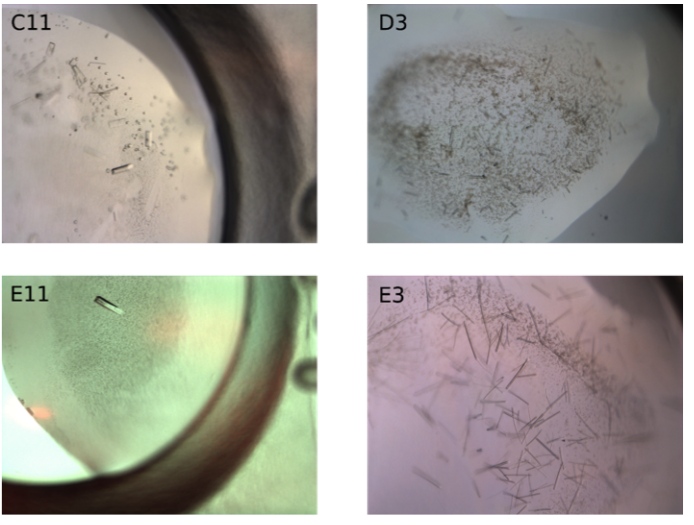

Figure 4.8: Pm TRAP crystals from formed four different conditions

region can hinder protein-protein contacts. This is detrimental for the formation of ordered crystal lattice [122, 123].

4.5 Summary

In this chapter, the path from cloning to diffracting quality crystals is described. The basis of discussion is the membrane protein overexpression, solubilization, purification and crystallization of sialic acid TRAP transporters orthologues. This work is described in more detail in Paper 3.

Chapter 5

Crystal Structure of a Sialic Acid Transporter

5.1 *Proteus mirabilis* SiaT

In Paper 4, the first crystal structure of a sialic acid transporter from *P. mirabilis* (Pm)SiaT is the first sialic acid is reported. Pm SiaT was determined at 1.95 Å and captured in an outward-facing conformation [124]. Moreover, the structure shows bound β-anomeric sialic acid and two sodium ions. The first sodium ion is bound to the conserved Na2 site while the second sodium ion is located in a novel sodium binding site designated Na3 site. In agreement with the predicted *in silico* topology, Pm SiaT is composed of 13 membrane spanning helices (Tm0 and Tm1-Tm12) with the N-terminal located in the periplasm while the C-terminal is located in the cytoplasm. The domain core structure (Tm1-Tm5) and (Tm6-Tm10) has inverted repeat motif which resembles the same fold as vSGLT [67], LeuT [68]) and BetP [125](Figure 5.1).

The expression, solubilization and purification of Pm SiaT shares the same principle as the expression, solubilization and purification of the TRAP transporters in GFP fusion. After IMAC purification the GFP tag was cleaved and then loaded onto HisTrap column. After SEC, the purified Pm SiaT was crystallized through hanging drop diffusion method. The Pm SiaT crystal structure was determined by the SAD method for experimental phasing.

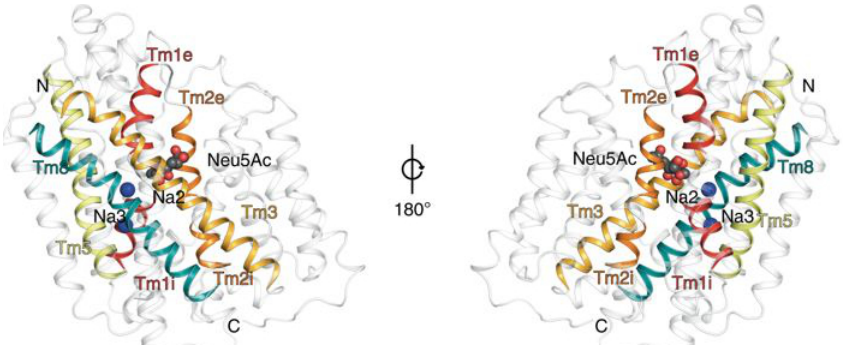

Figure 5.1: The overall structure of *P. mirabilis* sialic acid transporter adopting an outward facing conformation. Sialic acid is bound in the middle of the transporter. (Reproduced from Wahlgren et al [124])

5.2 SiaT sialic acid binding site

That SiaT binds a β-anomeric sialic acid supports the hypothesis that the transport system includes a mutarotase that converts the scavenged α-anomeric sialic acid to its β-anomeric form prior to its transport across the cytoplasmic membrane.

