UiT
THE ARCTIC UNIVERSITY OF NORWAY

Faculty of Science and Technology

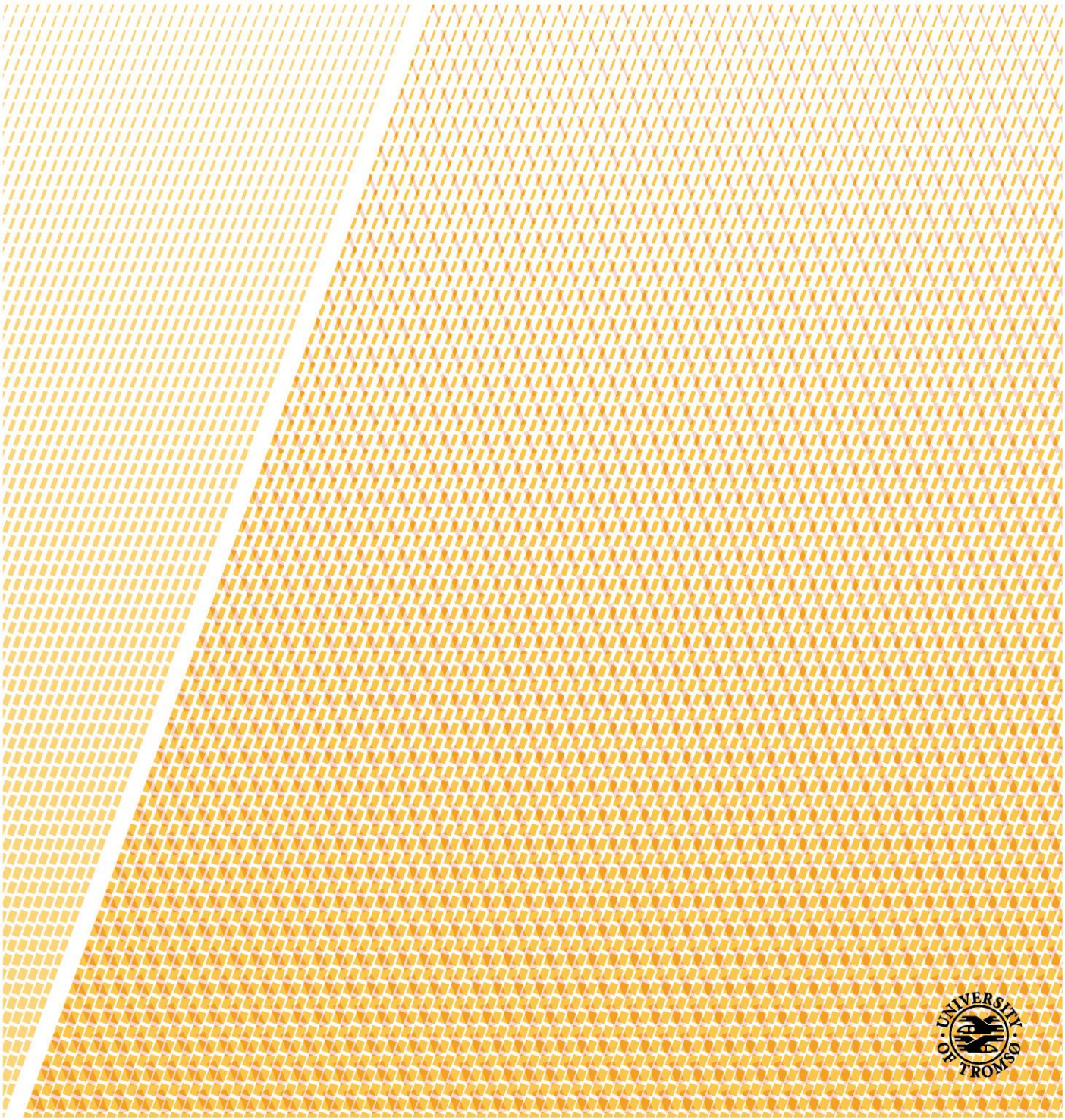
Along the path of bacterial nonulosonic acids

A study of the bio- and in vitro synthesis of sialic acid related compounds

—

Marie-Josée Haglund Halsør

A dissertation for the degree of Philosophiae Doctor – June 2019



Along the path of nonulosonic acids

A study of the bio- and in vitro synthesis of sialic acid related compounds

Marie-Josée Haglund Halsør

A dissertation for the degree of Philosophiae Doctor

FACULTY OF SCIENCE AND TECHNOLOGY

DEPARTMENT OF CHEMISTRY

June 2019

"There is a single light of science and to brighten it anywhere is to brighten it everywhere."

- Unsourced, credited to Isaac Asimov.

Preface

"*Why?*", and later "*How?*". Those two questions are what led me to research, without doubt. I've asked them (aloud or not) every day for as long as I can remember, about practically everything. The other thing is being amazed by Nature. The diversity of every aspect and how it all functions as one, somehow. My favorite as a child were the documentaries by "le Commandant Cousteau" (the sharks!), and my dream was to be an oceanographer. I pursued that dream up until my first year of university, when I discovered biochemistry. I had already grown a liking for chemistry, and it was the only discipline that answered the "biological whys and hows" without going into physics. Biochemistry studies and does, both trying to unravel Nature's secrets and building its own means to do so. It also uses the knowledge to improve human living conditions, at least in theory. I was sold, and here I am. It is at least as much fun as I thought it would be.

I began this project with only general knowledge about glycosylation, having specialized in enzymology and bioinformatics during my studies. I knew that sugar modifications were, roughly, adding another level of diversity to proteins and cell surfaces, but I had greatly underestimated the scale of it and the consequences of that. If nothing else, I am happy that I got to learn about it. Nature is still amazing, if you're wondering.

I thought that I would spend my time doing enzymology with a little modelling on the side, especially looking forward to the kinetics. The latter is my one and only regret, and I fully intend to rectify that in the future. However, I got to do crystallization, genomics and metabolomics. I got to try to make sense of existing data, provide some new, and prepare for further studies. I got to try, fail, try again, fail again, give up and do something else. I learned so much that I would like to do it all over again, so that I could do everything better, and try a thousand other things.

I cannot talk about this experience without mentioning Tromsø. It is such a special place that it is an experience in itself. The arctic climate of course, but also living in a post-card every hour of every day. The people, the mix of local and international that creates such a special atmosphere. There is everything and everyone, in the middle of nowhere. And the whales, next best thing to sharks.

Doing a PhD was just fantastic. Fun in a brain-racking, nerve wracking, incredibly fulfilling way. An exercise in perseverance. A study of the response to environmental stress. But most of all, an opportunity to explore the "whys" and "hows".

Acknowledgments

The work presented here was carried out from August 2013 to January 2019 at the Norwegian Structural Biology Centre, Department of Chemistry, Faculty of Science and Technology at the UiT-The Arctic University of Norway. UiT as well as the NorzymeD project from the Research Council of Norway (Grant number 221568) provided the financial support. I also would like to acknowledge the Norwegian Graduate School in Structural Biology (Biostruct) for their training and additional financial support.

None of this would have been possible without my supervisors, Arne O. Smalås and Inger Lin U. Ræder. I thank you for the opportunity to work on such an interesting project, and for your support. Arne, I greatly appreciated your encouragement through the years. As for Inger Lin, you have been with me from day one, first as co-supervisor and then as main supervisor. I thank you so much for willing to step up and bring your energy towards finishing this work. You have helped me at every stage of this journey, and at every level. You have my deepest gratitude.

It would also have been much different without my co-supervisors Bjørn Altermark and Ulli Rothweiler. You both have different fields of expertise and personalities, but you are alike in the way you always think two steps ahead. Your ability to analyze and interpret data is amazing, and I hope to develop this quality to your levels. I am also grateful for constant support, from always making time for me to the origami Yoda that appeared on my bench and driving me to the ER when I dislocated my knee.

My deepest appreciation goes to Jostein Johansen for his tremendous help with trying to setup the AGE assay, and the LC-MS. I appreciated your patience and willingness to teach. I also enjoyed doing the chemistry lab with you. I would also like to give my warmest thanks to the rest of the “Sia group”, Man Kumari and Tor Olav. I really enjoyed working with you guys, and your help has been invaluable. Your company was greatly appreciated too, especially on normal working hours such as ten in the evening and during the week ends.

I cannot name evenings and weekends without thinking of my fellow “downstairs people”, thought most of you have changed office locations. Thanks to my office mates Netsanet, Kjersti, Sunniva and Miriam, I enjoyed “living” with you all. Susann, I miss the morning spin sessions. Bjarte, I am so happy I got to hang with you at Svalbard, it was excellent. Dilip, Kazi, Amuda, Kristel and Titti, I have fun memories with each of you that I will never forget. Miriam, Yvonne, and Adele, you are both co-workers and friends, the people I grew closest to at work.

I would also like to acknowledge the “Norstruct” and “Chemistry” people. I have never been in a department were everyone is so close, which makes for a wonderful work environment (and great parties). Hana-Kirsti for always asking how I do, despite not being involved in my project. Jennifer, Marteen and Anastasia for being so enthusiastic and energetic organizers. Anastasia thanks again for calling the ER when I dislocated my knee (the other time).

Not all help comes from the workplace, though. Sometimes “normality takes too much effort”, and I am glad that I had the opportunity to pursue my career as an adventurer at Imladris with such great companions. The list of names is way too long because of the honorific titles, but you know who you are (alternatively, make a Int check). I would also like to mention the members of the Waffle Army, and my friends from France.

Finally, I want to acknowledge my team. Mathilda, Felicia, thank you for your patience. Vegard, you are “supercute”, just stop sleeping on my face. Gaute, words cannot express how grateful I am to share your life. You are my home.

Marie-Josée H. Halsør

Tromsø, June 2019

Table of Contents

Abstract

Nonulosonic acids (NulOs) are sugars expressed on the surface of bacterial cells as well as vertebrates and are involved in a multitude of physiological and pathogenic processes. Their biosynthesis is directed by gene clusters that vary in both gene content and organization, leading to a diverse family of compounds. The exploration of this diversity is relatively recent, helped by the amount of genetic data generated by new technologies. It is however limited by the availability of the compounds themselves and the means to produce them. The presented work describes, for the first part, the study of NulO biosynthesis gene clusters in the fish pathogens *A. salmonicida*, *A. wodanis* and *M. viscosa* in relation to the type of NulO they produce. The key enzyme of the biosynthesis, NeuB, is considered in terms of amino-acid sequence, active-site geometry, and substrate specificity. The second part deals with the bioprospecting of *N*-acetyl-2-glucosamine epimerases (AGEs) for the *in vitro* synthesis of *N*-acetylneuraminic acid, the most abundant NulO in nature. Two isolates of the cyanobacterium *Nostoc* sp. (KVJ10 and KV20) were targeted. *Nostoc* sp. KVJ20 was sequenced, and the AGEs from both strains (nAGE10 and nAGE20) were cloned, expressed, and purified. The crystal structure of nAGE10 was solved, revealing a new (and most likely biologically relevant) dimeric organization of AGEs. Together, these studies participate in the monumental task of understanding the determinants of NulO diversity as well as creating the means for the study of these compounds in general.

List of Papers

Paper I

Analysis of nonulosonic acid biosynthetic gene clusters in *Aliivibrio salmonicida*, *Aliivibrio wodanis* and *Moritella viscosa*. (2019) Halsør MJH, Altermark B, Ræder ILU. Manuscript

Paper II

Draft genome sequence of *Nostoc* sp. KVJ20. (2019) Halsør MJH, Liaimer A, Ræder ILU, Pandur S, Smalås AO, Altermark B. Manuscript submitted to Microbiology Resource Annoucements.

Paper III

The crystal structure of the N-acetyl-D-glucosamine 2-epimerase from *Nostoc* sp. KVJ10 reveals the true dimer. (2019) Halsør MJH, Rothweiler U, Altermark B, Ræder ILU. Acta Cryst. (2019). D75, 90–100

Abbreviations and acronyms

Ac, NAc, OAc	Acetyl, *N*-acetyl, *O*-acetyl
Aci	Acinetaminic acid; 3,5,7,9-tetradeoxy-L-*glycero*-L-*altro*-non-2-ulosonic acid
AGE(s)	*N*-acetyl(-D)-glucosamine-2-epimerase(s)
ATP	Adenosine 5'-triphosphate
CMP	Cytidine 5'-monophosphate
Fus	Fusaminic acid; 3,5,7,9-tetradeoxy-D-*glycero*-L-*gluco*-non-2-ulosonic acid
GDP	Guanosine 5'-diphosphate
GlcNAc	*N*-acetyl-D-glucosamine
Leg	Legionaminic acid; 3,5,7,9-tetradeoxy-D-*glycero*-D-*galacto*-non-2-ulosonic acid
ManNAc	*N*-acetyl-D-mannosamine
NAB	Nonulosonic Acid Biosynthesis
NAL	*N*-acylneuraminate lyase (*N*-acetylneuraminic acid aldolase)
Neu	Neuraminic acid; 3,5-dideoxy-D-*glycero*-D-*galacto*-non-2-ulosonic acid
NulO(s)	Non-2-ulosonic acid(s)
PEP	Phosphoenolpyruvate
Pse	Pseudaminic acid; 3,5,7,9-tetradeoxy-L-*glycero*-L-*manno*-non-2-ulosonic acid
SDR	Short-chain dehydrogenase/reductase
UDP	Uridine 5'-diphosphate
UiT	University of Tromsø-The Arctic University of Norway

1 Introduction

Living cells are composed of four major classes of molecules, each providing a level of information and regulation which direct cell structure and function. DNA contains the information necessary to synthetize all other cell components, and proteins, which result from its interpretation, form the bulk of the cell structure and effector machinery. The two remaining components, lipids and carbohydrates, possess a greater diversity both in structure and function, the latter being the most diverse of all. The carbohydrate components of cells, referred to as glycans, include both mono-, oligo-, and polysaccharides, which can be either free or attached to proteins (glycoproteins) or lipids (glycolipids) [1].

Bacterial glycans are a major component of all elements of the bacterial cell wall, from the inner membrane to the lipopolysaccharide of Gram-negative bacteria, but also of capsules, pili, flagella and biofilms [2]. The structures that they form enable cells to, for example, resist various osmotic conditions, or prevent desiccation. They also give the cell wall its negative charge. In addition to their more structural functions, their location at the periphery of the cell places them in an ideal position for participating in the many interactions that take place with the immediate surroundings, be it another cell, a host, or the outside environment. In pathogenic bacteria, cell wall glycans house many virulence factors such as lipid A, O-, F-, K- and H-antigens, and nonulosonic acids, which are the subject of this thesis.

1.1 The sugars of the bacterial cell surface

Prokaryotic and eukaryotic cells are delimited by a cytoplasmic membrane coated by different structures, all containing various glycans amongst which nonulosonic acids can be found. This section gives an overview of these structures within bacteria.

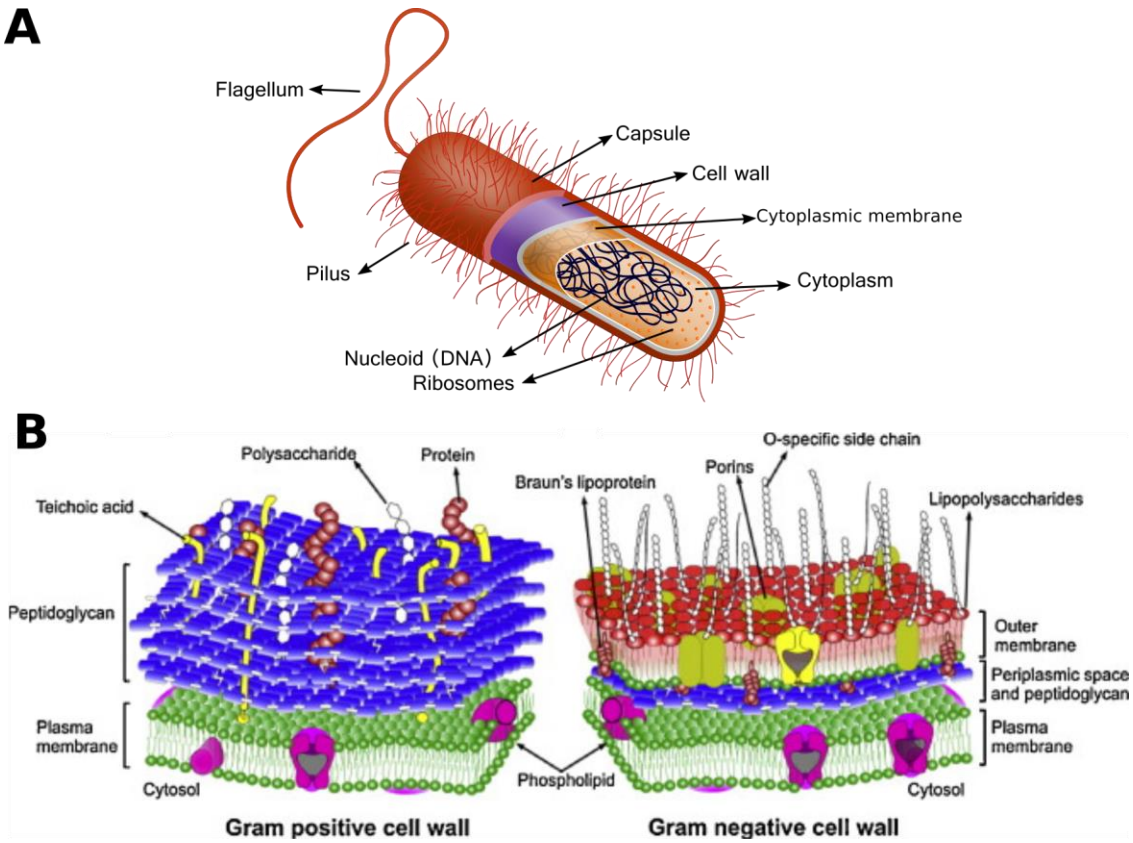


Figure 1. Bacterial cell surface. *A: Schemativ representation of a bacterial cell. Adapted from the image by Ali Zifan, distributed under the CC BY-SA 4.0 license. B: Cell wall of typical gram-positive and gram-negative bacteria. Adapted from [3].*

1.1.1 Bacterial membranes and lipopolysaccharide

Every cell is separated from the environment by a semi-permeable barrier composed of one or several layers (see Figure 1). The innermost layer is the cytoplasmic membrane (CM), also known as the plasma membrane, which is common to all cells. It surrounds the cytoplasm and serves as anchor to the cytoskeleton as well as the components that form the outer layers coating it (see Figure 1B). In gram-negative bacteria and mycobacteria, a second membrane is present, called the outer membrane (OM) [2]. In this case, the CM is often referred to as the inner membrane, which has a different composition than the single membrane of gram-positive bacteria.

The CM is composed of lipids and proteins, some of which are glycosylated. Phospholipids are the main lipid component, but various phosphate-free lipids are also present, and membrane composition varies across species, even within the same bacteria type [4]. Sterols, which are ubiquitous to eukaryotic membranes, are seldom found in prokaryotic ones. The proteins of the CM vary according to whether the outer membrane is present or not, serving as anchor to other structures.

As mentioned earlier, gram-negative bacteria and mycobacteria have an outer membrane. It is separated from the cytoplasmic membrane by a thin layer of peptidoglycan with periplasmic space on either side of it (see next section). It is formed by an asymmetric lipid bilayer composed of phospholipids on the inner leaflet and lipid A on the outer one.

The inner leaflet of the outer membrane faces the periplasm and is linked to the peptidoglycan by lipoproteins [5]. The outer leaflet is characterized by its main component, lipid A, which carries various glycan chains a thus forms the bacterial lipopolysaccharide (LPS). The glycan chains are usually strain specific, and the terminal part of the LPS is referred to as O-antigen due to its antigenic properties. Lipooligosaccharides (LOS) lack the O-antigen, consisting only of lipid A and the core oligosaccharide. The sterol-like lipids hopanoids interact with lipid A and play a role in ordering the membrane [6].

The glycans associated to the plasma and outer membranes are various and depend on the type of membrane considered. The CM of single membrane prokaryotes carries numerous glycoconjugates, but they are mostly associated with the different extracellular structures composing their cell surfaces and will be treated in the corresponding sections [2]. For gram-negative bacteria, glycoconjugates seems to absent from the CM [7]. The glycans from the OM are mainly associated with the LPS, although some are, as for the CM of gram-positive bacteria, corresponding to other structures [2]. Nonulosonic acid residues can be found as terminal residues in the LPS of numerous bacteria.

1.1.2 Peptidoglycan

The cell wall of bacteria contains a layer of peptidoglycan, also known as murein, that differs according to their type [8]. As mentioned earlier, gram-negative bacteria have a thin peptidoglycan layer between to membranes while gram-positive bacteria have a thicker peptidoglycan outside their single, plasma membrane (see Figure 1B). This difference is the determinant between the “Gram” classification, which is based on the results obtained by Gram staining [9]. The peptidoglycan is stained with crystal violet dye and subsequently partially decolorized using alcohol or acetone: thin peptidoglycans are decolorized first, thus distinguishing between the two cell wall structures. It is to be noted that some bacteria, such as mycoplasma, stain negative on the Gram test even though they only have one membrane, because they lack a peptidoglycan layer.

The peptidoglycan is organized in a mesh and consists in linear sugar chains crosslinked by interacting oligopeptide chains. Each chain is a polymer of *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc), with the oligopeptide attached to the latter. The composition of the oligopeptide and the length of glycan chains vary across species [10]. In gram-positive bacteria, the peptidoglycan also contains a range of glycans such as teichoic acids or the polymers produced by mcyobacteria [2].

1.1.3 Glycocalyx and S-layer

The glycocalyx often refers to the coating found on the apical surface of epithelial cells of animals, but also encompasses the bacterial capsule and slime layer. The main difference between the two is their degree of structure and anchorage to the underlying cell wall structures, with the capsule the "tighter" of the two. The glycans that compose them are different, but both capsular and slime layer polysaccharides, the latter called exopolysaccharides, show a high degree of diversity. Capsular polysaccharides from pathogenic bacteria are often designated as the K-antigen. Nonulosonic acids are often found in capsules, of which they can be the main component.

The S-layer is distinct from the slime layer: it is a crystalline, monomolecular layer of proteins or glycoproteins, which find anchor in either the peptidoglycan (and teichoic acids) or the LPS [11]. It is usually, when present, the outermost layer of the bacterial cell surface, and can also contain nonulosonic acids.

1.1.4 Pili and flagella

Bacterial pili and flagella are protein appendages that protrude from the bacterial cell, as shown in Figure 1A. The term fimbriae is often used to designate pili that do not participate in bacterial conjugation, which are also categorized as F-antigens. In a similar manner, flagella, or more precisely the flagellins that composes them, are designated as H-antigens.

Flagella are known as the surface component conferring bacterial motility, although they perform other functions [12]. They are anchored in the CM by their basal bodies, which extends to either the OM or the peptidoglycan layer (gram-negative/positive bacteria). The filament portion is a polymer of flagellin that is linked to the basal body by a hook region, and the glycosylated part of the flagellum [13].

In contrast to flagella which share a common structure, there are several types of pili, which reflects the different functions they fullfill [14]. Theie filaments can however be described as polymers of pilin units, with pilin representing a class of proteins. As for flagellin, pilins can be glycosylated, and nonulosonic acids can be found associated to them.

1.2 Structure of bacterial nonulosonic acids

Nonulosonic acids (NulOs) are ketoaldonic acids, meaning that they contain both a carboxylic acid and a ketone group [15]. The latter is carried by the second carbon on the sugar chain, and thus ulosonic acids are also α-keto acids. The "non-" prefix indicates that they have a backbone formed by nine carbon atoms. Their closed structure, for which an example is presented in Figure 2, is that of a pyranose (formed by the carbons C2 to C6) carrying a glycerol chain (at C6). NulOs contain at least 6 asymmetric centers (C2, and C4-C8) which are used to determine their absolute configuration. The center at C2 distinguishes between the α and β isomers, while the C4-C7 stretch and C8 are compared to hexoses and glycerol, respectively. For neuraminic acid (Neu, see Figure 2), also known as sialic acid and the most studied of NulOs, its absolute configuration is (α/β)-D-*glycero*-D-*galacto*. In addition to the closed structure, it has been shown that several acyclic forms are present in solution [16].

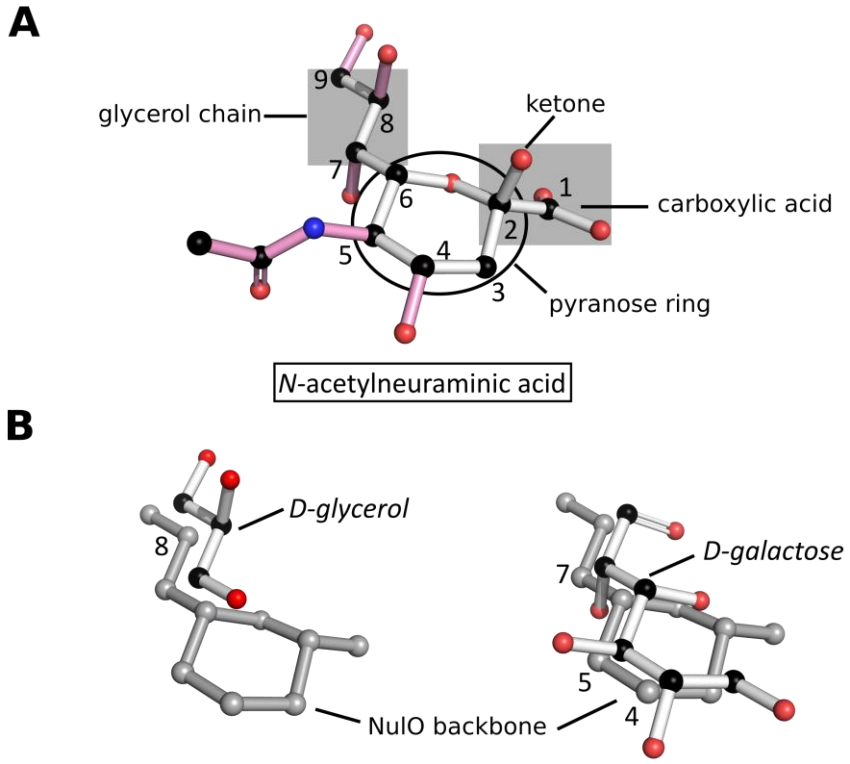


Figure 2. Structure and configuration of N-acetylneuraminic acid (Neu). *A: Structure. The nine carbons forming the Neu backbone are numbered and represented by black spheres and linked by white bonds. The functional groups area (C1-C2) and the glycerol chain (C7-C9) are indicated by grey rectangles, the pyranose ring by a black circle. The substituents carried by Neu are indicated by pink bonds. B: Superposition of D-glycerol and D-galactose onto the Neu backbone (grey). Neu carries its substituents in a manner adopting a D-glycero-D-galacto configuration. For each molecule, the concerned asymmetric centres are numbered (with the exception of C6 for the superposition with galactose, for visibility).*

The varying nature and orientation of substituents on both the pyranose ring and the glycerol chain is responsible for the large diversity of this group of compounds. While established classification exists, bacterial NulOs can be grouped in three main families, as presented in Figure 3. The discrimination is based on the type of gene cluster that govern their biosynthesis, which will be presented in the next section. The first family is composed of Neu related compounds, designated as 3,5-dideoxy-D-*glycero*-D-*galacto*-non-2-ulosonic acids. Its representative member, *N*-acetylneuraminic acid (Neu5Ac), carries an *N*-acetyl group at C5, in equatorial position. The two other families consist of 3,5,7,9-tetradeoxy-non-2-ulosonic acids, which are differentiated by their absolute configurations. Pseudaminic acids (Pse) have the L-*glycero*-L-*manno* configuration, while legionaminic acids (Leg) are more diverse. The family takes its name from its representative compound, legionaminic acid, which has the D-*glycero*-D-*galacto* configuration seen for Neu, but four other isomers have been identified (Figure 3B) [17-21]. Recently, a new isomer most likely having the D-*glycero*-L-*gluco* configuration was identified and designated as fusaminic acid (Fus, see Figure 3A), but its biosynthesis pathway has not been investigated yet [22].

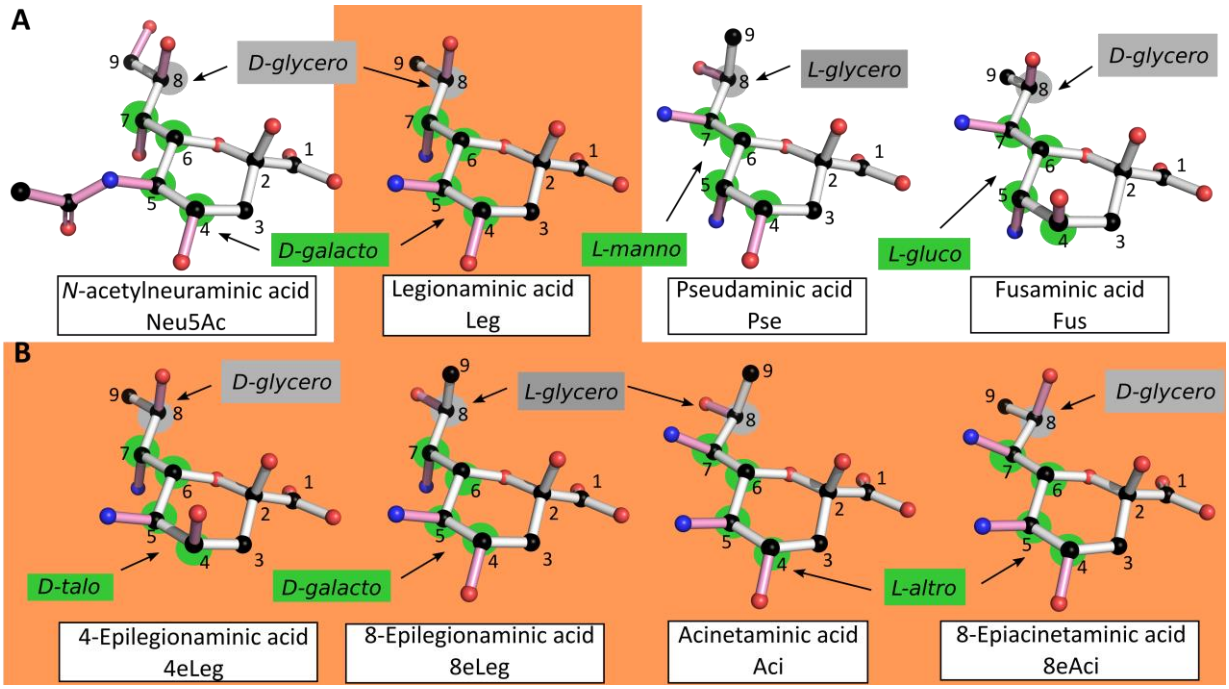


Figure 3. Structure of known NulOs. *The common, nine-carbon backbone of NulOs is represented with white bonds, and the carbons are numbered. The absolute configuration of each isomer is indicated in either gray or green depending on the concerned chiral centers, which are marked by disks of the corresponding color. A: The three main NulOs families, according to their synthesis pathways, and the newly identified fusaminic acid (Fus). Neuraminic acid (Neu) is represented carrying an N-acetyl group in C5 (Neu5Ac), since it is the most commonly found species. Legionaminic (Leg), pseudaminic (Pse) and Fus carry N-linked groups in C5 and C7, which were omitted for clarity leaving only the nitrogen atom. B: Isomers from the legionaminic acid synthesis pathway presenting different absolute configurations. As for Leg, N-linked groups carried at the C5 and C7 positions are represented by the nitrogen atom only. The orange background serves to highlight the common synthesis pathway of Leg isomers.*

These families represent above 40 different compounds that have been so far characterized in different bacteria, for which an overview (as of the end of 2018) is presented in Table 1. Their diversity lies in the variety of substituents that can decorate C5, C7, and C8, but also C9 and C4. *N*-acetylation at C5 and/or C7 is the most common form for both Neu and tetradeoxy- NulOs, followed by the substitution of one of the acetyl groups by an acetamide and the presence of *O*-acetylation at C8. However, a relatively small set of bacteria has been investigated so far compared to the range of species that might be producing NulOs, and the most widespread modifications might be different [23].

Table 1. Structurally identified bacterial NulOs

Monosaccharide	Organism	Reference(s)
Neu5Ac[1,2]	*Escherichia coli*	[24-26]
	Neisseria meningitidis	[27, 28]
	Hafnia alvei	[29]
	Campylobacter jejuni	[30-35]
	Aliivibrio salmonicida	[36]
	Moritella viscosa	[37]
	Rhizobium meliloti	[38]
	Fusobacterium nucleatum	[39]
Neu5Ac7(8)OAc	*Neisseria meningitidis*	[27]
Neu5Ac7(9)OAc	*Salmonella enterica*	[40]
	Aliivibrio salmonicida	[36]
	Moritella viscosa	[37]
Neu5Ac9OAc	*Rhizobium meliloti*	[38]
Neu5Ac7OAc8OAc	*Neisseria meningitidis*	[27]
Neu5Ac7Ac9Ac	*Escherichia coli*	[41]
Pse5Ac7Ac	*Escherichia coli*	[42]
	Campylobacter jejuni	[34, 35, 43, 44]
	Campylobacter coli	[43, 45]
	Proteus vulgaris	[46]
	Helicobacter pylori	[44, 47]
	Pseudoalteromonas atlantica	[48]
	Rhizobium sp.	[49]
	Cellulophaga fucicola	[50]
	Piscirickettsia salmonis	[51]
	Acinetobacter baumannii	[52, 53]
Pse5Ac7Ac9Ac	*Aeromonas caviae*	[44]
Pse5Ac7Hb[1]	*Pseudomonas aeruginosa*	[54, 55]

	Shigella boydii	[54, 55]
	Sinorhizobium fredii	[56]
Pse5Ac7Hb[4]	*Kribella spp.*	[57, 58]
Pse5Ac7Dhb[1,4]	*Actinoplanes utahensis*	[59]
Pse5Ac7Fm[1]	*Pseudomonas aeruginosa*	[60, 61]
	Pseudoalteromonas distincta	[62]
Pse7Fm	*Pseudomonas aeruginosa*	[63]
Pse5Hb7Fm	*Pseudomonas aeruginosa*	[54, 63]
Pse5Am7Ac	*Vibrio cholerae*	[64]
	Campylobacter jejuni	[43, 44]
Pse5Ac7Am	*Campylobacter jejuni*	[34, 35]
Pse5Ac7Gra[1]	*Vibrio vulnificus*	[65]
Pse5Am7Gra	*Tannerella forsythia*	[66][5]
Pse5Gra7Gra	*Campylobacter jejuni*	[43][5]
Pse5Ac7Ac8OAc	*Campylobacter jejuni*	[44]
Pse5Am7Ac8GlnAc[1]	*Campylobacter jejuni*	[44]
Pse5Ac7,9O(S-Pyr) [1]	*Pseudomonas glareae*	[67]
Pse5Gra(OMe)$_2$7Am[1]	*Campylobacter jejuni*	[35]
Pse5Gra(OMe)$_2$7Ac	*Campylobacter jejuni*	[35]
Leg5Ac7Ac	*Campylobacter jejuni*	[35]
	Vibrio alginolyticus	[19]
	Pseudomonas fluorescens	[68][3]
	Acinetobacter baumanii	[69][3]
	Vibrio parahaemolyticus	[70]
Leg5Am7Ac	*Campylobacter jejuni*	[35]
	Campylobacter coli	[45]
	Pseudomonas fluorescens	[68][3]
	Aliivibrio salmonicida	[71, 72][3]
Leg5Hb7Ac	*Acinetobacter baumanii*	[69][3]
Leg5Ac7AcAla[1]	*Vibrio parahaemolyticus*	[70]
	Vibrio vulnificus	[73]
Leg5Ac7Ala	*Escherichia coli*	[74]
Leg5AmMe7Ac	*Campylobacter jejuni*	[35]
	Campylobacter coli	[45]
Leg5GluMe7Ac[1]	*Clostridium botulinum*	[75]
Leg5Am7Ac8OAc	*Legionella pneumophilia*	[76][3]
	Pseudomonas fluorescens	[68][3]
4eLeg5Ac7Ac	*Shewanella japonica*	[77]
4eLeg5Ac7Ac8OAc	*Legionella pneumophila*	[18][3]

4eLeg5Am7Ac	*Legionella pneumophila*	[78]
4eLeg5Am7Ac8OAc	*Legionella pneumophila*	[78]
8eLeg5Ac7Ac	*Pseudomonas aeruginosa*	[17][3]
	Escherichia coli	[79-81]
	Providencia stuartii	[82]
	Acinetobacter baumanii	[83]
8eLeg5Am7Ac	*Morganella morganii*	[84]
8eLeg5Ac7Am8OAc	*Shewanella putrefaciens*	[85]
8eLeg5Hb7Ac	*Salmonella arizonae*	[86][3]
	Yersinia ruckerii	[87][3]
Aci5Ac7Ac	*Acinetobacter baumanii*	[20]
8eAci5Ac7Ac	*Acinetobacter baumanii*	[21]
Fus5Ac7Ac	*Fusobacterium nucleatum*	[22]

[1] Ac=acetyl, Am= acetimidoyl, Fm=formyl, Hb=hydroxybutyryl, Dhb=dihydroxybutyryl, GlnAc=*N*-acetylglucosamine, Me=methyl, Gra=glyceroyl, S-Pyr=(S)-1-carboxyethylidene, Ala=alanyl, Glu=glutamyl

[2] Unless mentioned otherwise, the groups are *N*-linked (*i.e.* Neu5Ac represents Neu5NAc, and Leg5AmMeAc represents Leg5NAmNMe7NAc)

[3] The ultimate configuration was determined by Tsvetkov *et al.*, using GLC-MS and NMR on synthetized compounds [19]. The data was compared to that of natural compounds previously published.

[4] derivate with additional modifications at C4

[5] name corrected by Friedrich *et al*. [88] .

1.3 *nab* gene clusters and NulO biosynthesis

De novo NulO biosynthesis is directed by *nab* gene clusters (for *nonulosonic acid biosynthesis*) along three different pathways leading to neuraminic, legionaminic and pseudaminic acids, respectively (see Figure 3). Acinetaminic acids (Aci) are synthetized by a set of additional genes affixed to the Leg-pathway cluster, and the gene cluster responsible for the synthesis of Fus has not been investigated yet [20]. Every *nab* cluster contains homologous genes that are responsible for the main biosynthesis steps [23]. Amongst them, the NeuB (NAB-2) homologs catalyze the critical reaction in NulO biosynthesis, the condensation of a hexosamine precursor with phosphoenolpyruvate (PEP) to form the corresponding α-ketoacid [89-91]. NeuB is the only NAB protein that is conserved across all three pathways and has been used for investigating the evolutionary history of NulOs [23].

A major hurdle in the study of the genes and enzymes that participate in NulO biosynthesis is their denominations. *Nab* gene clusters have been identified as part of other loci such as capsule biosynthesis, while the enzyme names have been assigned either by homology or after pathway determination. As a

result, each enzyme has multiple names according to which organism and/or pathway it belongs to, as well as its similarity to other enzymes of related function. Gene names tend to be different from the enzyme names, when they are annotated at all. An overview of the main NAB enzymes (NAB1-3) is given in Table 2.

Table 2. Nomenclature of main NAB enzymes

Pathway	description	EC number	Names
Neu	UDP-*N*-acetylglucosamine 2-epimerase	3.2.1.183	NeuC, NeuC1, NAB3
	N-acetylneuraminate synthase	2.5.1.56	NeuB, NeuB1, NAB2
	N-acetylneuraminate cytidylyltransferase	2.7.7.43	NeuA, NeuA1, NAB1
Leg	hydrolyzing UDP-*N,N'*-diacetylbacillosamine 2-epimerase	3.2.1.184	NeuC, NeuC2, NAB3, PtmD, LegG
	N,N'-diacetyllegionaminate synthase	2.5.1.101	NeuB, NeuB2, NAB2, PtmC, LegI
	CMP-*N,N'*-diacetyllegionaminic acid synthase	2.7.7.82	NeuA, NeuA2, NAB1, PtmB, LegF
Pse	pseudaminic acid synthase	2.5.1.97	NeuB, NeuB3, NAB2, PseI
	pseudaminic acid cytidylyltransferase	2.7.7.81	NeuA, NeuA3, NAB3, PseF

1.3.1 Neu pathway

The *nab* gene cluster directing the synthesis of bacterial Neu products was first identified in *Escherichia coli* and *Neisseria meningitidis* serogroup B, as a part of their capsule biosynthetic region which is presented in Figure 4 [92, 93]. The *neuA-C* homologs (named *siaA-C* or *cssA-C* in *N. meningitidis*) encode the three core NAB enzymes NeuA, NeuB and NeuC, leading to the synthesis of CMP-Neu5Ac [94-97]. Other organisms contain an additional gene coding for an *O*-acetyltransferase (NeuD family or non-homologous proteins) responsible for diacetylated Neu products [98, 99]. The only other known modification of Neu is a tri-acetylated product that was identified in one strain of *E. coli* which genome has not been released [41].

The Neu biosynthesis pathway leads to derivatives of CMP-Neu5Ac from the precursor UDP-*N*-acetylglucosamine (UDP-GlcNAc, see Figure 4). There seems to be two alternatives depending on whether the coding sequence for the NeuD *O*-acetyltransferase homolog is present. Without NeuD, the synthesis of CMP-Neu5Ac is performed in three steps: first the epimerization of UDP-GlcNAc to *N*-acetylmannosamine (ManNAc) by NeuC, then the condensation of ManNAc and phosphoenolpyruvate (PEP) to Neu5Ac by NeuB, and last the activation of Neu5Ac by NeuA, leading to CMP-Neu5Ac. Organisms which genome codes for NeuD have an additional step of *O*-acetylation that has been shown

to happen before activation, and may be the first step in Neu biosynthesis [100, 101]. They also possess a bi-functional NeuA with an *O*-acetyl-esterase activity [102].