Sialic acid is bound in the core of the structure interacting with eight key residues from the four helices (Tm1-Tm3 and Tm6) (Figure 5.2a). The acetyl moiety of Neu5Ac located close to nonpolar residues Phe78 (Tm2), Gly81 (Tm2) and Phe 243 (Tm6) while the hydroxyl group of its glycerol tail interacts with Gln82 (Tm2) and Thr58 (Tm1). The carboxylate region of Neu5Ac interacts with Ser60, Thr63 and forms a salt bridge with Arg135. Consistent with the structural finding with the substrate-binding protein SiaP from the TRAP transporters, a highly conserved arginine residue is associated with the Neu5Ac specificity recognition [57, 59]. Additionally, the presence of basic residue is a common feature of sugar-binding protein [67] [126]. Mutation studies on the key residues along the substrate-binding site (Thr58Ala, Ser60Ala, Thr63Ala, Gln82Asp and Arg135Glu) rendered the transporter inactive with the exception of Thr58Ala, the least conserved

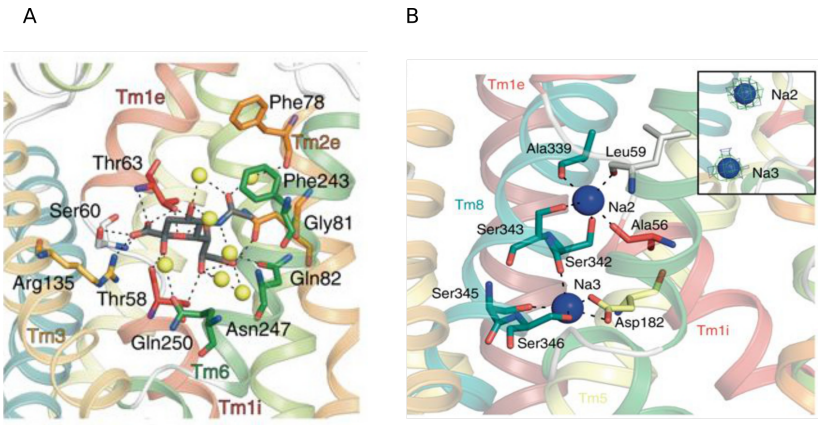

Figure 5.2: *P. mirabilis* SiaT structure. A. The sialic acid binding site showing sialic acid in coordination with key residues. B. The second sodium binding site. (Reproduced from Wahlgren et al.[124])

residue in the binding site which has twice the uptake of the wild type protein.

5.3 The sodium-binding site

There are two sodium ions bound to the structure. The Na2 site is 7 Å away from the substrate-binding site and it is located between Tm1 and Tm8. The sodium ion is coordinated by residues Ala56 and Leu59 (segment of Tm1) and Ser342, Ser343 and Ala339 (Tm8). Transport assays on mutants Ser324Ala and Ser343Ala indicate loss of uptake function.

The novel Na3 site that is 6.5 Å and 14 Å away from the Na2 site and the substrate-binding site respectively. The sodium ion on Na3 site coordinates with residues Ser342 (Tm8), Ser345, Ser346 and Asp182 (Tm5) (Figure 5.2b). Mutations on key interacting residues on the Na3 site indicate varied degree of transport. This signifies that mutations in the Na2 site rendered the transporter inactive while a mutation in the Na3 site is not as detrimental.

5.4 Alternating access mechanism

An alternating access mechanism for SiaT is proposed by constructing an inward-facing model of SiaT based on the inward-facing structure of vSGLT [126] . In this mechanism, the outer gate closes over the binding pocket as Tm10 swings towards Tm1e and Tm2 leading to movement of Tm9 closer to Tm3. During the process, the extracellular loop helices (Elh7a-Elh7b) bend inwards towards Tm1e. The inward facing movement is predicted to involve the movement of the intracellular loop/helix(Ilh0) from its closed position, stabilized by salt bridges between conserved two conserved Arg residues(Ilh0) with (Tm4i-Tm5 loop). Finally, the inner gate opens as the intracellular loop (Ilh0) swings radially outward, away from the center and simultaneously Tm8 and Tm9 rearrange by moving away from the inner cavity.

The first crystal structure of the sialic acid transporter Pm SiaT elucidates its function. The binding of a β-anomeric sialic acid by SiaT is consistent with the involvement of a mutarotase in converting α-anomeric to β-anomeric conformation before traversing the cytoplasmic membrane. Furthermore, SiaT interpreted as an outward facing structure provides valuable insight into the alternating access transport mechanism employed by the SSS family. One important aspect of the SiaT structure determination is the discovery of a novel sodium binding site that is implicated in enhancing the substrate-binding affinity and has a role in upholding structural stability of the outward-facing conformation

5.5 Summary

In this chapter, the first crystal structure of a sialic acid transporter, *P. mirabilis* SiaT was described in an outward-facing conformation and a novel second sodium binding site was presented.

Chapter 6

Concluding Remarks

This thesis presents work aimed at structural studies of enzymes or transporters associated with the mechanism of sialic acid utliziation in bacteria. As this work illustrates, the path to structure determination is paved with many bottlenecks and hurdles. The purification of large quantities of high quality recombinant protein is a key step in X-ray crystallography. **Paper 1** describes an entirely automated pipeline for optimizing the cloning, expression and purification of several enzymes involved in sialic acid catabolic and sialylation pathways. This approach led to the large-scale production of FnNanK that was then used for crystallization.

In **Paper 2**, the crystal structure of apo FnNanK was determined at 2.2 Å resolution. FnNanK displays motifs intrinsic to the ROK superfamily but lacks the distinctive zinc-binding region that has been implicated in stabilizing the active site of this enzyme. For future studies, an FnNanK structure bound to a substrate and ATP analogues will elucidate the conformational changes that take place upon substrate binding. It would also be interesting to compare the substrate binding mechanism of FnNanK to other Nan kinase ortholgues known to have the zinc-binding region. Different substrate analogues can also be co-crystallized with FnNanK to gain more insight into the mechanism substrate binding. Since (to my knowledge) it is still unclear if deleting FnNanK compromises the bacteria's pathogenicity, or if its function is salvaged by another mechanism in the pathway, it would be important to perform these studies before deciding if this enzyme may be interesting as an antibacterial drug target.

In **Paper 4**, the first crystal structure of a sialic acid transporter was reported at 1.9 Å resolution. Pm SiaT belongs to the SSS family of membrane transporters that employs the alternating access mechanism when transporting sialic acid into the cytoplasm. Since this structure captured the outwards facing conformation, it would be interesting to determine Pm SiaT structures that adopt an inward-facing conformation and an occluded conformation, in order to understand better the transport mechanism through the transporter. Moreover, the structure reveals a novel second sodium-binding site that is suggested to enhance the substrate binding affinity and to stabilize the outward facing conformation. It would be interesting to compare Pm SiaT with other SiaT structures that also have a second second sodium-binding site in order to better understand its cellular function.

In **Paper 3**, we described the process of membrane protein production, detergent screening, purification and crystallization that started from the broad-based screening of four TRAP transporter orthologues. This work funneled down to Pm TRAP yielding diffraction quality crystals which diffracted to low resolution. To push this project on towards structure determination, however, other crystallization methods may need to be be tried, such as lipidic cubic phase crystallization or the use of bicelle crystallization protocols. Another strategy to explore is to try single-particle cryo-electron microscopy as structural determination method. The PmTRAP construct with a GFP moiety is around 100 kDa in size and this is close to the minimum size where this method can usefully be applied. When in complex with the substrate-binding subunit SiaP. The total size reaches 135 kDa, which may be suitable for cryo-EM. Moreover, conditions may be found where PmTRAP forms a stable higher oligomeric state, in which case the size of the complex would not be the limiting factor for possible structural determination. Regardless of the structural method used, the purification of high quality, homogenous and stable protein remains the focal point for any structural methods. The work presented here may thus serve as the groundwork for future exploration.