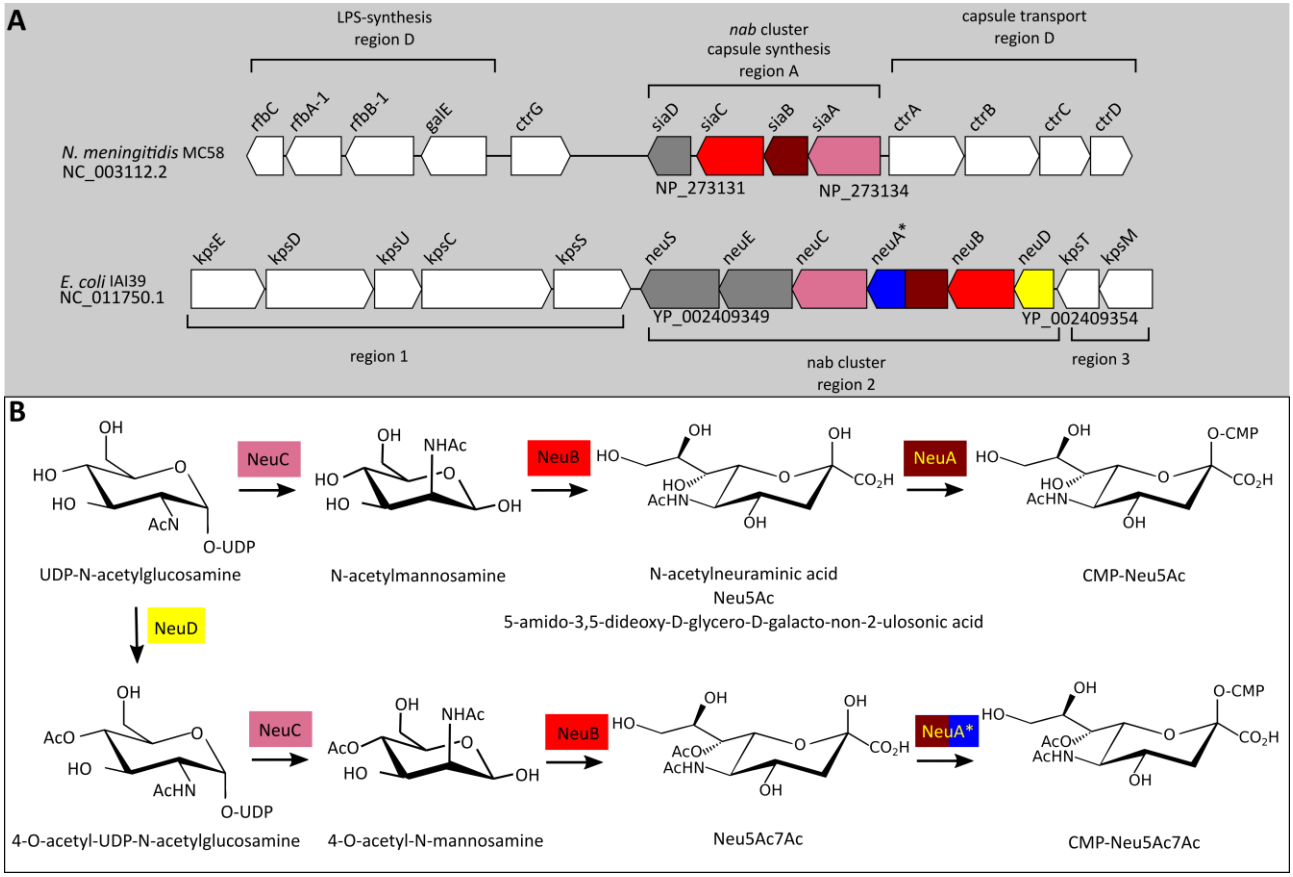


Figure 4. Alternative Neu biosynthesis pathways and biosynthesis clusters. *A: Capsule synthesis clusters in N. meningitidis MC58 (genome sequence accession number NC_003112.2) and E. coli IAI39 (NC_011750.1). They contain genes for the synthesis (colored according to the activity of protein products) of Neu, its transfer onto glycoproteins (siaD/neuS), and its polymerization in the case of E. coli (neuE). The accession numbers for the protein sequences coded by the first and last genes of each nab cluster are indicated below their respective locations. B: Biosynthesis pathways of Neu5Ac and Neu5Ac7Ac. Activated neuraminic acid (CMP-Neu) is biosynthesized from UDP-N-acetylglucosamine (UDP-GlcNAc) along two alternate routes, producing either mono- or di-acetylated Neu. For organisms without the neuD gene, UDP-GlcNAc is directly epimerized to N-acetylmannosamine (ManNAc) by the UDP-GlcNAc-2-epimerase NeuC and homologous proteins. Organisms with NeuD first synthetize O-acetylated-GlcNAc which then epimerized by NeuB to form O-acetylated-ManNAc. The ManNAc-based compounds are then condensed with phosphoenolpyruvate (PEP) by the Neuraminic acid synthase NeuB to form neuraminic acids, which are activated by the CMP-Neu synthetase NeuA. In the case of NeuD containing organisms, the NeuA homolog is a bifunctional enzyme with esterase activity (NeuA*). Adapted from [101].*

The enzymes of the Neu pathway have been characterized in several species, and the NeuA and NeuB crystal structures from *N. meningitidis* are available [103, 104]. NeuB is the most studied enzyme of all the NAB pathways and has been characterized in the most organisms [36, 37, 94, 105-109]. It is known as *N*-acetylneuraminate synthase and is classified as a transferase (EC 2.5.1.56). An inhibitor has recently been synthetized [110]. The UDP-*N*-acetylglucosamine 2-epimerase NeuC is a glycosidase (EC 3.2.1.183) which hydrolyses UDP-GlcNAc to free ManNAc, not to be confused with its "non-hydrolyzing" counterpart (isomerase, EC 5.1.3.4) producing UDP-ManNAc [95, 111, 112]. NeuA, also known as *N*-acetylneuraminate synthetase or cytidylyltransferase, is a nucleotidyltransferase (EC 2.7.7.43) responsible for the activation of Neu5Ac by attaching a CMP group on the oxygen carried by C2 [113, 114]. The *O*-acetyl-esterase domain (enzyme activity corresponding to EC 3.1.1.53) of the bifunctional enzymes is thought to regulate the *O*-acetylation level of capsular glycans [102]. As for the other enzymes in the pathway, they have been characterized in several bacterial organisms [113, 115-120].

1.3.2 Leg pathway

Legionaminic acid and its derivates can be synthetized from either UDP- or GDP-GlcNAc, but the pathway using GDP-linked intermediates seems to be the biologically relevant one [121, 122]. Biosynthesis from UDP-GlcNAc goes via the synthesis of UDP-*N,N′*-diacetylbacillosamine (UDP-Bac2Ac4Ac), which is known as a component of bacterial glycoproteins [123]. The genes responsible for this part are the *pglD-F* homologs, which code for an acetyltransferase (PglD), an aminotransferase (PglE), and a dehydratase (PglF), respectively. PglF is responsible for the removal of the -OH group at C6 and the oxidation of the -OH at C4 of UDP-GlcNAc. The C4 ketone is replaced by an amino group via the action of PglE, and an acetyl is attached to it by PglD [124]. The pathway employing GDP-linked intermediates, presented in Figure 5, proceeds in a similar fashion using a different set of dehydratase/aminotransferase/acetyltransferase [122]. They are coded by the *legB*, *legC* and *legH* genes, respectively. In both cases, the resulting compound is processed by a hydrolyzing homolog of the NeuC epimerase to give 2,4-diacetamido-2,4,6-trideoxy-D-mannopyranose which is then condensed and activated by the NeuB and NeuA homologs, respectively.

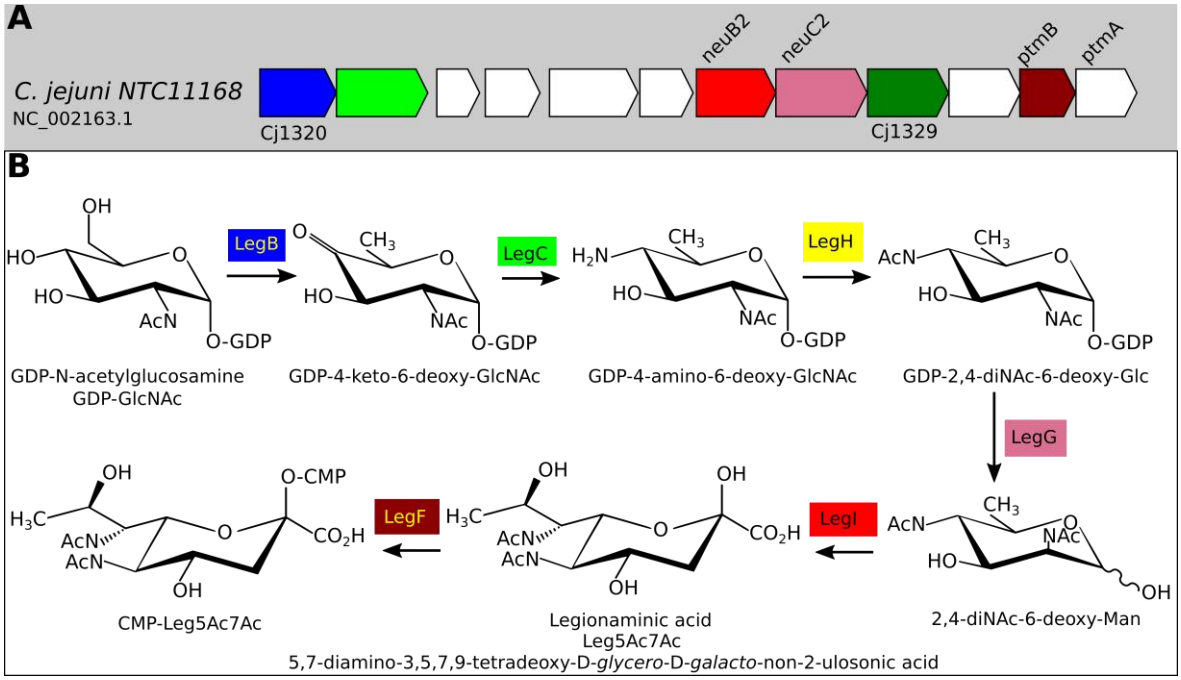


Figure 5. Legionaminic acid biosynthesis pathway from GDP-N-acetylglucosamine in C. jejuni. *A: Leg biosynthesis cluster [125, 126]. The transferase identified as LegH (coded by Cj1298) during pathway elucidation is not coded in the vicinity of the cluster, which contains several putative transferase sequences that have not been investigated (genes Cj1321-Cj1325). B: Leg biosynthesis from GDP-N-acetylglucosamine (GlcNAc). The dehydratase LegB, aminotransferase LegC and N-acetyltransferase LegH (or another transferase within the cluster) successively modify GDP-GlcNAc, replacing the OH group in C4 by an N-acetyl group. The glucosamine derivate is then epimerized by the NeuC homolog LegG to 2,4-di-N-acetyl-6-deoxy-mannosamine, which is condensed with PEP by the NeuB homolog LegI and thereafter activated by the NeuA homolog LegF (also known as PtmB). Adapted from [122].*

Both LegB and PglF are NDP-*N*-glucosamine 4,6-dehydratases (EC 4.2.1.135 for PglF) characterized in *Campylobacter jejuni*, coded by the non-homologous Cj1319 and Cj1120c genes, respectively [122, 124, 127]. The crystal structure of the latter has recently been published [128]. In a similar manner, the aminotransferases LegC and PglE are non-homologous enzymes. PglE is a UDP-*N*-acetylbacillosamine transaminase (EC 2.6.1.34) for which the crystal structure is available, and LegC a pyridoxal-phosphate (PLP) -dependent GDP-4-amino-4,6-dideoxy-α-D-*N*-acetylglucosamine aminotransferase [129, 130]. PglD is a UDP-*N*-acetylbacillosamine *N*-acetyltransferase (EC 2.3.1.203), while LegH is reported as a GDP-4-amino-4,6-dideoxy-α-D-*N*-acetylglucosamine *N*-acetyltransferase. The crystal structure of PglD is available [131]. The resulting compound is NDP-2,4-diacetamido-2,4,6-trideoxy-α-D-glucopyranose (or NDP-*N,N'*-diacetylbacillosamine). As mentioned earlier, the second part of the pathway is carried out by Neu homologs [121, 122]. They are all known by several names. The NeuC homolog (or NAB3, NeuC2, PtmD, and LegG) is categorized as a hydrolyzing UDP-*N,N'*-diacetylbacillosamine 2-epimerase (EC 3.2.1.184). The NeuB homolog is the *N,N'*-diacetyllegionaminate synthase (EC 2.5.1.101), also known as NAB2, NeuB2, PtmC, or LegI. At last, the NeuA homolog (NAB1, NeuA2, PtmB, LegF) is the CMP-*N,N'*-diacetyllegionaminic acid synthase (EC 2.7.7.82).

While bacterial Neu variants that have been identified so far only differ by their degree of *O*-acetylation, Leg compounds are a much more diverse family (see Table 1). For most of the Leg derivates, identified as a part of the structural characterization of bacterial surface elements, genomic information has not been released and the manner of their biosynthesis remains unclear. The biosynthesis clusters have however been located for a few of them, which are presented in Figure 6. They seem to be composed of a set of six core genes, with the three Neu homologs and the dehydratase/aminotransferase pair responsible for the first steps of the synthesis. The last common gene does not correspond to the acetyltransferase LegH that was characterized for the pathway in *C. jejuni*, but codes for (putative) homologs of the NTP-transferase PtmE (colored dark green in Figure 6), which catalyzes the synthesis of GDP-GlcNAc [122]. The sequence coding for the aforementioned acetyltransferase can be replaced by other kinds of acyltransferases, which lead to different substituents at the C7 position of Leg.

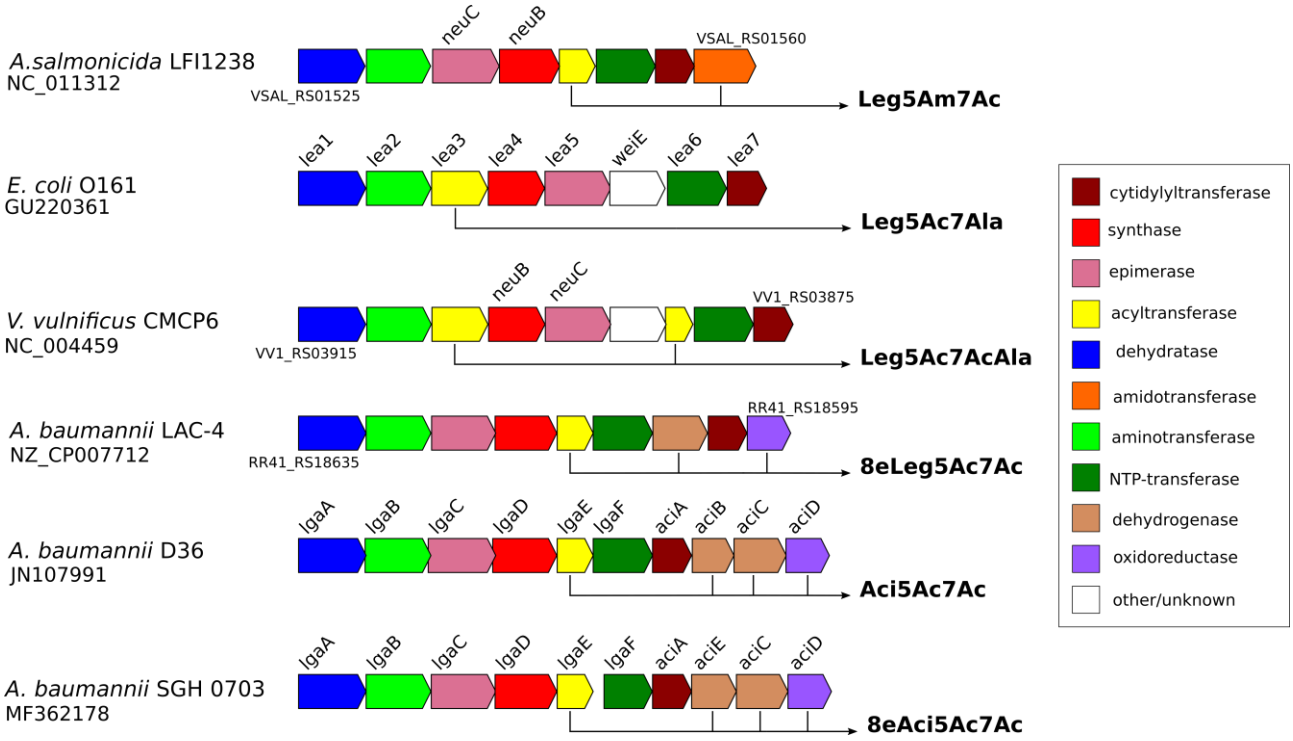


Figure 6. Legionaminic acid biosynthesis clusters producing Leg derivates. *For each strain, the accession number for its genomic sequence is indicated under its name. The genes of each cluster are labelled as they are in public databases and colored according to (putative) function as indicated by the legend on the right. Locus tags are given for the first and last gene of clusters where no gene names were assigned. Coding sequences responsible for extra modification of Leg and the corresponding NulO structure are indicated by black arrows. A. salmonicida LFI1238 [71, 72, 132, 133]. E. coli GU220361 [74]. V. vulnificus CMCP6 [73]. A. baumanii LAC-4 [83, 134]. A. baumanii D36 [20, 135]. A. baumanii SGH 0703 [21].*

The Leg biosynthesis cluster from *Aliivibrio salmonicida* LFI1238 has not been studied yet, but its genome has been published [133]. The cluster encodes a putative *N*-acetyltransferase homologous to

NeuD (in yellow in Figure 6), as well as a putative *N*-acetyl amidotransferase (locus VSAL_RS01560, in orange), and is known to produce Leg5Am7Ac [19, 71, 72].

A pathway has been proposed for the synthesis of Leg5Ac7Ala in *E. coli* O161, involving the putative alanyl transferase coded by *lea3* (shown in yellow)[74]. The cluster also contains an unknown sequence designated as the PLP-dependent transferase *weiE*, which may be involved in the synthesis of amino sugars [136]. Whether this sequence is essential to the synthesis of amino derivatives of Leg has not been investigated. The cluster from *Vibrio vulnificus*, which produces Leg5Ac7AcAla, contains a similar sequence (VV1_RS03905, referred to as *nab4* by the authors) as well as an additional putative *N*-acetyltransferase (VV1_RS03885, *nab5*) with an acetyltransf_3 domain (pfam13302) thought to add the extra acetyl group on the alanyl fragment [73].

Pathways for the synthesis of 8eLeg5Ac7Ac and Aci5Ac7Ac have also been proposed [20]. They involve the dehydrogenation of CMP-Leg5Ac7Ac by the MviM-like dehydrogenases ElaA (coded by RR41_RS18605, *elaA*, shown in light brown in Figure 6) and AciB, leading to compounds carrying a ketone group at C8 for 8eLeg and a double bond between C7 and C8 for Aci. Reduction of the ketone by the short-chain reductase ElaC (RR41_RS18595, *elaC*, colored light purple) gives 8eLeg5Ac7Ac. For Aci, it is proposed that AciB contains a reductase domain upon which action the D-*glycero*-L-*altro* isomer of Leg is formed, that is 8eAci5Ac7Ac. This seems a little particular when considering that *Acinetobacter baumanii* SGH 0703, which produces 8eAci5Ac7Ac, has a set of genes that does not include *aciB*, but another SDR coded by *aciE* as well as the two remaining homologs of the *aci* genes, *aciC* and *aciD* [21]. The proteins they code are supposed to perform the same as ElaA and C, albeit on the Aci substrate.

1.3.3 Pse pathway

The Pse biosynthesis pathway, presented in Figure 7, was determined in *Helicobacter pylori* and *C. jejuni* and follows a route similar to that of the Leg pathway [137]. It uses UDP-GlcNAc as starting block and involves a set of dehydratase/aminotransferase/acetyltransferase leading to the synthesis of the Pse precursor (UDP-2,4-diacetamido-2,4,6-trideoxy-β-L-altropyranose). The first enzyme of the pathway, PseB, catalyzes the 4,6-dehydratation as well as C5 epimerization of UDP-GlcNAc. The ketone in C4 is then replaced by an amino group as in the Leg pathway, but with a different (axial) orientation. The reaction, performed by the aminotransferase PseC, is followed by *N*-acetylation at this position catalyzed by the PseH *N*-acetyltransferase. The Pse pathway does not involve epimerization by the NeuC homolog, which coding sequence is replaced by that of the hydrolase PseG within the cluster. PseG removes the UDP group from the Pse precursor that is then processed by the NeuB and NeuA homologs to give activated Pse.

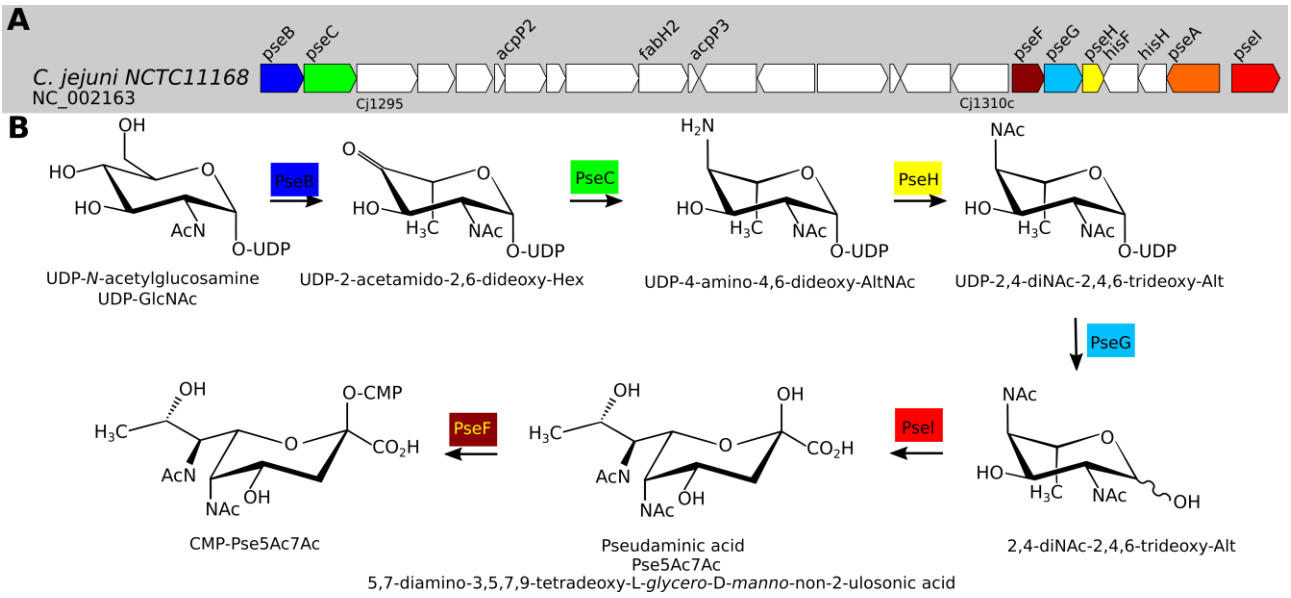


Figure 7. Pseudaminic acid biosynthesis pathway in C. jejuni NCTC 11168. *A: Pse biosynthetic genes within the C. jejuni genome [125, 126]. Genes for which no names have been assigned in the deposited sequence are indicated the first and last loci of the corresponding region. Genes involved in the biosynthesis of Pse are colored according to their function. B: Pse biosynthesis from UDP-N-acetylglucosamine (UDP-GlcNAc). The dehydratase PseB, aminotransferase PseC, N-acetyltransferase PseH and hydrolase PseH are responsible for the synthesis of the Pse precursor, 2,4-diacetamido-2,4,6-trideoxy-β-L-altropyranose from UDP-GlcNAc. The precursor is condensed with PEP by the NeuB homolog PseI, which is then activated by the NeuA homolog PseF, leading to CMP-Pse5Ac7Ac. The activated NulO can be modified by the amidotransferase PseA to give Pse5Ac7Am [35]. Adapted from [137].*

As mentioned earlier, the enzymes responsible for the synthesis of the Pse precursor belong to the dehydratase, aminotransferase, acetyltransferase families. While the Leg pathway also involves such enzymes in the same order, they are not homologs [127]. For example, PseB (or FlaA1) is a configuration-inverting UDP-*N*-acetylglucosamine 4,6-dehydratase (EC 4.2.1.115) where LegB is configuration retaining in addition to accepting GDP-linked substrates [127, 138, 139]. PseC is a UDP-4-amino-4,6-dideoxy-*N*-acetyl-beta-L-altrosamine transaminase (EC 2.6.1.92), and PseH (or FlmH, FlaG1) a UDP-4-amino-4,6-dideoxy-N-acetyl-beta-L-altrosamine *N*-acetyltransferase (EC 2.3.1.202) [127, 137]. The crystal structures for all three enzymes are available [140-143]. PseG is a UDP-2,4-diacetamido-2,4,6-trideoxy-beta-L-altropyranose hydrolase (EC 3.6.1.57), also with an available structure [144, 145]. The pseudaminic acid synthase PseI (NeuB3, EC 2.5.1.97) and the cytidylyltransferase PseF (NeuA3, EC 2.7.7.81) have been characterized [91, 137].

Identified Pse modifications are even more diverse than what has been found for Leg compounds, with substituents such as formyl or glyceroyl groups (see Table 1). Considering that Pse biosynthetic genes seem to be spread within regions coding for glycosylation and/or surface components, the available information about the proteins and genes responsible for the various modifications observed remains limited.

1.4 Integration in bacterial glycans and biological roles

After synthesis, activated NulOs are expressed on the surface of the bacterial cell where they influence the survival capabilities of the bacteria, through the modulation of adhesion and motility as well as providing protection against the environment. This is particularly relevant in the case of pathogenic bacteria, with surface NulOs serving both as antigens and means to evade host immune systems.

1.4.1 NulOs in cell surface components

The CMP-NulOs obtained by either *de novo* biosynthesis or from the environment are incorporated within different structures of the outer cell surface depending on the species. The lipopolysaccharide (LPS) structures of Gram-negative bacteria house the greatest variety of NulOs, which form diverse antigens. The LPS of *E. coli* O171 and O104 contain di- and tri-acetylated Neu, respectively [41, 146]. Several derivatives of Pse are present in *Pseudomonas aeruginosa*, and Leg was originally identified from the LPS of *Legionella pneumophilia* [54, 76].

Several bacteria are known to produce a polysialic acid capsule, which is composed of homo- or heteropolymers of Neu5Ac with different linkages [24, 25, 27]. In *E. coli* K1, polysialic acid capsule synthesis is directed by the cluster presented in Figure 4, with NeuF responsible for the elongation of

polysialic acid and NeuE a putative glycosyltransferase involved in the transfer onto the acceptor [147]. The Aci derivatives are found in the capsule of *A. baumanii* [20, 21, 148]. NulOs are also incorporated in glycoproteins, via the sequential O-glycosylation system. Flagellin subunits are modified with Leg and/or Pse in several species, and the pilin from *Pseudomonas aeruginosa* contains Pse [43, 47, 75, 149]. In *Tannerella forsythia*, Pse and Leg are found in glycoproteins of the S-layer [66, 88].

1.4.2 Role in survival and pathogenicity

Bacterial NulOs are involved in numerous processes relating to both cell integrity and communication with its environment, such as a host [73, 150]. The most striking example of the latter is the ability of pathogenic bacteria to evade the host immune system by coating their surface with NulOs structurally similar to- or directly from- host sialic acids [151, 152]. Depending on the type of sialylconjugate that is emulated, different parts of the host immune response are affected [153-155]. In a paradoxal way, the fact that Leg- and Pse-related compounds are exclusively synthetized by bacteria gives them antigenic properties. They also make bacteria sensitive to bacteriophages that recognize nonulosonic acid modifications [156, 157].

Aside from molecular mimicking, NulOs influence bacterial motility through flagellin glycosylation, which has been shown to be essential for flagella formation in several species [47, 158, 159]. Whether NulOs from the LPS are also involved in motility regulation remains to be investigated [160, 161]. In addition to motility, host colonization is affected by biofilm formation as well as similar multi-cell structures, again involving NulOs [162, 163]. It is important to note that the role of NulOs in pathogenicity is not limited to bacterial NulOs. Sialic acids from vertebrates are recognized by a number of pathogen proteins, not necessarily from bacteria [164]. The most famous example is the recognition and removal of host sialic acids by hemagglutinins and sialidases (neuraminidases) from influenza viruses. In bacteria, adhesion to host tissues can be mediated by sialic acid-specific adhesins [165].

1.4.3 NulOs in *A. salmonicida*, *M. viscosa*, and *A. wodanis*

A. salmonicida, *Aliivibrio wodanis*, and *Moritella viscosa* are fish pathogens causing cold water vibriosis and winter ulcer, diseases that particularly affecting farmed fish [166, 167]. In particular, *M. viscosa* and *A. wodanis* have been shown to co-infect Atlantic salmon and influence each other [168, 169]. Both *A. salmonicida* and *M. viscosa* can produce mono- (Neu5Ac) and di- (Neu5Ac7(9)Ac) acetylated Neu variants, and Leg5Am7Ac has been detected in the LPS of *A. salmonicida* [19, 37, 71, 101, 170]. Their genomes contain each a Neu a cluster for which the NeuB homologs have been characterized, as well as a Leg (*A. salmonicida*, shown in Figure 6) and a Pse (*M. viscosa*) cluster [36, 37, 133, 168]. While the presence of NulOs in *A. wodanis* has not been investigated yet, its genome contains a cluster that is similar to that of *V. vulnificus*, which produces Leg5Ac7AcAla [73].

1.5 NulO synthesis using GlcNAc epimerases

1.5.1 NulOs in research and industry

NulOs produced by pathogenic bacteria are involved in evading the host immune system by mimicking host sugars, but they can also act as antigens due to their unique structures. In addition, host sialic acids are the target of neuraminidases of bacterial and viral origin, a key factor in the spreading of pathogens such as influenza viruses in host tissues [171]. Because of this, they are prime targets for the synthesis of antibiotics based on their structures [172-174]. Applications of NulO analogues are not limited to antibiotics, however. Neu analogues can be used to regulate sialic acid synthesis in mammals, and a method for labelling of Pse-producing bacteria using an analogue of its precursor has been developed [175, 176]. With the importance of NulOs in pathogenicity also comes the necessity to unravel their biological roles and understand their metabolism, catabolism, as well as the regulation mechanisms that influence them.

In vertebrates, sialic acid is involved in numerous physiological processes [177]. Its presence in breastmilk, with its potential role in brain development, prompts its inclusion in infant formulas [178]. Polysialic acid could be used to promote axon repair, and its poor immunogenicity also makes it a good candidate for drug delivery systems [179, 180]. As a marker, upregulation of sialylation is a cancer indicator and serum sialic acid can indicate excessive alcohol consumption independent of liver disease [181, 182].

Considering their biological relevance, the demand for NulOs compounds is high, higher than their availability, and several strategies are pursued for their synthesis. Early methods for the production of Neu were focused on purification from natural sources, but they have been abandoned due to the low final yields [183, 184]. The current "biological" methods use whole-cell approaches with engineered bacteria [185-187]. They have been used for the synthesis of compounds from all three families as well as polysialic acid, with promising results. Various chemical synthesis methods have also been used and are still developed today [188-191]. They are however often complicated and resource hungry, and a variation consists in using *N*-acetylneuraminic acid aldolase, also known as *N*-acylneuraminate lyase (NAL, EC 4.1.3.3), for the condensation of chemically produced ManNAc with pyruvate [192-194]. Another chemo-enzymatic synthesis method using macrophomate synthase has recently been proposed [195]. The last strategy uses purified enzymes for the *in vitro* enzymatic synthesis of Neu compounds. The native enzymes from the NAB pathways are used mainly for the goal of their elucidation, while NAL enzymes coupled with *N*-acetyl-glucosamine 2-epimerases (AGEs, EC: 5.1.3.8) are more commonly used for synthesis purposes [122, 137, 196-202]. The AGE/NAL pair is used as soluble or immobilized enzymes, and some whole-cell methods also use them [202-205].

1.5.2 The AGE/NAL coupled reaction

The reactions catalyzed the *N*-acylneuraminate lyase and *N*-acetyl-glucosamine 2-epimerases were first described in the 1960s, while the coupled reaction (see Figure 8) was first used for Neu5Ac synthesis 30 years later [196, 206, 207]. The equilibrium of each of the reactions favors the opposite direction from Neu5Ac synthesis, but, compared to other methods of production, enzymatic synthesis still offers high yields at a relatively low cost [198, 208]. The use of AGEs and NALs instead of the metabolic enzymes is also cost-related. NAL uses pyruvate as co-factor while NeuB uses PEP, which almost 20 times more expensive. The use of AGEs allows synthesis from GlcNAc instead of ManNAc, three times cheaper, even though they require at least catalytic concentrations of ATP as allosteric regulator [197, 202]. UDP-GlcNAc, the NeuC substrate, is even more expensive than ManNAc.

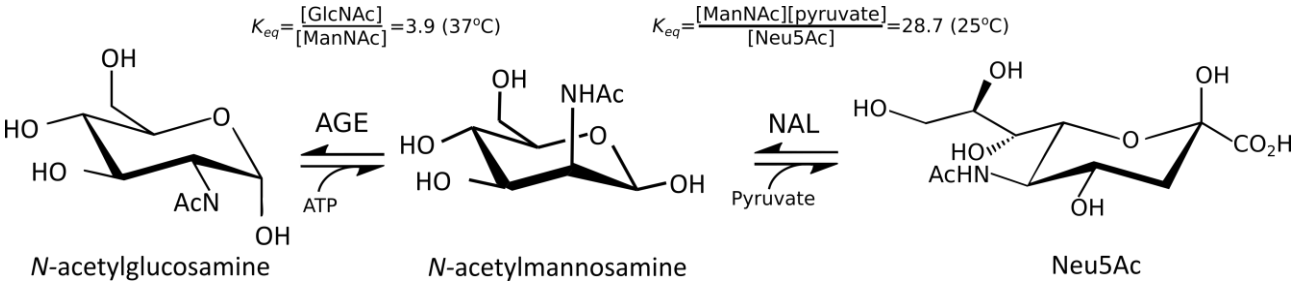


Figure 8. Neu5Ac synthesis using AGE and NAL. *The equilibrium constants determined by [207] and [196] are indicated above each reaction, and the co-factors necessary for ManNAc and Neu5Ac synthesis as a part of the corresponding half arrow.*

Several studies on the parameters of both reactions as well as the synthesis process have been performed, and a kinetic model representing the system has been proposed [208]. As mentioned earlier, the equilibriums are in the favor of the formation of GlcNAc and ManNAc, respectively; the rate-limiting step was shown to be the condensation reaction [202]. In order to push the equilibrium towards the synthesis of Neu5Ac, an excess of pyruvate is used. However, pyruvate inhibits AGEs and it has a similar pK_a to that of Neu5Ac, which makes their separation by ion-exchange chromatography difficult [196, 209]. Increasing the amount of ManNAc available for the condensation by using an excess of GlcNAc also shifts the equilibrium towards Neu5Ac, and both ManNAc and GlcNAc can easily be separated from Neu5Ac. NAL is however inhibited by GlcNAc [194]. Optimization is then achieved by managing the ratios of substrates and enzymes [196, 202, 208].

A critical factor of this process is the enzymes themselves, motivating the characterization of both AGEs and NALs in multiple organisms in search of candidates suitable for Neu5Ac synthesis. For example, NALs with higher catalytic efficiency would raise the speed of the rate-limiting step of the reaction. One such candidate was characterized recently [210]. Enzymes that are less sensitive to inhibition would also be an advantage and using AGEs that do not require ATP would considerably reduce the cost of the reaction. In this regard, the AGEs from rat and *Anabaena variabilis* are good candidates, requiring only a fourth of the ATP concentration needed by other enzymes [202, 211]. In addition to this are considerations on temperature, buffer, and pH requirements.

The characterization of new AGEs and NALs also allows for a further understanding of their mechanisms and requirements, opening the possibility for their engineering. Structural determinations are a valuable source of information, and numerous structures are available for both NAL wild-types and mutants, including one from *A. salmonicida* [212]. For comparison, only 2 structures are available for AGEs, and several aspects of their properties remain unclear [213, 214]. No AGE structure exists in complex with either substrate, inhibitor nor nucleotide, thus providing no definite proof so as to how they interact with each other.

1.6 GlcNAc epimerases: characteristics and bioprospecting

1.6.1 Reaction mechanism and regulation of AGEs

AGEs belong to the enzyme class of isomerases acting on carbohydrates and derivatives (EC: 5.1.3.-). They catalyze the reversible epimerization of ManNAc to GlcNAc via deprotonation/reprotonation, with an acid/base catalytic pair [207, 214-216]. In the current accepted model, the acid moiety abstracts the proton carried by the C2 position of the substrate in open form, leading to an enolate intermediate. The base donates a new proton to it on the opposite face of C2, thus inverting the configuration at this position. The residues involved in those steps have been shown to be a Glu and Arg for the AGE from *Pedobacter heparinus* (PhGn2E), its reaction mechanism presented in Figure 9 [216].

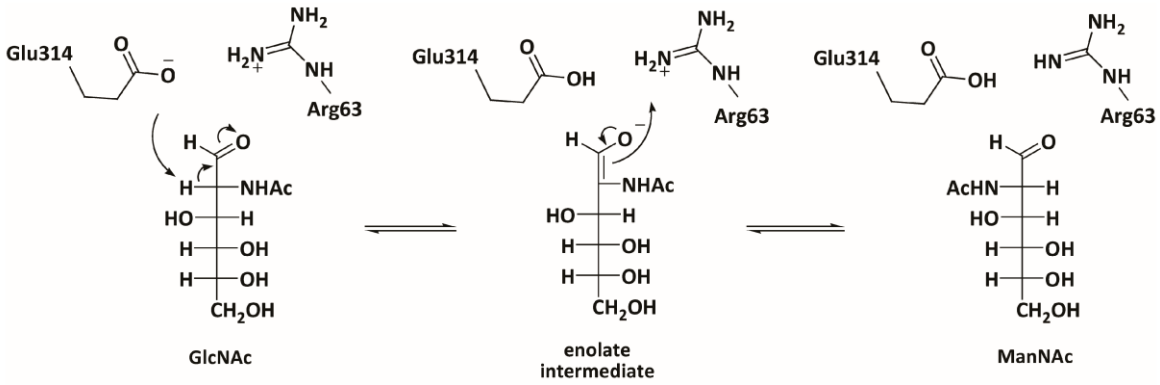


Figure 9. Model for the reaction mechanism of PhGn2E. *The deprotonation/reprotonation mechanism was determined by deuterium exchange experiments, and the involved residues (Glu314 and Arg63) were identified by mutagenesis. Figure taken from [216].*

Activity of AGEs is modulated mostly by the presence of nucleotides, in particular ATP [202, 211, 217]. It is not consumed during the reaction but can be essential and enhances AGE activity up to 20-fold. The manner of its regulation mechanism is not known, and the connection between cellular ATP levels and AGE activity has not been investigated yet.

AGEs are inhibited by pyruvate and Neu5Ac, as well as renin, although it can also be activated by the latter [218]. For a time, they were identified as renin-binding proteins (RnBP), until their cloning revealed their true nature [219, 220]. This raised the question of whether AGEs were involved in the renin-angiotensin system (RAS), which regulates blood pressure, due to their ability to specifically inhibit renin and their coexistence in renal tissues [221, 222]. However, it was shown that mice lacking RnBP are normotensive and that neither renin expression nor activity was affected [223]. Investigations on their role in mammals suggest that they may participate in sialic acid catabolism [224]. As for nucleotide regulation, the mechanim(s) behind AGE inhibition are yet unclear.

1.6.2 Biochemical properties of characterized AGEs

The first AGE to be characterized, pAGE, was isolated from hog kidney [207]. It was used to determine the equilibrium constant of the reaction (K_{eq}=3.9 in favor of GlcNAc formation), the affinity of pAGE for its substrates (K_m=3.4 mM for GlcNAc and 3.3 mM for ManNAc), and its ATP requirement (K_m=1.1 mM for the epimerization of GlcNAc to ManNAc). Further studies on the role of ATP revealed it to act as a non-essential allosteric activator, and it was shown that other nucleotides could replace it, albeit with different affinities and effects on enzyme activity [211, 217]. The protein was later found to elute as a dimer, with a monomer mass around 45 kDa and three cysteine residues involved in dimer interface [219, 225, 226]. It is inhibited by renin, pyruvate, and Neu5Ac [196, 219, 227]. Other mammalian AGEs have also been characterized, with the human enzyme (hAGE) showing an absolute requirement for ATP [211, 220]. The rat enzyme (rAGE) was shown to have an especially high affinity for ATP and was likewise activated by ADP.

The first bacterial AGE to be characterized was from *Synechocystis* sp., with properties similar to that of the porcine and human enzymes [199]. A turning point was reached when the enzyme from the cyanobacterium *Anabaena* sp. was characterized, revealing a 4-fold higher activity coupled with a 100-fold lower dependency on ATP [202, 214]. It was also activated by AMPPNP, the non-hydrolysable analog of ATP, confirming previous observations that ATP is indeed not used in the reaction. The potential of AnaAGE for use in Neu synthesis prompted further studies on several cyanobacterial AGEs, and the enzyme from *Anabaena variabilis* (AvaAGE) was shown to be less sensitive to pyruvate while having similar activity and ATP affinity as AnaAGE [227]. Bacterial AGEs from other families have also been studied [216, 218].

In order to identify the residues critical to AGE function and regulation, numerous mutagenesis studies have been performed [214, 215, 228-232]. From these studies, several residues critical to AGE activity, inhibition, and ATP-binding could be identified. Such residues are the C380 of hAGE or the active site residues of AnaAGE. Two putative sites for ATP-binding were proposed, but lack of AGE structures complexed with ligands prevents the definitive identification of the residues involved [202, 218, 229-231]. This holds true for the catalytic residues as well, which are suspected to be a Glu/Arg pair although His residues are essential for AGE activity [214, 216].

1.6.3 Structure of AGEs and the AGE superfamily

As mentioned earlier, only two AGE crystal structures are available as of the end of 2018 [213, 214]. They are from pAGE and AnaAGE, with the latter using the former as template for molecular replacement. Figure 10 presents an overview of the pAGE and Ana AGE structures, with the AGE monomers consisting of 12 helices folding as a $(\alpha/\alpha)_6$-barrel. The faces of the AGE barrel can be designed as "front" and "back", according to the length of the loops that join each pair of helices. The front face is then the one where the loops are long, whereas the back face has short loops. Both crystals belong to the same space group ($P2_12_12_1$), and the unit cell is composed of two monomers interacting via their back faces.

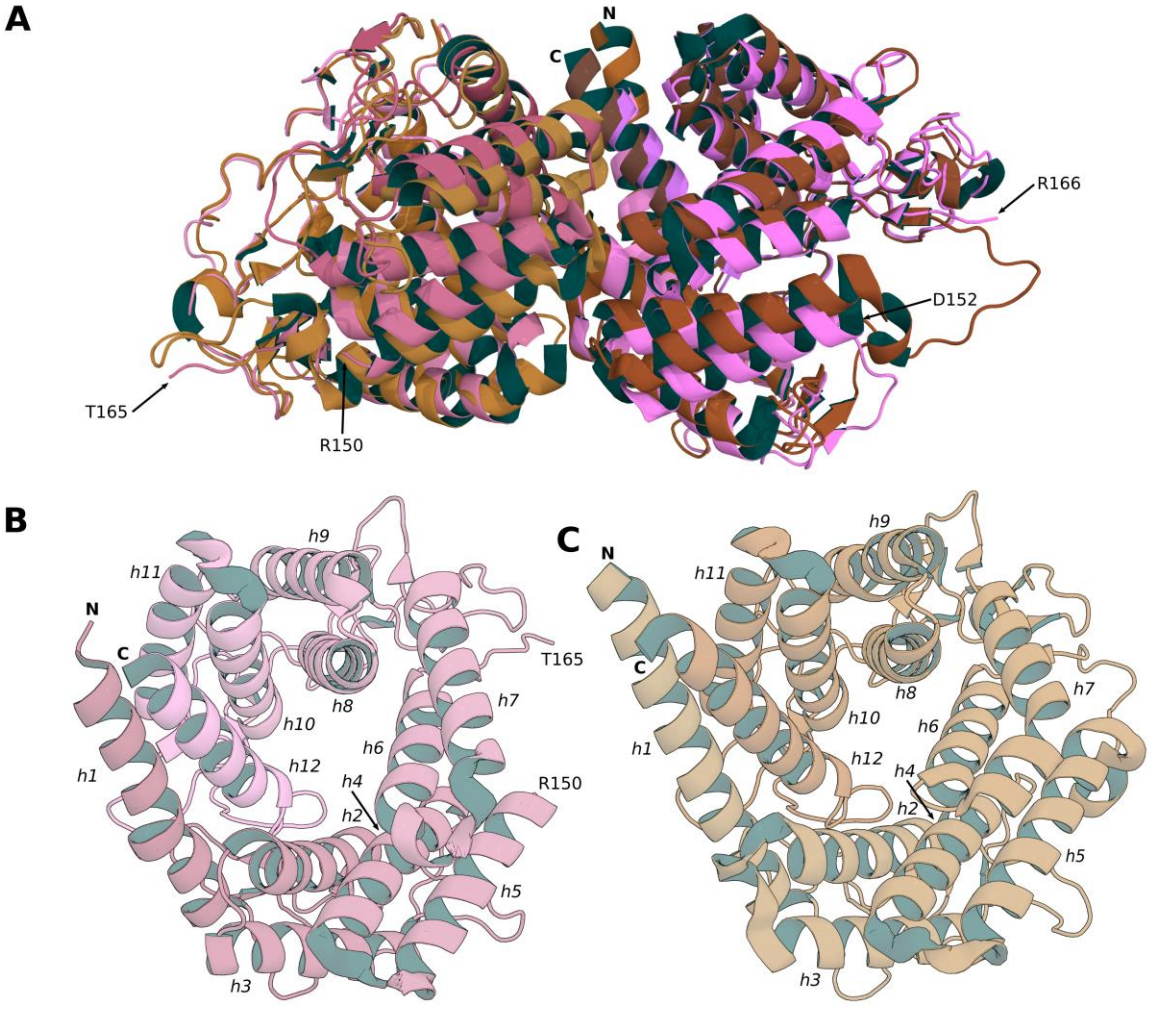


Figure 10. Crystal structures of pAGE and AnaAGE. *A: AGE unit cell (polymer only) for pAGE and AnaAGE. The AnaAGE protein chains, colored in a pink palette along the residue count, is superimposed onto that of pAGE, which is represented in a brown palette. The superimposition was obtained by structural alignment of the Cα. The missing loop of AnaAGE is indicated by black arrows, and the corresponding residues are labelled. Only the N- and C-terminal residues of pAGE are visible, they are indicated. B,C: Back face of the AnaAGE (B) and pAGE (C) monomers. A-helices are numbered along the amino-acid sequence, and the residues surrounding the missing loop of AnaAGE are labelled. The AnaAGE and pAGE structures correspond to the PDB IDs 2GZ6 and 1FP03, respectively [213, 214].*

The $(\alpha/\alpha)_6$-barrel fold of AGEs is common amongst sugar metabolizing enzymes, even though their sequences are diverse [213, 214]. In particular, AGEs form, together with cellobiose 2-epimerases (CE; EC 5.1.3.11) and *yihS*-encoded sulfoquinovose isomerases (EC 5.3.1.31), the AGE superfamily (Pfam accession No. PF07221) [233, 234]. The structural comparison of pAGE and YihS from *Salmonella enterica*, which is in complex with mannose, showed that active site residues were conserved between the two enzymes [235].