Bibliography

1. Kallmeyer, J., Pockalny, R., Adhikari, R. R., Smith, D. C. & D'Hondt, S. Global distribution of microbial abundance and biomass in subseafloor sediment. *Proc Natl Acad Sci U S A* **109,** 16213–6 (2012).
2. Zaman, S. B. *et al.* A Review on Antibiotic Resistance: Alarm Bells are Ringing. *Cureus* **9,** e1403 (2017).
3. Davies, J. & Davies, D. Origins and evolution of antibiotic resistance. *Microbiol Mol Biol Rev* **74,** 417–33 (2010).
4. Davies, J. Origins and evolution of antibiotic resistance. *Microbiologia* **12,** 9–16 (1996).
5. Bengtsson-Palme, J., Kristiansson, E. & Larsson, D. G. J. Environmental factors influencing the development and spread of antibiotic resistance. *FEMS Microbiol Rev* **42,** 68–80 (2018).
6. Organisation, W. H. Fact Sheet: Antibiotics Resistance (2017).
7. Ochman, H., Lawrence, J. G. & Groisman, E. A. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405,** 299–304 (2000).
8. Casadevall, A. & Pirofski, L. Host-pathogen interactions: the attributes of virulence. *J Infect Dis* **184,** 337–44 (2001).
9. Casadevall, A. & Pirofski, L. A. Host-pathogen interactions: basic concepts of microbial commensalism, colonization, infection, and disease. *Infect Immun* **68,** 6511–8 (2000).
10. Vimr, E. in *Protein-Carbohydrate Interactions in Infectious Disease* (ed Carole, B.) chap. 8 (2006).

11. Vimr, E. R., Kalivoda, K. A., Deszo, E. L. & Steenbergen, S. M. Diversity of microbial sialic acid metabolism. *Microbiol Mol Biol Rev* **68,** 132–53 (2004).

12. Lundblad, A. Gunnar Blix and his discovery of sialic acids. Fascinating molecules in glycobiology. *Ups J Med Sci* **120,** 104–12 (2015).

13. Angata, T. & Varki, A. Chemical diversity in the sialic acids and related alpha-keto acids: an evolutionary perspective. *Chem Rev* **102,** 439–69 (2002).

14. Vimr, E. R. Unified theory of bacterial sialometabolism: how and why bacteria metabolize host sialic acids. *ISRN Microbiol* **2013,** 816713 (2013).

15. Rutishauser, U. Polysialic acid in the plasticity of the developing and adult vertebrate nervous system. *Nat Rev Neurosci* **9,** 26–35 (2008).

16. Tiralongo, J. in *Sialobiology:Structure, Biosynthesis and Function* (ed Tiralongo, J.) chap. 1 (2013).

17. Chen, H. Y. & Varki, A. O-acetylation of GD3: an enigmatic modification regulating apoptosis? *J Exp Med* **196,** 1529–33 (2002).

18. Kazatchkine, M. D., Fearon, D. T. & Austen, K. F. Human alternative complement pathway: membrane-associated sialic acid regulates the competition between B and β1 H for cell-bound C3b. *J Immunol* **122,** 75–81 (1979).

19. Bevilacqua, M. P. & Nelson, R. M. Selectins. *J Clin Invest* **91,** 379–87 (1993).

20. Lasky, L. A. Selectins: interpreters of cell-specific carbohydrate information during inflammation. *Science* **258,** 964–9 (1992).

21. Lanoue, A., Batista, F. D., Stewart, M. & Neuberger, M. S. Interaction of CD22 with alpha2,6-linked sialoglycoconjugates: innate recognition of self to dampen B cell autoreactivity? *Eur J Immunol* **32,** 348–55 (2002).

22. Haines-Menges, B. L., Whitaker, W. B., Lubin, J. B. & Boyd, E. F. Host Sialic Acids: A Delicacy for the Pathogen with Discerning Taste. *Microbiol Spectr* **3,** MBP–0005–2014 (2015).

23. Münster-Kühnel, A. K. in *Sialobiology: Structure, Biosynthesis and Function* (ed Tiralongo, J.) chap. 3 (2013).

24. Tanner, M. E. The enzymes of sialic acid biosynthesis. *Bioorg Chem* **33,** 216–28 (2005).

25. Petit, D., Teppa, E., Cenci, U., Ball, S. & Harduin-Lepers, A. Reconstruction of the sialylation pathway in the ancestor of eukaryotes. *Scientific Reports* **8,** 2946 (2018).

26. Almagro-Moreno, S. & Boyd, E. F. Insights into the evolution of sialic acid catabolism among bacteria. *BMC Evol Biol* **9,** 118 (2009).