1.6.4 Cyanobacterial sources for AGEs and NGS technology

As mentioned above, it seems that AGEs from cyanobacterial sources are the best suited to the synthesis of Neu5Ac. Cyanobacteria are a very diverse group found in all kinds of environments. They are able to perform oxygenic photosynthesis, the only prokaryotes to do so, playing a significant role in the oxygenation of the Earth's surface and atmosphere [236]. They are able to fix carbon and nitrogen, as well as produce secondary metabolites that can be used for a wide range of applications such as bio-fuels, biocides or pharmaceuticals [237-241].

For these reasons, cyanobacteria are a target for bioprospecting endeavors around the world [242-245]. The success of these studies largely depends on the technologies that have been developed in the last two decades, in particular next-generation sequencing (NGS). Indeed, up until 2005 the main way to obtain genetic data was by the Sanger method [246]. This method exploits the inability of DNA polymerase to extend a nascent DNA chain from a dideoxy base (ddNTP), because it has no 3'-hydroxyl group [247]. Using a fraction of ddNTP, the reaction leads to a range of DNA fragments of various sizes, each ending by the chosen base [248]. The reaction is performed in parallel for each nucleotide, and the fragments are separated according to size by electrophoresis on acrylamide gel. The original was improved by several advancements over the years, such has the use of capillaries instead of gels, fluorescent terminators and polymerase chain reaction [249, 250].

Several NGS methods are available, which can be classified in two groups according to whether they use single-molecule or PCR-amplified DNA templates. They are described and illustrated in a comprehensive way in the review by Metzker *et al*. [251]. For single-molecule DNA sequencing, the most commonly used technique is sequencing-by-synthesis, where fluorescent nucleotides are excited as they are incorporated by DNA polymerase [252]. The sequencing methods requiring amplified templates include cyclic reversible termination (CRT), sequencing by ligation (SBL), single-nucleotide addition (pyrosequencing), and real-time sequencing [251].

CRT uses reversible terminators, that is modified nucleotides triggering the termination of synthesis, but for which the 3'-blocking group can be removed, thus allowing further elongation by DNA polymerases (as opposition to ddNTPs, which are "irreversible terminators"). In this process, the

templates and primers are immobilized on solid support. Primers are then extended by one base, because all present nucleotides are terminators. After washing away the free nucleotides, the newly incorporated base is identified by the color of its fluorophore. Both the fluorescent tag and the 3'-blocking group are then removed, and a new cycle of extension can be performed (after washing away the cleavage products).

Pyrosequencing also uses DNA polymerase, but instead of using a mix of terminator nucleotides, it uses a single type of dNTP in limiting amounts for the extension reaction [253, 254]. Incorporation is not detected by fluorescent nucleotides, but by following the release of inorganic pyrophosphate using quantitative bioluminescence. The extension reaction is paused after incorporation, and the excess nucleotide is either washed or degraded. A new cycle can then be performed with a new dNTP. In the case of real-time sequencing, the incorporation of fluorescent nucleotides (for which the tag does not interfere with the polymerization reaction) is measured in real time, while they are retained in the active site of an immobilized DNA polymerase. The method heavily reposes on the detection volume, which has to be localized around the DNA polymerase so as to avoid the presence of other fluorescent dNTPs.

For SBL, fluorescent oligonucleotide probes containing either one or two interrogation bases and a few universal and degenerate bases are used in a cyclic ligation reaction. Probes are incubated with the template/primer pair for hybridization. DNA ligase is then added to the reaction mix in order to join the hybridized probe to the primer. After washing, the identity of the probe is revealed by fluorescence imaging. As for the nucleotides used in CRT, the probes are regenerated by removing their fluorescent tag and, in this case, regenerating the groups necessary for further ligation reactions. Several variations of this method exist, with different combinations of primers and probes.

Independently of the method used, the improvement was such that it is now common to sequence entire genomes as well as metagenomes as a part of other studies rather than stand-alone projects. The data thus generated can be used for different purposes.

2 Aims of the study

This study is a part of a bigger project that aims to explore NulO biosynthesis in the *A. salmonicida* and *M. viscosa* fish pathogens. The characterization of the enzymes composing their NAB pathways is a major aim for the group, and the NeuB homologs from the Neu pathway have been characterized [36, 37]. The cloning, expression and characterization of acetyltransferases was the primary aim of this study, however the difficulties in cloning and expressing those proteins lead to the termination of this part of the project for the time being. Instead, it was chosen to focus on bioprospecting sources for cyanobacterial AGEs. Indeed, the ability to efficiently synthetize Neu from GlcNAc presents several advantages and opens multiples possibilities for the project. In this direction, the NAL from *A. salmonicida* has been characterized, and only the AGE remains for us to perform the coupled reaction necessary for Neu5Ac synthesis [212]. The two *Nostoc* sp. strains KVJ10 and KVJ20 were chosen as sources for AGEs, the former with its genome already sequenced and annotated. Again, the final aim was the characterization of the protein products.

The other aim of the study was to map the kind of NulOs produced by these bacteria, taking advantage of a database of local strains from the same species. Considering the diversity of *nab* clusters within species of the *Vibrionaceae* family, it is reasonable to expect multiple unique clusters even within a set of closely related organisms [23, 255]. Those clusters might direct the biosynthesis of previously undescribed NulOs. In addition to this, the analysis of the protein coding sequences might provide us with new targets for further characterization.

3 Summary of Papers

3.1 Paper I

Analysis of nonulosonic acid biosynthetic gene clusters in *Aliivibrio salmonicida* and *Moritella viscosa*.

Halsør MJH, Altermark B, Ræder ILU.

Exploration of the diversity of NulOs in bacteria is at its beginnings, and new bacterial NulOs keep being identified. The sequencing of numerous bacterial genomes presents a source of data concerning their *nab* gene clusters, and their analysis indicates which type of NulO might be produced. Considering the highly variable nature of *nab* clusters, it may also reveal new genes leading to previously unidentified compounds. This paper presents the study of *nab* gene clusters in a set of *A. salmonicida* and *M. viscosa* strains, with the identification of three unique clusters for which NulO structure was predicted. The NulO content of *A. salmonicida* LFI1238, *M. viscosa* 06/09/139, *A. wodanis* 06/09/139, *A. salmonicida* R8-63 and *Vibrio* B9-25K2, the two latter with a new *nab* cluster, was released and labelled with 1, 2-diamino-4, 5-methyleneoxybenzene (DMB). The quinoxalines were analyzed by LC-MS, revealing an unidentified peak at m/z = 530 for that of *A. salmonicida* R8-63 and verifying the presence of $PseAc_2$ for *Vibrio* B9-25K2. MS/MS analysis of the $PseAc_2$ compounds from *M. viscosa* and *Vibrio* showed that they do not share the same configuration. Considering the unknown coding sequences in the *nab* cluster of *Vibrio*, it may produce the first epimer from the Pse pathway to be reported. In addition to this, the sequences for NeuB homologs were investigated for the presence of specificity regions, which discriminate between Neu, Leg, and Pse substrates. A critical residue for the specificity of the Pse pathway was thus identified.

3.2 Paper II

Draft genome of the Self-Cleaning Cyanobacterium *Nostoc* sp. KVJ20.
Halsør MJH, Liaimer A, Ræder ILU, Pandur S, Smalås AO and Altermark B.

The cyanobacterium *Nostoc* sp. KVJ20 was isolated from the symbiotic organs of the liverwort *Blasia pusilla L.*, which is found as a weed in a plant-school on the island of Kvaløya, Northern Norway. Its "self-cleaning" properties as well as the potential applications for its cellular components motivated the full-length sequencing of its genome. It totals 9,183,495 bp distributed amongst 425 contigs forming 332 scaffolds, with 41.69% GC. 7,676 genes were identified, with 7,210 protein coding sequences and 104 RNA genes. Several biosynthesis clusters for secondary metabolites that may confer its antibiotic properties were identified.

3.3 Paper III

The crystal structure of the *N*-acetyl-D-glucosamine 2-epimerase from Nostoc sp. KVJ10 reveals the true dimer.

Halsør MJH, Rothweiler U, Altermark B, and Ræder ILU.

AGEs catalyze the interconversion of GlcNAc and ManNAc. They are used industrially to synthetize sialic acid, and enzymes from cyanobacterial sources possess the best kinetic parameters of all characterized AGEs. This paper presents the crystal structure of the AGE from the local cyanobacteria *Nostoc* sp. KVJ10 (called nAGE10), which differs from previously published AGE structures in that the dimer organization is different while the monomer folds in an identical fashion as a $(\alpha/\alpha)_6$-barrel. The new dimer involves the faces of the monomer where the helices are linked by long loops ("front" faces), while the previous organization was "back-to-back". The "front-to-front" dimer has better parameters in terms of interaction surfaces and allows a better explanation of observations made earlier from biochemical experiments. In addition to this, the nAGE10 dimer, which was crystallized in a different space group than that of previous structures, can be generated by symmetry within those structures, while the "back-to-back" dimer can't. Taken together, these results strongly suggest that the new dimer represents the biological organization of AGEs.

4 Results and Discussion

4.1 Diversity of *nab* clusters in *Vibrionaceae*

With the goal of exploring NulO biosynthesis in *A. salmonicida* and *M. viscosa*, we investigated in paper I the genomes from a set of strains from these bacteria. We also included strains of other species such as *A. wodanis*, which is known to co-infect together with *M. viscosa¸ Vibrio*, and *Photobacterium¸* the latter as a negative control since it is known to not produce NulOs (see Table 1, paper I). In order to do this, we first identified potential *nab* clusters by searching for NeuB homolog protein sequences, the only enzyme of NulO synthesis that is common to all pathways. The queries used were the NeuB sequences from the published *A. salmonicida* LFI1238 (2 sequences, Neu and Leg pathways), *A. wodanis* (1 sequence, Leg pathway) and *M. viscosa* (2 sequences, Neu and Pse pathways) 06/09/139 strains, covering at least one sequence per pathway [133, 168].

The results revealed that all the strains apart from the *Photobacterium* and *Vibrio anguillarum* ones possessed at least a NeuB homolog sequence, as was expected. Amongst the 8 *M. viscosa* strains, 6 of them possessed sequences identical to both homologs from the *M. viscosa* query, the two remaining (strains F57 and NVI5482) having one sequence identical to each of the query homologs, respectively (Neu pathway homolog for F57 and Pse pathway homolog for NVI5482). Within the four *A. salmonicida* strains, two possessed sequences identical to both homologs from the *A. salmonicida* LFI1238 query (strains R8-68 and R8-70). The other two strains contained unique NeuB sequences, closest to the NeuB homologs from the Leg pathway from either the *A. salmonicida* (strain MR17-77, 97% sequence identity) or *A. wodanis* (strain R8-63, 97% sequence identity) queries. The last sequence in the set, from *Vibrio* B9-25K2, was closest to the NeuB homolog from *M. viscosa* from the Pse pathway, although the sequence identity was only 57%, against identities above 90% for the rest of the sequences investigated. Considering this, a wider database search was performed using this sequence, leading to closer sequences (around 80% sequence identity) from other *Vibrio* species, but still corresponding to the Pse pathway according to their annotation.

Once the NeuB homologs were identified, they were used to investigate the presence and content of *nab* clusters at their locations within the target genomes. All of them indeed presented *nab* clusters (see Figure 3, paper I), supporting the efficiency of using NeuB homolog similarity searches for their detection. We were able to assign a putative function to most of the coding sequences by searching public databases for similar sequences that were annotated as well as conserved domains. We were thus able to identify a putative cytidylyltransferase non-homologous to NeuA for *Vibrio* B9-25K2. The comparison of gene content and organization of each cluster revealed that in the case of identical NeuB sequences, the corresponding *nab* clusters were also identical although this was only true for our set of sequences. On the other side, non-identical NeuB sequences, even if they were highly similar, meant that

the clusters were different. This is especially obvious in the case of *A. salmonicida* R8-63 (similar to *A. wodanis*) and MR17-77 (similar to *A. salmonicida* LFI1238), which NeuB sequences share 97% identity with that of their closest homologs but still have different clusters.

The cluster analysis confirmed the information gained from the comparison of NeuB sequences and gave further insight so as to what the final NulO produced by each strain may be, at least in part. All the *M. viscosa* strains except of NVI5482 are expected to produce NeuAc and $NeuAc_2$, the latter due to the presence of an *N*-acetyltransferase within the cluster. For Pse, only the F57 should not be able to produce $PseAc_2$. The *A. salmonicida* strains R8-68 and R8-70 should produce NeuAc, $NeuAc_2$, and LegAmAc. It is expected that *A. salmonicida* MR17-77 and R8-63 produce compounds similar to $LegAc_2$ and Leg5Ac7AcAla, respectively. In the case of *Vibrio* B9-25K2, it might produce $PseAc_2$ unless the unknown protein domains/sequences within its NAB cluster are responsible for major modifications.

The hypotheses concerning NulO structure were investigated further by analyzing the NulO content of these bacteria (labelled with DMB) using mass spectrometry. *A. salmonicida* LFI1238, *M. viscosa* and *A. wodanis* were also included both as positive controls and in order to verify the presence of Pse and Leg5Ac7AcAla in the last two. Compounds with masses corresponding to the expected NulOs from *A. salmonicida*, *M. viscosa* and *A. wodanis* were found, thus confirming their presence. Tandem mass spectrometry was performed for the compounds corresponding to LegAmAc and $PseAc_2$, as it had not been done previously. *Vibrio* B9-25K2 seems to produce a compound that, while corresponding to the mass of $PseAc_2$, is different from that produced by *M. viscosa*, the respective compounds having different retention times and fragmentation spectra. *A. salmonicida* R8-63 produces a NulO of higher mass than Leg5Ac7AcAla which has not been previously reported. *A. salmonicida* MR17-77 could not be grown at the time of the experiment and was therefore not analyzed.

4.2 Determinants for NeuB substrate specificity

The three pathways for the biosynthesis of NulOs are distinguished by different *nab* gene clusters coding for enzymes specific to each of them. The ones that carry the same activity are however often homologs, the most conserved being the *N*-acetylneuraminic acid synthase NeuB, which catalyzes the condensation reaction between the hexosamine precursor of each pathway and phosphoenolpyruvate (PEP) to give Neu, Leg, or Pse, respectively. In order to investigate the residues involved in substrate specificity, we compared in paper I the structure of the NeuB active site from *N. meningitidis* (nmNeuB) to the sequences of the homologs from *C. jejuni* NCTC11168 (cjNeuB1-3), the only species known to produce all three types of NulOs [103, 256]. Its NeuB homologs, while from different pathways, are from the same organism and thus share a higher sequence identity than sequences from different pathways and species.

Two regions within the active site could be described, a strictly conserved "catalytic region" comprising of the PEP binding site as well as the residues binding the substituents at the C1 position of the sugar substrate and a "specificity region" binding the rest of the hexose, which varied between isozymes (see Figure 2, paper I). In nmNeuB, this region is composed of residues Q55, N74, Y186, H236, D247 and R314 which directly interact with the reduced ManNAc (rManNAc) ligand. R314 interacts with the NAc group at C2, Q55 and D247 with the -OH at C3, the latter also interacting with the -OH at C5. N74 is in contact with the -OH groups at C5 and C6. Three water molecules are also involved; they are coordinated by N74 and R314. Comparison with the *C. jejuni* NeuB model from the same pathway (cjNeuB1) showed that all those positions were conserved between the two enzymes (Figure 2A, paper I). Within 4Å of the ligand (near PEP and the Mn^{2+}), only C182 is replaced, by a threonine residue.

In the model corresponding to the Leg pathway homolog (cjNeuB2, see Figure 2B, paper I) the region in proximity of C6 has a different geometry, with the positions corresponding to N74, I79 and M83 occupied by other residues (Tyr, Gln and Val, respectively). The loop carrying N74 has a completely different sequence and geometry in cjNeuB2, none of its residues appearing to be involved with substrate binding. This reflects the difference between the substrates of the two enzymes, the Leg precursor 2,4-diNAc-2,4,6-trideoxy-mannose lacking the -OH group present at C6 of ManNAc. The only other difference between the Neu and Leg precursors is the C4 position, occupied by -OH in ManNAc and a NAc group in the latter (but sharing the same configuration). The sites are almost strictly conserved at this location, except for M83 (replaced by Val). In nmNeuB, the OH of rManNAc only interacts with a water molecule, which position could be taken by the NAc group of the Leg precursor (the M83V mutation providing extra space). The Q55 and D247 positions are conserved, as expected considering that the C3 position is identical for all NeuB substrates.

The NeuB homolog from the Pse pathway (cjNeuB3) accepts 2,4-diNAc-2,4,6-trideoxy-altrose as a substrate, which differs from ManNAc in several more aspects than the Leg precursor. While both lack an OH group at C6, the configuration of their chiral centers at C4 and C5 are different, and it is reasonable to assume that the active sites of their respective NeuB homologs would be different at this location (as well as being different from cjNeuB1). Indeed, the M83 position of nmNeuB, conserved in cjNeuB1 and Val in cjNeuB2, is occupied by a phenylalanine in cjNeuB3. Similarly, The I79 position is occupied by a leucine residue (Ile in cjNeuB1 and Gln in cjNeuB2). The third position of interest, occupied by N74 in nmNeuB, was carried by a loop which was different in cjNeuB2 (but conserved in geometry, if not sequence, between nmNeuB and cjNeuB1). This is also the case for cjNeuB3, for which the loop is different in both sequence and geometry. The second difference between the Leg and Pse precursors is their configuration at C2, with the NAc group in an orientation similar to that of rManNAc for the Leg precursor but different for the latter. This difference is reflected within the NeuB active sites, with the G133, N184, T285, and A289 positions of nmNeuB replaced by Phe, Ser, Leu and Val residues in cjNeuB3.

Those positions are strictly conserved between nmNeuB and cjNeuB1, while they are occupied by a Gly, Thr, Asp and Val in cjNeuB2. Again, the residues involved with specificity at C3 are conserved.

Our goal was to explore NulO biosynthesis in *A. salmonicida*, *A. wodanis* and *M. viscosa*, and thus we constructed a NeuB alignment with sequences from those organisms (as well as the *Vibrio* B9-25K2 strain, because it had a unique NeuB sequence), presented in Figure 1 from paper I. Its analysis, also considering the pathway specific alignments previously published, confirmed that the positions highlighted by the active site comparison were indeed of interest. Most notably, the G133 position of nmNeuB, conserved in cjNeuB1 and cjNeuB2 but replaced by Phe in cjNeuB3, was also conserved according to pathway for our set of *Vibrionaceae* NeuB sequences. Sequences identified as Pse-pathway homologs have Phe while the other ones have Gly, suggesting that this position may be critical for the specificity at the C2 position of NulO substrates. Specificity at C4-C6 involves the same set of residues and is dependent on both the nature of the substituents in C4 and C6 and the configuration C4 and C5 chiral centers. Leg and Pse substrates carry an NAc group in C4 while Neu substrates carry an OH group, but may have an OAc if NeuD containing organisms use 4-*O*-acetylated-ManNAc as a substrate [36]. This hypothesis is strengthened by the local geometry of the nmNeuB active site, with the water molecule as interaction partner for the OH at C4. As for Leg compounds, it is reasonable to imagine this water replaced by a substituent at C4. Considering the configuration at C4 and C5, we know that the Neu and Leg substrates have the same absolute configuration, but not the Pse precursor. It would thus make sense that the NeuB residues interacting with those positions are more conserved across the Neu and Leg pathways than they are with the Pse pathway. Taken together, these observations make it difficult to assign specific positions to each pathway with our limited set of sequences. They however explain the sequence variations within the concerned region and highlight its importance.

4.3 Sequencing of *Nostoc* sp. KVJ20

The study of NulOs requires that they are made readily available, which is not currently the case. Our group is pursuing the possibility of Neu synthesis via the AGE/NAL coupled reaction, using the NAL from *A. salmonicida* [212]. Considering the promising qualities of cyanobacterial AGEs, we screened genetic data from environmental samples for their presence. The data was produced by the Microorganisms and Plants research group at the Department of Arctic and Marine Biology at UiT. The group has been studying the role of secondary metabolites in *Nostoc* spp., with an emphasis on regulation of cell differentiation and cell-cell communication [257-259]. As a part of their work, they have collected isolates from the liverwort *Blasia pusilla* L. which occurs naturally in the region surrounding Tromsø and has a symbiotic relationship with nitrogen-fixing cyanobacteria. They have classified these isolates into 25 genotypes, showing the high genetic diversity amongst the symbionts of *B. pusilla* L. [259].

PCR screening for AGE coding sequences was performed and two target strains, *Nostoc* sp. KVJ10 and KVJ20 were thus obtained. Both were used for sequencing, but only the data pertaining to KVJ20 was of sufficient quality. The draft genome sequence of *Nostoc* sp. KVJ20 is reported in paper II.

The genome totals 9,183,495 bp distributed amongst 425 contigs forming 332 scaffolds, with 41.69% GC. 7,676 genes were identified, with 7,210 protein coding sequences and 104 RNA genes. The genome was found to be closest to that of *Nostoc punctiforme* PCC73102 [260]. The *Nostoc* genus is a focus of the Microorganisms and Plants research group at UiT- The Arctic University of Norway, and this particular strain presents an interest for them because of its "self-cleaning" ability. As such, paper II focuses on the genes involved in the synthesis of secondary metabolites likely to confer antibiotic properties to the strain rather than the one AGE coding sequence the genome contains. Nevertheless, the sequence was cloned and the AGE, referred to as nAGE20, was expressed and purified (data not shown). It was capable of Neu synthesis *in vitro* but crystallization attempts were unsuccessful, and those results were not included in any paper.

4.4 Characterization of nAGE10 from *Nostoc* sp. KVJ10

Paper III reports the crystal structure of AGE from *Nostoc* sp. KVJ10 (nAGE10), another target cyanobacterial strain from bioprospecting. The crystals were obtained using the hanging-drop vapor diffusion method and diffracted to a resolution of 1.7Å. The structure was solved by molecular replacement (R=0.176) using the AGE structure from *Anabaena variabilis* (AnaAGE, PDB ID: 2GZ6) as template, the two proteins sharing 90% sequence identity [214]. As expected, the tertiary structure of nAGE10 is very similar to that of AnaAGE, with a $(\alpha/\alpha)_6$-barrel fold where the helices are joined by long loops on one face (front) and short ones on the other (back). The r.m.s.d. between the two structures is 0.8, but the crystals belong to different space groups ($P4_22_12$ for nAGE10 and $P2_12_12_1$ for AnaAGE). The unit cell of nAGE10 contains a monomer versus a dimer for AnaAGE, which could not be generated by symmetry for the nAGE10 crystal. Another favorable dimeric organization was found instead, involving the mate issued from the *y-1, x+1, -z* operation. The difference between the two dimers is that they involve different faces of the monomer: the nAGE10 dimer is "front-to-front" while the one presented in the AnaAGE structure is "back-to-back". The front-to-front dimer could be generated for AnaAGE (symmetry operator x-1, y, z) and could have been chosen as unit cell instead. The fact the front-to-front organization can be found regardless of the space group of the crystal whereas the back-to-back one is only found for one of them indicates that the former might be biological unit. This hypothesis was strengthened by the comparison of the interfaces for between assemblies for AnaAGE as well as the structure of the porcine enzyme (pAGE) it is based on [213]. Interface area, solvation energy and interaction numbers all suggest that the front-to-front organization is more favorable than the back-to-

back one for both enzymes. Analysis of the conservation of interface residues in sequence and distribution for all three AGEs revealed that the front interface is conserved across the species while the back interface is not, further supporting our hypothesis.

Aside from the considerations about the dimer organization of AGEs, the nAGE10 structure provided new insights into the structure of the AGE binding site. We compared it with that of the other AGE structures (AnaAGE and pAGE) as well as the structures of the other members of the AGE superfamily, which share a common fold despite low sequence identity [213, 214, 235, 261]. The active site of nAGE10 contains a chloride ion which is also present in that of the cellobiose epimerase from *Rhodothermus marinus* (RmCE; PDB entries 3WKG-I). The "chloride location" is occupied by waters in the other structures, but they either have a B-factor that is low compared to that of neighboring waters (AnaAGE and pAGE), or the coordinating glycine (G370 of nAGE10) that is conserved in AGEs and RmCE is replaced by another residue (cysteine or aspartate). Taken together, these observations suggest that AGEs may have a conserved chloride in their active site.

In addition to the chloride, there is an ethylene glycol molecule in the active site of nAGE10 which, when compared to the sites of ligand-bound RmCE, is roughly occupying the positions taken by the end of the mannose chain. Because of this, we can assume that the RmCE substrates mimic the position of AGE substrates in acceptable fashion, thus making it possible to discuss the position of active sites residues relative to them. For example, measuring the distance between the putative AGE catalytic residues and the "substrate" revealed that while the arginine position (R68 in nAGE10) is close to the epimerization target, the glutamic acid position (E309) is almost 8Å away from it [216]. A closer candidate for the glutamic acid partner would be E243, which is only about 5Å from the C2 position of the hexose. Aside from catalytic residues, it seems that the AGE binding pocket for the acetyl group carried by GlcNAc/ManNAc is formed by F116, I177 and A173 of nAGE10. Those conserved AGE positions are occupied by bulkier residues in the enzyme from *Pedobacter heparinus*, which is not capable of epimerizing GlcNAc [216]. At last, W368 might be involved in substrate specificity at the C4 position of the sugar substrate. This is especially interesting when considering that this position corresponds to the C7 of NulOs, which commonly carries substituents.

While nAGE10 could not be crystallized in the presence of ATP and the putative ATP-binding loop (between the H5 and H6 helices) could not be resolved, the new dimer organization implies that ATP binds at the dimer interface. This drastically changes the way we look at AGE regulation by nucleotides, which may involve the oligomeric state of the enzyme. In addition to the H5/H6 loop, the glycine-rich fragment located at the C-terminus of AGEs, in the H11/H12 loop has also been proposed as putative ATP-binding site [231]. This loop is also located at the dimer interface, in a position opposite to that of the H5/H6 loop in pAGE, with both loops pointing towards each other. Considering that their orientation might be similar in other AGEs makes it possible that both sites are involved.

Apart from the information provided by the structural characterization, it is worth noting that nAGE10, which was expressed as a His-tagged protein without chaperones, did not form inclusion bodies. It could be purified with high yields in one step by affinity chromatography. It is also active, and Neu5Ac synthesis was performed by the AGE/NAL coupled reaction using the NAL from *A. salmonicida*. The reaction could be performed both in the absence and presence of ATP, although only a fraction of Neu5Ac could be synthetized without.

5 Future perspectives

Considering that each paper treats a different subject, they cannot be easily discussed in relation to one another. They are, however, all within the spectrum of NulO studies, dealing with its genetic, metabolic and technical aspects. Papers II and III deal with the necessary preparation steps for NulO studies, from the bioprospecting and characterization angles. Paper I adds to the current knowledge about NulO diversity as well as investigates the structural details governing NeuB specificity. Together, these papers illustrate the range of topics that need to be taken into consideration when exploring somewhat uncharted scientific territory.

The analysis of *nab* clusters within the chosen set of sequences revealed the unique clusters from *Vibrio* B9-25K2 as well as *A. salmonicida* R8-63 and MR17-77. The compounds from *Vibrio* B9-25K2 and *A. salmonicida* R8-63 most likely correspond to previously undescribed derivates of Pse and Leg, respectively. They await structural determination, which should provide valuable information on the function of the uncharacterized genes from their respective clusters. Whether *A. salmonicida* MR17-77 produces $LegAc_2$ remains to be revealed.

The annotation of the clusters also provided interesting targets for further studies, such as the putative cytidylyltransferase from *Vibrio* B9-25K2 that is non-homologous to NeuA and has a second domain of unknown function. Another is the *N*-acyltransferase from *A. salmonicida* R8-63. It is similar to the putative alanyl transferase from *V. vulnificus* but may well transfer another residue, considering the difference in m/z ratios of the final NulOs [73]. These two and the other uncharacterized proteins from these clusters would make good targets for characterization studies and ones involving deletion mutants.

Concerning the NeuB homologs, a whole set of mutagenesis targets were given. These will allow the investigation of the requirements for NeuB substrate specificity, which has not been explored before. In particular, the positions corresponding to the N74, I79, M83 and G133 residues of nmNeuB seem of importance. Other possible experiments include sequence comparison studies using large datasets in order to spot further pathway-specific trends and gather evolutionary data.

Lastly, the information gained by the crystallization of nAGE10 gives new leads of investigation into the structural relationship between AGEs and nucleotides, and their role in substrate binding. New data concerning the structure of the AGE active site, with the identification of putative buried chloride and NAc binding pocket, also provides new targets for mutagenesis. Those are the positions corresponding to the F116, I177, A173, W368 and G370 residues of nAGE10.

The works presented provide further insight into the field of NulO study, opening promising possibilities to strengthen the current knowledge within this particular branch of glycobiology.

Works cited

1. **Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W. and Etzler, M.E.** (2009) *Essentials of Glycobiology*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY).
2. **Whitfield, C., Szymanski, C.M. and Aebi, M.** (2017) Eubacteria. In: Varki, A., Cummings, R.D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Darvill, A.G., Taroh, K., Packer, N. H., Prestegard, J. H., Schnaar, R. L., Seeberger, P. H. (eds) *Essentials of Glycobiology*, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY).
3. **Tripathi, P., Beaussart, A., Andre, G., Rolain, T., Lebeer, S., Vanderleyden, J., Hols, P. and Dufrêne, Y.F.** (2012) Towards a nanoscale view of lactic acid bacteria. *Micron*. **43** (12): 1323-1330.
4. **Sohlenkamp, C. and Geiger, O.** (2015) Bacterial membrane lipids: diversity in structures and pathways. *FEMS Microbiology Reviews*. **40** (1): 133-159.
5. **Dramsi, S., Magnet, S., Davison, S. and Arthur, M.** (2008) Covalent attachment of proteins to peptidoglycan. *FEMS Microbiology Reviews*. **32** (2): 307-320.
6. **Sáenz, J.P., Grosser, D., Bradley, A.S., Lagny, T.J., Lavrynenko, O., Broda, M. and Simons, K.** (2015) Hopanoids as functional analogues of cholesterol in bacterial membranes. *Proceedings of the National Academy of Sciences*. **112** (38): 11971.
7. **Hopf, P.S., Ford, R.S., Zebian, N., Merkx-Jacques, A., Vijayakumar, S., Ratnayake, D., Hayworth, J. and Creuzenet, C.** (2011) Protein Glycosylation in *Helicobacter pylori*: Beyond the Flagellins? *PloS one*. **6** (9): e25722.
8. **Litzinger, S. and Mayer, C.** (2010) The Murein Sacculus. In: König H., Claus H., Varma A. (eds) *Prokaryotic Cell Wall Compounds*. Springer, Berlin, Heidelberg
9. **Coico, R.** (2006) Gram Staining. *Current Protocols in Microbiology*. **00** (1): A.3C.1-A.3C.2.
10. **Vollmer, W., Blanot, D. and De Pedro, M.A.** (2008) Peptidoglycan structure and architecture. *FEMS Microbiology Reviews*. **32** (2): 149-167.
11. **Sleytr, U.B., Schuster, B., Egelseer, E.-M. and Pum, D.** (2014) S-layers: principles and applications. *FEMS Microbiology Reviews*. **38** (5): 823-864.
12. **Altegoer, F., Schuhmacher, J., Pausch, P. and Bange, G.** (2014) From molecular evolution to biobricks and synthetic modules: a lesson by the bacterial flagellum. *Biotechnology and Genetic Engineering Reviews*. **30** (1): 49-64.
13. **Merino, S. and Tomás, J.M.** (2014) Gram-negative flagella glycosylation. *International journal of molecular sciences*. **15** (2): 2840-2857.
14. **Proft, T. and Baker, E.N.** (2008) Pili in Gram-negative and Gram-positive bacteria — structure, assembly and their role in disease. *Cellular and Molecular Life Sciences*. **66** (4): 613.
15. **McNaught, A.D.** (1996) Nomenclature of carbohydrates (IUPAC Recommendations 1996). *Pure and Applied Chemistry*. **68** (10): 1919-2008.
16. **Klepach, T., Carmichael, I. and Serianni, A.S.** (2008) 13C-Labeled *N*-Acetyl-neuraminic Acid in Aqueous Solution: Detection and Quantification of Acyclic Keto, Keto Hydrate, and Enol Forms by 13C NMR Spectroscopy. *Journal of the American Chemical Society*. **130** (36): 11892-11900.
17. **Knirel, Y.A., Vinogradov, E.V., Shashkov, A.S., Dmitriev, B.A., Kochetkov, N.K., Stanislavsky, E.S. and Mashilova, G.M.** (1987) Somatic antigens of *Pseudomonas aeruginosa*. The structure of the O-specific polysaccharide chain of the lipopolysaccharide from *P. aeruginosa* O13 (Lanyi). *European Journal of Biochemistry*. **163** (3): 627-637.
18. **Knirel, Y.A., Moll, H., Helbig, J.H. and Zahringer, U.** (1997) Chemical characterization of a new 5,7-diamino-3,5,7,9-tetradeoxynonulosonic acid released by mild acid hydrolysis of the *Legionella pneumophila* serogroup 1 lipopolysaccharide. *Carbohydrate Research*. **304** (1): 77-79.
19. **Tsvetkov, Y.E., Shashkov, A.S., Knirel, Y.A. and Zahringer, U.** (2001) Synthesis and identification in bacterial lipopolysaccharides of 5,7-diacetamido-3,5,7,9-tetradeoxy-D-*glycero*-D-*galacto*- and -D-*glycero*-D-*talo*-non-2-ulosonic acids. *Carbohydrate Research*. **331** (3): 233-237.
20. **Kenyon, J.J., Marzaioli, A.M., De Castro, C. and Hall, R.M.** (2015) 5,7-di-*N*-acetyl-acinetaminic acid: A novel non-2-ulosonic acid found in the capsule of an *Acinetobacter baumannii* isolate. *Glycobiology*. **25** (6): 644-654.

21. **Kenyon, J.J., Notaro, A., Hsu, L.Y., De Castro, C. and Hall, R.M.** (2017) 5,7-Di-*N*-acetyl-8-epiacinetaminic acid: A new non-2-ulosonic acid found in the K73 capsule produced by an *Acinetobacter baumannii* isolate from Singapore. *Scientific Reports*. **7** (1): 11357.
22. **Vinogradov, E., St Michael, F. and Cox, A.D.** (2018) Structure of the LPS O-chain from Fusobacterium nucleatum strain ATCC 23726 containing a novel 5,7-diamino-3,5,7,9-tetradeoxy-L-*gluco*-non-2-ulosonic acid presumably having the D-*glycero*-L-*gluco* configuration. *Carbohydrate Research*. **468**: 69-72.
23. **Lewis, A.L., Desa, N., Hansen, E.E., Knirel, Y.A., Gordon, J.I., Gagneux, P., Nizet, V. and Varki, A.** (2009) Innovations in host and microbial sialic acid biosynthesis revealed by phylogenomic prediction of nonulosonic acid structure. *Proceedings of the National Academy of Sciences of the United States of America*. **106** (32): 13552-13557.
24. **McGuire, E.J. and Binkley, S.B.** (1964) The Structure and Chemistry of Colominic Acid. *Biochemistry*. **3**: 247-251.
25. **Egan, W., Liu, T.Y., Dorow, D., Cohen, J.S., Robbins, J.D., Gotschlich, E.C. and Robbins, J.B.** (1977) Structural studies on the sialic acid polysaccharide antigen of *Escherichia coli* strain Bos-12. *Biochemistry*. **16** (16): 3687-3692.
26. **Dutton, G.G., Parolis, H. and Parolis, L.A.** (1987) The structure of the neuraminic acid-containing capsular polysaccharide of *Escherichia coli* serotype K9. *Carbohydrate Research*. **170** (2): 193-206.
27. **Bhattacharjee, A.K., Jennings, H.J., Kenny, C.P., Martin, A. and Smith, I.C.** (1975) Structural determination of the sialic acid polysaccharide antigens of *Neisseria meningitidis* serogroups B and C with carbon 13 nuclear magnetic resonance. *Journal of Biological Chemistry*. **250** (5): 1926-1932.
28. **Bhattacharjee, A.K. and Jennings, H.J.** (1976) Determination of the linkages in some methylated, sialic acid-containing, meningococcal polysaccharides by mass spectrometry. *Carbohydrate Research*. **51** (2): 253-261.
29. **Gamian, A., Romanowska, E., Dabrowski, U. and Dabrowski, J.** (1991) Structure of the O-specific, sialic acid containing polysaccharide chain and its linkage to the core region in lipopolysaccharide from *Hafnia alvei* strain 2 as elucidated by chemical methods, gas-liquid chromatography/mass spectrometry, and 1H NMR spectroscopy. *Biochemistry*. **30** (20): 5032-5038.
30. **Moran, A.P., Rietschel, E.T., Kosunen, T.U. and Zahringer, U.** (1991) Chemical characterization of *Campylobacter jejuni* lipopolysaccharides containing *N*-acetylneuraminic acid and 2,3-diamino-2,3-dideoxy-D-glucose. *Journal of Bacteriology*. **173** (2): 618-626.
31. **Aspinall, G.O., McDonald, A.G., Raju, T.S., Pang, H., Moran, A.P. and Penner, J.L.** (1993) Chemical structures of the core regions of *Campylobacter jejuni* serotypes O:1, O:4, O:23, and O:36 lipopolysaccharides. *European Journal of Biochemistry*. **213** (3): 1017-1027.
32. **Aspinall, G.O., McDonald, A.G., Pang, H., Kurjanczyk, L.A. and Penner, J.L.** (1994) Lipopolysaccharides of *Campylobacter jejuni* serotype O:19: structures of core oligosaccharide regions from the serostrain and two bacterial isolates from patients with the Guillain-Barre syndrome. *Biochemistry*. **33** (1): 241-249.
33. **Shin, J.E., Ackloo, S., Mainkar, A.S., Monteiro, M.A., Pang, H., Penner, J.L. and Aspinall, G.O.** (1997) Lipo-oligosaccharides of *Campylobacter jejuni* serotype O:10. Structures of core oligosaccharide regions from a bacterial isolate from a patient with the Miller-Fisher syndrome and from the serotype reference strain. *Carbohydrate Research*. **305** (2): 223-232.
34. **McNally, D.J., Hui, J.P., Aubry, A.J., Mui, K.K., Guerry, P., Brisson, J.R., Logan, S.M. and Soo, E.C.** (2006) Functional characterization of the flagellar glycosylation locus in *Campylobacter jejuni* 81-176 using a focused metabolomics approach. *Journal of Biological Chemistry*. **281** (27): 18489-18498.
35. **Logan, S.M., Hui, J.P., Vinogradov, E., Aubry, A.J., Melanson, J.E., Kelly, J.F., Nothaft, H. and Soo, E.C.** (2009) Identification of novel carbohydrate modifications on *Campylobacter jejuni* 11168 flagellin using metabolomics-based approaches. *FEBS Journal*. **276** (4): 1014-1023.
36. **Gurung, M.K., Raeder, I.L., Altermark, B. and Smalas, A.O.** (2013) Characterization of the sialic acid synthase from *Aliivibrio salmonicida* suggests a novel pathway for bacterial synthesis of 7-*O*-acetylated sialic acids. *Glycobiology*. **23** (7): 806-819.

37. **Berg, T.O., Gurung, M.K., Altermark, B., Smalas, A.O. and Raeder, I.L.** (2015) Characterization of the *N*-acetylneuraminic acid synthase (NeuB) from the psychrophilic fish pathogen *Moritella viscosa*. *Carbohydrate Research*. **402**: 133-145.
38. **Defives, C., Bouslamti, R., Derieux, J.C., Kol, O. and Fournet, B.** (1989) Characterization of sialic acids containing lipopolysaccharide from *Rhizobium meliloti* M 11 S. *FEMS Microbiology Letters*. **57** (2): 203-207.
39. **Vinogradov, E., St Michael, F., Homma, K., Sharma, A. and Cox, A.D.** (2017) Structure of the LPS O-chain from *Fusobacterium nucleatum* strain 10953, containing sialic acid. *Carbohydrate Research*. **440-441**: 38-42.
40. **Gamian, A., Jones, C., Lipinski, T., Korzeniowska-Kowal, A. and Ravenscroft, N.** (2000) Structure of the sialic acid-containing O-specific polysaccharide from Salmonella enterica serovar Toucra O48 lipopolysaccharide. *European Journal of Biochemistry*. **267** (11): 3160-3167.
41. **Gamian, A., Romanowska, E., Ulrich, J. and Defaye, J.** (1992) The structure of the sialic acid-containing Escherichia coli O104 O-specific polysaccharide and its linkage to the core region in lipopolysaccharide. *Carbohydrate Research*. **236**: 195-208.
42. **Staaf, M., Weintraub, A. and Widmalm, G.** (1999) Structure determination of the O-antigenic polysaccharide from the enteroinvasive *Escherichia coli* O136. *European Journal of Biochemistry*. **263** (3): 656-661.
43. **Thibault, P., Logan, S.M., Kelly, J.F., Brisson, J.R., Ewing, C.P., Trust, T.J. and Guerry, P.** (2001) Identification of the carbohydrate moieties and glycosylation motifs in Campylobacter jejuni flagellin. *Journal of Biological Chemistry*. **276** (37): 34862-34870.
44. **Schirm, M., Schoenhofen, I.C., Logan, S.M., Waldron, K.C. and Thibault, P.** (2005) Identification of unusual bacterial glycosylation by tandem mass spectrometry analyses of intact proteins. *Analytical Chemistry*. **77** (23): 7774-7782.
45. **McNally, D.J., Aubry, A.J., Hui, J.P., Khieu, N.H., Whitfield, D., Ewing, C.P., Guerry, P., Brisson, J.R., Logan, S.M. and Soo, E.C.** (2007) Targeted metabolomics analysis of *Campylobacter coli* VC167 reveals legionaminic acid derivatives as novel flagellar glycans. *Journal of Biological Chemistry*. **282** (19): 14463-14475.
46. **Kondakova, A.N., Perepelov, A.V., Bartodziejska, B., Shashkov, A.S., Senchenkova, S.N., Wykrota, M., Knirel, Y.A. and Rozalski, A.** (2001) Structure of the acidic O-specific polysaccharide from *Proteus vulgaris* O39 containing 5,7-diacetamido-3,5,7,9-tetradeoxy-L-*glycero*-L-*manno*-non-2-ulosonic acid. *Carbohydrate Research*. **333** (3): 241-249.
47. **Schirm, M., Soo, E.C., Aubry, A.J., Austin, J., Thibault, P. and Logan, S.M.** (2003) Structural, genetic and functional characterization of the flagellin glycosylation process in *Helicobacter pylori*. *Molecular Microbiology*. **48** (6): 1579-1592.
48. **Perepelov, A.V., Shashkov, A.S., Torgov, V.I., Nazarenko, E.L., Gorshkova, R.P., Ivanova, E.P., Gorshkova, N.M. and Widmalm, G.** (2005) Structure of an acidic polysaccharide from the agar-decomposing marine bacterium *Pseudoalteromonas atlantica* strain IAM 14165 containing 5,7-diacetamido-3,5,7,9-tetradeoxy-L-*glycero*-L-*manno*-non-2-ulosonic acid. *Carbohydrate Research*. **340** (1): 69-74.
49. **Le Quere, A.J., Deakin, W.J., Schmeisser, C., Carlson, R.W., Streit, W.R., Broughton, W.J. and Forsberg, L.S.** (2006) Structural characterization of a K-antigen capsular polysaccharide essential for normal symbiotic infection in *Rhizobium* sp. NGR234: deletion of the rkpMNO locus prevents synthesis of 5,7-diacetamido-3,5,7,9-tetradeoxy-non-2-ulosonic acid. *Journal of Biological Chemistry*. **281** (39): 28981-28992.
50. **Perepelov, A.V., Shashkov, A.S., Tomshich, S.V., Komandrova, N.A. and Nedashkovskaya, O.I.** (2007) A pseudoaminic acid-containing O-specific polysaccharide from a marine bacterium *Cellulophaga fucicola*. *Carbohydrate Research*. **342** (10): 1378-1381.
51. **Vinogradov, E., Frimmelova, M. and Toman, R.** (2013) Chemical structure of the carbohydrate backbone of the lipopolysaccharide from *Piscirickettsia salmonis*. *Carbohydrate Research*. **378**: 108-113.
52. **Kenyon, J.J., Marzaioli, A.M., Hall, R.M. and De Castro, C.** (2014) Structure of the K2 capsule associated with the KL2 gene cluster of *Acinetobacter baumannii*. *Glycobiology*. **24** (6): 554-563.

53. **Senchenkova, S.N., Shashkov, A.S., Shneider, M.M., Arbatsky, N.P., Popova, A.V., Miroshnikov, K.A., Volozhantsev, N.V. and Knirel, Y.A.** (2014) Structure of the capsular polysaccharide of *Acinetobacter baumannii* ACICU containing di-*N*-acetylpseudaminic acid. *Carbohydrate Research*. **391**: 89-92.
54. **Knirel, Y.A., Vinogradov, E.V., L'Vov V, L., Kocharova, N.A., Shashkov, A.S., Dmitriev, B.A. and Kochetkov, N.K.** (1984) Sialic acids of a new type from the lipopolysaccharides of *Pseudomonas aeruginosa* and *Shigella boydii*. *Carbohydrate Research*. **133** (2): C5-8.
55. **Knirel, Y.A., Vinogradov, E.V., Shashkov, A.S., Kochetkov, N.K., L'Vov, V. and Dmitriev, B.A.** (1985) Identification of 5-acetamido-3,5,7,9-tetradeoxy-7-[(R)-3-hydroxybutyramido]-L- *glycero*-L-*manno*-nonulosonic acid as a component of bacterial polysaccharides. *Carbohydrate Research*. **141** (2): C1-3.
56. **Gil-Serrano, A.M., Rodriguez-Carvajal, M.A., Tejero-Mateo, P., Espartero, J.L., Menendez, M., Corzo, J., Ruiz-Sainz, J.E. and Buendi, A.C.A.M.** (1999) Structural determination of a 5-acetamido-3,5,7, 9-tetradeoxy-7-(3-hydroxybutyramido)-L-*glycero*-L-*manno*-nonulos onic acid-containing homopolysaccharide isolated from *Sinorhizobium fredii* HH103. *Biochemical Journal*. **342 Pt 3**: 527-535.
57. **Shashkov, A.S., Tul'skaya, E.M., Streshinskaya, G.M., Senchenkova, S.N., Avtukh, A.N. and Evtushenko, L.I.** (2009) New cell wall glycopolymers of the representatives of the genus *Kribbella.* *Carbohydrate Research*. **344** (16): 2255-2262.
58. **Tul'skaya, E.M., Streshinskaya, G.M., Shashkov, A.S., Senchenkova, S.N., Avtukh, A.N., Baryshnikova, L.M. and Evtushenko, L.I.** (2011) Novel teichulosonic acid from cell walls of some representatives of the genus *Kribbella. Carbohydrate Research*. **346** (13): 2045-2051.
59. **Shashkov, A.S., Streshinskaya, G.M., Kozlova, Y.I., Tul'skaya, E.M., Senchenkova, S.N., Arbatskii, N.P., Bueva, O.V. and Evtushenko, L.I.** (2012) Teichulosonic acid, an anionic polymer of a new class from the cell wall of *Actinoplanes utahensis* VKM Ac-674(T). *Biochemistry (Moscow)*. **77** (5): 511-517.
60. **Knirel, Y.A., Shashkov, A.S., Dmitriev, B.A. and Kochetkov, N.K.** (1984) Structural studies of the *Pseudomonas aeruginosa* immunotype 1 antigen, containing the new sugar constituents 2-acetamido-2-deoxy-D-galacturonamide and 2-deoxy-2-formamido-D-galacturonic acid. *Carbohydrate Research*. **133** (2): C12-14.
61. **Knirel, Y.A., Kocharova, N.A., Shashkov, A.S. and Kochetkov, N.K.** (1986) The structure of the *Pseudomonas aeruginosa* immunotype 6 O-antigen: isolation and identification of 5-acetamido-3,5,7,9-tetradeoxy-7-formamido-L-*glycero*-L-*manno*- nonuloson ic acid. *Carbohydrate Research*. **145** (2): C1-4.
62. **Muldoon, J., Shashkov, A.S., Senchenkova, S.N., Tomshich, S.V., Komandrova, N.A., Romanenko, L.A., Knirel, Y.A. and Savage, A.V.** (2001) Structure of an acidic polysaccharide from a marine bacterium *Pseudoalteromonas distincta* KMM 638 containing 5-acetamido-3,5,7,9-tetradeoxy-7-formamido-L-*glycero*-L-*manno*-nonulosonic acid. *Carbohydrate Research*. **330** (2): 231-239.
63. **Knirel, Y.A., Kocharova, N.A., Shashkov, A.S., Dmitriev, B.A., Kochetkov, N.K., Stanislavsky, E.S. and Mashilova, G.M.** (1987) Somatic antigens of *Pseudomonas aeruginosa*. The structure of O-specific polysaccharide chains of the lipopolysaccharides from *P. aeruginosa* O5 (Lanyi) and immunotype 6 (Fisher). *European Journal of Biochemistry*. **163** (3): 639-652.
64. **Kenne, L., Lindberg, B., Schweda, E., Gustafsson, B. and Holme, T.** (1988) Structural studies of the O-antigen from *Vibrio cholerae* O:2. *Carbohydrate Research*. **180** (2): 285-294.
65. **Vinogradov, E., Wilde, C., Anderson, E.M., Nakhamchik, A., Lam, J.S. and Rowe-Magnus, D.A.** (2009) Structure of the lipopolysaccharide core of *Vibrio vulnificus* type strain 27562. *Carbohydrate Research*. **344** (4): 484-490.
66. **Posch, G., Pabst, M., Brecker, L., Altmann, F., Messner, P. and Schäffer, C.** (2011) Characterization and Scope of S-layer Protein O-Glycosylation in *Tannerella forsythia*. **286** (44): 38714-38724.
67. **Kokoulin, M.S., Kalinovsky, A.I., Romanenko, L.A. and Mikhailov, V.V.** (2018) 5-Acetamido-3,5-dideoxy-L-*glycero*-L-*manno*-non-2-ulosonic acid-containing O-polysaccharide from marine bacterium *Pseudomonas glareae* KMM 9500T. *Carbohydrate Research*. **461**: 19-24.

68. **Knirel, Y.A., Helbig, J.H. and Zahringer, U.** (1996) Structure of a decasaccharide isolated by mild acid degradation and dephosphorylation of the lipopolysaccharide of *Pseudomonas fluorescens* strain ATCC 49271. *Carbohydrate Research*. **283**: 129-139.
69. **Haseley, S.R. and Wilkinson, S.G.** (1997) Structural studies of the putative O-specific polysaccharide of *Acinetobacter baumannii* O24 containing 5,7-diamino-3,5,7,9-tetradeoxy-L-*glycero*-D-*galacto*-nonulosonic acid. *European Journal of Biochemistry*. **250** (2): 617-623.
70. **Hashii, N., Isshiki, Y., Iguchi, T., Hisatsune, K. and Kondo, S.** (2003) Structure and serological characterization of 5,7-diamino-3,5,7,9-tetradeoxy-non-2-ulosonic acid isolated from lipopolysaccharides of *Vibrio parahaemolyticus* O2 and O-untypable strain KX-V212. *Carbohydrate Research*. **338** (10): 1055-1062.
71. **Edebrink, P., Jansson, P.E., Bogwald, J. and Hoffman, J.** (1996) Structural studies of the *Vibrio salmonicida* lipopolysaccharide. *Carbohydrate Research*. **287** (2): 225-245.
72. **Bøgwald, J. and Hoffman, J.** (2006) Structural studies of the O-antigenic oligosaccharide from *Vibrio salmonicida* strain C2 isolated from Atlantic cod, *Gadus morhua* L. *Carbohydrate Research*. **341** (11): 1965-1968.
73. **McDonald, N.D., DeMeester, K.E., Lewis, A.L., Grimes, C.L. and Boyd, E.F.** (2018) Structural and functional characterization of a modified legionaminic acid involved in glycosylation of a bacterial lipopolysaccharide. *Journal of Biological Chemistry*.
74. **Li, X., Perepelov, A.V., Wang, Q., Senchenkova, S.N., Liu, B., Shevelev, S.D., Guo, X., Shashkov, A.S., Chen, W., Wang, L. and Knirel, Y.A.** (2010) Structural and genetic characterization of the O-antigen of *Escherichia coli* O161 containing a derivative of a higher acidic diamino sugar, legionaminic acid. *Carbohydrate Research*. **345** (11): 1581-1587.
75. **Twine, S.M., Paul, C.J., Vinogradov, E., McNally, D.J., Brisson, J.R., Mullen, J.A., McMullin, D.R., Jarrell, H.C., Austin, J.W., Kelly, J.F. and Logan, S.M.** (2008) Flagellar glycosylation in Clostridium botulinum. *FEBS Journal*. **275** (17): 4428-4444.
76. **Knirel, Y.A., Rietschel, E.T., Marre, R. and Zahringer, U.** (1994) The structure of the O-specific chain of *Legionella pneumophila* serogroup 1 lipopolysaccharide. *European Journal of Biochemistry*. **221** (1): 239-245.
77. **Nazarenko, E.L., Perepelov, A.V., Shevchenko, L.S., Daeva, E.D., Ivanova, E.P., Shashkov, A.S. and Widmalm, G.** (2011) Structure of the O-Specific polysaccharide from *Shewanella japonica* KMM 3601 containing 5,7-Diacetamido-3,5,7,9-tetradeoxy-D-*glycero*-D-*talo*-non-2-ulosonic acid. *Biochemistry (Moscow)*. **76** (7): 791-796.
78. **Knirel, Y.A., Senchenkova, S.N., Kocharova, N.A., Shashkov, A.S., Helbig, J.H. and Zahringer, U.** (2001) Identification of a homopolymer of 5-acetamidino-7-acetamido-3,5,7,9-tetradeoxy-D-*glycero*-D-*talo*-nonulosonic acid in the lipopolysaccharides of *Legionella pneumophila* Non-1 serogroups. *Biochemistry (Moscow)*. **66** (9): 1035-1041.
79. **Knirel, Y.A., Senchenkova, S.y.N., Shashkov, A.S., Shevelev, S.D., Perepelov, A.V., Bin, L., Feng, L. and Wang, L.** (2009) First Isolation and Identification of a Di-*N*-Acyl Derivative of 5,7-Diamino-3,5,7,9-tetradeoxy-L-*glycero*-D-*galacto*-non-2-ulosonic (8-Epilegionaminic) Acid. *Advanced Science Letters*. **2** (3): 384-387.
80. **Perepelov, A.V., Wang, Q., Liu, B., Senchenkova, S.N., Feng, L., Shashkov, A.S., Wang, L. and Knirel, Y.A.** (2009) Structure of the O-polysaccharide of *Escherichia coli* O61, Another *E. coli* O-antigen that contains 5,7-Diacetamido-3,5,7,9-tetradeoxy-l-*glycero*-D-*galacto*- non-2-ulosonic (Di-*N*-acetyl-8-epilegionaminic) acid. *Journal of Carbohydrate Chemistry*. **28** (7-8): 463-472.
81. **Perepelov, A.V., Liu, B., Senchenkova, S.N., Shashkov, A.S., Shevelev, S.D., Feng, L., Wang, L. and Knirel, Y.A.** (2010) Structure of the O-antigen and characterization of the O-antigen gene cluster of *Escherichia coli* O108 containing 5,7-diacetamido-3,5,7,9-tetradeoxy-L-*glycero*-D-*galacto*-non-2-ulosonic (8-epilegionaminic) acid. *Biochemistry (Moscow)*. **75** (1): 19-24.
82. **Shashkov, A.S., Kocharova, N.A., Zatonsky, G.V., Blaszczyk, A., Knirel, Y.A. and Rozalski, A.** (2007) Structure of the O-antigen of *Providencia stuartii* O20, a new polysaccharide containing 5,7-diacetamido-3,5,7,9-tetradeoxy-l-*glycero*-d-*galacto*-non-2-ulosonic acid. *Carbohydrate Research*. **342** (3-4): 653-658.
83. **Vinogradov, E., Maclean, L., Xu, H.H. and Chen, W.** (2014) The structure of the polysaccharide isolated from *Acinetobacter baumannii* strain LAC-4. *Carbohydrate Research*. **390**: 42-45.

84. **Kilcoyne, M., Shashkov, A.S., Senchenkova, S.A., Knirel, Y.A., Vinogradov, E.V., Radziejewska-Lebrecht, J., Galimska-Stypa, R. and Savage, A.V.** (2002) Structural investigation of the O-specific polysaccharides of *Morganella morganii* consisting of two higher sugars. *Carbohydrate Research*. **337** (18): 1697-1702.
85. **Shashkov, A.S., Torgov, V.I., Nazarenko, E.L., Zubkov, V.A., Gorshkova, N.M., Gorshkova, R.P. and Widmalm, G.** (2002) Structure of the phenol-soluble polysaccharide from *Shewanella putrefaciens* strain A6. *Carbohydrate Research*. **337** (12): 1119-1127.
86. **Vinogradov, E.V., Shashkov, A.S., Knirel, Y.A., Kochetkov, N.K., Dabrowski, J., Grosskurth, H., Stanislavsky, E.S. and Kholodkova, E.V.** (1992) The structure of the O-specific polysaccharide chain of the lipopolysaccharide of *Salmonella arizonae* O61. *Carbohydrate Research*. **231**: 1-11.
87. **Beynon, L.M., Richards, J.C. and Perry, M.B.** (1994) The structure of the lipopolysaccharide O antigen from *Yersinia ruckeri* serotype 01. *Carbohydrate Research*. **256** (2): 303-317.
88. **Friedrich, V., Janesch, B., Windwarder, M., Maresch, D., Braun, M.L., Megson, Z.A., Vinogradov, E., Goneau, M.-F., Sharma, A., Altmann, F., Messner, P., Schoenhofen, I.C. and Schäffer, C.** (2017) *Tannerella forsythia* strains display different cell-surface nonulosonic acids: biosynthetic pathway characterization and first insight into biological implications. *Glycobiology*. **27** (4): 342-357.
89. **Vimr, E.R., Aaronson, W. and Silver, R.P.** (1989) Genetic analysis of chromosomal mutations in the polysialic acid gene cluster of *Escherichia coli* K1. *Journal of Bacteriology*. **171** (2): 1106-1117.
90. **Luneberg, E., Zetzmann, N., Alber, D., Knirel, Y.A., Kooistra, O., Zahringer, U. and Frosch, M.** (2000) Cloning and functional characterization of a 30 kb gene locus required for lipopolysaccharide biosynthesis in *Legionella pneumophila. International Journal of Medical Microbiology*. **290** (1): 37-49.
91. **Chou, W.K., Dick, S., Wakarchuk, W.W. and Tanner, M.E.** (2005) Identification and characterization of NeuB3 from *Campylobacter jejuni* as a pseudaminic acid synthase. *Journal of Biological Chemistry*. **280** (43): 35922-35928.
92. **Edwards, U., Müller, A., Hammerschmidt, S., Gerardy-Schahn, R. and Frosch, M.** (1994) Molecular analysis of the biosynthesis pathway of the α-2,8 polysialic acid capsule by *Neisseria meningitidis* serogroup B. *Molecular Microbiology*. **14** (1): 141-149.
93. **Vimr, E.R., Steenbergen, S.M. and Cieslewicz, M.J.** (1995) Biosynthesis of the polysialic acid capsule in *Escherichia coli* K1. *Journal of Industrial Microbiology*. **15** (4): 352-360.
94. **Vann, W.F., Tavarez, J.J., Crowley, J., Vimr, E. and Silver, R.P.** (1997) Purification and characterization of the *Escherichia coli* Kl *neuB* gene product *N*-acetylneuraminic acid synthetase. *Glycobiology*. **7** (5): 697-701.
95. **Vann, W.F., Daines, D.A., Murkin, A.S., Tanner, M.E., Chaffin, D.O., Rubens, C.E., Vionnet, J. and Silver, R.P.** (2004) The NeuC protein of *Escherichia coli* K1 is a UDP *N*-acetylglucosamine 2-epimerase. *Journal of Bacteriology*. **186** (3): 706-712.
96. **Vann, W.F., Silver, R.P., Abeijon, C., Chang, K., Aaronson, W., Sutton, A., Finn, C.W., Lindner, W. and Kotsatos, M.** (1987) Purification, properties, and genetic location of *Escherichia coli* cytidine 5'-monophosphate *N*-acetylneuraminic acid synthetase. *Journal of Biological Chemistry*. **262** (36): 17556-17562.
97. **Odile, B.H., Heike, C., Ying, J., Julia, S.B., Holly, B.B., Keith, A.J., Craig, C., Rory, C., Jan, T.P., Wendell, D.Z., Carl, E.F., David, S.S., Ian, F., Matthias, F., Julian, P., Ulrich, V., Michael, A.Q., Stephen, D.B. and Martin, C.J.M.** (2013) Description and Nomenclature of *Neisseria meningitidis* Capsule Locus. *Emerging Infectious Disease journal*. **19** (4): 566.
98. **Lewis, A.L., Hensler, M.E., Varki, A. and Nizet, V.** (2006) The Group B Streptococcal Sialic Acid *O*-Acetyltransferase Is Encoded by *neuD*, a Conserved Component of Bacterial Sialic Acid Biosynthetic Gene Clusters. *Journal of Biological Chemistry*. **281** (16): 11186-11192.
99. **Claus, H., Borrow, R., Achtman, M., Morelli, G., Kantelberg, C., Longworth, E., Frosch, M. and Vogel, U.** (2004) Genetics of capsule *O*-acetylation in serogroup C, W-135 and Y meningococci. *Molecular Microbiology*. **51** (1): 227-239.
100. **Lewis, A.L., Nizet, V. and Varki, A.** (2004) Discovery and characterization of sialic acid *O*-acetylation in group B *Streptococcus. Proceedings of the National Academy of Sciences of the United States of America*. **101** (30): 11123-11128.

101. **Gurung, M.K., Raeder, I.L.U., Altermark, B. and Smalås, A.O.** (2013) Characterization of the sialic acid synthase from *Aliivibrio salmonicida* suggests a novel pathway for bacterial synthesis of 7-*O*-acetylated sialic acids. *Glycobiology*. **23** (7): 806-819.
102. **Lewis, A.L., Cao, H., Patel, S.K., Diaz, S., Ryan, W., Carlin, A.F., Thon, V., Lewis, W.G., Varki, A., Chen, X. and Nizet, V.** (2007) NeuA sialic acid *O*-acetylesterase activity modulates *O*-acetylation of capsular polysaccharide in group B *Streptococcus. Journal of Biological Chemistry*. **282** (38): 27562-27571.
103. **Gunawan, J., Simard, D., Gilbert, M., Lovering, A.L., Wakarchuk, W.W., Tanner, M.E. and Strynadka, N.C.J.** (2005) Structural and Mechanistic Analysis of Sialic Acid Synthase NeuB from K K in Complex with Mn^{2+}, Phosphoenolpyruvate, and *N*-Acetylmannosaminitol. **280** (5): 3555-3563.
104. **Mosimann, S.C., Gilbert, M., Dombroswki, D., To, R., Wakarchuk, W. and Strynadka, N.C.J.** (2001) Structure of a Sialic Acid-activating Synthetase, CMP-acylneuraminate Synthetase in the Presence and Absence of CDP. *Journal of Biological Chemistry*. **276** (11): 8190-8196.
105. **Blacklow, R.S. and Warren, L.** (1962) Biosynthesis of sialic acids by *Neisseria meningitidis. Journal of Biological Chemistry*. **237**: 3520-3526.
106. **Komaki, E., Ohta, Y. and Tsukada, Y.** (1997) Purification and Characterization of *N*-Acetylneuraminate Synthase from *Escherichia coli* K1-M12. *Bioscience, Biotechnology, and Biochemistry*. **61** (12): 2046-2050.
107. **Suryanti, V., Nelson, A. and Berry, A.** (2003) Cloning, over-expression, purification, and characterisation of *N*-acetylneuraminate synthase from *Streptococcus agalactiae. Protein Expression and Purification*. **27** (2): 346-356.
108. **Sundaram, A.K., Pitts, L., Muhammad, K., Wu, J., Betenbaugh, M., Woodard, R.W. and Vann, W.F.** (2004) Characterization of *N*-acetylneuraminic acid synthase isoenzyme 1 from *Campylobacter jejuni. The Biochemical journal*. **383** (Pt 1): 83-89.
109. **Hao, J., Balagurumoorthy, P., Sarilla, S. and Sundaramoorthy, M.** (2005) Cloning, expression, and characterization of sialic acid synthases. *Biochemical and Biophysical Research Communications*. **338** (3): 1507-1514.
110. **Liu, F., Lee, H.J., Strynadka, N.C.J. and Tanner, M.E.** (2009) Inhibition of *Neisseria meningitidis* Sialic Acid Synthase by a Tetrahedral Intermediate Analogue. *Biochemistry*. **48** (39): 9194-9201.
111. **Kawamura, T., Kimura, M., Yamamori, S. and Ito, E.** (1978) Enzymatic formation of uridine diphosphate *N*-acetyl-D-mannosamine. *Journal of Biological Chemistry*. **253** (10): 3595-3601.
112. **Murkin, A.S., Chou, W.K., Wakarchuk, W.W. and Tanner, M.E.** (2004) Identification and Mechanism of a Bacterial Hydrolyzing UDP-*N*-Acetylglucosamine 2-Epimerase. *Biochemistry*. **43** (44): 14290-14298.
113. **Warren, L. and S Blacklow, R.** (1962) The biosynthesis of cytidine 5'-monophospho-*N*-acetylneuraminic acid by an enzyme from *Neisseria meningitidis*. **237**: 3527-3534.
114. **Mizanur, R.M. and Pohl, N.L.** (2008) Bacterial CMP-sialic acid synthetases: production, properties, and applications. *Applied Microbiology and Biotechnology*. **80** (5): 757.
115. **Haft, R.F. and Wessels, M.R.** (1994) Characterization of CMP-*N*-acetylneuraminic acid synthetase of group B streptococci. *Journal of Bacteriology*. **176** (23): 7372.
116. **Tullius, M.V., Munson, R.S., Wang, J. and Gibson, B.W.** (1996) Purification, Cloning, and Expression of a Cytidine 5′-Monophosphate *N*-Acetylneuraminic Acid Synthetase from *Haemophilus ducreyi. Journal of Biological Chemistry*. **271** (26): 15373-15380.
117. **Bravo, I.G., Barrallo, S., Ferrero, M.A., RodrÍGuez-Aparicio, L.B., MartÍNez-Blanco, H. and Reglero, Á.** (2001) Kinetic properties of the acylneuraminate cytidylyltransferase from *Pasteurella haemolytica* A2. *Biochemical Journal*. **358** (3): 585.
118. **Yu, H., Ryan, W., Yu, H. and Chen, X.** (2006) Characterization of a Bifunctional Cytidine 5′-Monophosphate *N*-acetylneuraminic acid Synthetase Cloned from *Streptococcus Agalactiae. Biotechnology Letters*. **28** (2): 107-113.
119. **Mizanur, R.M. and Pohl, N.L.** (2007) Cloning and characterization of a heat-stable CMP-*N*-acylneuraminic acid synthetase from *Clostridium thermocellum. Applied Microbiology and Biotechnology*. **76** (4): 827-834.

120. **Kajiwara, H., Mine, T., Miyazaki, T. and Yamamoto, T.** (2011) A CMP-*N*-acetylneuraminic Acid Synthetase Purified from a Marine Bacterium, *Photobacterium leiognathi* JT-SHIZ-145. *Bioscience, Biotechnology, and Biochemistry*. **75** (1): 47-53.
121. **Glaze, P.A., Watson, D.C., Young, N.M. and Tanner, M.E.** (2008) Biosynthesis of CMP-N,N′-Diacetyllegionaminic Acid from UDP-*N,N′*-Diacetylbacillosamine in *Legionella pneumophila. Biochemistry*. **47** (10): 3272-3282.
122. **Schoenhofen, I.C., Vinogradov, E., Whitfield, D.M., Brisson, J.R. and Logan, S.M.** (2009) The CMP-legionaminic acid pathway in *Campylobacter*: biosynthesis involving novel GDP-linked precursors. *Glycobiology*. **19** (7): 715-725.
123. **Nothaft, H. and Szymanski, C.M.** (2010) Protein glycosylation in bacteria: sweeter than ever. *Nature Reviews Microbiology*. **8**: 765.
124. **Olivier, N.B., Chen, M.M., Behr, J.R. and Imperiali, B.** (2006) In Vitro Biosynthesis of UDP-*N,N'*-Diacetylbacillosamine by Enzymes of the *Campylobacter jejuni* General Protein Glycosylation System. *Biochemistry*. **45** (45): 13659-13669.
125. **Parkhill, J., Wren, B.W., Mungall, K., Ketley, J.M., Churcher, C., Basham, D., Chillingworth, T., Davies, R.M., Feltwell, T., Holroyd, S., Jagels, K., Karlyshev, A.V., Moule, S., Pallen, M.J., Penn, C.W., Quail, M.A., Rajandream, M.A., Rutherford, K.M., van Vliet, A.H.M., Whitehead, S. and Barrell, B.G.** (2000) The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature*. **403**: 665.
126. **Gundogdu, O., Bentley, S.D., Holden, M.T., Parkhill, J., Dorrell, N. and Wren, B.W.** (2007) Re-annotation and re-analysis of the *Campylobacter jejuni* NCTC11168 genome sequence. *BMC Genomics*. **8** (1): 162.
127. **Schoenhofen, I.C., McNally, D.J., Vinogradov, E., Whitfield, D., Young, N.M., Dick, S., Wakarchuk, W.W., Brisson, J.R. and Logan, S.M.** (2006) Functional characterization of dehydratase/aminotransferase pairs from *Helicobacter* and *Campylobacter*: enzymes distinguishing the pseudaminic acid and bacillosamine biosynthetic pathways. *Journal of Biological Chemistry*. **281** (2): 723-732.
128. **Riegert, A.S., Thoden, J.B., Schoenhofen, I.C., Watson, D.C., Young, N.M., Tipton, P.A. and Holden, H.M.** (2017) Structural and Biochemical Investigation of PglF from *Campylobacter jejuni* Reveals a New Mechanism for a Member of the Short Chain Dehydrogenase/Reductase Superfamily. *Biochemistry*. **56** (45): 6030-6040.
129. **Vijayakumar, S., Merkx-Jacques, A., Ratnayake, D.B., Gryski, I., Obhi, R.K., Houle, S., Dozois, C.M. and Creuzenet, C.** (2006) Cj1121c, a novel UDP-4-keto-6-deoxy-GlcNAc C-4 aminotransferase essential for protein glycosylation and virulence in *Campylobacter jejuni*. *Journal of Biological Chemistry*. **281** (38): 27733-27743.
130. **Riegert, A.S., Young, N.M., Watson, D.C., Thoden, J.B. and Holden, H.M.** (2015) Structure of the external aldimine form of PglE, an aminotransferase required for *N,N'*-diacetylbacillosamine biosynthesis. *Protein Science*. **24** (10): 1609-1616.
131. **Rangarajan, E.S., Ruane, K.M., Sulea, T., Watson, D.C., Proteau, A., Leclerc, S., Cygler, M., Matte, A. and Young, N.M.** (2008) Structure and Active Site Residues of PglD, an *N*-Acetyltransferase from the Bacillosamine Synthetic Pathway Required for N-Glycan Synthesis in *Campylobacter jejuni*. *Biochemistry*. **47** (7): 1827-1836.
132. **Tsvetkov, Y.E., Shashkov, A.S., Knirel, Y.A. and Zähringer, U.** (2001) Synthesis and NMR spectroscopy of nine stereoisomeric 5,7-diacetamido-3,5,7,9-tetradeoxynon-2-ulosonic acids. *Carbohydrate Research*. **335** (4): 221-243.
133. **Hjerde, E., Lorentzen, M.S., Holden, M.T., Seeger, K., Paulsen, S., Bason, N., Churcher, C., Harris, D., Norbertczak, H., Quail, M.A., Sanders, S., Thurston, S., Parkhill, J., Willassen, N.P. and Thomson, N.R.** (2008) The genome sequence of the fish pathogen *Aliivibrio salmonicida* strain LFI1238 shows extensive evidence of gene decay. *BMC Genomics*. **9**: 616.
134. **Ou, H.-Y., Kuang, S.N., He, X., Molgora, B.M., Ewing, P.J., Deng, Z., Osby, M., Chen, W. and Xu, H.H.** (2015) Complete genome sequence of hypervirulent and outbreak-associated *Acinetobacter baumannii* strain LAC-4: epidemiology, resistance genetic determinants and potential virulence factors. *Scientific Reports*. **5**: 8643.

135. **Kenyon, J.J., Holt, K.E., Pickard, D., Dougan, G. and Hall, R.M.** (2014) Insertions in the OCL1 locus of *Acinetobacter baumannii* lead to shortened lipooligosaccharides. *Research in Microbiology*. **165** (6): 472-475.
136. **He, X.M. and Liu, H.-w.** (2002) Formation of Unusual Sugars: Mechanistic Studies and Biosynthetic Applications. *Annual Review of Biochemistry*. **71** (1): 701-754.
137. **Schoenhofen, I.C., McNally, D.J., Brisson, J.R. and Logan, S.M.** (2006) Elucidation of the CMP-pseudaminic acid pathway in *Helicobacter pylori*: synthesis from UDP-*N*-acetylglucosamine by a single enzymatic reaction. *Glycobiology*. **16** (9): 8C-14C.
138. **Creuzenet, C., Schur, M.J., Li, J., Wakarchuk, W.W. and Lam, J.S.** (2000) FlaA1, a New Bifunctional UDP-GlcNAc C6Dehydratase/ C4 Reductase from *Helicobacter pylori*. *Journal of Biological Chemistry*. **275** (45): 34873-34880.
139. **Morrison, J.P., Schoenhofen, I.C. and Tanner, M.E.** (2008) Mechanistic studies on PseB of pseudaminic acid biosynthesis: A UDP-*N*-acetylglucosamine 5-inverting 4,6-dehydratase. *Bioorganic Chemistry*. **36** (6): 312-320.
140. **Ishiyama, N., Creuzenet, C., Miller, W.L., Demendi, M., Anderson, E.M., Harauz, G., Lam, J.S. and Berghuis, A.M.** (2006) Structural Studies of FlaA1 from *Helicobacter pylori* Reveal the Mechanism for Inverting 4,6-Dehydratase Activity. *Journal of Biological Chemistry*. **281** (34): 24489-24495.
141. **Schoenhofen, I.C., Lunin, V.V., Julien, J.P., Li, Y., Ajamian, E., Matte, A., Cygler, M., Brisson, J.R., Aubry, A., Logan, S.M., Bhatia, S., Wakarchuk, W.W. and Young, N.M.** (2006) Structural and functional characterization of PseC, an aminotransferase involved in the biosynthesis of pseudaminic acid, an essential flagellar modification in *Helicobacter pylori*. *Journal of Biological Chemistry*. **281** (13): 8907-8916.
142. **Song, W.S., Nam, M.S., Namgung, B. and Yoon, S.-i.** (2015) Structural analysis of PseH, the *Campylobacter jejuni N*-acetyltransferase involved in bacterial O-linked glycosylation. *Biochemical and Biophysical Research Communications*. **458** (4): 843-848.
143. **Ud-Din, A.I., Liu, Y.C. and Roujeinikova, A.** (2015) Crystal Structure of *Helicobacter pylori* Pseudaminic Acid Biosynthesis *N*-Acetyltransferase PseH: Implications for Substrate Specificity and Catalysis. *PloS one*. **10** (3): e0115634.
144. **Rangarajan, E.S., Proteau, A., Cui, Q., Logan, S.M., Potetinova, Z., Whitfield, D., Purisima, E.O., Cygler, M., Matte, A., Sulea, T. and Schoenhofen, I.C.** (2009) Structural and Functional Analysis of *Campylobacter jejuni* PseG: A UDP-sugar Hydrolase from the Pseudaminic Acid Biosynthesis Pathway. *Journal of Biological Chemistry*. **284** (31): 20989-21000.
145. **Liu, F. and Tanner, M.E.** (2006) PseG of Pseudaminic Acid Biosynthesis: A UDP-sugar Hydrolase as a Masked Glycosyltransferase. *Journal of Biological Chemistry*. **281** (30): 20902-20909.
146. **Ali, T., Weintraub, A. and Widmalm, G.** (2006) Structural determination of the O-antigenic polysaccharide from the Shiga toxin-producing *Escherichia coli* O171. *Carbohydrate Research*. **341** (11): 1878-1883.
147. **Andreishcheva, E.N. and Vann, W.F.** (2006) Gene Products Required for De Novo Synthesis of Polysialic Acid in *Escherichia coli* K1. *Journal of Bacteriology*. **188** (5): 1786.
148. **Kenyon, J.J., Kasimova, A.A., Notaro, A., Arbatsky, N.P., Speciale, I., Shashkov, A.S., De Castro, C., Hall, R.M. and Knirel, Y.A.** (2017) *Acinetobacter baumannii* K13 and K73 capsular polysaccharides differ only in K-unit side branches of novel non-2-ulosonic acids: di-*N*-acetylated forms of either acinetaminic acid or 8-epiacinetaminic acid. *Carbohydrate Research*. **452**: 149-155.
149. **Castric, P., Cassels, F.J. and Carlson, R.W.** (2001) Structural characterization of the *Pseudomonas aeruginosa* 1244 pilin glycan. *Journal of Biological Chemistry*. **276** (28): 26479-26485.
150. **Vimr, E.R., Kalivoda, K.A., Deszo, E.L. and Steenbergen, S.M.** (2004) Diversity of microbial sialic acid metabolism. *Microbiology and Molecular Biology Reviews*. **68** (1): 132-153.
151. **Severi, E., Hood, D.W. and Thomas, G.H.** (2007) Sialic acid utilization by bacterial pathogens. *Microbiology*. **153** (Pt 9): 2817-2822.
152. **Khatua, B., Roy, S. and Mandal, C.** (2013) Sialic acids siglec interaction: a unique strategy to circumvent innate immune response by pathogens. *Indian Journal of Medical Research*. **138** (5): 648-662.

153. **Czuchry, D., Desormeaux, P., Stuart, M., Jarvis, D.L., Matta, K.L., Szarek, W.A. and Brockhausen, I.** (2015) Identification and Biochemical Characterization of the Novel α2,3-Sialyltransferase WbwA from Pathogenic *Escherichia coli* Serotype O104. *Journal of Bacteriology*. **197** (24): 3760.
154. **Bloch, S., Zwicker, S., Bostanci, N., Sjöling, Å., Boström, E.A., Belibasakis, G.N. and Schäffer, C.** (2018) Immune response profiling of primary monocytes and oral keratinocytes to different *Tannerella forsythia* strains and their cell surface mutants. *Molecular Oral Microbiology*. **33** (2): 155-167.
155. **Stephenson, H.N., Mills, D.C., Jones, H., Milioris, E., Copland, A., Dorrell, N., Wren, B.W., Crocker, P.R., Escors, D. and Bajaj-Elliott, M.** (2014) Pseudaminic acid on *Campylobacter jejuni* flagella modulates dendritic cell IL-10 expression via Siglec-10 receptor: a novel flagellin-host interaction. *The Journal of Infectious Diseases*. **210** (9): 1487-1498.
156. **Kwiatkowski, B., Boschek, B., Thiele, H. and Stirm, S.** (1982) Endo-*N*-Acetylneuraminidase associated with bacteriophage particles. **43**. 1982. 697-704.
157. **Javed, M.A., van Alphen, L.B., Sacher, J., Ding, W., Kelly, J., Nargang, C., Smith, D.F., Cummings, R.D. and Szymanski, C.M.** (2015) A receptor-binding protein of *Campylobacter jejuni* bacteriophage NCTC 12673 recognizes flagellin glycosylated with acetamidino-modified pseudaminic acid. *Molecular Microbiology*. **95** (1): 101-115.
158. **Goon, S., Kelly, J.F., Logan, S.M., Ewing, C.P. and Guerry, P.** (2003) Pseudaminic acid, the major modification on *Campylobacter* flagellin, is synthesized via the *Cj1293* gene. *Molecular Microbiology*. **50** (2): 659-671.
159. **Hitchen, P., Brzostek, J., Panico, M., Butler, J.A., Morris, H.R., Dell, A. and Linton, D.** (2010) Modification of the *Campylobacter jejuni* flagellin glycan by the product of the *Cj1295* homopolymeric-tract-containing gene. *Microbiology*. **156** (Pt 7): 1953-1962.
160. **Lindhout, T., Lau, P.C.Y., Brewer, D. and Lam, J.S.** (2009) Truncation in the core oligosaccharide of lipopolysaccharide affects flagella-mediated motility in *Pseudomonas aeruginosa* PAO1 via modulation of cell surface attachment. **155** (10): 3449-3460.
161. **Post, D.M.B., Yu, L., Krasity, B.C., Choudhury, B., Mandel, M.J., Brennan, C.A., Ruby, E.G., McFall-Ngai, M.J., Gibson, B.W. and Apicella, M.A.** (2012) O-antigen and Core Carbohydrate of *Vibrio fischeri* Lipopolysaccharide: Composition and Analysis of their Role in *Euprymna scolopes* Light Organ Colonization. *Journal of Biological Chemistry*. **287** (11): 8515-8530.
162. **Anderson, G.G., Goller, C.C., Justice, S., Hultgren, S.J. and Seed, P.C.** (2010) Polysaccharide Capsule and Sialic Acid-Mediated Regulation Promote Biofilm-Like Intracellular Bacterial Communities during Cystitis. *Infection and Immunity*. **78** (3): 963.
163. **Lubin, J.B., Lewis, W., Gilbert, N., Boyd, E. and Lewis, A.** (2014) Sialic acid-like molecules in motility, biofilm formation, and systemic pathogenesis by the seafood-borne pathogen *Vibrio vulnificus* (LB144). *FASEB Journal*. **28** (1_supplement): LB144.
164. **Varki, A. and Gagneux, P.** (2012) Multifarious roles of sialic acids in immunity. *Annals of the New York Academy of Sciences*. **1253** (1): 16-36.
165. **Lehmann, F., Tiralongo, E. and Tiralongo, J.** (2006) Sialic acid-specific lectins: occurrence, specificity and function. *Cellular and Molecular Life Sciences*. **63** (12): 1331-1354.
166. **Bruno, D.W., Noguera, P.A. and Poppe, T.T.** (2013) Bacterial Diseases. In: *A Colour Atlas of Salmonid Diseases*. pp. 73-98, Springer, Dordrecht.
167. **Lunder, T., Evensen, Ø., Holstad, G. and Håstein, T.** (1995) 'Winter ulcer' in the Atlantic salmon *Salmo salar*. Pathological and bacteriological investigations and transmission experiments. *Diseases of Aquatic Organisms*. **23**: 39-49.
168. **Hjerde, E., Karlsen, C., Sorum, H., Parkhill, J., Willassen, N.P. and Thomson, N.R.** (2015) Co-cultivation and transcriptome sequencing of two co-existing fish pathogens *Moritella viscosa* and *Aliivibrio wodanis*. *BMC Genomics*. **16**: 447.
169. **Karlsen, C., Vanberg, C., Mikkelsen, H. and Sorum, H.** (2014) Co-infection of Atlantic salmon (*Salmo salar*), by *Moritella viscosa* and *Aliivibrio wodanis*, development of disease and host colonization. *Veterinary Microbiology*. **171** (1-2): 112-121.

170. **Bogwald, J. and Hoffman, J.** (2006) Structural studies of the O-antigenic oligosaccharide from *Vibrio salmonicida* strain C2 isolated from Atlantic cod, *Gadus morhua* L. *Carbohydrate Research*. **341** (11): 1965-1968.
171. **Shtyrya, Y.A., Mochalova, L.V. and Bovin, N.V.** (2009) Influenza virus neuraminidase: structure and function. *Acta naturae*. **1** (2): 26-32.
172. **Chan, T.-H., Xin, Y.-C. and von Itzstein, M.** (1997) Synthesis of Phosphonic Acid Analogues of Sialic Acids (Neu5Ac and KDN) as Potential Sialidase Inhibitors. *The Journal of Organic Chemistry*. **62** (11): 3500-3504.
173. **Gulati, S., Schoenhofen, I.C., Whitfield, D.M., Cox, A.D., Li, J., St. Michael, F., Vinogradov, E.V., Stupak, J., Zheng, B., Ohnishi, M., Unemo, M., Lewis, L.A., Taylor, R.E., Landig, C.S., Diaz, S., Reed, G.W., Varki, A., Rice, P.A. and Ram, S.** (2015) Utilizing CMP-Sialic Acid Analogs to Unravel *Neisseria gonorrhoeae* Lipooligosaccharide-Mediated Complement Resistance and Design Novel Therapeutics. *PLoS Pathogens*. **11** (12): e1005290.
174. **Vavricka, C.J., Muto, C., Hasunuma, T., Kimura, Y., Araki, M., Wu, Y., Gao, G.F., Ohrui, H., Izumi, M. and Kiyota, H.** (2017) Synthesis of Sulfo-Sialic Acid Analogues: Potent Neuraminidase Inhibitors in Regards to Anomeric Functionality. *Scientific Reports*. **7** (1): 8239.
175. **Heise, T., Pijnenborg, J.F.A., Büll, C., van Hilten, N., Kers-Rebel, E.D., Balneger, N., Elferink, H., Adema, G.J. and Boltje, T.J.** (2019) Potent Metabolic Sialylation Inhibitors Based on C-5-Modified Fluorinated Sialic Acids. *Journal of Medicinal Chemistry*. **62** (2): 1014-1021.
176. **Andolina, G., Wei, R., Liu, H., Zhang, Q., Yang, X., Cao, H., Chen, S., Yan, A., Li, X.D. and Li, X.** (2018) Metabolic Labeling of Pseudaminic Acid-Containing Glycans on Bacterial Surfaces. *ACS Chemical Biology*. **13** (10): 3030-3037.
177. **Varki, A.** (2008) Sialic acids in human health and disease. *Trends in molecular medicine*. **14** (8): 351-360.
178. **Wang, B., Brand-Miller, J., McVeagh, P. and Petocz, P.** (2001) Concentration and distribution of sialic acid in human milk and infant formulas. *The American Journal of Clinical Nutrition*. **74** (4): 510-515.
179. **El Maarouf, A., Petridis, A.K. and Rutishauser, U.** (2006) Use of polysialic acid in repair of the central nervous system. *Proceedings of the National Academy of Sciences of the United States of America*. **103** (45): 16989-16994.
180. **Zhang, T., She, Z., Huang, Z., Li, J., Luo, X. and Deng, Y.** (2014) Application of sialic acid/polysialic acid in the drug delivery systems. *Asian Journal of Pharmaceutical Sciences*. **9** (2): 75-81.
181. **Pearce, O.M.T. and Läubli, H.** (2015) Sialic acids in cancer biology and immunity. *Glycobiology*. **26** (2): 111-128.
182. **Romppanen, J., Punnonen, K., Anttila, P., Jakobsson, T., Blake, J. and Niemela, O.** (2002) Serum sialic acid as a marker of alcohol consumption: effect of liver disease and heavy drinking. *Alcoholism: Clinical and Experimental Research*. **26** (8): 1234-1238.
183. **Martin, J.E., Tanenbaum, S.W. and Flashner, M.** (1977) A facile procedure for the isolation of *N*-acetylneuramic acid from edible bird's-nest. *Carbohydrate Research*. **56** (2): 423-425.
184. **Koketsu, M., Juneja, L.R., Kawanami, H., Kim, M. and Yamamoto, T.** (1992) Preparation of *N*-acetylneuraminic acid from delipidated egg yolk. *Glycoconjugate Journal*. **9** (2): 70-74.
185. **Liu, F., Aubry, A.J., Schoenhofen, I.C., Logan, S.M. and Tanner, M.E.** (2009) The Engineering of Bacteria Bearing Azido-Pseudaminic Acid-Modified Flagella. *ChemBioChem*. **10** (8): 1317-1320.
186. **Lin, B. and Tao, Y.** (2016) Production of Sialic Acid and Its Derivatives by Metabolic Engineering of *Escherichia coli*. In: *Quality Living Through Chemurgy and Green Chemistry*. pp. 301-318, Springer Berlin Heidelberg, Berlin, Heidelberg.
187. **Hassan, M.I., Lundgren, B.R., Chaumun, M., Whitfield, D.M., Clark, B., Schoenhofen, I.C. and Boddy, C.N.** (2016) Total Biosynthesis of Legionaminic Acid, a Bacterial Sialic Acid Analogue. *Angewandte Chemie International Edition in English*. **55** (39): 12018-12021.
188. **Chen, X. and Varki, A.** (2010) Advances in the biology and chemistry of sialic acids. *ACS Chemical Biology*. **5** (2): 163-176.