27. Vimr, E., Lichtensteiger, C. & Steenbergen, S. Sialic acid metabolism's dual function in Haemophilus influenzae. *Mol Microbiol* **36,** 1113–23 (2000).

28. Almagro-Moreno, S. & Boyd, E. F. Sialic acid catabolism confers a competitive advantage to pathogenic vibrio cholerae in the mouse intestine. *Infect Immun* **77,** 3807–16 (2009).

29. Honma, K., Ruscitto, A., Frey, A. M., Stafford, G. P. & Sharma, A. Sialic acid transporter NanT participates in Tannerella forsythia biofilm formation and survival on epithelial cells. *Microb Pathog* **94,** 12–20 (2016).

30. Vimr, E. & Lichtensteiger, C. To sialylate, or not to sialylate: that is the question. *Trends Microbiol* **10,** 254–7 (2002).

31. Roy, S., Douglas, C. W. & Stafford, G. P. A novel sialic acid utilization and uptake system in the periodontal pathogen Tannerella forsythia. *J Bacteriol* **192,** 2285–93 (2010).

32. Stafford, G., Roy, S., Honma, K. & Sharma, A. Sialic acid, periodontal pathogens and Tannerella forsythia: stick around and enjoy the feast! *Mol Oral Microbiol* **27,** 11–22 (2012).

33. Bouchet, V. *et al.* 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,** 8898–903 (2003).

34. Severi, E., Hood, D. W. & Thomas, G. H. Sialic acid utilization by bacterial pathogens. *Microbiology* **153,** 2817–22 (2007).

35. McGuckin, M. A., Linden, S. K., Sutton, P. & Florin, T. H. Mucin dynamics and enteric pathogens. *Nat Rev Microbiol* **9,** 265–78 (2011).

36. Sakarya, S., Gokturk, C., Ozturk, T. & Ertugrul, M. B. Sialic acid is required for nonspecific adherence of Salmonella enterica ssp. enterica serovar Typhi on Caco-2 cells. *FEMS Immunol Med Microbiol* **58,** 330–5 (2010).

37. Jeong, H. G. *et al.* The Capability of Catabolic Utilization of N-Acetylneuraminic Acid, a Sialic Acid, Is Essential for Vibrio vulnificus Pathogenesis. *Infect. Immun.* **77,** 3209–3217 (2009).

38. Lewis, W. G., Robinson, L. S., Gilbert, N. M., Perry, J. C. & Lewis, A. L. Degradation, foraging, and depletion of mucus sialoglycans by the vagina-adapted Actinobacterium Gardnerella vaginalis. *J Biol Chem* **288,** 12067–79 (2013).

39. Vimr, E. R. & Troy, F. A. Identification of an inducible catabolic system for sialic acids (nan) in Escherichia coli. *J Bacteriol* **164,** 845–53 (1985).

40. Condemine, G., Berrier, C., Plumbridge, J. & Ghazi, A. Function and expression of an N-acetylneuraminic acid-inducible outer membrane channel in Escherichia coli. *J Bacteriol* **187,** 1959–65 (2005).

41. Wirth, C. *et al.* NanC crystal structure, a model for outer-membrane channels of the acidic sugar-specific KdgM porin family. *J Mol Biol* **394,** 718–31 (2009).

42. Severi, E., Hood, D. W. & Thomas, G. H. Sialic acid utilization by bacterial pathogens. *Microbiology* **153,** 2817–22 (2007).

43. Steenbergen, S. M., Jirik, J. L. & Vimr, E. R. YjhS (NanS) is required for Escherichia coli to grow on 9-O-acetylated N-acetylneuraminic acid. *J Bacteriol* **191,** 7134–9 (2009).

44. Varki, A. Diversity in the sialic acids. *Glycobiology* **2,** 25–40 (1992).

45. North, R. A. *et al.* "Just a spoonful of sugar...": import of sialic acid across bacterial cell membranes. *Biophys Rev* **10,** 219–227 (2018).

46. Thomas, G. H. Sialic acid acquisition in bacteria-one substrate, many transporters. *Biochem Soc Trans* **44,** 760–5 (2016).

47. Weyand, S. *et al.* The alternating access mechanism of transport as observed in the sodium-hydantoin transporter Mhp1. *J Synchrotron Radiat* **18,** 20–3 (2011).