189. **Norimura, Y., Yamamoto, D. and Makino, K.** (2017) Synthesis of sialic acid derivatives based on chiral substrate-controlled stereoselective aldol reactions using pyruvic acid oxabicyclo[2.2.2]octyl orthoester. *Organic and Biomolecular Chemistry*. **15** (3): 640-648.
190. **Carter, J.R. and Kiefel, M.J.** (2018) A new approach to the synthesis of legionaminic acid analogues. *RSC Advances*. **8** (62): 35768-35775.
191. **Gintner, M., Yoneda, Y., Schmölzer, C., Denner, C., Kählig, H. and Schmid, W.** (2019) A versatile de novo synthesis of legionaminic acid and 4-epi-legionaminic acid starting from d-serine. *Carbohydrate Research*. **474**: 34-42.
192. **Comb, D.G. and Roseman, S.** (1960) The sialic acids. I. The structure and enzymatic synthesis of *N*-acetylneuraminic acid. *Journal of Biological Chemistry*. **235**: 2529-2537.
193. **Simon, E.S., Bednarski, M.D. and Whitesides, G.M.** (1988) Synthesis of CMP-NeuAc from *N*-acetylglucosamine: generation of CTP from CMP using adenylate kinase. *Journal of the American Chemical Society*. **110** (21): 7159-7163.
194. **Mahmoudian, M., Noble, D., Drake, C.S., Middleton, R.F., Montgomery, D.S., Piercey, J.E., Ramlakhan, D., Todd, M. and Dawson, M.J.** (1997) An efficient process for production of *N*-acetylneuraminic acid using *N*-acetylneuraminic acid aldolase. *Enzyme and Microbial Technology*. **20** (5): 393-400.
195. **Stallforth, P., Matthies, S., Adibekian, A., Gillingham, D.G., Hilvert, D. and Seeberger, P.H.** (2012) *De novo* chemoenzymatic synthesis of sialic acid. *Chemical Communications*. **48** (98): 11987-11989.
196. **Kragl, U., Gygax, D., Ghisalba, O. and Wandrey, C.** (1991) Enzymatic Two-Step Synthesis of *N*-Acetyl-neuraminic Acid in the Enzyme Membrane Reactor. *Angewandte Chemie International Edition in English*. **30** (7): 827-828.
197. **Maru, I., Ohnishi, J., Ohta, Y. and Tsukada, Y.** (1998) Simple and large-scale production of *N*-acetylneuraminic acid from *N*-acetyl-D-glucosamine and pyruvate using *N*-acyl-D-glucosamine 2-epimerase and *N*-acetylneuraminate lyase. *Carbohydrate Research*. **306** (4): 575-578.
198. **Maru, I., Ohnishi, J., Ohta, Y. and Tsukada, Y.** (2002) Why is sialic acid attracting interest now? complete enzymatic synthesis of sialic acid with *N*-acylglucosamine 2-epimerase. *Journal of Bioscience and Bioengineering*. **93** (3): 258-265.
199. **Tabata, K., Koizumi, S., Endo, T. and Ozaki, A.** (2002) Production of *N*-acetyl- D -neuraminic acid by coupling bacteria expressing *N*-acetyl- D -glucosamine 2-epimerase and *N*-acetyl- D -neuraminic acid synthetase. *Enzyme and Microbial Technology*. **30** (3): 327-333.
200. **Lee, J.O., Yi, J.K., Lee, S.G., Takahashi, S. and Kim, B.G.** (2004) Production of *N*-acetylneuraminic acid from *N*-acetylglucosamine and pyruvate using recombinant human renin binding protein and sialic acid aldolase in one pot. *Enzyme and Microbial Technology*. **35** (2-3): 121-125.
201. **Wang, T.-H., Chen, Y.-Y., Pan, H.-H., Wang, F.-P., Cheng, C.-H. and Lee, W.-C.** (2009) Production of *N*-acetyl-D-neuraminic acid using two sequential enzymes overexpressed as double-tagged fusion proteins. *BMC biotechnology*. **9**: 63-63.
202. **Lee, Y.-C., Chien, H.-C.R.C. and Hsu, W.-H.** (2007) Production of *N*-acetyl-D-neuraminic acid by recombinant whole cells expressing *Anabaena* sp. CH1 *N*-acetyl-D-glucosamine 2-epimerase and *Escherichia coli N*-acetyl-D-neuraminic acid lyase. *Journal of Biotechnology*. **129** (3): 453-460.
203. **Wang, T.-H., Chen, Y.-Y., Pan, H.-H., Wang, F.-P., Cheng, C.-H. and Lee, W.-C.** (2009) Production of *N*-acetyl-D-neuraminic acid using two sequential enzymes overexpressed as double-tagged fusion proteins. *BMC biotechnology*. **9** (1): 63.
204. **Hu, S., Chen, J., Yang, Z., Shao, L., Bai, H., Luo, J., Jiang, W. and Yang, Y.** (2010) Coupled bioconversion for preparation of *N*-acetyl-D-neuraminic acid using immobilized *N*-acetyl-D-glucosamine-2-epimerase and *N*-acetyl-D-neuraminic acid lyase. *Applied Microbiology and Biotechnology*. **85** (5): 1383-1391.
205. **Kao, C.-H., Chen, Y.-Y., Wang, L.-R. and Lee, Y.-C.** (2018) Production of *N*-acetyl-d-neuraminic Acid by Recombinant Single Whole Cells Co-expressing *N*-acetyl-d-glucosamine-2-epimerase and *N*-acetyl-d-neuraminic Acid Aldolase. *Molecular Biotechnology*. **60** (6): 427-434.
206. **Comb, D.G. and Roseman, S.** (1958) Composition and Enzymatic Synthesis of *N*-acetylneuraminic Acid (Sialic Acid). *Journal of the American Chemical Society*. **80** (2): 497-499.

207. **Ghosh, S. and Roseman, S.** (1965) The Sialic Acids. V. *N*-Acyl-D-Glucosamine 2-Epimerase. *Journal of Biological Chemistry*. **240**: 1531-1536.
208. **Zimmermann, V., Hennemann, H.G., Daußmann, T. and Kragl, U.** (2007) Modelling the reaction course of *N*-acetylneuraminic acid synthesis from *N*-acetyl-D-glucosamine - New strategies for the optimisation of neuraminic acid synthesis. *Applied Microbiology and Biotechnology*. **76** (3): 597-605.
209. **Augé, C., David, S. and Gautheron, C.** (1984) Synthesis with immobilized enzyme of the most important sialic acid. *Tetrahedron Letters*. **25** (41): 4663-4664.
210. **Wang, S.-l., Li, Y.-l., Han, Z., Chen, X., Chen, Q.-j., Wang, Y. and He, L.-s.** (2018) Molecular Characterization of a Novel *N*-Acetylneuraminate Lyase from a Deep-Sea Symbiotic *Mycoplasma. Marine Drugs*. **16** (3).
211. **Takahashi, S., Hori, K., Takahashi, K., Ogasawara, H., Tomatsu, M. and Saito, K.** (2001) Effects of nucleotides on *N*-acetyl-d-glucosamine 2-epimerases (renin-binding proteins): comparative biochemical studies. *Journal of Biochemistry*. **130** (6): 815-821.
212. **Gurung MK, Altermark B, Helland R, Smalås AO, Ræder ILU.** (2019) Features and Structure of a Novel Cold Active *N*-Acetylneuramininate Lyase. *PLoS One* . **14** (6): e0217713
213. **Itoh, T., Mikami, B., Maru, I., Ohta, Y., Hashimoto, W. and Murata, K.** (2000) Crystal structure of *N*-acyl-D-glucosamine 2-epimerase from porcine kidney at 2.0 Å resolution. *Journal of Molecular Biology*. **303** (5): 733-744.
214. **Lee, Y.-C., Wu, H.-M., Chang, Y.-N., Wang, W.-C. and Hsu, W.-H.** (2007) The Central Cavity from the (α/α)6 Barrel Structure of *Anabaena* sp. CH1 *N*-Acetyl-d-glucosamine 2-Epimerase Contains Two Key Histidine Residues for Reversible Conversion. *Journal of Molecular Biology*. **367** (3): 895-908.
215. **Takahashi, S., Takahashi, K., Kaneko, T., Ogasawara, H., Shindo, S., Saito, K. and Kawamura, Y.** (2001) Identification of Functionally Important Cysteine Residues of the Human Renin-Binding Protein as the Enzyme *N*-Acetyl-D-Glucosamine 2-Epimerase. *Journal of Biochemistry*. **129**: 529-535.
216. **Wang, S.-Y., Laborda, P., Lu, A.-M., Duan, X.-C., Ma, H.-Y., Liu, L. and Voglmeir, J.** (2016) *N*-acetylglucosamine 2-Epimerase from *Pedobacter heparinus*: First Experimental Evidence of a Deprotonation/Reprotonation Mechanism. *Catalysts*. **6** (12): 212.
217. **Datta, A.** (1970) Regulatory role of adenosine triphosphate on hog kidney *N*-acetyl-D-glucosamine 2-epimerase. *Biochemistry*. **9** (17): 3363-3370.
218. **Sola-Carvajal, A., Sánchez-Carrón, G., García-García, M.I., García-Carmona, F. and Sánchez-Ferrer, Á.** (2012) Properties of BoAGE2, a second *N*-acetyl-d-glucosamine 2-epimerase from *Bacteroides ovatus* ATCC 8483. *Biochimie*. **94** (1): 222-230.
219. **Maru, I., Ohta, Y., Murata, K. and Tsukada, Y.** (1996) Molecular cloning and identification of *N*-acyl-D-glucosamine 2-epimerase from porcine kidney as a renin-binding protein. *Journal of Biological Chemistry*. **271** (27): 16294-16299.
220. **Takahashi, S., Takahashi, K., Kaneko, T., Ogasawara, H., Shindo, S. and Kobayashi, M.** (1999) Human renin-binding protein is the enzyme *N*-acetyl-D-glucosamine 2-epimerase. *Journal of Biochemistry*. **125** (2): 348-353.
221. **Takahashi, S., Ohsawa, T., Miura, R. and Miyake, Y.** (1983) Purification of High Molecular Weight (11MW) Renin from Porcine Kidney and Direct Evidence that the HMW Renm Is a Complex of Renin with Renin Binding Protein (RnBP). *The Journal of Biochemistry*. **93** (1): 265-274.
222. **Inoue, H., Takahashi, S., Fukui, K. and Miyake, Y.** (1991) Leucine zipper motif in porcine renin-binding protein (RnBP) and its relationship to the formation of an RnBP-renin heterodimer and an RnBP homodimer. *Journal of Biological Chemistry*. **266** (18): 11896-11900.
223. **Schmitz, C., Gotthardt, M., Hinderlich, S., Leheste, J.-R., Gross, V., Vorum, H., Christensen, E.I., Luft, F.C., Takahashi, S. and Willnow, T.E.** (2000) Normal Blood Pressure and Plasma Renin Activity in Mice Lacking the Renin-binding Protein, a Cellular Renin Inhibitor. *Journal of Biological Chemistry*. **275** (20): 15357-15362.
224. **Luchansky, S.J., Yarema, K.J., Takahashi, S. and Bertozzi, C.R.** (2003) GlcNAc 2-epimerase can serve a catabolic role in sialic acid metabolism. *Journal of Biological Chemistry*. **278** (10): 8035-8042.

225. **Takahashi, S., Miura, R. and Miyake, Y.** (1988) Characterization of hog kidney renin-binding protein: interconversion between monomeric and dimeric forms. *Biochemistry international*. **16** (6): 1053-1060.
226. **Takahashi, S., Inoue, H., Fukui, K. and Miyake, Y.** (1994) Structure and function of renin binding protein. *Kidney International*. **46** (6): 1525-1527.
227. **Klermund, L., Groher, A. and Castiglione, K.** (2013) New *N*-acyl-D-glucosamine 2-epimerases from cyanobacteria with high activity in the absence of ATP and low inhibition by pyruvate. *Journal of Biotechnology*. **168** (3): 256-263.
228. **Takahashi, S., Takahashi, K., Kaneko, T., Ogasawara, H., Shindo, S., Saito, K. and Kobayashi, M.** (1999) Identification of Cysteine-380 as the Essential Residue for the Human *N*-Acetyl-D-Glucosamine 2-Epimerase (Renin Binding Protein). *The Journal of Biochemistry*. **126** (4): 639-642.
229. **Takahashi, S., Ogasawara, H., Takahashi, K., Hori, K., Saito, K. and Mori, K.** (2002) Identification of a Domain Conferring Nucleotide Binding to the *N*-Acetyl-D-Glucosamine 2-Epimerase (Renin Binding Protein). *Journal of Biochemistry*. **131**: 605-610.
230. **Takahashi, S., Ogasawara, H., Hiwatashi, K., Hata, K., Hori, K., Koizumi, Y. and Sugiyama, T.** (2005) Amino acid residues conferring the nucleotide binding properties of *N*-acetyl-D-glucosamine 2-epimerase (renin binding protein). *Biomedical Research*. **26** (3): 117-121.
231. **Liao, H.F., Kao, C.H., Lin, W.D., Hsiao, N.W., Hsu, W.H. and Lee, Y.C.** (2012) *N*-Acetyl-D-glucosamine 2-epimerase from *Anabaena* sp. CH1 contains a novel ATP-binding site required for catalytic activity. *Process Biochemistry*. **47** (6): 948-952.
232. **Klermund, L., Riederer, A., Hunger, A. and Castiglione, K.** (2016) Protein engineering of a bacterial *N*-acyl-D-glucosamine 2-epimerase for improved stability under process conditions. *Enzyme and Microbial Technology*. **87-88**: 70-78.
233. **Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L., Studholme, D.J., Yeats, C. and Eddy, S.R.** (2004) The Pfam protein families database. *Nucleic Acids Research*. **32** (Database issue): D138-141.
234. **Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L., Tate, J. and Punta, M.** (2014) Pfam: the protein families database. *Nucleic Acids Research*. **42** (Database issue): D222-230.
235. **Itoh, T., Mikami, B., Hashimoto, W. and Murata, K.** (2008) Crystal structure of YihS in complex with D-mannose: structural annotation of *Escherichia coli* and *Salmonella enterica* yihS-encoded proteins to an aldose-ketose isomerase. *Journal of Molecular Biology*. **377** (5): 1443-1459.
236. **Hamilton, T.L., Bryant, D.A. and Macalady, J.L.** (2016) The role of biology in planetary evolution: cyanobacterial primary production in low-oxygen Proterozoic oceans. *Environmental Microbiology*. **18** (2): 325-340.
237. **Parmar, A., Singh, N.K., Pandey, A., Gnansounou, E. and Madamwar, D.** (2011) Cyanobacteria and microalgae: A positive prospect for biofuels. *Bioresource Technology*. **102** (22): 10163-10172.
238. **Zehr, J.P.** (2011) Nitrogen fixation by marine cyanobacteria. *Trends in Microbiology*. **19** (4): 162-173.
239. **Haque, F., Banayan, S., Yee, J. and Chiang, Y.W.** (2017) Extraction and applications of cyanotoxins and other cyanobacterial secondary metabolites. *Chemosphere*. **183**: 164-175.
240. **Zhang, A., Carroll, A.L. and Atsumi, S.** (2017) Carbon recycling by cyanobacteria: improving CO2 fixation through chemical production. *FEMS Microbiology Letters*. **364** (16).
241. **Kultschar, B. and Llewellyn, C.** (2018) *Secondary Metabolites in Cyanobacteria*.
242. **Brito, Â., Gaifem, J., Ramos, V., Glukhov, E., Dorrestein, P.C., Gerwick, W.H., Vasconcelos, V.M., Mendes, M.V. and Tamagnini, P.** (2015) Bioprospecting Portuguese Atlantic coast cyanobacteria for bioactive secondary metabolites reveals untapped chemodiversity. *Algal Research*. **9**: 218-226.
243. **Montalvao, S., Demirel, Z., Devi, P., Lombardi, V., Hongisto, V., Perala, M., Hattara, J., Imamoglu, E., Tilvi, S.S., Turan, G., Dalay, M.C. and Tammela, P.** (2016) Large-scale bioprospecting of cyanobacteria, micro- and macroalgae from the Aegean Sea. *N Biotechnol*. **33** (3): 399-406.

244. **Hamilton, E.F., Element, G., van Coeverden de Groot, P., Engel, K., Neufeld, J.D., Shah, V. and Walker, V.K.** (2019) Anadromous Arctic Char Microbiomes: Bioprospecting in the High Arctic. *Front Bioeng Biotechnol*. **7**: 32.
245. **Ojit, S.K., Gunapati, O. and Tiwari, O.** (2012) New Record of Potential Cyanobacteria from Indian Region Falling Indo-Burma Biodiversity Hotspots (North-East Region of India) and Partial Characterization for Value Additions. *Philippine Journal of Science*. **141** (1): 57-66.
246. **Mardis, E.R.** (2011) A decade's perspective on DNA sequencing technology. *Nature*. **470**: 198.
247. **Atkinson, M.R., Deutscher, M.P., Kornberg, A., Russell, A.F. and Moffatt, J.G.** (1969) Enzymatic synthesis of deoxyribonucleic acid. XXXIV. Termination of chain growth by a 2',3'-dideoxyribonucleotide. *Biochemistry*. **8** (12): 4897-4904.
248. **Sanger, F., Nicklen, S. and Coulson, A.R.** (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*. **74** (12): 5463-5467.
249. **Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R., Heiner, C., Kent, S.B.H. and Hood, L.E.** (1986) Fluorescence detection in automated DNA sequence analysis. *Nature*. **321** (6071): 674-679.
250. **Gyllensten, U.B.** (1989) PCR and DNA sequencing. *Biotechniques*. **7** (7): 700-708.
251. **Metzker, M.L.** (2010) Sequencing technologies — the next generation. *Nature Reviews Genetics*. **11**: 31-46.
252. **Thompson, J.F. and Milos, P.M.** (2011) The properties and applications of single-molecule DNA sequencing. *Genome biology*. **12** (2): 217-217.
253. **Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlén, M. and Nyrén, P.** (1996) Real-Time DNA Sequencing Using Detection of Pyrophosphate Release. *Analytical Biochemistry*. **242** (1): 84-89.
254. **Ronaghi, M., Uhlén, M. and Nyrén, P.** (1998) A Sequencing Method Based on Real-Time Pyrophosphate. *Science*. **281** (5375): 363.
255. **Lewis, A.L., Lubin, J.B., Argade, S., Naidu, N., Choudhury, B. and Boyd, E.F.** (2011) Genomic and metabolic profiling of nonulosonic acids in Vibrionaceae reveal biochemical phenotypes of allelic divergence in *Vibrio vulnificus*. *Applied and Environmental Microbiology*. **77** (16): 5782-5793.
256. **Linton, D., Karlyshev, A.V., Hitchen, P.G., Morris, H.R., Dell, A., Gregson, N.A. and Wren, B.W.** (2000) Multiple *N*-acetyl neuraminic acid synthetase (neuB) genes in Campylobacter jejuni: identification and characterization of the gene involved in sialylation of lipo-oligosaccharide. *Mol Microbiol*. **35** (5): 1120-1134.
257. **Liaimer, A., Jenke-Kodama, H., Ishida, K., Hinrichs, K., Stangeland, J., Hertweck, C. and Dittmann, E.** (2011) A polyketide interferes with cellular differentiation in the symbiotic cyanobacterium *Nostoc punctiforme. Environmental Microbiology Reports*. **3** (5): 550-558.
258. **Liaimer, A., Helfrich, E.J.N., Hinrichs, K., Guljamow, A., Ishida, K., Hertweck, C. and Dittmann, E.** (2015) Nostopeptolide plays a governing role during cellular differentiation of the symbiotic cyanobacterium *Nostoc punctiforme*. *Proceedings of the National Academy of Sciences*. **112** (6): 1862.
259. **Liaimer, A., Jensen, J.B. and Dittmann, E.** (2016) A genetic and chemical perspective on symbiotic recruitment of cyanobacteria of the genus *Nostoc* into the host plant *Blasia pusilla* L. *Frontiers in Microbiology*. **7** (NOV).
260. **Ekman, M., Picossi, S., Campbell, E.L., Meeks, J.C. and Flores, E.** (2013) A *Nostoc punctiforme* sugar transporter necessary to establish a Cyanobacterium-plant symbiosis. *Plant Physiology*. **161** (4): 1984-1992.
261. **Fujiwara, T., Saburi, W., Matsui, H., Mori, H. and Yao, M.** (2014) Structural insights into the epimerization of β-1,4-Linked Oligosaccharides Catalyzed by Cellobiose 2-Epimerase, the Sole Enzyme Epimerizing non-Anomeric Hydroxyl groups of Unmodified Sugars. *Journal of Biological Chemistry*. **289** (6): 3405-15.

Paper I

Analysis of nonulosonic acid biosynthetic gene clusters in *Aliivibrio salmonicida* and *Moritella viscosa*

Marie-Josée Haglund Halsør, Bjørn Altermark, and Inger Lin Uttakleiv Ræder*

The Norwegian Structural Biology Centre (NorStruct), Department of Chemistry, UiT – The Arctic University of Norway, 9037 Tromsø, Norway. *Correspondence e-mail: inger.l.rader@uit.no

Abstract

Nonulosonic acid (NulO) biosynthesis in bacteria is directed by *nab* gene clusters that can lead to neuraminic, legionaminic or pseudaminic acids. Analysis of the gene content from a set of *Aliivibrio salmonicida* and *Moritella viscosa* strains reveals the existence of several unique *nab* clusters, which can be used to guide tandem mass spectrometry studies. NulOs from several strains were thus identified.

Introduction

Bacterial nonulosonic acids (NulOs) are 9-carbon α-keto acids found on the cell surface, as parts of the capsule, flagella, lipopolysaccharide and pili [1-6]. Neuraminic acids (Neu), a family of dideoxynonulosonic acids better known as sialic acids, are most widely spread in vertebrates in which they were first discovered [7, 8]. The most common bacterial NulOs are tetradeoxynonulosonic acids, although neuraminic acids are also found, and one instance of a trideoxy-compound has been reported recently [1, 9]. The various tetradeoxy-NulOs are somewhat classified according to their absolute configuration. Legionaminic acid (Leg) is a 5,7-diamino-3,5,7,9-tetradeoxy-D-*glycero*-D-*galacto*-non-2-ulosonic acid, while pseudaminic acid (Pse) is the L-*glycero*-L-*manno* isomer [2, 3, 10, 11]. The L-*glycero*-D-*galacto* and L-*glycero*-D-*talo* isomers are designated as 8- and 4-epilegionaminic acids, respectively [12, 13]. The L-*glycero*-L-*altro* isomer is known as Acinetaminic acid (Aci), and the D-*glycero*-L-*altro* isomer as its 8-epimer [14, 15]. Most recently, the (D/L)-*glycero*-L-*gluco* isomer was identified as fusaminic acid (Fus), like its trideoxy-counterpart [9, 16].

NulOs from pathogenic bacteria play a central role in host-pathogen interactions. Their structure mimics that of host neuraminic acids, thus enabling them to bind to the Siglecs (Sialic acid-binding immunoglobulin-type lectins) of host immune cells, which mediate the inhibition of immune cell activation [17, 18]. As a critical component of flagella and pili they indirectly affect host colonization, but also bacterial motility in general [19-21]. In addition to synthetizing NulOs themselves, bacteria can utilize host NulOs to evade the immune system [22, 23].

NulO biosynthesis is directed by *nab* gene clusters [24-26]. Neu, Leg and Pse are synthetized by different pathways (Nonulosonic Acid Biosynthesis, or NAB), although each contains homologous genes that are responsible for the main biosynthesis steps [27]. Aci is synthetized by a set of additional genes affixed to the Leg cluster, and the gene cluster responsible for the synthesis of Fus has not been investigated yet [14]. The composition and organization of *nab* clusters varies not only between pathways and species, but also within species [27, 28]. Only one gene, coding for the *N*-acetylneuraminate synthase NeuB, is conserved across both pathways and species. The NeuB homologs catalyze the key reaction of NulO biosynthesis, the condensation of a hexosamine precursor with pyruvate leading to either Neu, Leg, or Pse.

Aliivibrio salmonicida (formerly *Vibrio salmonicida*) is a psychrophilic fish pathogen causing cold water vibriosis, a disease that used to seriously impact the output of salmon aquaculture before the use of vaccines [29-31]. It is known to produce Neu5Ac and Neu5Ac7(9)Ac, and the presence of 8eLeg5Am7Ac has been detected in its lipopolysaccharide [32-34].

Moritella viscosa is the main agent in causing the winter ulcer disease in salmon and cod, with *Aliivibrio wodanis* as a co-pathogen [35-37]. As for *A. salmonicida*, Neu5Ac and Neu5Ac7(9)Ac were shown to be produced by *M. viscosa* [38]. The NulO content of *A. wodanis* has not been investigated so far, but its *nab* gene cluster appears similar to that of *Vibrio vulnificus* CMCP6, which produces the Leg variant Leg5Ac7AcAla [39]. The diversity of *nab* clusters within *Vibrionaceae* has already been shown, with an emphasis on *Vibrio vulnificus* [28].

This study investigates the presence and composition of nab gene clusters in a set of strains *from A. salmonicida* and *M. viscosa*, as well as *A. wodanis* and a few other members of the *Vibrionaceae* family. The strains were screened for the presence of NeuB coding sequences which were used to locate *nab* clusters in their genomes as well as study NeuB substrate specificity. Each cluster was analyzed in terms of gene content and organization, using sequence homology as a basis for functional annotation. From this, hypotheses pertaining to the nature of the NulO(s) produced by each strain could made. They were verified by analyzing the NulO content of the concerned organisms using mass spectrometry.

Materials and Methods

Bacterial strains and genome sequences

The genome sequences for *A. salmonicida* LFI1238, *M. viscosa* 06/09/139 and *A. wodanis* 06/09/139 are published and were obtained from public databases [40, 41]. Sequences for the other *Aliivibrio* strains, as well as for the *Photobacterium* strain were provided by Espen Robertsen, via the MetaRep database [42-44]. The *M. viscosa* genomes were sequenced as a project and were provided by Terje Klemetsen

[45]. They are available through the European Nucleotide Archive (ENA) under the PRJEB16014 study. The genome sequence for *V. anguillarum* NB10 was provided by Kåre Olav Holm [46]. The source information for the genomes used in this study is summarized in **Table 1**.

The strains for this study were isolated from *Eurythenes gryllus* (Amphipoda) and *Porifera indet* [47]. *Vibrio* B9-25 was isolated from *Halichondria* sp. and *P. phosphoreum* SP005 was isolated from the gut of *Onogadus argentatus*. *V. anguillarum* NB10 was isolated from *Gadhus morhua* [48, 49].

Table 1. Bacterial sources of genome sequences.

Organism	Strain
Aliivibrio salmonicida	LFI1238[1], B9-25K2[2], MR17-77[2], R8-63[2], R8-68[2], R8-70[2]
Aliivibrio wodanis	06/09/139[3]
Vibrio anguillarum	NB10[4]
Photobacterium	SP005[2]
Moritella viscosa	06/09/139[3], K58[5], K56[5], Vvi7[5], Vvi11[5], NVI5482[5], NVI4917[5], LFI5006[5], F57[5]

[1] Accession numbers: NC_011310.1 to 011316.1 [41].

[2] Obtained from the MetaRep database [42].

[3] Accession numbers: LN554846.1 to LN554851.1 [40].

[4] [46]

[5] [45]

Genome analysis, *nab* cluster assignment, and sugar prediction

The target genomes were, if necessary, annotated using the GePan pipeline of the Galaxy server at the University of Tromsø [50, 51]. Glimmer v3.2 was used for gene prediction and BLASTp for database search within Bacteria [52, 53].

The protein sequences for the NeuB homologs from *A. salmonicida* LFI1238 (accession numbers WP_012551408.1 and WP_012549051.1), *M. viscosa* 06/09/139 (WP_045111757.1 and WP_045111735.1), and *A. wodanis* 06/09/139 (WP_045100955.1) were used as query for a similarity search against the set of target genomes using the BLAST suite. [53]. Unique NeuB sequences were thereafter aligned with MUSCLE, using default settings [54, 55].

The positive NeuB hits were located on the target genomes and the surrounding area was investigated for the presence of putative *nab* gene clusters, using the Artemis software for viewing [56]. The cluster regions were determined by the gene content and direction of the relevant areas. Gene names were assigned according to protein sequence similarity to known *nab* genes, using the most comprehensive (to the authors) denominations.

Identified clusters were analyzed for the function of each coding sequence composing them by performing sequence similarity searches within the non-redundant protein sequence database (BLAST) as well as domain searches within the Conserved domain Database [57].

Homology modelling of *C. jejuni* NeuB homologs

The sequences for the NeuB homologs of *C. jejuni* NCTC11168 (WP_002858213.1, WP_002864265.1, and WP_002870258.1) were retrieved from public databases and used as targets for homology modelling using the NeuB homolog from *N. meningitidis* (PDB IDs: 1XUU, 1XUZ) as a template [58, 59]. The modelling was performed on the SWISS-MODEL server [60].

Nonulosonic content release and derivatization

Cultures (15mL) of *A. salmonicida* LFI1238, *M. viscosa* 06/09/139, and *Vibrio* B9-25K2 were grown in liquid LB media containing 2.5% NaCl (48h, 12°C, 200 rpm). 3 mL pellets were harvested and washed with dH_2O, before 0.1 µL phenylmethane sulfonyl fluoride (PMSF) was added. After a 15 minutes incubation on ice, 200 µL acetic acid (2M) and 2 µL butylated hydroxytoluene (BHT, 1%) was added. The samples were thereafter incubated for 3 hours at 80°C and spun down for 10 minutes at 13000 rpm. The supernatant was collected and filtered (Amicon 10K spin column) in order to remove large molecules. The filtrate was dried for 2 hours using a speed-vac, and the resulting samples were stored at -20°C until use.

Samples were resuspended in 10 µL dH_2O before performing the labelling reaction with 1,2-diamino-4,5-methylenedioxybenzene (DMB, from TaKaRa) according to the manufacturer's instructions [61-63]. The reaction mixtures were incubated in the dark at 50°C for 2.5 hours.

Mass spectrometry analyses

The quinoxaline (Q) content of the samples described in the previous section was analyzed by HPLC-MS/MS using the procedure described by Gurung *et al.* [32]. Water with 0.1% formic acid (A) and acetonitrile with 0.1% formic acid (B) were used for the HPLC-MS elution gradient presented in **Table 2**, at a flow rate of 400 µL/min. Tandem mass spectrometry was performed on samples containing compounds corresponding to masses equivalent to that of LegAmAcQ (*m/z*$[M+H]^+$ = 450.19887323) and $LegAc_2Q$/ $PseAc_2Q$ (*m/z*$[M+H]^+$ = 451.1823402). The scan range was *m/z* 350-550.

Table 2. Elution gradient for HPLC-MS

Time (minutes)	A (%)	B (%)
0	95	5
1	95	5
8	75	25
8.01	10	90
9	10	90
9.01	95	5
10	95	5

Graphical output generation

Gene clusters were rendered in SVG graphics using scripts written by the author. The scripts are available through the python-bioinformatics repository on GitHub [64]. Molecular structures were obtained from PubChem and modified in Molview and Pymol [65, 66]. All figures were prepared using Inkscape [67].

Results and discussion

NeuB sequences comparison

The two unique NeuB protein sequences from *A. salmonicida* LFI1238 (representing the Neu and Leg pathways) and *M. viscosa* (representing the Neu and Pse pathways), as well as the one from *A. wodanis* were used as queries for conducting a sequence similarity search within the set of target genomes, with the aim of detecting potential *nab* clusters. Out of the 15 targets, 10 had at least one sequence identical to one of the queries, and 3 had a similar sequence (an alignment of the unique NeuB sequences is presented in Figure 1). As expected, no NeuB coding sequences were found in either *P. phosphoreum* or *V. anguillarum*.

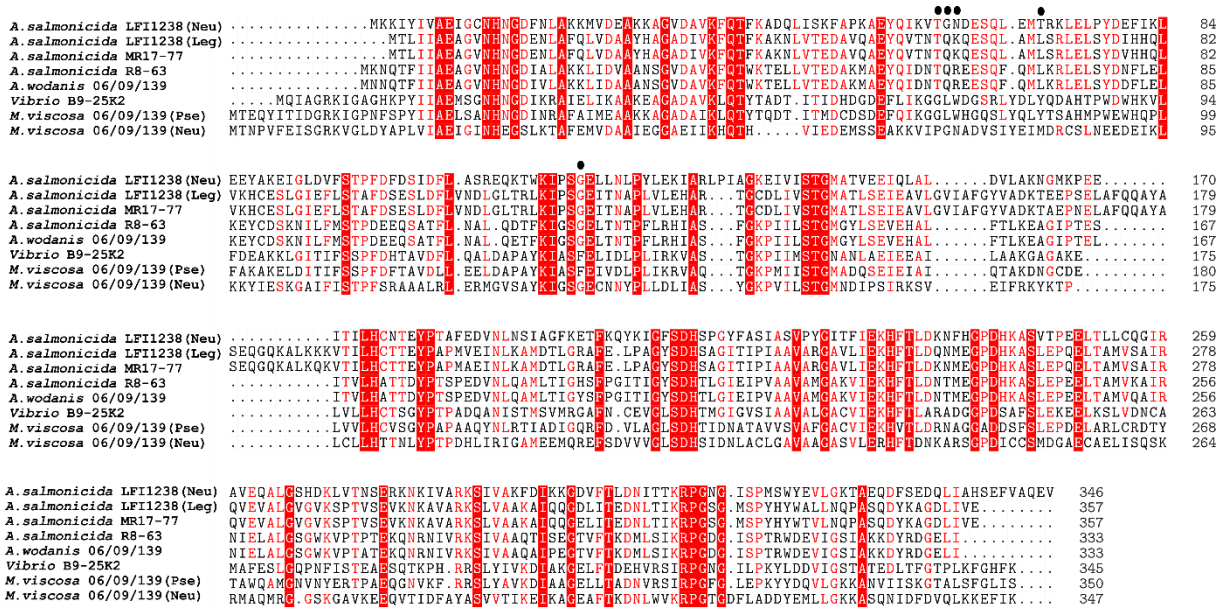


Figure 1. Multiple alignment of unique NeuB sequences from Vibrionaceae. *The sequences from A. salmonicida LFI1238, R8-63, and MR17-77 were aligned with that of M. viscosa and A. wodanis 06/09/139 as well as Vibrio B9-25K2. Residues putatively involved in substrate specificity are indicated by black circles.*

Six of the eight targeted *M. viscosa* strains have two NeuB coding sequences identical to the ones from the 06/09/139 strain. The two remaining strains contain each one sequence, identical to either the Neu-pathway homolog (strain F57) or the Pse-pathway homolog (strain NVI-5482). This is consistent with the evolutionary relations between *M. viscosa* genomes according to which those two strains form their own clade [45]. For the *Aliivibrio* strains, two strains (R8-68 and R8-70) have sequences identical two both homologs from LFI1238. The MR17-77 and R8-63 strains each have one hit that is almost identical (97% sequence identity for both) to the Leg-pathway homolog of *A. salmonicida* LFI1238 and *A. wodanis*,

respectively. Again, this is in accordance with phylogenetic data [43, 44]. The NeuB sequence from *Vibrio* sp. B9-25K2 is most similar (54% sequence identity) to the *M. viscosa* Pse-pathway homolog. Closer sequences (above 80% identity) were found amongst *Vibrionaceae*, also coding for pseudaminic acid synthases (data not shown).

The possibility of assigning NAB pathways solely based on the sequence of the NeuB homologs raises the question of what the determinants which discriminate between the pathways are. The genome of *Campylobacter jejuni* strains contain genes coding for NeuB variants for all three pathways and was therefore chosen for studying the pathway-related characteristics of NeuB homologs [68]. The NCTC 11168 strain is able to produce Neu5Ac, Pse5Ac7Ac, and Leg5Ac7Ac, as well as derivatives of the Leg/Pse compounds (assuming that the acetamidino- derivatives are synthetized from the diacetylated compounds) [69]. Their sequences were modelled onto the structure of the NeuB homolog (Neu pathway) from *Nesseria meningitidis* (nmNeuB, PDB ID: 1XUU), and the structure of their active sites was compared with its PEP-bound form (PDB ID: 1XUZ) [58].

The results, presented in Figure2, show that while the active residues are strictly conserved between the Neu-pathway enzymes, they differ for the Leg- and Pse- pathway ones, as is expected. Indeed, the different pathways indicate that while NeuB from the Neu acts on ManNAc for the production of Neu5Ac, the substrates leading to diacetylated Pse and Leg compounds are 2,4-diacetamido-2,4,6-trideoxyaltrose and 2,4-diacetamido-2,4,6-trideoxymannose, respectively [70-72]. Neither Leg nor Pse substrates carry a -OH group at C6, which explains the variations observed for residues N74 and M83 of nmNeuB. It is interesting to note that the OH group carried by C4 is interacting with a water molecule, which is probably not present in the cjNeuB2/cjNeuB3 active sites due to the N-acetyl group at this position, thus compensating for the additional bulk without changing the overall geometry of the active site. Its orientation is different in the Leg and Pse substrates (same orientation as the -OH of rMAnNAc for the Leg substrate), which justifies the different residues for cjNeuB2 and cjNeuB3 at this location. The F129 residue of cjNeuB3 is clashing with the N-acetyl group carried by C2, which makes sense since its orientation is different in the Pse substrate.

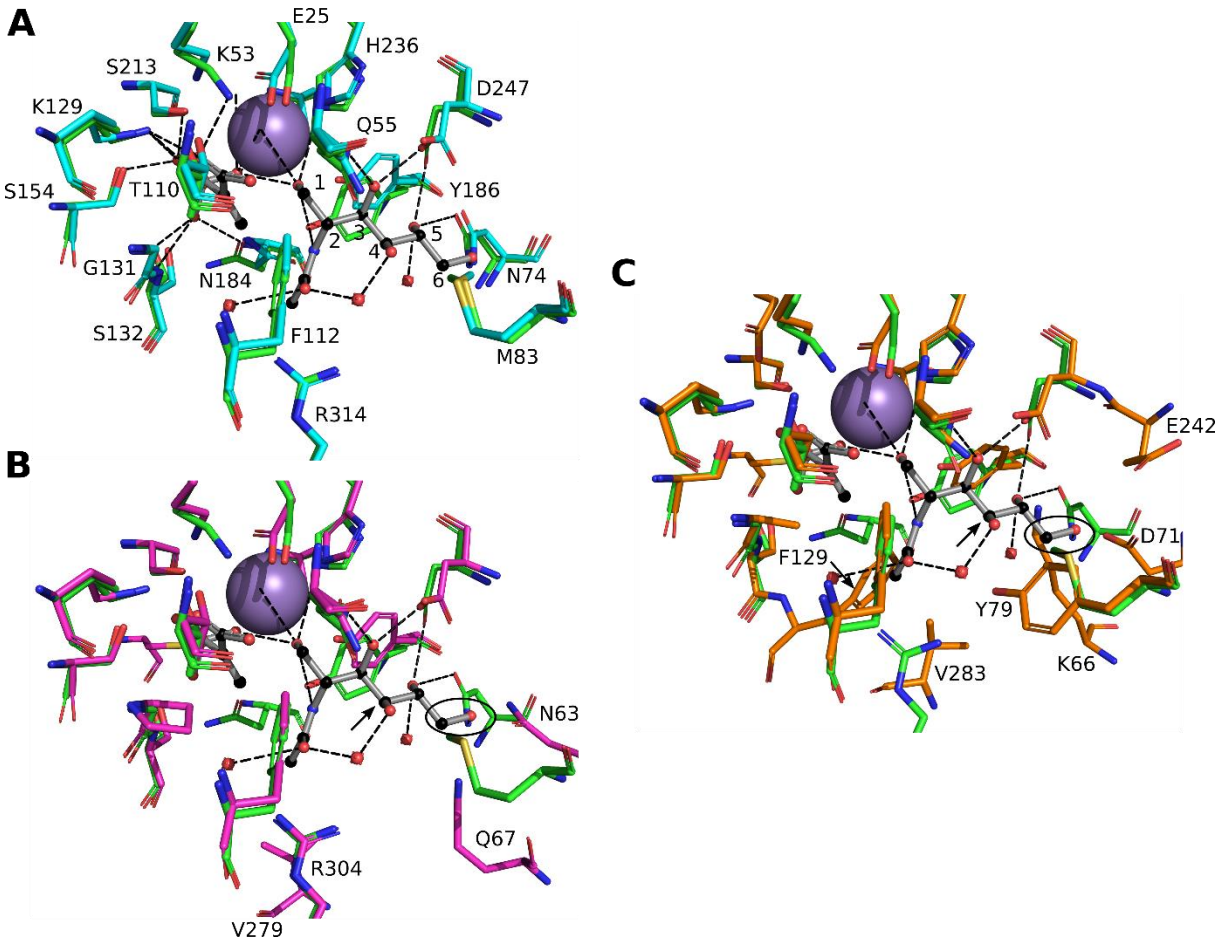


Figure 2. Comparison of NeuB active sites of C. jejuni and N. meningitidis. *Green: NeuB from Nesseria meningitidis (1xuz). Ball and stick representations: reduced ManNAc (rMAnNAc, right) and phosphoenol pyruvate (PEP, left). Purple sphere: Mn^{2+}. A: Neu pathway. Comparison with the model for NeuB1 of C. jejuni NCTC 11168. The carbon backbone of rManNAc is numbered, and the active site residues of N. meningitidis are indicated. B: Leg pathway. Comparison with the model for NeuB2 of C. jejuni NCTC 11168. C: Pse-pathway. Comparison with the model for NeuB3 of C. jejuni NCTC 11168. For the Leg and Pse pathways, the C4 position of rManNAc (where there should be a second N-acetyl group) is indicated by an arrow. The C9 position, which does not carry a -OH group in Leg/Pse, is circled. Residues of NeuB2/NeuB3 that differ from that of N. meningitidis are labeled.*

These results are also consistent with the proposed mechanism of NeuB enzymes, which depends on the area of the active site that houses PEP and the C1 of rMAnNAc. This area is conserved in all three isozymes, while the area binding the rest of the sugar molecule is less so, allowing for specificity at C2 via position 133 (F129 in NeuB3) and at C4-C6 via the S2 loop and helix H4 (positions 70-85). These observations hold true for our set of sequences, were the corresponding residues (indicated by black circles in Figure) clearly distinguish the Pse-pathway sequences from the other ones, notably at the position corresponding to that of F129 of cjNeuB3 from *N. meningitidis*, which is either occupied similarly by a phenylalanine (Pse-pathway), or a glycine (Neu/Leg pathways). The distinction between

sequences from the Neu and Leg pathway (specificity at C4-C6) is less clear, with some of the concerned residues conserved across the pathways and/or distinct between the two Neu-pathway homologs.

The same determinants may discriminate between isozymes from the same pathway but producing different sugars. The NeuB from *A. salmonicida* from the Neu pathway is thought to accept 4-acetylated ManNAc as a natural substrate rather than the monoacetylated form even though it is capable of utilizing it, which is consistent with the flexibility at the C4 position where the active site water molecule would compensate for the lack of NAc group [32]. Its other NeuB homolog is responsible for the synthesis of Leg [33]. The configuration at C4 for ManNAc2 would then be identical to that of LegAc2, since the substituents are in the same orientation, while C6 remains as it is in Neu5Ac, carrying an -OH group. In any case, it seems that *A. salmonicida* MR17-77 has the same specificity as *A. salmonicida* LFI1238, and *A. salmonicida* R8-63 as *A. wodanis* (the sequences they are closest to), since their sequences are identical within the putative specificity regions. In the case of the NeuB homologs from the Pse pathway, it is difficult to say if the differences within the "specificity region" are due to a different substrate and/or related to the bacterial species.

nab cluster identification and analysis

The genome regions surrounding each NeuB homolog hit were investigated for the presence of putative *nab* clusters. The results, presented in Figure 3, reveal that hits (or queries) with identical NeuB sequences had identical clusters (identical gene composition, sequences and organization). This is especially interesting in the case of the Leg pathway cluster from *A. salmonicida* LFI1238, where the pseudogene sequence for NeuA is conserved with that of *A. salmonicida* R8-68 and R8-70. Furthermore, where the NeuB sequences are not identical, the clusters are different, even if the sequences are highly similar. Indeed, the Leg clusters from *A. salmonicida* MR17-77 and R8-63, for which the NeuB sequences share 97% identity with that of *A. salmonicida* LFI1238 and *A. wodanis*, respectively, have a different architecture composed of different genes. When homologs are present, their sequences are similar, but not identical. It is worth noting that the Leg cluster of *A. salmonicida* R8-63 is closest to that of *A. wodanis* for its first part (up to the second aminotransferase), but then shows more similarity with that of *A. salmonicida* LFI1238 (PtmE and NeuA sequences).

In order to confirm if this trend is observed for other species, we investigated the set of NeuB sequences similar to that of *Vibrio* B9-25K2, leading to the conclusion that when NeuB sequences are identical, the corresponding clusters might share a similar architecture with the same number of proteins coding for the same homologs in the same order, but they are not necessarily all identical in protein sequence (data not shown).

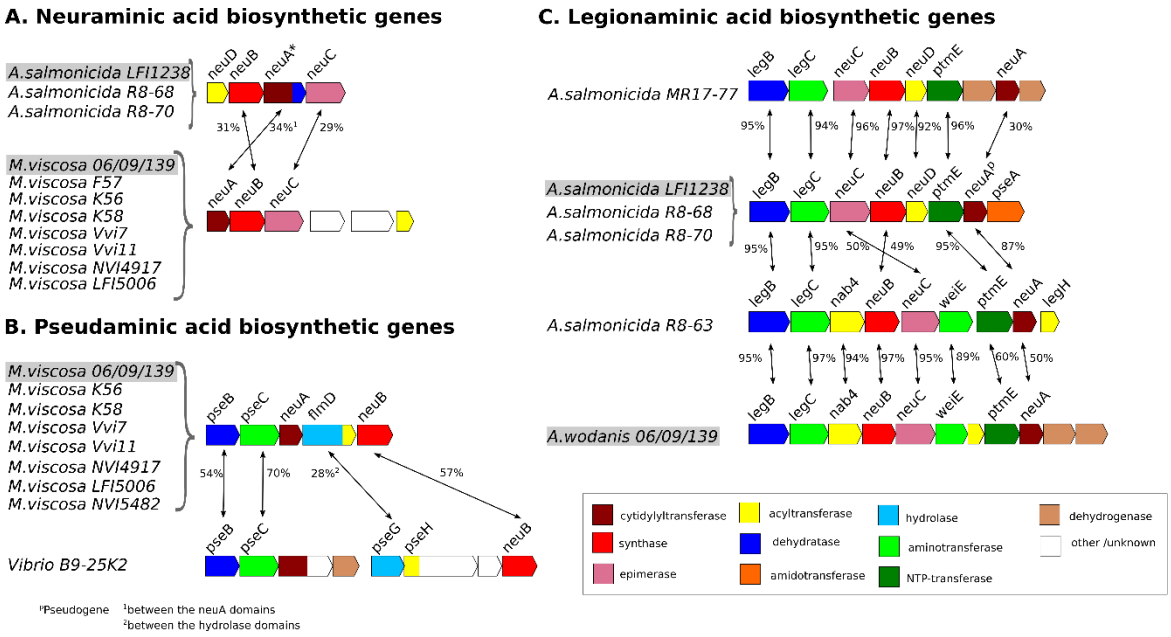


Figure 3. Unique nab clusters amongst Aliivibrio, Vibrio, and Moritella viscosa strains. *The translated nucleotide sequence of each gene of each cluster was compared globally to the protein sequences coded by the nab clusters of A. salmonicida LFI1238, A. wodanis 06/09139, and M. viscosa 06/09/139. Each gene is colored according to the function of the encoded protein, as stated in the figure legend. The percentage of sequence identity between homologs is indicated by black arrows, with the value next to it.*

The cluster of *Vibrio* B9-25K2, for which the NeuB sequence was most similar to that of Pse pathway homolog from *M. viscosa*, contains homologs of sequences coding for PseB, PseC and PseG, which confirms its appartenance to this pathway. It also contains a homolog of PseH as a domain in a bi-functional protein for which the second domain is of unknown function, but that is associated to lipid metabolism. A conserved domains analysis of the cluster coding sequences revealed that the bi-functional enzyme encoded after PseC contains a putative cytidylyltransferase domain similar to the SpsF spore coat protein, indicating that it may be replacing the NeuA homolog absent from this cluster. The other coded domain is that of an aldo-keto reductase. Following this gene is a sequence coding for a dehydrogenase. The PseH homolog is followed by a gene coding for a hypothetical protein, with a domain similar to that of methylmalonyl CoA epimerase. The clusters corresponding to the closest NeuB sequences from public databases also contain the hypothetical protein and the bi-functional PseH homolog, both with around 60% sequence identity compared to that of *Vibrio* B9-25K2 (data not shown). A likewise search for the bi-functional cytidylyltransferase yielded a different set of strains sharing similar architectures, but none for which the NulO content has been investigated (data not shown).

From the analysis of the NeuB homologs, we were able to hypothetize that the NeuBs from *A. salmonicida* MR17-77 and R8-63 most likely produces Leg compounds, with the one from MR17-77 producing the same as that of *A. salmonicida* LFI1238 ($LegAc_2$) and R8-63 the same as that of *A. wodanis* (LegAcAlaAc or similar, since the sugar in *A. wodanis* has not been identified yet).The NeuB from *Vibrio* B9-25K2 seems to produce a Pse related NulO, but the lack of close sequences with identified NulO content in the corresponding organisms prevented a more precise hypothesis. The results from the cluster analysis are in agreement with these suggestions, and they suggest that *Vibrio* B9-25K2 may produce a diacetylated Pse on account of the presence of a PseH homolog. The presence of enzymes of unknown function might indicate further modifications of produced compound. For the *Aliivibrio* strains with close NeuB sequences, they most likely ultimately produce different compounds since the composition of their clusters is not identical albeit NeuB homologs seemingly producing the same compound.

Analysis of bacterial NulO content by LC-MS

The NulO content of *A. salmonicida* LFI1238 and R8-63, *M. viscosa* and *A. wodanis* 06/09/139, and *Vibrio* sp. B9-25K2 was released by mild acid hydrolysis and labelled with DMB (*A. salmonicida* MR17-77 could not be grown at the time of the experiment). Separation of the quinoxalines by LC-MS led to the detection of several potential NulOs, as presented in Figure 4. As expected, the MS spectra for *A. salmonicida* LFI1238 contained peaks at masses corresponding to NeuAcQ (m/z = 426), $NeuAc_2Q$ (m/z = 468), and LegAmAcQ (m/z = 450). The *M. viscosa* sample showed peaks at masses corresponding to NeuAcQ, $NeuAc_2Q$, and $PseAc_2Q$ (m/z = 451), with the latter also present in the *Vibrio* B9-25K2 sample but with a different retention time. The samples from *A. salmonicida* R8-63 and *A. wodanis* had peaks at m/z= 530 and 522, respectively. While the former does not correspond to any identified NulO by this method, the latter can be assigned to the LegAcAlaAcQ compound observed in *Vibrio vulnificus* [39]. Those results verify the hypotheses made after analysis of the NeuB homologs and the *nab* clusters concerning the type of sugar produced by each cluster. A similar approach has already been used successfully in several species and the method appears robust, although it is limited to detecting compounds similar (by either mass or gene cluster) to previously identified sugars [27, 39, 73].

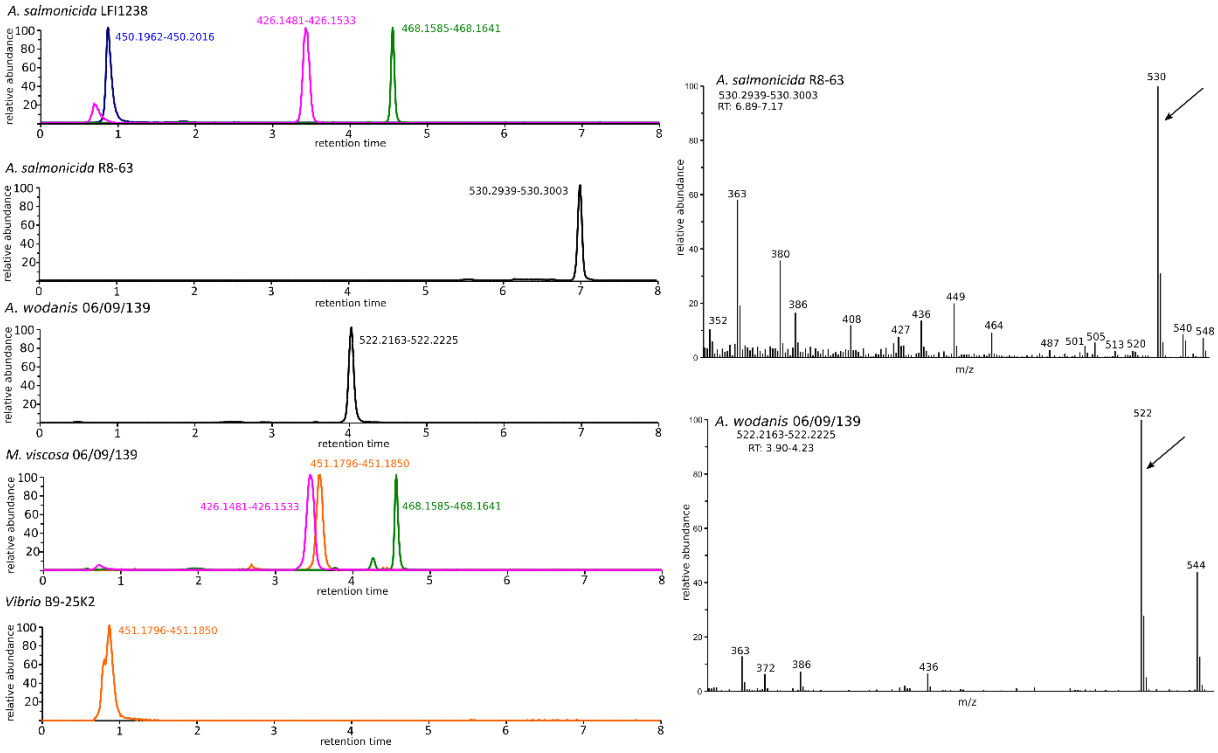


Figure 4. LC-MS analysis of bacterial NulO content derivatized with DMB. *Left: FTMS+ pESI full MS spectra. The spectra are represented for each species as a function of relative abundance versus retention time (in minutes). When several mass ranges are considered, the spectra are combined in a single graph using different colors for each curve (blue, pink, green, and orange). The mass range (m/z) corresponding to each peak is indicated next to it. Right: Mass spectra for the peaks from A. salmonicida R8-63 and A. wodanis.*

Tandem mass spectrometry

While tandem mass spectrometry has been previously performed on the quinoxalinones of Neu compounds, no standards are available for Leg and Pse derivatives with the exception of LegAcAlaAc which was identified after this study was performed [32, 38, 39, 74]. MS/MS was therefore performed for peaks at masses corresponding to that of LegAmAcQ (m/z[M+H]$^+$ = 450.19887323), and PseAc$_2$Q (m/z[M+H]$^+$ = 451.1823402). The LegAcAlaAcQ compound was not identified at the time of the experiment, and masses related to it were not investigated. The results, presented in Figure 5, show that the compounds from *A. salmonicida* (putative Leg5Am7Ac) and *M. viscosa* (putative Pse5Ac7Ac) share similar fragmentation patterns. Interestingly, the compound from *Vibrio* B9-25K2 has a different spectrum from that of *M. viscosa*, despite both corresponding to PseAc$_2$. This suggests that either the position of the N-acetyl groups is different, or that some of the asymmetrical carbons have a different geometry, thus producing an epimer of Pse. No reports of a di-N-acetylated NulO carrying groups at positions other than C5 and C7 exist yet, and the only pathway known to produce epimers of its NulO is

the Leg pathway [11-14]. The determination of the absolute configuration of $PseAc_2$ from *Vibrio* B9-25K2, as well as studies targeting the proteins of its cluster should provide the necessary information.

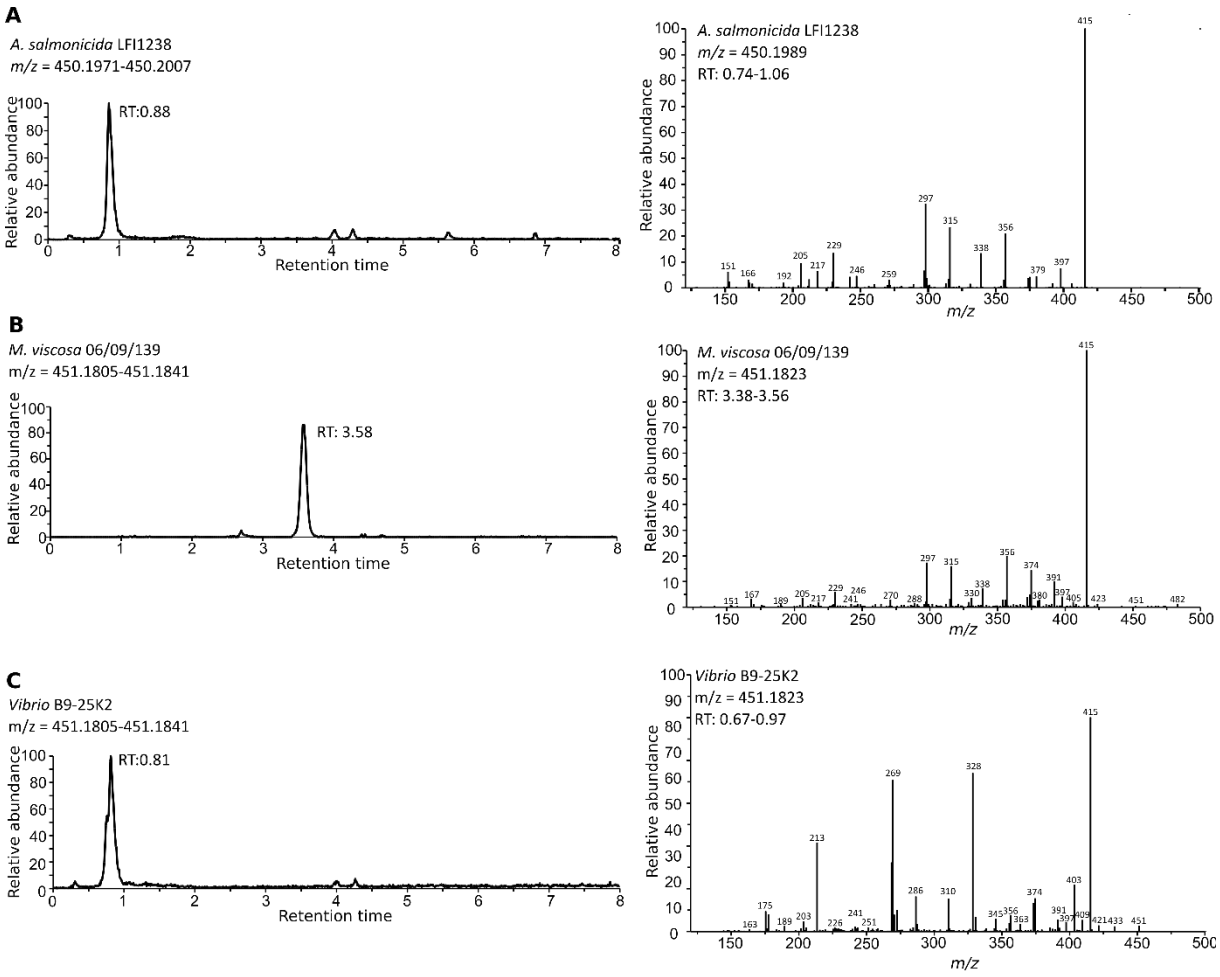


Figure 5. Mass spectrometry analysis of putative LegAcAm and $PseAc_2$ compounds. *A. Analysis for A. salmonicida LFI1238. Left: FTMS+ pESI full MS spectra at m/z= 450.1971-450.2007. Right: ITMS + cESI with full MS2 spectra of mass 450.1989 at time 0.74-1.06 minutes. B. Analysis for M. viscosa 06/09/139. Left: FTMS+ pESI full MS spectra at m/z= 451.1805-451.1841. Right: ITMS + cESI with full MS2 spectra of mass 451.1823 at time 3.38-3.56 minutes. C. Analysis for Vibrio B9-25K2. Left: FTMS+ pESI full MS spectra at m/z= 451.1805-451.1841. Right: ITMS + cESI with full MS2 spectra of mass 451.1823 at time 0.67-1.06 minutes.*

The parent ions for all compounds are either weak or absent, as is expected from molecules rich in alcohol groups. The base peak, at m/z = 415, can result from the loss of either two water molecules from the $PseAc_2$ compounds or one water and one ammonia from LegAmAc. The fragmentation route proposed by Klein *et al.*, which involves the formation of a ring between the C4 and C8 of the NulO moiety, is consistent with the observed peaks (see Figure 6). Indeed, the ring formation corresponds to the loss of a water molecule gives a compound at m/z= 433 for $PseAc_2$ (432 for LegAmAc), where a small

peak is observed for *Vibrio* B9-25K2. This route leads, when the ring substituents are removed in C5-C7, to a compound with m/z= 297 which loses the ring to form the compound at m/z = 229. The compound from *Vibrio* B9-25K2 most likely forms another ring, due to a different proximity of the groups either because it is an epimer of Pse5Ac7Ac, and/or because the N-acetyl groups are carried by other positions.

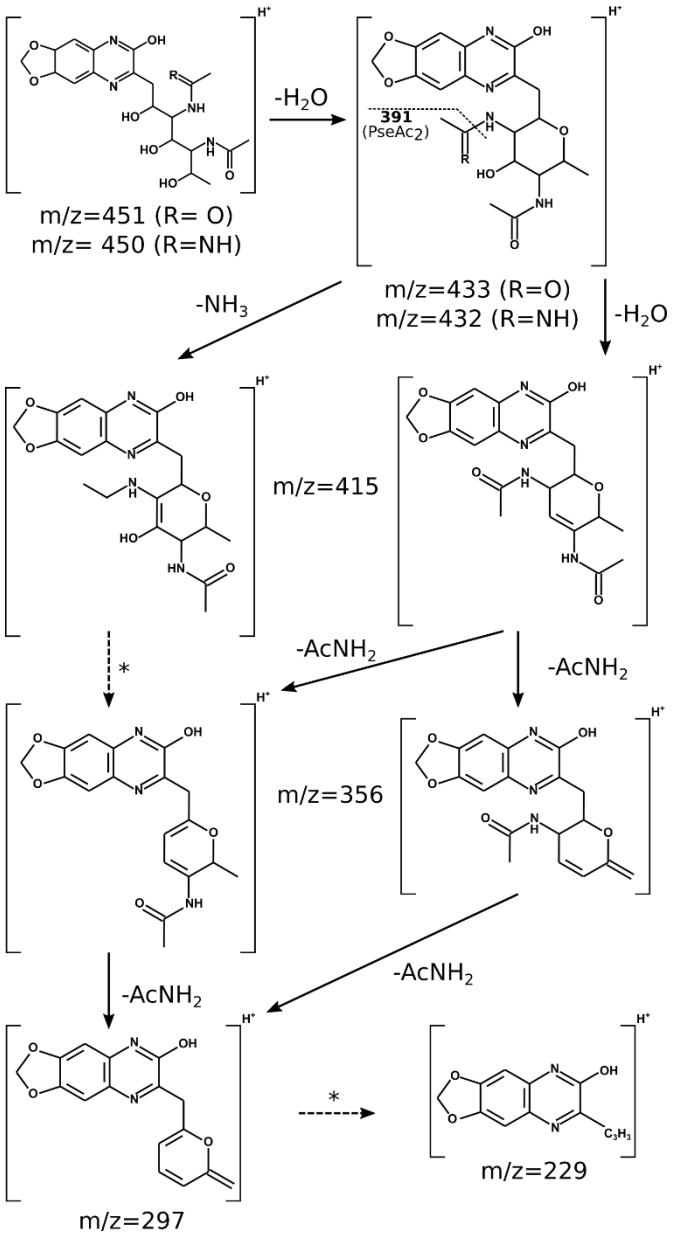


Figure 6. Proposed fragmentation route of LegAmAc (A. salmonicida) and PseAc₂ (M. viscosa). *The major peaks of the MS2 spectra (ITMS + cESI with full MS2) for both compounds were assigned to fragments of corresponding mass, according to the fragmentation route of Neu quinoxalines described by Klein et al. [74]. Multiple steps are marked with dashed arrows and asterisks (*).*

Conclusion

This study investigated the NulO content from a set of bacteria from the *Vibrionaceae* family, with an emphasis on *A. salmonicida* and *M. viscosa*. We showed that by comparing the sequences for their NeuB homologs, new *nab* clusters could be identified with non-identical NeuB sequences. We also determined substrate specificity regions, in particular a critical position occupied by either a glycine (Neu and Leg pathways) or a phenylalanine (Pse pathway) which discriminates between NAc orientation at the C2 position of NeuB substrates. Further study of NeuB specificity not only opens for a better prediction of its substrates and products, but also the tailoring of already characterized enzymes for the production of various NulOs.

Going further, the study of the *nab* clusters associated with NeuB sequences of interest allows for better hypotheses concerning NulO structure as well as the detection of new NAB proteins previously unidentified, such as the non-NeuA homolog cytidylyltransferase from *Vibrio* B9-25K2. Mass spectrometry analysis of the NulO content of the target organisms both confirmed the presence of predicted sugars, but also revealed a new compound in *A. salmonicida* R8-63 for which the corresponding quinoxaline has a *m/z* of 530, which awaits identification.

Taken together, the results confirm that the methodology used in this paper is a robust approach for screening genomic data for NulO producing organisms and the identification of new protein and/or monosaccharide targets. In the case of *Vibrio* B9-25K2, which seems to produce a variant of $PseAc_2$ with a different configuration and which cluster presents several unknown coding sequences, further studies might uncover a new branch of the Pse pathway.

Acknowledgments

The authors would like to thank Espen Robertsen and Terje Klemetsen for their help in retrieving the target genome sequences, as well as Jostein Johansen for his help with the LC-MS experiments.

Author Contributions

B. Altermark proposed the targets. M.-J.H. Halsør obtained the sequences, compared them, analysed the gene clusters, predicted the sugar structures, performed homology modelling and mass spectrometry analyses. I.L.U. Ræder and M.-J.H. Halsør performed the mild acid hydrolysis and derivatization of the samples. B. Altermark, I.L.U. Ræder and M.-J.H. Halsør designed the study and wrote the manuscript. All authors read and approved the final manuscript.

References

1. **Barry, G.T.** (1958) Colominic acid, a polymer of *N*-acetylneuraminic acid. *Journal of Experimental Medecine*. **107**(4): 507-21.
2. **Knirel, Y.A., E.V. Vinogradov, L. L'Vov V, N.A. Kocharova, A.S. Shashkov, B.A. Dmitriev, and N.K. Kochetkov** (1984) Sialic acids of a new type from the lipopolysaccharides of *Pseudomonas aeruginosa* and *Shigella boydii*. *Carbohydrate Research*. **133**(2): C5-8.
3. **Knirel, Y.A., E.T. Rietschel, R. Marre, and U. Zahringer** (1994) The structure of the O-specific chain of *Legionella pneumophila* serogroup 1 lipopolysaccharide. *European Journal of Biochemistry*. **221**(1): 239-45.
4. **Castric, P., F.J. Cassels, and R.W. Carlson** (2001) Structural characterization of the *Pseudomonas aeruginosa* 1244 pilin glycan. *Journal of Biological Chemistry*. **276**(28): 26479-85.
5. **Thibault, P., S.M. Logan, J.F. Kelly, J.R. Brisson, C.P. Ewing, T.J. Trust, and P. Guerry** (2001) Identification of the carbohydrate moieties and glycosylation motifs in *Campylobacter jejuni* flagellin. *Journal of Biological Chemistry*. **276**(37): 34862-70.
6. **Logan, S.M., J.F. Kelly, P. Thibault, C.P. Ewing, and P. Guerry** (2002) Structural heterogeneity of carbohydrate modifications affects serospecificity of *Campylobacter* flagellins. *Molecular Microbiology*. **46**(2): 587-97.
7. **Blix, G.** (1936) Über die Kohlenhydratgruppen des Submaxillrismucins [Concerning the carbohydrate groups of submaxillary mucin]. *Hoppe-Seyler's Zietschrift für Physiologische Chemie*. **240**(1-2): 43-54.
8. **Klenk, E.** (1941) Neuraminsäure, das Spaltprodukt eines neuen Gehirnlipoids [Neuraminic acid, the cleavage product of a new brain lipoid]. *Hoppe-Seyler's Zietschrift für Physiologische Chemie*. **268**: 50-58.
9. **Vinogradov, E., F. St Michael, and A.D. Cox** (2017) The structure of the LPS O-chain of *Fusobacterium nucleatum* strain 25586 containing two novel monosaccharides, 2-acetamido-2,6-dideoxy-l-altrose and a 5-acetimidoylamino-3,5,9-trideoxy-gluco-non-2-ulosonic acid. *Carbohydrate Research*. **440-441**: 10-15.
10. **Knirel, Y.A., J.H. Helbig, and U. Zahringer** (1996) Structure of a decasaccharide isolated by mild acid degradation and dephosphorylation of the lipopolysaccharide of *Pseudomonas fluorescens* strain ATCC 49271. *Carbohydrate Research*. **283**: 129-39.
11. **Tsvetkov, Y.E., A.S. Shashkov, Y.A. Knirel, and U. Zahringer** (2001) Synthesis and identification in bacterial lipopolysaccharides of 5,7-diacetamido-3,5,7,9-tetradeoxy-D-glycero-D-galacto- and -D-glycero-D-talo-non-2-ulosonic acids. *Carbohydrate Research*. **331**(3): 233-7.
12. **Knirel, Y.A., E.V. Vinogradov, A.S. Shashkov, B.A. Dmitriev, N.K. Kochetkov, E.S. Stanislavsky, and G.M. Mashilova** (1987) Somatic antigens of Pseudomonas aeruginosa. The structure of the O-specific polysaccharide chain of the lipopolysaccharide from P. aeruginosa O13 (Lanyi). *European Journal of Biochemistry*. **163**(3): 627-37.
13. **Knirel, Y.A., H. Moll, J.H. Helbig, and U. Zahringer** (1997) Chemical characterization of a new 5,7-diamino-3,5,7,9-tetradeoxynonulosonic acid released by mild acid hydrolysis of the Legionella pneumophila serogroup 1 lipopolysaccharide. *Carbohydrate Research*. **304**(1): 77-9.
14. **Kenyon, J.J., A.M. Marzaioli, C. De Castro, and R.M. Hall** (2015) 5,7-di-N-acetyl-acinetaminic acid: A novel non-2-ulosonic acid found in the capsule of an Acinetobacter baumannii isolate. *Glycobiology*. **25**(6): 644-54.
15. **Kenyon, J.J., A. Notaro, L.Y. Hsu, C. De Castro, and R.M. Hall** (2017) 5,7-Di-N-acetyl-8-epiacinetaminic acid: A new non-2-ulosonic acid found in the K73 capsule produced by an Acinetobacter baumannii isolate from Singapore. *Scientific Reports*. **7**(1): 11357.
16. **Vinogradov, E., F. St Michael, and A.D. Cox** (2018) Structure of the LPS O-chain from Fusobacterium nucleatum strain ATCC 23726 containing a novel 5,7-diamino-3,5,7,9-tetradeoxy-L-gluco-non-2-ulosonic acid presumably having the D-glycero-L-gluco configuration. *Carbohydrate Research*. **468**: 69-72.

17. **Stephenson, H.N., D.C. Mills, H. Jones, E. Milioris, A. Copland, N. Dorrell, B.W. Wren, P.R. Crocker, D. Escors, and M. Bajaj-Elliott** (2014) Pseudaminic acid on Campylobacter jejuni flagella modulates dendritic cell IL-10 expression via Siglec-10 receptor: a novel flagellin-host interaction. *The Journal of Infectious Diseases*. **210**(9): 1487-98.
18. **Khatua, B., S. Roy, and C. Mandal** (2013) Sialic acids siglec interaction: a unique strategy to circumvent innate immune response by pathogens. *Indian Journal of Medical Research*. **138**(5): 648-62.
19. **Ud-Din, A.I.M.S. and A. Roujeinikova** (2018) Flagellin glycosylation with pseudaminic acid in Campylobacter and Helicobacter: prospects for development of novel therapeutics. *Cellular and Molecular Life Sciences*. **75**(7): 1163-1178.
20. **Zebian, N., A. Merkx-Jacques, P.P. Pittock, S. Houle, C.M. Dozois, G.A. Lajoie, and C. Creuzenet** (2016) Comprehensive analysis of flagellin glycosylation in Campylobacter jejuni NCTC 11168 reveals incorporation of legionaminic acid and its importance for host colonization. *Glycobiology*. **26**(4): 386-97.
21. **Merino, S., E. Aquilini, K.M. Fulton, S.M. Twine, and J.M. Tomas** (2015) The polar and lateral flagella from Plesiomonas shigelloides are glycosylated with legionaminic acid. *Frontiers in Microbiology*. **6**: 649.
22. **Gulati, S., I.C. Schoenhofen, D.M. Whitfield, A.D. Cox, J. Li, F. St. Michael, E.V. Vinogradov, J. Stupak, B. Zheng, M. Ohnishi, M. Unemo, L.A. Lewis, R.E. Taylor, C.S. Landig, S. Diaz, G.W. Reed, A. Varki, P.A. Rice, and S. Ram** (2015) Utilizing CMP-Sialic Acid Analogs to Unravel Neisseria gonorrhoeae Lipooligosaccharide-Mediated Complement Resistance and Design Novel Therapeutics. *PLoS Pathogens*. **11**(12): e1005290.
23. **Thomas, G.H.** (2016) Sialic acid acquisition in bacteria-one substrate, many transporters. *Biochemical Society Transactions*. **44**(3): 760-5.
24. **Silver, R.P., W.F. Vann, and W. Aaronson** (1984) Genetic and molecular analyses of Escherichia coli K1 antigen genes. *Journal of Bacteriology*. **157**(2): 568-75.
25. **Vimr, E.R., W. Aaronson, and R.P. Silver** (1989) Genetic analysis of chromosomal mutations in the polysialic acid gene cluster of Escherichia coli K1. *Journal of Bacteriology*. **171**(2): 1106-17.
26. **Linton, D., A.V. Karlyshev, P.G. Hitchen, H.R. Morris, A. Dell, N.A. Gregson, and B.W. Wren** (2000) Multiple N-acetyl neuraminic acid synthetase (neuB) genes in Campylobacter jejuni: identification and characterization of the gene involved in sialylation of lipo-oligosaccharide. *Mol Microbiol*. **35**(5): 1120-34.
27. **Lewis, A.L., N. Desa, E.E. Hansen, Y.A. Knirel, J.I. Gordon, P. Gagneux, V. Nizet, and A. Varki** (2009) Innovations in host and microbial sialic acid biosynthesis revealed by phylogenomic prediction of nonulosonic acid structure. *Proceedings of the National Academy of Sciences of the United States of America*. **106**(32): 13552-7.
28. **Lewis, A.L., J.B. Lubin, S. Argade, N. Naidu, B. Choudhury, and E.F. Boyd** (2011) Genomic and metabolic profiling of nonulosonic acids in *Vibrionaceae* reveal biochemical phenotypes of allelic divergence in *Vibrio vulnificus*. *Applied and Environmental Microbiology*. **77**(16): 5782-93.
29. **Egidius, E., R. Wiik, K. Andersen, K.A. Hoff, and B. Hjeltnes** (1986) *Vibrio salmonicida* sp. nov., a New Fish Pathogen. *International Journal of Systematic Bacteriology*. **36**(4): 518-520.
30. **Urbanczyk, H., J.C. Ast, M.J. Higgins, J. Carson, and P.V. Dunlap** (2007) Reclassification of *Vibrio fischeri, Vibrio logei, Vibrio salmonicida* and *Vibrio wodanis* as *Aliivibrio fischeri* gen. nov., comb. nov., *Aliivibrio logei* comb. nov., *Aliivibrio salmonicida* comb. nov. and *Aliivibrio wodanis* comb. nov. *International Journal of Systematic and Evolutionary Microbiology*. **57**(Pt 12): 2823-9.
31. **Kashulin, A., N. Seredkina, and H. Sorum** (2017) Cold-water vibriosis. The current status of knowledge. *Journal of Fish Diseases*. **40**(1): 119-126.
32. **Gurung, M.K., I.L. Raeder, B. Altermark, and A.O. Smalas** (2013) Characterization of the sialic acid synthase from *Aliivibrio salmonicida* suggests a novel pathway for bacterial synthesis of 7-*O*-acetylated sialic acids. *Glycobiology*. **23**(7): 806-19.
33. **Bøgwald, J. and J. Hoffman** (2006) Structural studies of the O-antigenic oligosaccharide from *Vibrio salmonicida* strain C2 isolated from Atlantic cod, *Gadus morhua* L. *Carbohydrate Research*. **341**(11): 1965-8.

34. **Edebrink, P., P.E. Jansson, J. Bogwald, and J. Hoffman** (1996) Structural studies of the *Vibrio salmonicida* lipopolysaccharide. *Carbohydrate Research*. **287**(2): 225-45.
35. **Benediktsdottir, E., L. Verdonck, C. Sproer, S. Helgason, and J. Swings** (2000) Characterization of *Vibrio viscosus* and *Vibrio wodanis* isolated at different geographical locations: a proposal for reclassification of *Vibrio viscosus* as *Moritella viscosa* comb. nov. *International Journal of Systematic and Evolutionary Microbiology*. **50 Pt 2**: 479-88.
36. **Lunder, T., H. Sorum, G. Holstad, A.G. Steigerwalt, P. Mowinckel, and D.J. Brenner** (2000) Phenotypic and genotypic characterization of *Vibrio viscosus* sp. nov. and *Vibrio wodanis* sp. nov. isolated from Atlantic salmon (*Salmo salar*) with 'winter ulcer'. *International Journal of Systematic and Evolutionary Microbiology*. **50 Pt 2**: 427-50.
37. **Lunder, T., Ø. Evensen, G. Holstad, and T. Håstein** (1995) 'Winter ulcer' in the Atlantic salmon *Salmo salar*. Pathological and bacteriological investigations and transmission experiments. *Diseases of Aquatic Organisms*. **23**: 39-49.
38. **Berg, T.O., M.K. Gurung, B. Altermark, A.O. Smalas, and I.L. Raeder** (2015) Characterization of the *N*-acetylneuraminic acid synthase (NeuB) from the psychrophilic fish pathogen *Moritella viscosa. Carbohydrate Research*. **402**: 133-45.
39. **McDonald, N.D., K.E. DeMeester, A.L. Lewis, C.L. Grimes, and E.F. Boyd** (2018) Structural and functional characterization of a modified legionaminic acid involved in glycosylation of a bacterial lipopolysaccharide. *Journal of Biological Chemistry*.
40. **Hjerde, E., C. Karlsen, H. Sorum, J. Parkhill, N.P. Willassen, and N.R. Thomson** (2015) Co-cultivation and transcriptome sequencing of two co-existing fish pathogens *Moritella viscosa* and *Aliivibrio wodanis. BMC Genomics*. **16**: 447.
41. **Hjerde, E., M.S. Lorentzen, M.T. Holden, K. Seeger, S. Paulsen, N. Bason, C. Churcher, D. Harris, H. Norbertczak, M.A. Quail, S. Sanders, S. Thurston, J. Parkhill, N.P. Willassen, and N.R. Thomson** (2008) The genome sequence of the fish pathogen *Aliivibrio salmonicida* strain LFI1238 shows extensive evidence of gene decay. *BMC Genomics*. **9**: 616.
42. **Goll, J., D.B. Rusch, D.M. Tanenbaum, M. Thiagarajan, K. Li, B.A. Methe, and S. Yooseph** (2010) METAREP: JCVI metagenomics reports--an open source tool for high-performance comparative metagenomics. *Bioinformatics*. **26**(20): 2631-2.
43. **Purohit, A.A., J.A. Johansen, H. Hansen, H.K.S. Leiros, A. Kashulin, C. Karlsen, A. Smalås, P. Haugen, and N.P. Willassen** (2013) Presence of acyl-homoserine lactones in 57 members of the *Vibrionaceae* family. *Journal of Applied Microbiology*. **115**(3): 835-847.
44. **Klemetsen, T.** (2016) *An Insight into the Aliivibrio genus*. PhD dissertation, University of Tromsø, The Arctic University of Norway: Tromsø.
45. **Karlsen, C., E. Hjerde, T. Klemetsen, and N.P. Willassen** (2017) Pan genome and CRISPR analyses of the bacterial fish pathogen *Moritella viscosa*. *BMC Genomics*. **18**(1): 313.
46. **Holm, K.O., K. Nilsson, E. Hjerde, N.P. Willassen, and D.L. Milton** (2015) Complete genome sequence of *Vibrio anguillarum* strain NB10, a virulent isolate from the Gulf of Bothnia. *Standards in Genomic Sciences*. **10**: 60.
47. **Purohit, A.A., J.A. Johansen, H. Hansen, H.K. Leiros, A. Kashulin, C. Karlsen, A. Smalas, P. Haugen, and N.P. Willassen** (2013) Presence of acyl-homoserine lactones in 57 members of the *Vibrionaceae* family. *Journal of Applied Microbiology*. **115**(3): 835-47.
48. **Eiler, A. and S. Bertilsson** (2006) Detection and quantification of *Vibrio* populations using denaturant gradient gel electrophoresis. *Journal of Microbiological Methods*. **67**(2): 339-48.
49. **Rehnstam, A.S., A. Norqvist, H. Wolf-Watz, and A. Hagstrom** (1989) Identification of *Vibrio anguillarum* in fish by using partial 16S rRNA sequences and a specific 16S rRNA oligonucleotide probe. *Applied and Environmental Microbiology*. **55**(8): 1907-10.
50. **Kahlke, T.** (2013) *Analysis of the Vibrionaceae pan-genome*. PhD dissertation, University of Tromsø, the Arctic University of Norway: Tromsø.
51. **Robertsen, E.M., T. Kahlke, I.A. Raknes, E. Pedersen, E.K. Semb, M. Ernsten, L.A. Bongo, and N.P. Willassen**, *META-Pipe - Pipeline annotation, analysis and visualization of marine metagenomic sequence data*. 2016, arXiv.
52. **Delcher, A.L., K.A. Bratke, E.C. Powers, and S.L. Salzberg** (2007) Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics*. **23**(6): 673-9.

53. **Altschul, S.F., T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman** (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*. **25**(17): 3389-402.
54. **Edgar, R.C.** (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*. **32**(5): 1792-7.
55. **Edgar, R.C.** (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*. **5**: 113.
56. **Carver, T., S.R. Harris, M. Berriman, J. Parkhill, and J.A. McQuillan** (2012) Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics*. **28**(4): 464-9.
57. **Marchler-Bauer, A., Y. Bo, L. Han, J. He, C.J. Lanczycki, S. Lu, F. Chitsaz, M.K. Derbyshire, R.C. Geer, N.R. Gonzales, M. Gwadz, D.I. Hurwitz, F. Lu, G.H. Marchler, J.S. Song, N. Thanki, Z. Wang, R.A. Yamashita, D. Zhang, C. Zheng, L.Y. Geer, and S.H. Bryant** (2017) CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Research*. **45**(D1): D200-D203.
58. **Gunawan, J., D. Simard, M. Gilbert, A.L. Lovering, W.W. Wakarchuk, M.E. Tanner, and N.C.J. Strynadka** (2005) Structural and Mechanistic Analysis of Sialic Acid Synthase NeuB from *Neisseria meningitidis* in Complex with $Mn2^{+}$, Phosphoenolpyruvate, and *N*-Acetylmannosaminitol.*Journal of Biological chemistry*. **280**(5): 3555-3563.
59. **Gundogdu, O., S.D. Bentley, M.T. Holden, J. Parkhill, N. Dorrell, and B.W. Wren** (2007) Re-annotation and re-analysis of the *Campylobacter jejuni* NCTC11168 genome sequence. *BMC Genomics*. **8**(1): 162.
60. **Waterhouse, A., M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F.T. Heer, T.A P. de Beer, C. Rempfer, L. Bordoli, R. Lepore, and T. Schwede** (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*. **46**(W1): W296-W303.
61. **Nakamura, M., S. Hara, M. Yamaguchi, Y. Takemori, and Y. Ohkura** (1987) 1, 2-Diamino-4, 5-methylenedioxybenzene as a Highly Sensitive Fluorogenic Reagent for α-Keto Acids. *Chemical & Pharmaceutical Bulletin*. **35**(2): 687-692.
62. **Klein, A., S. Diaz, I. Ferreira, G. Lamblin, P. Roussel, and A.E. Manzi** (1997) New sialic acids from biological sources identified by a comprehensive and sensitive approach: liquid chromatography-electrospray ionization-mass spectrometry (LC-ESI-MS) of SIA quinoxalinones. *Glycobiology*. **7**(3): 421-32.
63. **Hara, S., M. Yamaguchi, Y. Takemori, K. Furuhata, H. Ogura, and M. Nakamura** (1989) Determination of mono-*O*-acetylated *N*-acetylneuraminic acids in human and rat sera by fluorometric high-performance liquid chromatography. *Anal Biochem*. **179**(1): 162-6.
64. **Halsør, M.-J.H.** (2018) GeneSVG. *GitHub repository*.
65. **Kim, S., P.A. Thiessen, E.E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Han, J. He, S. He, B.A. Shoemaker, J. Wang, B. Yu, J. Zhang, and S.H. Bryant** (2016) PubChem Substance and Compound databases. *Nucleic Acids Research*. **44**(D1): D1202-13.
66. **Schrödinger, L.**, *The PyMOL Molecular Graphics System, Version 1.8*. 2015.
67. **The Inkscape Contributors**. *www.inkscape.org*.
68. **Linton, D., A.V. Karlyshev, P.G. Hitchen, H.R. Morris, A. Dell, N.A. Gregson, and B.W. Wren** (2000) Multiple *N*-acetyl neuraminic acid synthetase (*neuB*) genes in *Campylobacter jejuni*: identification and characterization of the gene involved in sialylation of lipo-oligosaccharide. *Molecular Microbiology*. **35**(5): 1120-1134.
69. **Logan, S.M., J.P. Hui, E. Vinogradov, A.J. Aubry, J.E. Melanson, J.F. Kelly, H. Nothaft, and E.C. Soo** (2009) Identification of novel carbohydrate modifications on *Campylobacter jejuni* 11168 flagellin using metabolomics-based approaches. *FEBS Journal*. **276**(4): 1014-23.
70. **Chou, W.K., S. Dick, W.W. Wakarchuk, and M.E. Tanner** (2005) Identification and characterization of NeuB3 from *Campylobacter jejuni* as a pseudaminic acid synthase. *Journal of Biological Chemistry*. **280**(43): 35922-8.
71. **Schoenhofen, I.C., E. Vinogradov, D.M. Whitfield, J.R. Brisson, and S.M. Logan** (2009) The CMP-legionaminic acid pathway in *Campylobacter*: biosynthesis involving novel GDP-linked precursors. *Glycobiology*. **19**(7): 715-25.

72. **Sundaram, A.K., L. Pitts, K. Muhammad, J. Wu, M. Betenbaugh, R.W. Woodard, and W.F. Vann** (2004) Characterization of *N*-acetylneuraminic acid synthase isoenzyme 1 from *Campylobacter jejuni*. *The Biochemical journal*. **383**(Pt 1): 83-89.
73. **McNally, D.J., A.J. Aubry, J.P. Hui, N.H. Khieu, D. Whitfield, C.P. Ewing, P. Guerry, J.R. Brisson, S.M. Logan, and E.C. Soo** (2007) Targeted metabolomics analysis of *Campylobacter coli* VC167 reveals legionaminic acid derivatives as novel flagellar glycans. *Journal of Biological Chemistry*. **282**(19): 14463-75.
74. **Klein, A., S. Diaz, I. Ferreira, G. Lamblin, P. Roussel, and A.E. Manzi** (1997) New sialic acids from biological sources identified by a comprehensive and sensitive approach: liquid chromatography-electrospray ionization-mass spectrometry (LC-ESI-MS) of SIA quinoxalinones. *Glycobiology*. **7**(3): 421-432.