48. Severi, E. *et al.* 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,** 1173–85 (2005).

49. Severi, E., Hosie, A. H., Hawkhead, J. A. & Thomas, G. H. Characterization of a novel sialic acid transporter of the sodium solute symporter (SSS) family and in vivo comparison with known bacterial sialic acid transporters. *FEMS Microbiol Lett* **304,** 47–54 (2010).

50. Severi, E. *et al.* Sialic acid mutarotation is catalyzed by the Escherichia coli beta-propeller protein YjhT. *J Biol Chem* **283,** 4841–9 (2008).

51. Phansopa, C. *et al.* Structural and functional characterization of NanU, a novel high-affinity sialic acid-inducible binding protein of oral and gut-dwelling Bacteroidetes species. *Biochem J* **458,** 499–511 (2014).

52. Post, D. M., Mungur, R., Gibson, B. W. & Munson R. S., J. Identification of a novel sialic acid transporter in Haemophilus ducreyi. *Infect Immun* **73,** 6727–35 (2005).

53. Gruteser, N., Marin, K., Kramer, R. & Thomas, G. H. Sialic acid utilization by the soil bacterium Corynebacterium glutamicum. *FEMS Microbiol Lett* **336,** 131–8 (2012).

54. Marion, C., Aten, A. E., Woodiga, S. A. & King, S. J. Identification of an ATPase, MsmK, which energizes multiple carbohydrate ABC transporters in Streptococcus pneumoniae. *Infect Immun* **79,** 4193–200 (2011).

55. Mulligan, C., Fischer, M. & Thomas, G. H. Tripartite ATP-independent periplasmic (TRAP) transporters in bacteria and archaea. *FEMS Microbiol Rev* **35,** 68–86 (2011).

56. Allen, S., Zaleski, A., Johnston, J. W., Gibson, B. W. & Apicella, M. A. Novel sialic acid transporter of Haemophilus influenzae. *Infect Immun* **73,** 5291–300 (2005).

57. Fischer, M., Zhang, Q. Y., Hubbard, R. E. & Thomas, G. H. Caught in a TRAP: substrate-binding proteins in secondary transport. *Trends Microbiol* **18,** 471–8 (2010).

58. Johnston, J. W. *et al.* 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,** 855–65 (2008).

59. Gangi Setty, T., Cho, C., Govindappa, S., Apicella, M. A. & Ramaswamy, S. Bacterial periplasmic sialic acid-binding proteins exhibit a conserved binding site. *Acta Crystallogr D Biol Crystallogr* **70,** 1801–11 (2014).

60. Kelly, D. J. & Thomas, G. H. The tripartite ATP-independent periplasmic (TRAP) transporters of bacteria and archaea. *FEMS Microbiol Rev* **25,** 405–24 (2001).

61. Jenkins, G. A. *et al.* Sialic acid mediated transcriptional modulation of a highly conserved sialometabolism gene cluster in Haemophilus influenzae and its effect on virulence. *BMC Microbiol* **10,** 48 (2010).

62. Abramson, J. *et al.* Structure and mechanism of the lactose permease of Escherichia coli. *Science* **301,** 610–5 (2003).

63. Huang, Y., Lemieux, M. J., Song, J., Auer, M. & Wang, D. N. Structure and mechanism of the glycerol-3-phosphate transporter from *Escherichia coli*. *Science* **301,** 616–20 (2003).

64. Olson, M. E., King, J. M., Yahr, T. L. & Horswill, A. R. Sialic acid catabolism in Staphylococcus aureus. *J Bacteriol* **195,** 1779–88 (2013).

65. Ng, K. M. *et al.* Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature* **502,** 96–9 (2013).

66. Anba-Mondoloni, J., Chaillou, S., Zagorec, M. & Champomier-Verges, M. C. Catabolism of N-acetylneuraminic acid, a fitness function of the food-borne lactic acid bacterium Lactobacillus sakei, involves two newly characterized proteins. *Appl Environ Microbiol* **79,** 2012–8 (2013).

67. Faham, S. & Bowie, J. U. Bicelle crystallization: a new method for crystallizing membrane proteins yields a monomeric bacteriorhodopsin structure. *J Mol Biol* **316,** 1–6 (2002).