Paper II

Draft Genome of the Self-Cleaning Cyanobacterium *Nostoc* sp. KVJ20

Marie-Josée H. Halsør[1], Anton Liaimer[2], Seila Pandur[1], Inger L.U. Ræder[1], Arne O. Smalås[1] and Bjørn Altermark[1,*]

[1]NorStruct, Department of Chemistry, Faculty of Science and Technology, UiT The Arctic University of Norway.

[2]Department of Arctic and Marine Biology, Faculty of Biosciences, Fisheries and Economics, UiT The Arctic University of Norway.

*Corresponding author: E-mail: bjorn.altermark@uit.no

Data deposition: This project has been deposited at Genbank under accession number ERP003528.

Abstract

Nostoc sp. strain KVJ20 was isolated near Kvaløya, in Northern Norway. This cyanobacterium has a broad symbiotic competence and extracts from the isolate has inhibitory effects on cancer cell lines and microbes. The genome further reveal many genes encoding the production of secondary metabolites, glycans and biotechnologically relevant enzymes. The strain is highly interesting both for fundamental and applied science. The genome is in total 9,183,495 bp (425 contigs/332 scaffolds) and contains 7,676 genes (7,210 protein coding & 104 RNA genes). The GC percentage is 41.69%. Comparing operons containing non-ribosomal peptide synthases and polyketide synthases (NRPS and PKS) from *Nostoc* sp. KVJ20 with those of *Nostoc punctiforme* PCC73102 reveal differences, which may shed more light on their involvement in symbiotic competence, cell differentiation and the self-cleaning observed when grown in vitro.

Key words: secondary metabolites, NRPS, PKS, whole genome sequencing, bioprospecting, gene clusters.

Introduction

Nostoc sp. strain KVJ20 was isolated from the symbiotic organs of the liverwort *Blasia pusilla* L. found as a weed in a plant-school on the island of Kvaløya, Northern Norway. During the isolation process, we noted that the strain was unusually clean from the first steps of cultivation, and did not require additional treatments for bringing it into an axenic state. This was an indication that the organism may produce antibiotic compounds serving the function of self-cleaning. We had a similar experience with genetically similar *Nostoc* sp SKSF3 originated from soil. Metabolic profiling by MALDi-TOF showed similarity between the secondary metabolites sets produced by these two isolates (Liaimer, et al. 2016). In addition, the cell extracts of *Nostoc* sp. KVJ20 showed inhibitory effects on A2068 metastatic human melanoma cell line and MRC5 fibroblasts (Liaimer, et al. 2016). The insight into metabolic capacities of the strain, based on full genome sequencing, may therefore help in explaining antimicrobial and anticancer properties of this cyanobacterium.

In addition to its ability to exist in symbiosis with its natural liverwort host, *Nostoc* sp. KVJ20 successfully infected seedlings of angiosperm *Gunnera manicata* Linden, where it was housed intracellularly. The host plant, *G.manicata* Linden is a native of South America, and has no history of being grown above the Arctic Circle. Up to date, most of the genomes available in the databases for genus *Nostoc* represent free-living isolates. The only strain with equally broad symbiotic competence, whose genome is available, is *Nostoc punctiforme* PCC73102 (Ekman, et al. 2013). The traits responsible for symbiotic interactions between nitrogen fixing cyanobacteria and plants remain unknown. Thus, the genome of *Nostoc* sp. KVJ20 will add valuable information on the core features underlying symbiotic capacity in the genus. The observations on the bioactivities of *Nostoc* sp. KVJ20 and the broad symbiotic competence were therefore major reasons for choosing the strain for full genome sequencing.

The *Nostoc* genus represents a group of filamentous, nitrogen fixing, photosynthetic cyanobacteria. They are able to produce the four different cell types known to the phylum, and can be found as either independent colonies or endosymbionts in some plants. The formation and composition of filaments, as well as the symbiotic association, are highly dependent on cell differentiation, and several genes have been identified to play a significant role in those processes (Chapman, et al. 2008; Flores 2012; Meeks and Elhai 2002). Such genes can be identified by their ability to code for secondary metabolites specific to some cyanobacteria genera. One secondary metabolite, the ribosomal peptide nostopeptolide, was recently found to have a crucial role in the

differentiation of the hormogonium cell type in *Nostoc punctiforme* (Liaimer et al 2015). It is synthetized by the proteins of the nos gene cluster, which codes for a series of non-ribosomal peptide and polyketide synthases (NRPS and PKS). Similar clusters encode other secondary metabolites. The identification of genes coding for NRPSs and PKSs within the *Nostoc* sp. KVJ20 genome is one of our interests. Additionally, Cyanobacteria produce a variety of glycans and glycan modifying enzymes (Kehr and Dittmann 2015). These glycans can also have anti-cancer activity (Guo, et al. 2015) and the glycan modifying enzymes are of biotechnological interest.

Materials and Methods

Strain Isolation and DNA purification

The isolation of *Nostoc* sp. KVJ20 from the symbiotic organs of the liverwort *Blasia pusilla* was carried out as described earlier (West and Adams 1997). The strain is maintained at Department of Arctic and Marine Biology, Faculty of Biosciences, Fisheries and Economics, UiT The Arctic University of Norway. Genomic DNA from cultured *Nostoc* sp. KVJ20 were purified according to a published protocol with the modifications suggested by the group of Prof. John C. Meeks, University of California (http://microbiology.ucdavis.edu/meeks/xpro5.htm), (supplementary text 1).

Genome Sequencing, Assembly and Annotation

A Qubit Fluorometer (Thermo Fisher) together with the Qubit dsDNA High sensitivity kit was used for accurate DNA concentration measurements. For the library, Nextera DNA Library Prep kit was used. A Bioanalyzer (Agilent Technologies) was used together with Agilent High sensitivity DNA kit for fragment size assessment. The library was sequenced on a MiSeq machine (Illumina) at UiT The Artic University of Norway, using MiSeq reagents kit V3, 2 x 300 bp. Using CLC Genomics Workbench (QIAGEN), the reads were trimmed to remove remaining sequencing adaptors and indexes. Also, the 10 first nucleotides and the 20 last of each read were masked. All reads with more than 20 undetermined bases were discarded. Sequence areas with lower q-score than 20 were removed. 9,259,128 reads with an average read length of 208 bases were thereafter assembled. After manual inspection, eight low coverage contigs belonging to *Methylobacterium* sp. were discarded. The NCBI Prokaryotic Genome Annotation Pipeline was used for annotation, with GeneMarkS+ for gene identification (Borodovsky and Lomsadze 2014; Tatusova, et al. 2016). Genes, CDSs, tRNAs,

ncRNAs and repeat regions were annotated. The Whole Genome Shotgun (WGS) sequence is deposited in GenBank under accession number ERP003528.

Phylogenetic Tree, Secondary Metabolites and Gene Order

A phylogenetic tree based on aligned 16s rRNA genes, aligned using MUSCLE (Edgar 2004a, b) was used as input to the web-server "Phylogeny.fr" (Dereeper, et al. 2010; Dereeper, et al. 2008). A total of 1402 positions, not including gaps, was compared and used to reconstruct and analyze the phylogenetic relationships. By using the antiSMASH database and analyses tool (Weber, et al. 2015), complemented by additional Blast analyses (Altschul, et al. 1990), gene-clusters/operons coding genes involved in biosynthesis of NRPS, PKS, hybrid NRPS-PKS products as well as ribosomally synthesized and post-translationally modified peptides (RiPP) were predicted. To aid in visualization, a Python-script for creating vector-images of gene clusters was written. The script can be found at github (https://github.com/mariejhh/python-bioinformatics/wiki/GeneSVG).

Results and Discussion

Strain Phylogeny and Characteristics

The strain was remarkable clean after the first culturing and no special procedures had to be conducted in order to get the strain axenic. This is in strong contrast to what is expected when isolating cyanobacteria as bacterial contaminants are hard to avoid (Waterbury 2006).

Photos of *Nostoc* sp. KVJ20 at different life stages and the placement of *Nostoc* sp. KVJ20 relative to other strains within the *Nostocaceae* family is presented in Figure 1. By comparing the gene encoding 16s rRNA of several cyanobacteria with that of *Nostoc* sp. KVJ20 we found that the closest relative with a published complete genome is *Nostoc punctiforme* PCC 73102.

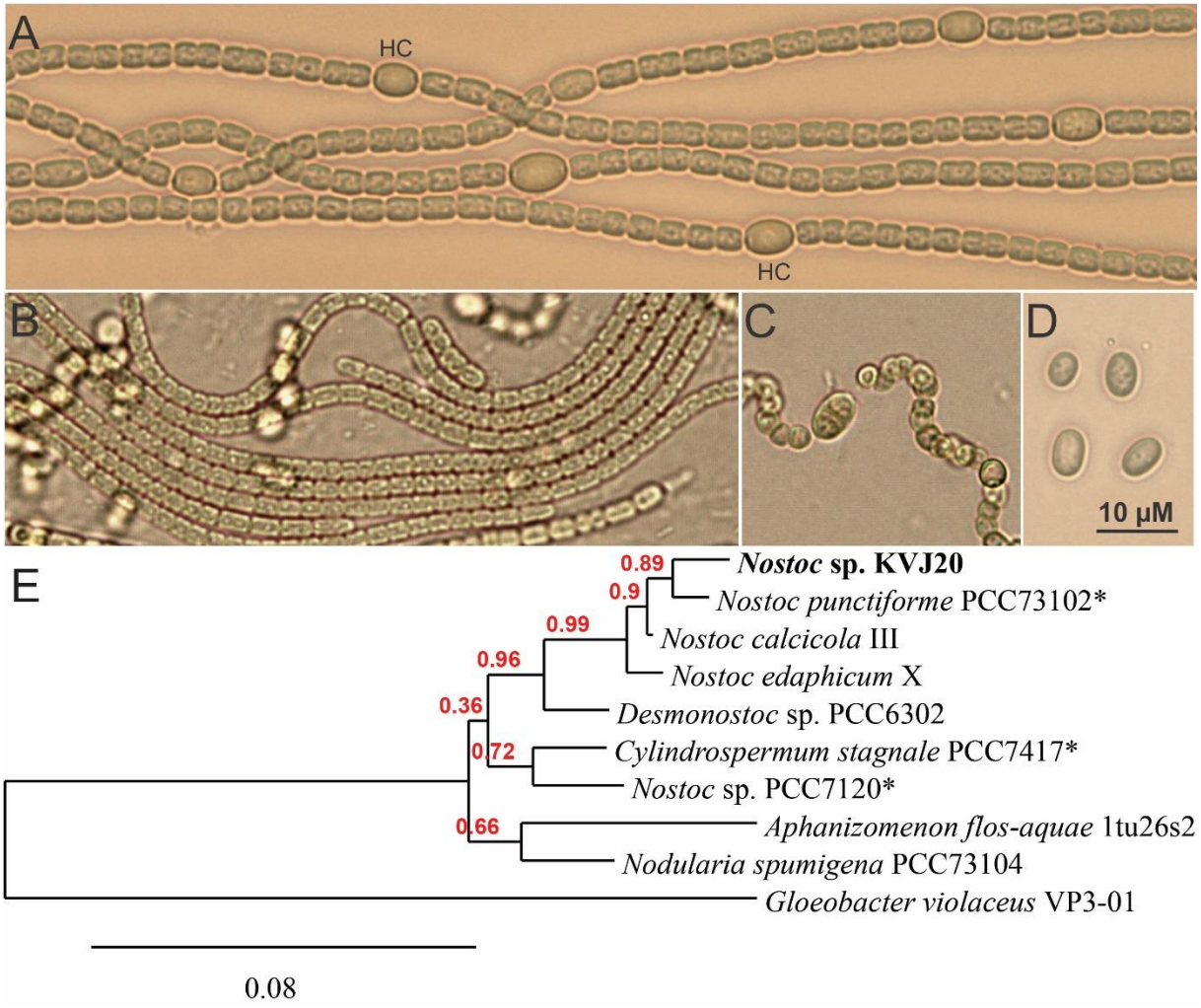


Figure 1. Nostoc sp. KVJ20 at different life stages. *(A) diazotrophically grown Nostoc sp. KVJ20 at exponential stage, with vegetative cells arranged in multicellular filaments interspersed by nitrogen-fixing cells- heterocysts. (B) motile filaments, hormogonia. (C) a filament differentiating resting cells, akinetes, and (D) mature resting cells, akinetes. The bar is 10 µM. (E) Phylogenetic tree highlighting the position of Nostoc sp. KVJ20 (in bold) relative to other strains within the Nostocaceae family. Strains names marked with an asterix (*) have been genome sequenced. The strains and their corresponding GenBank accession numbers for 16S rRNA genes are: Nostoc sp. KVJ20, LSSA00000000; Nostoc punctiforme PCC73102, AF027655.1; Nostoc calcicola III, AJ630447.1; Nostoc edaphicum X, AJ630449.1; Desmonostoc sp. PCC6302, HG004582.1; Cylindrospermum stagnale PCC7417, AJ133163.1; Nostoc sp. PCC7120, BA000019:2375734-2377222; Aphanizomenon flos-aquae 1tu26s2, AJ630443.1; Nodularia spumigena PCC73104, DQ185241.1. Confidence index values are shown in red and the bar represents the amount of nucleotide substitutions per site. Gloeobacter violaceus (FR798924.1) was used as an outgroup*

Genome Properties

The draft genome is composed of 425 contigs (332 scaffolds), giving a total size of 9,183,495 bp. After annotation, 4,443 genes were assigned to functional COG categories (Tatusov, et al. 2000). The properties and the statistics of the genome are summarized in Table 1.

Table 1. Properties and statistics of the Nostoc sp. KVJ20 genome.

Attribute	Value	% of Total
Genome size (bp)	9,183,495	100
DNA coding (bp)	7,004,057	76.27
DNA G+C (bp)	3,828,869	41.69
DNA scaffolds	332	100
Total genes	7,676	100
Protein coding genes	7,210	93.93
RNA genes	104	1.35
Pseudo genes	362	4.72
Genes assigned to COGs	4,443	57.88
CRISPR repeats	12	-
Coverage (x)	200	-

The fragmented nature of the genome suggests a high number of repeat sequences and transposable elements, which would make the assembly difficult. From the initial annotation, 70 sequences were found to code for transposases, 22 for integrases, and 12 CRISPR repeat regions were assigned. The presence of tandem repeats was investigated in each contig using Tandem Repeats Finder version 4.09 (Benson 1999). A total of 2,869 repeats were found in 298 contigs. In comparison, the genome of *Nostoc punctiforme* PCC 73102 contain 1,476 tandem repeats.

Secondary Metabolite Gene Clusters

Secondary metabolites are coded by NRPSs and PKSs. Originally 24 genes were annotated as coding for a NRPS or PKS, located on 10 different contigs. The translated sequences for these targets, as well as the coding sequences surrounding them, were subjected to a similarity search against the

protein database of NCBI. From this, gene clusters were identified (see Supplemental Table 1). The suggested product(s) of the clusters is a mixture of both known and putatively unknown novel compounds. They encode amongs other lantipeptides, microviridins, nostocyclopeptide and aeruginosin/banyaside (supplemental figure 1). Some of the secondary metabolites produced could be potent antimicrobial substances, which would explain the self-cleaning properties when culturing this strain.

The genome of the *Nostoc* sp. KVJ20 bacterium represents an interesting addition to the increasing number of sequenced cyanobacteria. Its symbiotic competence, the self-cleaning ability, the cellular differentiation mechanism and how these properties are connected to production of secondary metabolites from the identified gene clusters will be interesting to pursue further.

Supplementary Material

Supplementary data is presented after the references for this manuscript.

Acknowledgments

This work was supported by the Department of Chemistry, UiT The Arctic University of Norway and the NorZymeD project of the Research Council of Norway [grant number 221568]. Erik Hjerde (Dep. Of Chemistry, UiT The Arctic University of Norway) is greatly acknowledged for help during assembly.

Author Contributions

AL did the microbiological studies, purified genomic DNA and obtained the organism information; SP made the sequencing libraries and operated the MiSeq sequencer; MH, BA and AL assembled the sequencing data and completed the genome analysis; IR, MH, AL, AS and BA designed the study and wrote the manuscript. All authors read and approved the final manuscript.

Literature Cited

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403-410.

Benson G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acid Research* **27**: 573-580.

Borodovsky M, Lomsadze A. (2014). Gene identification in prokaryotic genomes, phages, metagenomes, and EST sequences with GeneMarkS suite. *Current Protocols in Microbiology* **32**: Unit 1E 7.

Chapman KE, Duggan PS, Billington NA, Adams DG. (2008). Mutation at different sites in the *Nostoc punctiforme cyaC* gene, encoding the multiple-domain enzyme adenylate cyclase, results in different levels of infection of the host plant *Blasia pusilla*. *Journal of Bacteriology* **190**: 1843-1847.

Dereeper A, Audic S, Claverie JM, Blanc G. (2010). BLAST-EXPLORER helps you building datasets for phylogenetic analysis. *BMC Evolutionary Biology* **10**: 8.

Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acid Research* **36**: W465-469.

Edgar RC. 2004a. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics **5**: 113.

Edgar RC. 2004b. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acid Research* **32**: 1792-1797.

Ekman M, Picossi S, Campbell EL, Meeks JC, Flores E. (2013). A *Nostoc punctiforme* sugar transporter necessary to establish a Cyanobacterium-plant symbiosis. *Plant Physiology* **161**: 1984-1992.

Flores E. (2012). Restricted cellular differentiation in cyanobacterial filaments. *Proceedings of the National Academy of Sciences of the United States* **109**: 15080-15081.

Guo M, Ding GB, Guo S, Li Z, Zhao L, Li K, Guo X. (2015). Isolation and antitumor efficacy evaluation of a polysaccharide from *Nostoc commune* Vauch. *Food & Function* **6**: 3035-3044.

Kehr JC, Dittmann E. (2015). Biosynthesis and function of extracellular glycans in cyanobacteria. *Life (Basel)* **5**: 164-180.

Liaimer A, Jensen JB, Dittmann E. (2016). A Genetic and Chemical Perspective on Symbiotic Recruitment of Cyanobacteria of the Genus *Nostoc* into the Host Plant *Blasia pusilla* L. *Frontiers in Microbiology* **7**: 1693.

Liaimer A, Helfrich EJ, Hinrichs K, Guljamow A, Ishida K, Hertweck C, Dittmann E. (2015). Nostopeptolide plays a governing role during cellular differentiation of the symbiotic cyanobacterium *Nostoc punctiforme*. *Proceedings of the National Academy of Sciences of the United States*. **112**:1862–7.

Meeks JC, Elhai J. (2002). Regulation of cellular differentiation in filamentous cyanobacteria in free-living and plant-associated symbiotic growth states. *Microbiology and Molecular Biology Reviews* **66**: 94-121.

Tatusov RL, Galperin MY, Natale DA, Koonin EV. (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acid Research* **28**: 33-36.

Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acid Research* **44**: 6614-6624.

Waterbury JB. (2006). The Cyanobacteria—Isolation, Purification and Identification. In: **Dworkin M. FS, Rosenberg E., Schleifer KH., Stackebrandt E.**, editor. *The Prokaryotes*. New York, NY: Springer. p. 1053-1073.

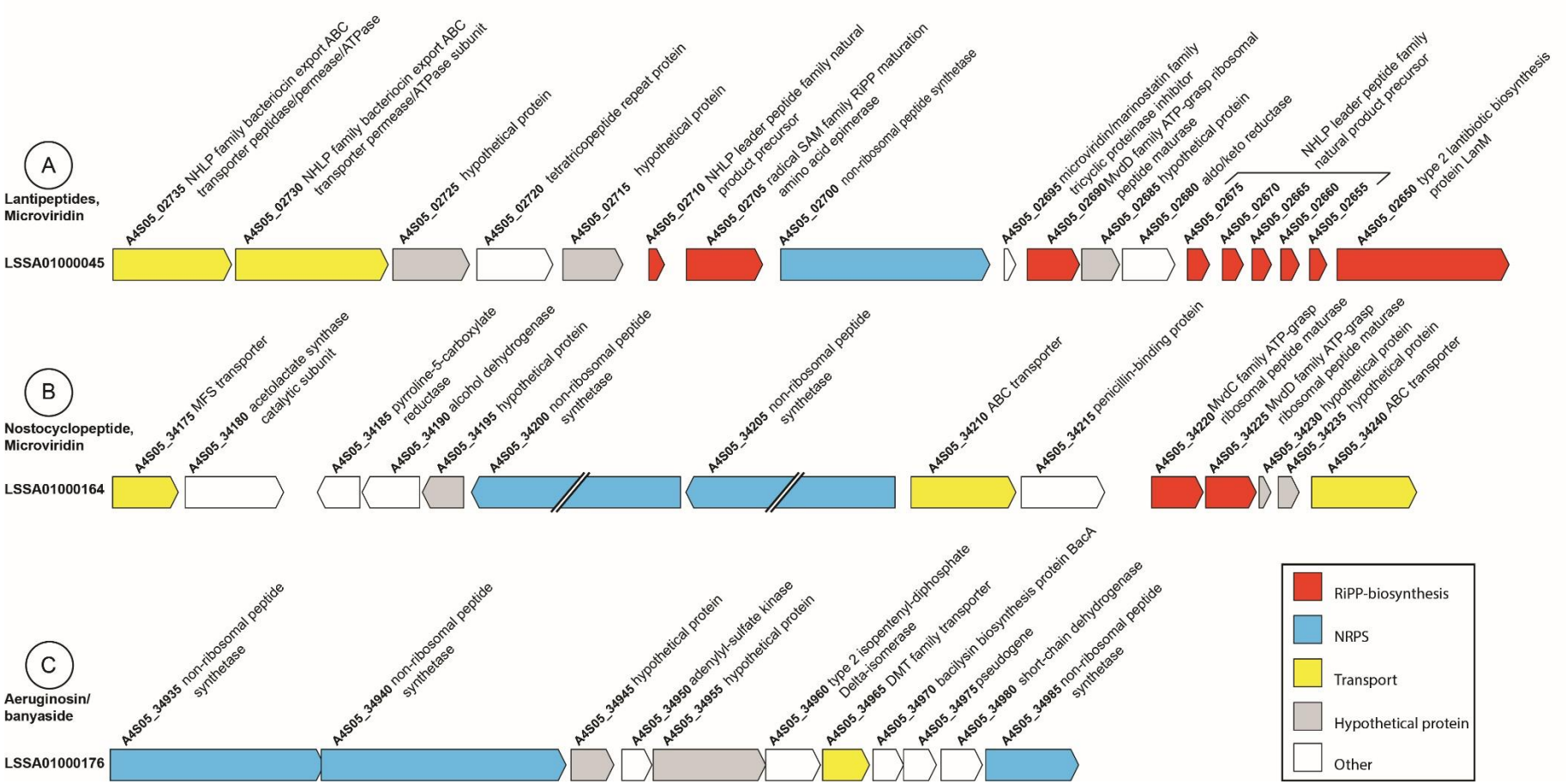
Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, Lee SY, Fischbach MA, Müller R, Wohlleben W, Breitling R, Takano E, Medema MH. (2015). antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acid Research* **43**: W237-243.

West NJ, Adams DG. (1997). Phenotypic and genotypic comparison of symbiotic and free-living cyanobacteria from a single field site. *Applied Environmental Microbiology* **63**: 4479-4484.

Supplementary Table 1. Summary of the predicted gene-clusters/operons coding genes involved in biosynthesis of non-ribosomally synthesised peptides (NRPS), polyketides (PKS), hybrid NRPS-PKS products as well as ribosomally synthesized and post-translationally modified peptides (RiPP). *Predictions are made based on data presented in (Liaimer et al. 2016). [abc]These clusters are visualized in Supplementary Figure 1.

Type of gene cluster	Contig accession number	Loci	Suggested product(s)	Homologous clusters, Organism: loci tags
NRPS-PKS	LSSA01000013.1	A4S05_21575-A4S05_21610	unknown	*N. punctiforme* PCC 73102: Npun_R6591-Npun_R6584
RiPP, NRPS[a]	LSSA01000045.1	A4S05_02650-A4S05_02735	lantipeptides, microviridin	
RiPP	LSSA01000061.1	A4S05_28035-A4S05_28095	bacteriocins	
RiPP	LSSA01000098.1	A4S05_30035-A4S05_30055	bacteriocins	
RiPP	LSSA01000102.1	A4S05_03330-A4S05_03365	bacteriocins	
NRPS, RiPP[b]	LSSA01000164.1	A4S05_34175-A4S05_34240	nostocyclopeptide* microviridin	*N. punctiforme* PCC 73102: Npun_F2181-Npun_F2193; *N*ostoc sp. ATCC 53789: AOO23328-AOO23334
PKS	LSSA01000165.1	A4S05_34290-A4S05_34320	heterocyst glycolipid	*N. punctiforme* PCC 73102: Npun_F0037-Npun_R0043
RiPP	LSSA01000173.1	A4S05_34670-A4S05_34700	bacteriocins	
NRPS[c]	LSSA01000176.1	A4S05_34935-A4S05_34985	aeruginosin/banyaside*	*Planktothrix agardhii* NIVA-CYA 126: CAM59601-CAM59606
PKS	LSSA01000184.1	A4S05_35385-A4S05_35455	heterocyst glycolipid	*N. punctiforme* PCC 73102: Npun_R1284-Npun_F1299

NRPS	LSSA01000197.1	A4S05_36095	unknown	*N. calcicola* FACHB389_22910
NRPS-PKS	LSSA01000332.1	A4S05_08280-A4S05_08315	unknown	*Scytonema millei* VB511283: QH73_014240-QH73_014315
RiPP, NRPS-PKS	LSSA01000336.1	A4S05_08535-A4S05_08660	bacteriocins. unknown	
RiPP	LSSA01000345.1	A4S05_10295-A4S05_10310	bacteriocins	
RiPP	LSSA01000352.1	A4S05_10950-A4S05_11010	bacteriocins	
RiPP	LSSA01000386.1	A4S05_14495-A4S05_14560	lantipeptides	
RiPP	LSSA01000401.1	A4S05_16960-A4S05_17015	bacteriocins	
NRPS	LSSA01000425.1	A4S05_02060-A4S05_02085	anabaenopeptin	*N. punctiforme* PCC 73102: Npun_F2460-Npun_F2466



Supplementary Figure 1. Graphical representation of gene organization of secondary metabolites encoding clusters containing non-ribosomal peptide synthetases (NRPS, blue color)) and/or ribosomally synthesized and post-translationally modified peptides (RiPP, red color). Transporter genes are marked in yellow, hypothetical genes in gray and other functions are left white. Accession numbers are in bold font. (A) lantipeptides, microviridin, (B) nostocyclopeptide, microviridin and (C) aeruginosin/banyaside.

http://microbiology.ucdavis.edu/meeks/xpro5.htm Transcript of web page accessed April 2018.

Experimental Protocols

Meeks Laboratory

Genomic DNA Isolation Procedure for Nostoc punctiforme *5/2000*

This procedure should yield high molecular weight (greater than 50 kb) genomic DNA that is clean enough for standard molecular manipulations such as enzyme digestions.

Reagents required:

5M NaCl
TE (10 mM Tris + 1mM EDTA, pH=8)
Lysozyme (20 mg/ml in TE, make up fresh)
0.5M EDTA, pH=8
20% SDS
Proteinase K (2 mg/ml in TE, make up fresh)
CTAB solution (10% CTAB + 0.7 M NaCl)
Phenol:chloroform (1:1, v/v)
Chloroform
3M sodium acetate (pH=5.2)
95% ethanol
70% ethanol

Procedure:

1. Pellet the culture by centrifuging at 1,000 x g (maximum rpm in clinacial centrifuge) for 5 min in a 15 or 50 ml plastic centrifuge tube. As much as 600 ug of Chl a (~ 6 x 10^8 cells) may be processed at one time.
2. Wash the pellet twice with 5M NaCl for removal of polysaccharides.
3. Resuspend by vortexing the pellet in 1.0 ml total volume with TE.
4. Add 1.0 ml of lysozyme solution. Invert gently, then incubate at 37C for one hour. Invert periodically duringincubation.
5. Add 0.5 ml of 0.5 M EDTA.
6. Add 1.0 ml of proteinase K solution and 100 ul of 20 % SDS. Invert gently to mix. Incubate at 37C for an hour, inverting gently several times during incubation.
7. Add 1/6 x vol (600 ul) of 5M NaCl, mix.
8. Add 1/8 x vol (450 ul) of CTAB solution, mix. Heat to 65C for 10 min.
9. Pellet the cell debris by centrifugation at 13, 000 x g for 5 min at room temperature. Decant the supernatant gently into a new tube. Be careful to avoid the disk of cell debris that sometimes forms on top of the liquid. Be sure the centrifuge is at room temperature and not cooled, as DNA will pellet at cooler temperatures as well.
10. Extract once with chloroform. Transfer aqueous phase to a new tube.
11. Precipitate the DNA with 2 x vol of 95% ethanol. Stringy DNA molecules should be macroscopically visible at this point. Pellet at 13,000 x g for 5 min.
12. Dissolve DNA pellet with 500 ml of TE and transfer to a microfuge tube. Extract with an equal volume of phenol:chloroform. Repeat extractions until the interface is clear. For highest yields of DNA back extract all organic phases.
13. Precipitate DNA by adding 1/10 x vol of 3 M sodium acetate and 2 x vol of 95% ethanol. Wash the pellet with 70% ethanol.
14. Dissolve final DNA pellet in desired volume of TE and digest with RNase if desired.

Paper III

STRUCTURAL BIOLOGY
ISSN 2059-7983

The crystal structure of the *N*-acetylglucosamine 2-epimerase from *Nostoc* sp. KVJ10 reveals the true dimer

Marie-Josée Haglund Halsør, Ulli Rothweiler, Bjørn Altermark and Inger Lin Uttakleiv Raeder*

The Norwegian Structural Biology Centre (NorStruct), Department of Chemistry, UiT – The Arctic University of Norway, 9037 Tromsø, Norway. *Correspondence e-mail: inger.l.rader@uit.no

Received 26 September 2018
Accepted 30 November 2018

Edited by M. Czjzek, Station Biologique de Roscoff, France

Keywords: *N*-acetylglucosamine 2-epimerase; AGE; sialic acid; crystal packing; ManNAc; GlcNAc; *N*-acetylmannosamine; *Nostoc* sp. KVJ10.

PDB reference: *N*-acetylglucosamine 2-epimerase, 6f04

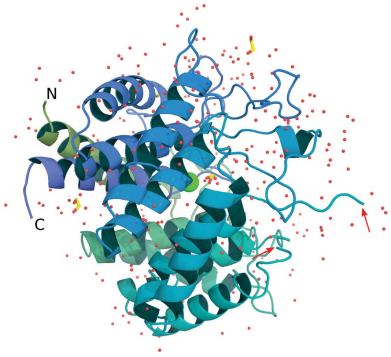
Supporting information: this article has supporting information at journals.iucr.org/d

N-Acetylglucosamine 2-epimerases (AGEs) catalyze the interconversion of *N*-acetylglucosamine and *N*-acetylmannosamine. They can be used to perform the first step in the synthesis of sialic acid from *N*-acetylglucosamine, which makes the need for efficient AGEs a priority. This study presents the structure of the AGE from *Nostoc* sp. KVJ10 collected in northern Norway, referred to as nAGE10. It is the third AGE structure to be published to date, and the first one in space group $P4_22_12$. The nAGE10 monomer folds as an $(\alpha/\alpha)_6$ barrel in a similar manner to that of the previously published AGEs, but the crystal did not contain the dimers that have previously been reported. The previously proposed 'back-to-back' assembly involved the face of the AGE monomer where the barrel helices are connected by small loops. Instead, a 'front-to-front' dimer was found in nAGE10 involving the long loops that connect the barrel helices at this end. This assembly is also present in the other AGE structures, but was attributed to crystal packing, even though the 'front' interface areas are larger and are more conserved than the 'back' interface areas. In addition, the front-to-front association allows a better explanation of the previously reported observations considering surface cysteines. Together, these results indicate that the 'front-to-front' dimer is the most probable biological assembly for AGEs.

1. Introduction

N-Acetylglucosamine 2-epimerases (AGEs; EC 5.1.3.8) catalyze the reversible epimerization of *N*-acetylmannosamine (ManNAc) to *N*-acetylglucosamine (GlcNAc) in both eukaryotic and prokaryotic cells (Ghosh & Roseman, 1965; Lee, Chien *et al.*, 2007). It seems that their role, at least in mammals, is to divert the metabolic flux away from sialic acid synthesis (Luchansky *et al.*, 2003). The reaction follows a deprotonation/reprotonation mechanism involving two key residues, which had been thought to be histidines, but which have recently been identified as a glutamine and an arginine (Lee, Wu *et al.*, 2007; Takahashi, Takahashi *et al.*, 2001; Wang *et al.*, 2016). AGE activity is greatly enhanced by the presence of nucleotides, in particular ATP, which serve as allosteric activators (Datta, 1970; Ghosh & Roseman, 1965; Tabata *et al.*, 2002; Takahashi, Hori *et al.*, 2001).

From an industrial point of view, AGEs can be used for the synthesis of *N*-acetylneuraminic acid (Neu5Ac), which is also known as sialic acid (Hu *et al.*, 2010; Kragl *et al.*, 1991; Lee *et al.*, 2004; Lee, Chien *et al.*, 2007; Maru *et al.*, 1996; Tabata *et al.*, 2002; Wang *et al.*, 2009). One current approach is a one-pot, coupled reaction with an *N*-acetylneuraminic acid lyase (NAL; EC 4.1.3.3), in which the AGE performs the (reverse) GlcNAc-to-ManNAc epimerization and the NAL performs



OPEN ACCESS

the condensation with pyruvate, thus producing Neu5Ac. However, there are several challenges related to the use of AGEs in this step, such as the unfavourable thermodynamic equilibrium, the requirement for ATP and the inhibition by both pyruvate and Neu5Ac (Datta, 1970; Ghosh & Roseman, 1965; Klermund *et al.*, 2013; Kragl *et al.*, 1991).

In order to optimize Neu5Ac production, the search for better AGEs is a current area of focus and enzymes from cyanobacterial sources seem to be the most promising. Indeed, the AGEs from *Anabaena* sp. CH1 (AnaAGE) and *A. variabilis* ATCC 29413 (AvaAGE) have a specific activity that is almost four times that of the porcine enzyme, despite having similar affinities for GlcNAc (Klermund *et al.*, 2013; Lee, Wu *et al.*, 2007). Other enzymes present interesting properties, such as the AGE from *Bacteroides ovatus* ATCC 8483 (BoAGE), which shows a much higher affinity for GlcNAc than ManNAc, and that from *Synechocystis* PCC 6803 (SynAGE), which has the lowest K_m values for both GlcNAc and ManNAc of all AGEs that have been characterized to date (Sola-Carvajal *et al.*, 2012; Tabata *et al.*, 2002).

In addition to bioprospecting for enzymes that are more suited for industrial purposes, the identification of the structural elements that govern the activity parameters of AGEs is a priority. Two AGEs had been crystallized prior to this study: the porcine enzyme (pAGE; PDB entry 1fp3) and that from *Anabaena* sp. CH1 (AnaAGE; PDB entry 2gz6), the latter of which used the former as a template for molecular replacement (Itoh *et al.*, 2000; Lee, Wu *et al.*, 2007). These studies revealed that the overall structure of AGEs is that of a $(\alpha/\alpha)_6$ barrel. Together with mutagenesis studies, this led to the proposal of a reaction mechanism involving two critical histidine residues as acid/base catalysts in the protonation/deprotonation of carbon C2 of GlcNAc/ManNAc. The enzyme from *Pedobacter heparinus* DSM2366 (PhGn2E) was used in a hydrogen/deuterium-exchange experiment, which confirmed this type of mechanism for the epimerization, albeit with a glutamine/arginine pair as catalysts (Wang *et al.*, 2016).

AGEs were found to be structurally similar to sulfoquinovose isomerases encoded by the *yihS* gene (EC 5.3.1.31) and to cellobiose 2-epimerases (CE; EC 5.1.3.11), despite a relatively low sequence identity (Denger *et al.*, 2014; Fujiwara *et al.*, 2013, 2014; Itoh *et al.*, 2008; Tyler & Leatherwood, 1967). Together, they form the AGE superfamily (Pfam accession No. PF07221), for which 13 structures are currently available (Bateman *et al.*, 2004; Finn *et al.*, 2014). Several of these structures have been solved as protein–ligand complexes, and structural comparison of the relatively conserved active sites helped to form the current hypothesis for the catalytic mechanism of AGEs (Itoh *et al.*, 2008).

In this study, we report the determination of the crystal structure of the AGE from the local strain *Nostoc* sp. KVJ10, which we refer to as nAGE10 (Liaimer *et al.*, 2016). We also demonstrate that the most probable biological assembly for the AGE dimer involves the opposite face of the barrel to that published previously. Finally, we reveal the presence of a putatively conserved chloride ion within the active site of AGE.

2. Materials and methods

2.1. Cloning, expression and purification

The gene coding for nAGE10, containing the coding sequence for a C-terminal hexahistidine tag and optimized for expression in *Escherichia coli*, was synthesized by GeneArt and subcloned into the pDEST14 expression vector using the Gateway cloning system (ThermoFisher Scientific). The protein was expressed in *E. coli* BL21 Star (DE3) cells (ThermoFisher Scientific), which were grown at 37°C to an OD_{600} of 0.6, brought to 20°C and induced with 0.5 m*M* isopropyl β-D-1-thiogalactopyranoside (IPTG; Sigma–Aldrich). The cells were incubated for approximately 16 h at 20°C and harvested by centrifugation at 6000*g* for 25 min. Bacterial cell pellets were resuspended in 50 m*M* Tris–HCl pH 7.5, 250 m*M* NaCl, 5 m*M* β-mercaptoethanol, 10 m*M* imidazole, along with half a tablet of protease-inhibitor cocktail (Roche) and 2 µl DNAse I. The cells were disrupted using a French press and the extract was centrifuged at 20 000*g* for 2 × 30 min to remove cell debris. The supernatant was filtered and purified by affinity chromatography using a 5 ml HisTrap column (GE Healthcare). The binding buffer consisted of 20 m*M* sodium phosphate pH 7.4, 0.5 *M* NaCl, 20 m*M* imidazole. The elution buffer consisted of 20 m*M* sodium phosphate pH 7.4, 0.5 *M* NaCl, 500 m*M* imidazole. The protein was eluted using a gradient from 5% to 80% elution buffer over 60 ml. After purification, fractions containing protein were assessed by SDS–PAGE. Pure fractions were pooled and dialyzed overnight in 50 m*M* Tris–HCl pH 7.5, 250 m*M* NaCl. The protein solutions were further dialyzed into storage buffer (20 m*M* HEPES pH 7.4, 115 m*M* NaCl). The oligomerization state of nAGE10 was determined by gel-filtration chromatography on a Superdex 75 10/300 GL column (GE Healthcare). The results are presented in the supporting information. The protein concentration was determined by measuring the absorbance at 280 nm using a theoretical extinction coefficient for the protein.

2.2. Differential scanning calorimetry

Differential scanning calorimetry (DSC) was used to determine the melting temperature of nAGE10. Thermal denaturation of nAGE10 was followed between 5 and 95°C using a heating/cooling rate of 1°C min^{-1}. A solution of nAGE10 at 1.5 mg ml^{-1} in storage buffer was used in the experiment. The results are presented in the supporting information.

2.3. Structure determination

The protein was concentrated to 10 mg ml^{-1} in storage buffer prior to crystallization. Screening for crystallization conditions was performed using the sitting-drop method (drop size of 200 + 200 nl). The initial screening was performed using in-house and commercially available screens at both room temperature and 4°C. From this, two initial hits were obtained (at both temperatures) with 0.1 *M* sodium acetate pH 5, 17.6% PEG 6000 or with a combination of 0.07 *M* sodium acetate

pH 5, 0.05 *M* calcium acetate and 12% PEG 3350. The optimized crystallization condition was 0.1 *M* sodium acetate pH 5, 0.1 *M* calcium acetate, 10% PEG 3350, 3% dextran. Crystals were obtained at room temperature after 1–3 days of incubation and were flash-cooled in liquid nitrogen. 20% ethylene glycol was used as a cryoprotectant. Crystallographic data were collected at the BESSY II photon source, Helmholtz-Zentrum Berlin, Germany. The images were integrated using *XDSapp* (Sparta *et al.*, 2016). The structure was solved by molecular replacement using the AnaAGE structure (PDB entry 2gz6) as a search model (Lee, Wu *et al.*, 2007). Refinement was performed using *PHENIX* and the *CCP*4 program *REFMAC*5 (Adams *et al.*, 2010; Murshudov *et al.*, 2011; Winn *et al.*, 2011). The waters were placed by *Coot* v.0.7.2 (Emsley *et al.*, 2010). The crystallographic data and model statistics are summarized in Table 1.

2.4. Analysis of interface areas

The surfaces of nAGE10, AnaAGE and pAGE were analyzed using the *PDBePISA* server in order to obtain parameters pertaining to the structural and chemical properties of possible assemblies (Krissinel & Henrick, 2007).

2.5. Sequence comparison

The amino-acid sequence of nAGE10 and the previously characterized AGEs were aligned using the *MUSCLE* multiple sequence alignment tool with default settings (Edgar, 2004*a*,*b*). Secondary-structure data for nAGE10 and pAGE were obtained from their respective PDB files. The graphical output was created using the *TeXshade* package for LaTeX (Beitz, 2000).

2.6. Enzyme-activity measurement

The activity of nAGE10 was assessed by coupling the epimerization to the condensation reaction catalyzed by the NAL from *Aliivibrio salmonicida* (AsNAL; EC 4.1.3.3; PDB entry 5afd; M. K. Gurung, B. Altermark, I. L. U. Rader, R. Helland & A. O. Smalas, unpublished work), resulting in the production of Neu5Ac. Its presence was detected using the thiobarbituric acid (TBA) assay (Warren, 1959). Samples consisting of 124 m*M* HEPES pH 8.0, either 15 m*M* GlcNAc or 15 m*M* ManNAc, 10 m*M* ATP, 15 m*M* pyruvate, 7 µg NAL and 10 µg AGE10 were incubated at room temperature for 1 h. The reactions were terminated by adding 137 µl sodium periodate (2.5 mg ml^{-1} sodium periodate in 57 m*M* H_2SO_4) followed by incubation at 37°C for 15 min with shaking. Arsenite solution (50 µl, 25 mg ml^{-1} in 0.5 *M* HCl) was added, resulting in a brown colour. The tubes were shaken until the brown colour disappeared and 100 µl TBA solution (71 mg ml^{-1} TBA pH 9.0) was then added. The samples were placed in boiling water for 7.5 min and then on ice for 5 min; they were then brought to room temperature for 5 min. Acidic *n*-butanol (5% HCl, 1 ml) was added and the samples were shaken for 10 min at room temperature. The tubes were spun down at 16 060*g* for 7 min and the absorbance of the upper layer was measured at 549 nm on a SpectraMax (Molecular Devices). A molar extinction coefficient of 57 000 M^{-1} cm^{-1} was used for concentration calculations (Warren, 1959).

Table 1
Data collection and processing.

Values in parentheses are for the outer shell.

PDB code	6f04
Data collection	
Synchrotron-radiation source	BESSY 14.1
Detector	PILATUS 6M
Wavelength (Å)	0.9184
No. of frames	1800
Oscillation range per frame (°)	0.1
Diffraction data	
Space group	$P4_22_12$
a, *b*, *c* (Å)	142.423, 142.423, 52.046
Protein molecules in asymmetric unit	1
Total No. of reflections	768654 (109502)
No. of unique reflections	59456 (9464)
Resolution range (Å)	50–1.7 (1.8–1.7)
Completeness (%)	99.9 (99.7)
$\langle I/\sigma(I)\rangle$	15.58 (1.49)
Observed *R* factor (%)	13.2 (143)
$CC_{1/2}$	99.9 (64.3)
Refinement	
Resolution limits (Å)	48.9–1.7
No. of used reflections	59447
Data completeness (%)	99.94
Percentage of free reflections	3.5
No. of protein atoms	3214
No. of heterogen atoms	17
No. of waters	318
R factor/R_{free}	0.179/0.209
Overall *B* factor from Wilson plot (Å^2)	21.61
R.m.s.d., bonds (Å)	0.007
R.m.s.d., angles (°)	0.915
Ramachandran statistics	
Favoured	381
Allowed	5
Outliers	0

2.7. Graphical output generation

Molecular representations were generated using *PyMOL* (v.1.8; Schrödinger) and sequence alignments were rendered with the *TeXshade* package for LaTeX (Beitz, 2000). Figure editing was performed in *Inkscape* (https://inkscape.org/).