68. Yamashita, A., Singh, S. K., Kawate, T., Jin, Y. & Gouaux, E. Crystal structure of a bacterial homologue of Na+/Cl–dependent neurotransmitter transporters. *Nature* **437,** 215–23 (2005).

69. Weyand, S. *et al.* Structure and molecular mechanism of a nucleobase-cation-symport-1 family transporter. *Science* **322,** 709–13 (2008).

70. Kendrew, J. C. *et al.* A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181,** 662–6 (1958).

71. Rupp, B. *Biomolecular Crystallography: Principles, Practice and Application to Structural Biology* (Garland Science, 2010).

72. McPherson, A. Introduction to protein crystallization. *Methods* **34,** 254–65 (2004).

73. Leslie, A. G. *et al.* Automation of the collection and processing of X-ray diffraction data – a generic approach. *Acta Crystallogr D Biol Crystallogr* **58,** 1924–8 (2002).

74. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **62,** 72–82 (2006).

75. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallographica Section D: Biological Crystallography* **69,** 1204–1214 (2013).

76. McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40,** 658–674 (2007).

77. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr* **66,** 22–5 (2010).

78. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60,** 2126–32 (2004).

79. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography* **67,** 355–367 (2011).

80. Brunger, A. T. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355,** 472–5 (1992).

81. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66,** 12–21 (2010).

82. Bairy, S. *et al.* 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 (2018).

83. Esposito, D., Garvey, L. A. & Chakiath, C. S. Gateway cloning for protein expression. *Methods Mol Biol* **498,** 31–54 (2009).

84. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41,** 207–34 (2005).

85. Mlynek, G. *et al.* The Center for Optimized Structural Studies (COSS) platform for automation in cloning, expression, and purification of single proteins and protein-protein complexes. *Amino Acids* **46,** 1565–82 (2014).

86. Ben Yehezkel, T. *et al.* Computer-aided high-throughput cloning of bacteria in liquid medium. *Biotechniques* **50,** 124–7 (2011).

87. Caing-Carlsson, R. *et al.* Crystal structure of N-acetylmannosamine kinase from Fusobacterium nucleatum. *Acta Crystallogr F Struct Biol Commun* **73,** 356–362 (2017).

88. Martinez, J. *et al.* Crystal structures of N-acetylmannosamine kinase provide insights into enzyme activity and inhibition. *J Biol Chem* **287,** 13656–65 (2012).

89. Titgemeyer, F., Reizer, J., Reizer, A. & Saier M. H., J. Evolutionary relationships between sugar kinases and transcriptional repressors in bacteria. *Microbiology* **140,** 2349–54 (1994).

90. Conejo, M. S., Thompson, S. M. & Miller, B. G. Evolutionary bases of carbohydrate recognition and substrate discrimination in the ROK protein family. *J Mol Evol* **70,** 545–56 (2010).

91. Mesak, L. R., Mesak, F. M. & Dahl, M. K. Bacillus subtilis GlcK activity requires cysteines within a motif that discriminates microbial glucokinases into two lineages. *BMC Microbiol* **4,** 6 (2004).

92. Schiefner, A. *et al.* The crystal structure of Mlc, a global regulator of sugar metabolism in Escherichia coli. *J Biol Chem* **280,** 29073–9 (2005).

93. Deisenhofer, J., Epp, O., Miki, K., Huber, R. & Michel, H. Structure of the protein subunits in the photosynthetic reaction centre of *Rhodopseudomonas viridis* at 3A resolution. *Nature* **318,** 618–24 (1985).

94. White, S. http://blanco.biomol.uci.edu/mpstruc/.

95. Doyle, D. A. *et al.* The structure of the potassium channel: molecular basis of K^+ conduction and selectivity. *Science* **280,** 69–77 (1998).

96. White, S. H. The progress of membrane protein structure determination. *Protein Sci* **13,** 1948–1949 (2004).

97. Moraes, I., Evans, G., Sanchez-Weatherby, J., Newstead, S. & Stewart, P. D. Membrane protein structure determination - the next generation. *Biochim Biophys Acta* **1838,** 78–87 (2014).

98. Smith, S. M. Strategies for the purification of membrane proteins. *Methods Mol Biol* **681,** 485–96 (2011).

99. Ohlendieck, K. Extraction of membrane proteins. *Methods Mol Biol* **244,** 283–93 (2004).

100. Wagner, S. *et al.* Tuning Escherichia coli for membrane protein overexpression. *Proc Natl Acad Sci U S A* **105,** 14371–6 (2008).

101. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front Microbiol* **5,** 172 (2014).

102. Bentley, W. E., Mirjalili, N., Andersen, D. C., Davis, R. H. & Kompala, D. S. Plasmid-encoded protein: the principal factor in the "metabolic burden" associated with recombinant bacteria. *Biotechnol Bioeng* **35,** 668–81 (1990).

103. Robinson, A. *Production of Membrane Proteins: Strategies for Expression and Isolation* (Wiley, 2011).

104. Xie, K. & Dalbey, R. E. Inserting proteins into the bacterial cytoplasmic membrane using the Sec and YidC translocases. *Nat Rev Microbiol* **6,** 234–44 (2008).

105. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* **177,** 4121–30 (1995).

106. Hsieh, J. M. *et al.* Bridging the gap: a GFP-based strategy for overexpression and purification of membrane proteins with intra and extracellular C-termini. *Protein Sci* **19,** 868–80 (2010).

107. Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol* **185,** 60–89 (1990).

108. Studier, F. W. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J Mol Biol* **219,** 37–44 (1991).

109. Dubendorff, J. W. & Studier, F. W. Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *J Mol Biol* **219,** 45–59 (1991).

110. Hjelm, A. *et al.* Optimizing E. coli-based membrane protein production using Lemo21(DE3) and GFP-fusions. *Methods Mol Biol* **1033,** 381–400 (2013).

111. Vera, A., Gonzalez-Montalban, N., Aris, A. & Villaverde, A. The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. *Biotechnol Bioeng* **96,** 1101–6 (2007).

112. Linke, D. Detergents: an overview. *Methods Enzymol* **463,** 603–17 (2009).

113. Le Maire, M., Champeil, P. & Moller, J. V. Interaction of membrane proteins and lipids with solubilizing detergents. *Biochim Biophys Acta* **1508,** 86–111 (2000).

114. Arnold, T. & Linke, D. The use of detergents to purify membrane proteins. *Curr Protoc Protein Sci* **53,** 4.8.1–4.8.30 (2008).

115. Neugebauer, J. M. Detergents: an overview. *Methods Enzymol* **182,** 239–53 (1990).

116. Seddon, A. M., Curnow, P. & Booth, P. J. Membrane proteins, lipids and detergents: not just a soap opera. *Biochim Biophys Acta* **1666,** 105–17 (2004).

117. Drew, D. E., von Heijne, G., Nordlund, P. & de Gier, J. W. Green fluorescent protein as an indicator to monitor membrane protein overexpression in *Escherichia coli*. *FEBS Lett* **507,** 220–4 (2001).

118. Drew, D. *et al.* A scalable, GFP-based pipeline for membrane protein overexpression screening and purification. *Protein Sci* **14,** 2011–7 (2005).

119. Drew, D., Lerch, M., Kunji, E., Slotboom, D. J. & de Gier, J. W. Optimization of membrane protein overexpression and purification using GFP fusions. *Nat Methods* **3,** 303–13 (2006).

120. Chae, P. S. *et al.* Maltose-neopentyl glycol (MNG) amphiphiles for solubilization, stabilization and crystallization of membrane proteins. *Nat Methods* **7,** 1003–8 (2010).

121. Kawate, T. & Gouaux, E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure* **14,** 673–81 (2006).

122. Loll, P. J. Membrane proteins, detergents and crystals: what is the state of the art? *Acta Crystallographica Section F* **70,** 1576–1583 (2014).

123. Wiener, M. C. A pedestrian guide to membrane protein crystallization. *Methods* **34,** 364–72 (2004).

124. Wahlgren, W. Y. *et al.* Substrate-bound outward-open structure of a Na^+-coupled sialic acid symporter reveals a new Na^+ site. *Nat Commun* **9,** 1753 (2018).

125. Ressl, S., Terwisscha van Scheltinga, A. C., Vonrhein, C., Ott, V. & Ziegler, C. Molecular basis of transport and regulation in the Na^+/betaine symporter BetP. *Nature* **458,** 47–52 (2009).

126. Faham, S. *et al.* The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na^+/sugar symport. *Science* **321,** 810–4 (2008).