3. Results and discussion

3.1. Structure of nAGE10

nAGE10 was crystallized using the hanging-drop vapour-diffusion method. Just as for pAGE (PDB entry 1fp3) and AnaAGE (PDB entry 2gz6), nAGE10 crystallized at a pH below 6. Attempts to either crystallize it at a higher pH or to increase the pH of the existing crystals were unsuccessful. Co-crystallization trials with GlcNAc, ManNAc or ATP, as well as a combination of the latter with hexosamine, did not result in crystallized complexes, although crystals of the apoenzyme were obtained from these experiments. Soaking did not affect the crystals. These results are similar to those described for previously published AGE structures, indicating that this behaviour may be inherent to AGEs rather than to the crystallization method. Temperature did not seem to play a

significant role in the crystallization process, as similar crystals were obtained at both room temperature and 4°C.

The nAGE10 crystal diffracted to 1.70 Å resolution, and the subsequent model, which was determined by molecular replacement using the AnaAGE structure (PDB entry 2gz6; 90% sequence identity), was refined to an R factor of 0.179 ($R_{\rm free}$ = 0.209). The r.m.s.d. between the two structures is 0.8 (Holm & Laakso, 2016). The refinement data are summarized in Table 1. In contrast to the pAGE and AnaAGE crystals, which both belonged to space group $P2_12_12_1$, the nAGE10 crystal belonged to space group $P4_22_12$.

The asymmetric unit contained one nAGE10 monomer, a chloride ion and three ethylene glycol molecules, as well as 318 waters (Fig. 1*a*). The structure of the monomer was resolved for the amino-acid sequence from Tyr3 to Leu391, with residues 157–164 missing. The fold is that of an $(\alpha/\alpha)_6$ barrel, with the helices connected by small loops at one end (the back) and long loops at the other end (the front). The nAGE10 barrel is composed of two concentric rings of six helices each, in which the sequence orientation is opposite. Helices H1, H3, H5, H7, H9 and H11 form the outer ring of the barrel, with the even-numbered helices as the inner ring (Fig. 1*b*). The nAGE10 barrel is about 50 Å in diameter and 30 Å in length.

3.2. Dimer assembly

Fig. 1(*c*) shows nAGE10 as a dimer, which was generated by twofold symmetry using the operation $y-1, x+1, -z$. The calculated interaction interface represents 9.1% of the accessible surface area, but this value is most likely to be underestimated owing to the disordered loop between residues 156 and 165 (indicated by red arrows in Fig. 1*c*). The interface is composed of 41 residues for each monomer, 35 of which participate in extensive interactions (Table 2). 11 hydrogen bonds involve the residue pairs Asp48/Arg234, Asp50/Arg167, Pro108/Thr166 (which only occurs once), Gln295/Lys367, Leu297/Gln301 and Trp299/Leu360. In addition, 53 pairs of residues form van der Waals (vdW) interactions. The proposed interactions involving residues Thr166 and Arg167 are probably influenced by the disorder of the aforementioned loop.

The nAGE10 dimer, which involves the front faces of each monomer, differs from the back-to-back association presented for AnaAGE and pAGE (Itoh *et al.*, 2000; Lee, Wu *et al.*, 2007). A front-to-front association for AnaAGE (symmetry operator $x-1, y, z$) is mentioned by Lee *et al.* (2007) and is attributed to crystal packing. However, the nAGE10 dimer can be superimposed onto that of AnaAGE assuming that the front-to-front packing in PDB entry 2gz6 is the biological dimer (instead of the back-to-back packing). pAGE, which has the same crystal packing as AnaAGE, also has a front-to-front dimer

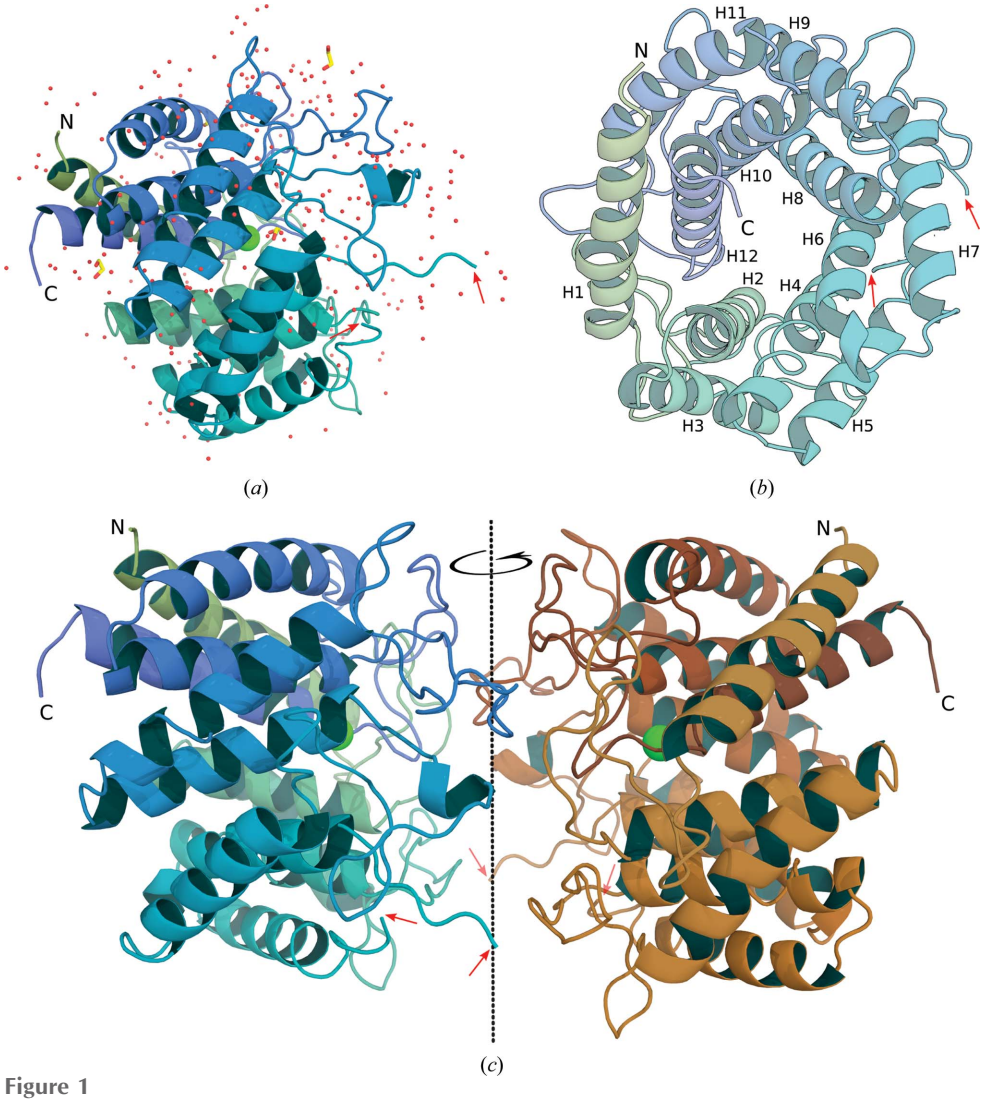


Figure 1
Structural overview of nAGE10. (*a*) The contents of the asymmetric unit of the nAGE10 structure, with the nAGE10 monomer shown from the side in a cartoon representation. It is coloured using a green–blue palette from the N-terminal to the C-terminal residues. Waters are shown as red, nonbonded spheres. Ethylene glycol molecules (three) are represented as yellow sticks and the buried chloride ion as a green sphere. (*b*) The nAGE10 monomer, shown from the front (the face where the α-helices that form the barrel are linked by short loops). Each helix is numbered (H1–H12) according to its placement in the amino-acid sequence. (*c*) The nAGE10 dimer. Generation of symmetry mates within 3 Å of nAGE10 reveals the presence of an additional monomer for which dimer association is probable involving the fronts of the monomers. Red arrows indicate the extremities of the disordered loop (residues 156–165).

Table 2
Interactions at the front dimer interface for nAGE10, pAGE and AnaAGE.

(*a*) nAGE10.

Hydrogen bonds		vdW interactions†	
1‡	2	1	2
D48	R234	C39, D48, F52	F231
D50	R167	Y47	F231, R234, H292, P293
P108	T166	T103, L109	T166, R167
R167	D50	K107, P108	T166
R234	D48	T166	T103, K107, P108, L109
Q295	K367	R167	T103, L109
L297	Q301	F231	Y47, F52, K367
W299	L360	R234	Y47, D48
Q301	L297	Y285, Y352	L297
L360	W299	H292	Y47
K367	Q295	P293	Y47, L360, L361, N362, L363
		P294, Q296	L361
		Q295	L361, K367
		L297	Y352, L360, L361, Y285, E298
		E298	L297, E298, Q301
		W299	L360, L361
		D300	L360
		Q301	L297, Q301
		N362, L363	P293
		K367	F231, Q295

(*b*) pAGE.

Hydrogen bonds		vdW interactions	
A1	B2	A1	B2
Y49	T304	Y49	L301, C239, T304
D50	C239	D50	C239, G238, R242
R113	P171	V97, V117	L170
P114	P171, A173	T109	G172
P171	R113	R113, P114	P171, G172
A173	P114	V115	L170, A173
C239	D50	L170	K116, V97, V115, K116
R242	Y49	P171	R113, P114
C302	T371	G172	T109, R113, P114, V115
W308	A369	A173	V115
A369	W308	G238	D50
T371	C302	C239, R242	Y49, D50
		L301	Y49
		C302	I370, T371, I372
		T308	I370
		L306	Y294, Y361, E307, M310, L312
		M310, L312, Y361	L306
		L370	C302, T304, L306, W308
		T371	C302
		I372	C302, T304
		P376	T304

Table 2 (continued)

(*c*) AnaAGE.

Hydrogen bonds		vdW interactions	
A1	B2	A1	B2
D47	F230, R233	C38, F51	F230
D49	R166	D40, N361, L362	P292
T165	P107	Y46	F230, R233, H291, P292
R166	D49, T102	D47	F230, R233
F230	D47	T102, L108	R166
R233	D47	T165	T102, E106, P107, L108, V109
L296	Q300	R166	T102, L108, V109
W298	L359	F230	Y46, F51, K366
Q300	L296	R233	Y46, D47
R355	E357	Y284, Y351	L296
E357	R355	H291	Y46
L359	W298	P292	Y46, L360, N361, L362
		P293	L360
		Q294	L360, K366, W367
		L296	Y284, E297, Q300, Y351, L359, L360
		E297	L296, E297, Q300
		W298	L359, L360
		D299	L359
		Q300	L296, E297, Q300
		N353	R355
		R355	N353, R355
		L359	L296, W298, D299
		L360	P292, P293, Q294, L296, W298
		K366	F230, Q294
		W367	Q294

† Between H atoms of monomers; within 4.5 Å. ‡ Monomers. In the cases of pAGE and AnaAGE, the chain is specified.

(Fig. 2). A comparison of the unit cell of nAGE10 with those of pAGE and AnaAGE reveals not only that the back-to-back assembly does not exist in the nAGE10 crystal, but also that either the front-to-front or the back-to-back assemblies can be used as the asymmetric unit for the pAGE and AnaAGE structures, as presented in Fig. 2.

A comparison of the assembly parameters calculated by the *PISA* method indicates that the front-to-front organization is more favourable than the back-to-back organization in terms of interface area, number of interactions and solvation energy (Krissinel & Henrick, 2007). For AnaAGE, the interface area is only 346.8 Å^2 for the back-to-back assembly, while it is 1386.3 Å^2 for the front-to-front complex (again, the missing residues may be part of the dimer interface, so the interface area for the front-to-front assembly is probably larger). The two interfaces of pAGE have similar areas (1227.2 Å^2 for the front and 1029 Å^2 for the back), but the solvation energy is much more favourable for the front-to-front assembly at −22.3 kcal mol^{-1}, compared with −1.2 kcal mol^{-1} for the back-to-back assembly. In terms of interactions, Itoh *et al.* (2000) reported that nine hydrogen bonds and 23 vdW contacts (<4.5 Å) occur between monomers in the back-to-back assembly. Using the same criteria, 13 hydrogen bonds and 52 vdW contacts can be found for the front-to-front dimer (Table 2). In the case of AnaAGE the front-to-front assembly is mentioned (as crystal packing), but only the interactions involving eight residues are described. Using the same criteria as the authors for identifying interactions and considering only hydrogen bonds leads to the involvement of 14 residues (Table 2). For example, the residues involved in the polar interactions that take place between Asp49 and Arg166, between Pro107 and Thr165 and between Trp298 and Leu359 below 3.5 Å are described, but not those between Asp47 and Arg233, between Leu296 and Gln300 and between Arg355 and Glu357 that also occur within the same distance (data not shown).

A look at the distribution of the interface residues, which is presented in Fig. 3, shows that the back interface areas of AnaAGE and pAGE are quite different (bottom surfaces in Figs. 3*a* and 3*c*). This is explained by the fact that the residues involved in this interface are not conserved between AnaAGE and pAGE (Fig. 3*e*). Indeed, the main interaction patch of pAGE, which involves the loops between H4 and H5 (residues

135–140) and between H6 and H7 (residues 194–198), is missing in AnaAGE. The shape of these loops is thus different and the monomers are further apart at this location in AnaAGE than they are in pAGE (data not shown). On the contrary, the residues involved in the front interface area are mostly conserved, with the exception of the loop between H5 and H6 (residues 170–173; Fig. 3*d*). This loop is also only resolved for pAGE, and it is likely that the interface parameters for the cyanobacterial AGEs, in particular the solvation energy, may be lower than calculated.

In addition to this, it was shown that cysteines are involved in the dimerization of pAGE (Takahashi *et al.*, 1988). The front interface contains three cysteines (residues 41, 239 and 302 of pAGE), which is the exact number of cysteines that were alkylated by the treatment with *N*-ethylmaleimide performed in that study. Those cysteines lose at least 55% of their solvent-accessible area upon dimer formation, with Cys239 and Cys302 involved in hydrogen bonds to the neighbouring monomer (Table 2). Their positions are identical to those in human AGE when mapping the hAGE sequence onto the pAGE structure (data not shown); mutation studies revealed the importance of Cys41 for the stability of hAGE, while Cys239 and Cys302 do not seem to play critical roles (Takahashi, Takahashi *et al.*, 2001; Takahashi, Takahashi, Kaneko, Ogasawara, Shindo, Saito *et al.*, 1999).

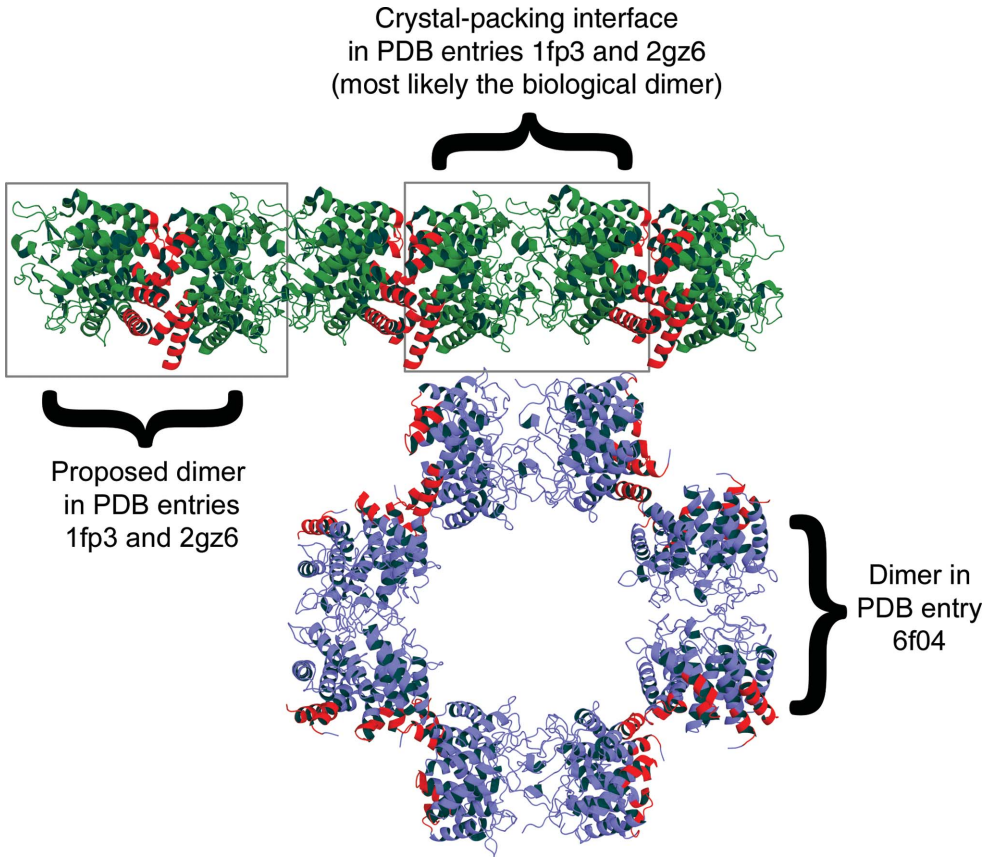


Figure 2
Comparison of the nAGE10 and pAGE dimers. Parts of the unit cell for pAGE (PDB entry 1fp3; green) and nAGE10 (PDB entry 6f04; blue) are shown to illustrate the different crystal packings of these structures. The residue ranges that are involved in the back-to-back interface of PDB entry 1fp3 and the corresponding ranges in nAGE10 are coloured red. PDB entry 2gz6 (AnaAGE) has the same crystal packing as PDB entry 1fp3 and was therefore omitted to avoid redundancy.

3.3. Dimer organization and ATP-binding site

One of the principal consequences of the front-to-front dimer is for the ATP-binding site. To date, two hypotheses have been put forward regarding the location of this site. One involves the H5/H6 loop and the other involves a glycine-rich fragment (residues 363–369 of AnaAGE) in the C-terminus (Lee, Chien *et al.*, 2007; Liao *et al.*, 2012; Sola-Carvajal *et al.*, 2012; Takahashi *et al.*, 2002, 2005). While the latter, which is based on sequence similarity to the motif of NTPases, is conserved across species, the evidence supporting the non-conserved H5/H6 loop is built on experiments using chimeric constructs, mutagenesis and ATP footprinting. A non-conserved binding region would also better explain the differences in nucleotide affinities that are observed across both species and nucleotide type (Lee, Chien *et al.*, 2007; Sola-Carvajal *et al.*, 2012; Takahashi, Hori *et al.*, 2001). This would mean that the H5/H6 loop is involved in both dimer interaction and ATP binding. However, the second putative site corresponds to part of the H11/H12 loop, which is also on the front face of the monomer in a position opposite to that of the H5/H6 loop. The loop points towards the centre of the barrel, and if the H5/H6 loop does the same, which is the case for the corresponding loop in pAGE, they might be in proximity to each other. This opens the possibility of both proposed sites being involved. Independently of which site is used, the localization of the ATP-binding site at the dimer interface is the determinant of its role in AGE activity, which may involve the oligomeric state of the enzyme. The inability to obtain crystals of the AGE–ATP complex either by co-crystallization or soaking could indicate the dissociation of the dimeric form upon ATP binding, as has been seen in different complexes (Nayar & Bhattacharyya, 1997; Ahmed *et al.*, 2015; Du *et al.*, 2014).

3.4. Active site and comparison with the AGE superfamily

The active site of nAGE10, which is presented in Fig. 4(*a*), contains an ethylene glycol molecule and a buried chloride

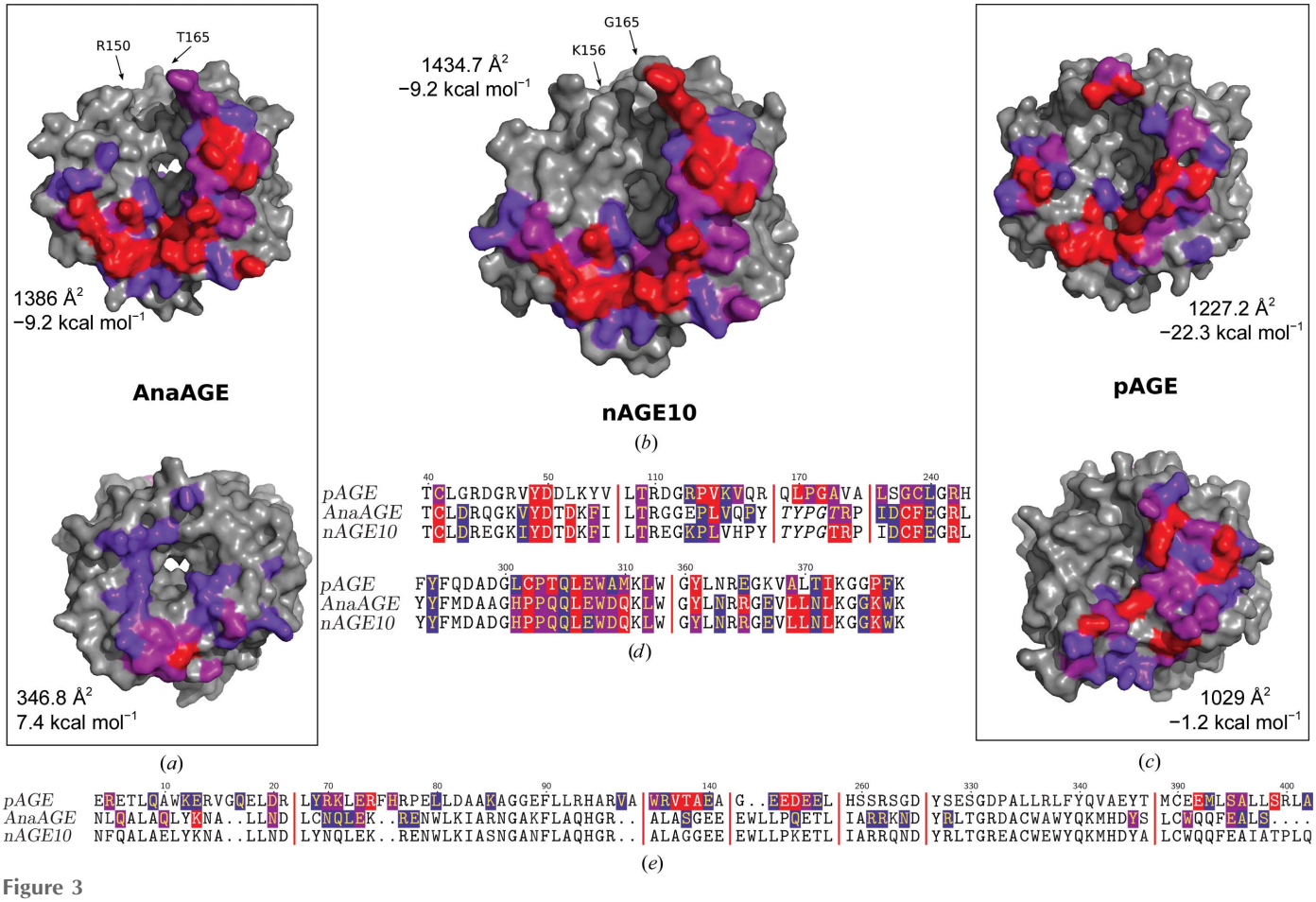


Figure 3
Dimer-interaction interfaces of AGEs as a function of buried area. The AGE monomers are represented as a surface and are coloured grey. Residues involved in dimer interactions, as calculated by *PISA*, are coloured purple (less than 30% buried), pink (less than 60% buried) and red (over 60% buried). The residues at each side of the missing loop are indicated by arrows and labelled. (*a*) AnaAGE. Top, front; bottom, back. (*b*) nAGE10 (front). (*c*) pAGE. Top, front; bottom, back. (*d*, *e*) Sequence alignment showing the interface residues using the same colour scheme as used for the surface representations. (*d*) Front. Residues which are unresolved in the crystal structures are shown in italics. (*e*) Back.

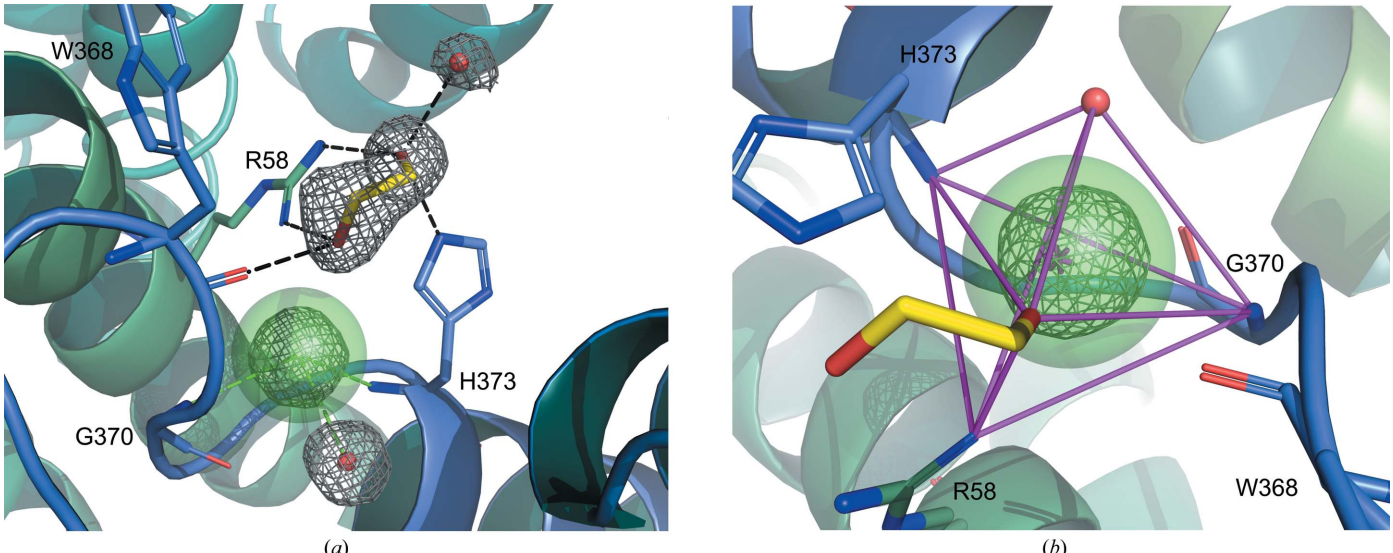


Figure 4
Structure of the active site of nAGE10. nAGE10 is shown in a cartoon representation and is coloured in a green–blue palette from the N-terminus to the C-terminus. Ethylene glycol (yellow) and active-site residues are shown as sticks, and waters and the chloride ion are shown as nonbonded spheres. (*a*) Active site. Electron density ($2F_o - F_c$ map) is shown around the waters, the chloride ion and the ethylene glycol molecule. Polar contacts for ethylene glycol are coloured black and coordination of the chloride ion is shown in green. (*b*) Coordination geometry for the buried chloride ion. The five coordinating atoms form a trigonal bipyramid (magenta) with the chloride at the centre of mass (magenta star).

ion. No ligand could be identified, although the enzyme was crystallized in the presence of ManNAc. The ethylene glycol interacts with the side chains of His373 and Arg58, as well as the main chain of Trp368 and a water molecule. The chloride ion is coordinated by the amine groups of His373 and the Gly370 main chain, which are within its first coordination sphere (3.4 Å), along with a water molecule (Carugo, 2014). However, the geometry of this three-atom coordination seems to be incorrect, as the chloride is completely outside the plane formed by these atoms. By investigating nearby atoms, the best geometry was obtained for a five-atom coordination, with the ethylene glycol (3.7 Å) and Arg58 (3.5 Å) as the two remaining partners (Fig. 4*b*). Other potential partners are the amine group of Phe372, the Gly370 carbonyl and the δ position of His373. The latter would mean that His373 is flipped, which has consequences for the interactions taking place within the active site.

The structures of pAGE and AnaAGE contain only waters within their active sites, and the active-site residues are conserved compared with those of nAGE10 (not shown). However, it is worth noting that for each of them there is a water molecule at the position where the chloride ion is found in the nAGE10 structure (water 389 in AnaAGE and water 521 in pAGE). The closest neighbour to both molecules is at 3.1 Å, and their *B* factors are below 3 Å^2, while those of the neighbouring water molecules range from 11.9 to 29.3 Å^2.

While none of the AGEs could be co-crystallized with ligands, several protein–ligand complex structures are available for other members of the AGE superfamily. A superimposition of the active sites of the cellobiose epimerase from *Rhodothermus marinus* (RmCE; PDB entry 3wki) and the isomerase YihS from *Salmonella enterica* (SeYihS; PDB entry 2zbl) is presented in Fig. 5 (Fujiwara *et al.*, 2014; Itoh *et al.*, 2008). It shows that the ethylene glycol molecule in the active site of nAGE10 is in proximity to the O5, C5, C6 and O6 positions of mannose, as well as the corresponding positions of cellobiitol. This placement is consistent with the suggested role of Arg63 of the epimerase PhGn2E as the agent that is responsible for the protonation and deprotonation of ManNAc (Wang *et al.*, 2016). Indeed, the corresponding position in AGE10 (Ar68) is only 2.9 Å from the oxygen of ethylene glycol that mimicks O5 of mannose (Fig. 5*b*). The other catalytic residue, Glu314 in PhGn2E (Glu309 in nAGE10), is surprisingly not in the vicinity of the ligands, with those in SeYihS (Gln320) and RmCE (Gln326) being more than 7 Å away from the C2 atom of their respective 'mannoses' (data not shown). Glu251 in SeYihS (Glu243 in nAGE10) is much closer: it is only 4.4 Å from the C2 atom.

Another interesting feature is the stacking interaction between Trp385 of RmCE and the glucose ring of cellobiitol (the same interaction is present with epilactose and glucosylmannose in the other RmCE structures; PDB entries 3wkg and 3wkh). This residue, which corresponds to Trp368 in nAGE10, is strictly conserved in AGEs but its function has not yet been investigated. Considering the position of the glucose relative to the mannose ring, it may be involved in the substrate-specificity differences observed by Wang *et al.* (2016) when using derivatives of glucosamine substituted at C4 (just like the glucose ring of cellobiitol is linked to C4 of the open mannose ring).

A structural comparison of the active sites within the AGE superfamily also reveals a putative binding site for the *N*-acetyl group of GlcNAc/ManNAc in AGEs. For RmCE (and SeYihS), the C2 position of the 'mannose' and the OH that it carries interact with the side chains of Tyr124 (Tyr111), His200 (His176) and Asn196 (Asn172). Those positions are not conserved in nAGE10, where they are replaced by Phe116, Ile177 and Ala173, respectively (Fig. 5*c*). This leaves a cavity that would be large enough to accommodate an *N*-acetyl group and would also explain why PhGn2E is not active on glucosamine, as it would not be retained in the active site. All three residues are either almost or strictly conserved in AGEs.

RmCE also possesses a chloride ion within its active site, at the same location as that in nAGE10, and the coordination by ethylene glycol seen in nAGE10 may not reflect what takes place in the presence of the substrate. The structure of another CE from *Ruminococcus albus* (RaCE; PDB entry 3vw5; Fujiwara *et al.*, 2013) contains a water (water 451) at the 'chloride

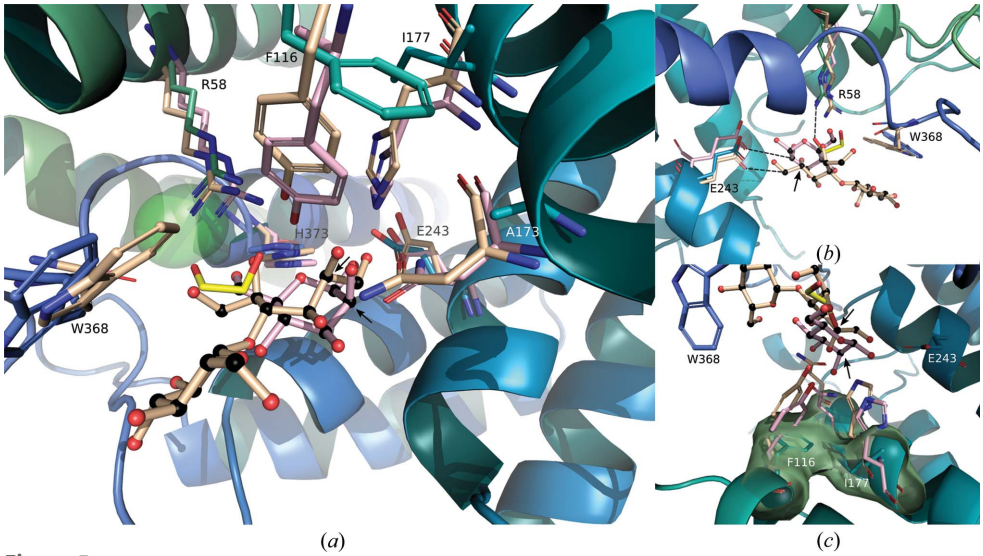


Figure 5
Comparison of active sites in the AGE family. The active sites of RmCE in complex with cellobiitol (wheat; PDB entry 3wki) and SeYihS in complex with mannose (light pink; PDB entry 2zbl) are superimposed onto the structure of nAGE10 (blues, with ethylene glycol in yellow). The C2 atoms of the mannose rings of cellobiitol and mannose are indicated by black arrows. Residues of nAGE10 are labelled. (*a*) Overview. Perspective was added for visualization purposes. (*b*) Detail of the catalytic residues and the interaction with Trp368. (*c*) Detail of the *N*-acetyl group binding pocket.

location', with a similar *B* factor to those of the neighbouring waters. Investigating the residues corresponding to coordinating residues in RmCE reveals that Gly387 is replaced by a cysteine (Cys371) in RaCE. In *yihS*-encoded proteins the spatial organization of waters differs from those in CEs and AGEs, despite their shared fold (PDB entries 2zbl, 2afa and 2rgk; Itoh *et al.*, 2008; SGX Research Center for Structural Genomics, unpublished work). The coordinating glycine is replaced in this case by an aspartate (Asp380 in SeYihS) that is conserved within the structures. This strengthens the hypothesis of a conserved chloride within the AGE active site and opens the possibility of studying its role by mutating the glycine position of AGEs.

3.5. Sequence similarity to characterized AGEs

The amino-acid sequence of nAGE10 was compared with those of other characterized AGEs from cyanobacteria, as well as those from human, pig, rat, *B. ovatus* and *P. heparinus* (Klermund *et al.*, 2013; Lee, Chien *et al.*, 2007; Maru *et al.*, 1996; Sola-Carvajal *et al.*, 2012; Tabata *et al.*, 2002; Takahashi, Takahashi, Kaneko, Ogasawara, Shindo, Saito & Kobayashi, 1999). Their sequences are presented in Fig. 6 as a multiple sequence alignment, which shows that the nAGE10 sequence is closest to that of *N. punctiforme* PCC 73102, with 96.14% sequence identity (Edgar, 2004*a*,*b*). AGEs are quite conserved amongst related species, with greater than 80% identity within mammalian sequences and greater than 90% in the Nostocaceae family (*Anabaena* and *Nostoc* genera). For the latter, sequences from *Anabaena* are grouped together along with that of *Nostoc* sp. PCC 7120, which is also known as *Anabaena* sp. PCC 7120 (Kaneko *et al.*, 2001). This is consistent with the currently accepted phylogenic distribution of *Nostoc* and *Anabaena* strains (Svenning *et al.*, 2005). The sequences for *P. heparinus*, *B. ovatus* and *Synechocystis* sp. stand out, which

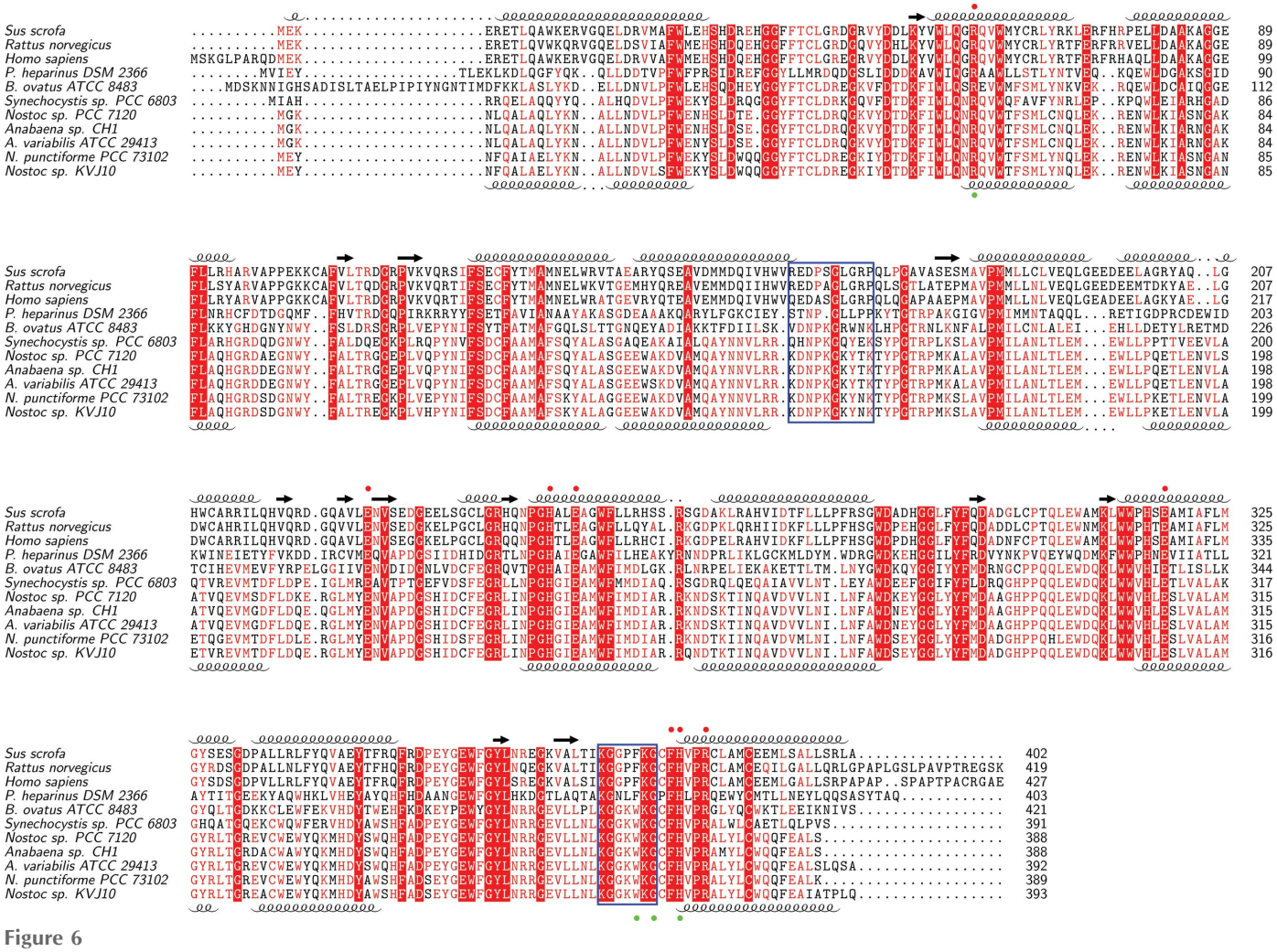


Figure 6
Sequence conservation amongst AGEs. The sequences (with accession numbers in parentheses) are from *Sus scrofa* (NP_999065.1), *Rattus norvegicus* (NP_112357.1), *Homo sapiens* (NP_002901.2), *P. heparinus* DSM2366 (ACU05446.1), *B. ovatus* ATCC 8483 (EDO12673.1), *Synechocystis* sp. PCC 6803 (BAA18210.1), *Nostoc* sp. PCC 7120 (WP_010997838.1), *Anabaena* sp. CH1 (ABG57043.1), *A. variabilis* ATCC 29413 (WP_011320279.1), *N. punctiforme* PCC 73101 (WP_012409471.1) and *Nostoc* sp. KVJ10 (NNBT01000060.1). Identical residues are shown in white on a red background and similar residues in red. α-Helices and β-strands correspond to those in the structures of the AGEs from pig (top; PDB entry 1fp3) and *Nostoc* sp. KVJ10 (bottom). Red circles (top) indicate residues that participate in catalysis in the AGEs from *P. heparinus* DSM 2366 and *Anabaena* sp. CH1 (Lee, Wu *et al.*, 2007; Wang *et al.*, 2016). Green circles (bottom) indicate residues that coordinate the chloride ion in *Nostoc* sp. KVJ10. The blue frames indicate the putative ATP-binding sites (Liao *et al.*, 2012; Wang *et al.*, 2016).

is expected considering that they do not belong to either of the two aforementioned groups.

3.6. Expression and purification

The AGE from *Nostoc* sp. KVJ10 was successfully expressed as a soluble, His-tagged protein in *E. coli* BL21 Star (DE3) cells. It could be purified in one step by affinity chromatography, giving a yield of purified protein of up to 40 mg per litre of culture when grown in LB medium or 87 mg per litre of culture when using TB medium (data not shown). Gel filtration showed that nAGE10 was present as a dimer, and a melting temperature of 72.3°C was determined by DSC. The SDS–PAGE gel and the chromatograms from the gel-filtration and DSC experiments are presented in the supporting information. Together with BoAGE and SynAGE, nAGE10 is the third report of a high-yielding AGE that does not form inclusion bodies when expressed without chaperones (Datta, 1970; Klermund *et al.*, 2013, 2015; Lee, Chien *et al.*, 2007; Maru *et al.*, 1996; Sola-Carvajal *et al.*, 2012; Tabata *et al.*, 2002).

3.7. Verification of nAGE10 activity

In order to verify that the purified nAGE10 was active, a one-pot, coupled reaction with AsNAL was performed and the production of Neu5Ac was detected by the TBA assay (Fig. 7). The reaction was performed at 25°C, according to the optimal working conditions for the NAL. The yield from the condensation reaction using ManNAc as a substrate ('No AGE'; grey bar) was defined as 100% for the purposes of comparison with the coupled reaction. The results show that the synthesis of Neu5Ac from GlcNAc is possible and that 40% of the maximum yield could be achieved under the assay conditions. It is worth noting that different conditions can give a much higher yield (data not shown). For the condensation reaction, the presence of active AGE (*i.e.* with ATP) reduces the yield starting from ManNAc by 28%, showing that nAGE10 is active in both directions. AGE will initially speed up the formation of GlcNAc; however, when the NAL uses more of the ManNAc then the production of ManNAc by the AGE will increase. The incubation times used will determine how severe this effect is. In the absence of ATP, only a fraction of Neu5Ac could be synthetized from GlcNAc, while the condensation reaction was not affected.

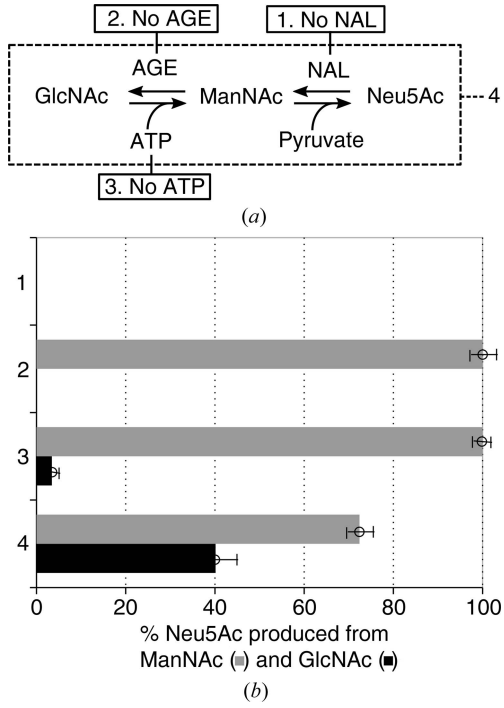


Figure 7
Synthesis of Neu5Ac by coupled epimerase–aldolase reactions. (*a*) Reaction scheme for the coupled reaction. Each reaction is performed using either GlcNAc or ManNAc as a substrate. 1: the reaction is performed in the absence of AsNAL and thus only the epimerization reaction occurs. 2: the reaction is performed in the absence of AGE and thus only the condensation reaction occurs. 3: the reaction is performed in the absence of ATP, affecting the epimerization reaction. 4: complete coupled reaction. (*b*) Detection of Neu5Ac by the TBA assay for reactions 1–4, using either GlcNAc (grey) or ManNAc (black) as a substrate. Production of Neu5Ac is expressed as percentage of the amount produced for the condensation reaction alone (2) using ManNAc as a substrate (grey bar). The reactions were performed at room temperature (25°C) with an incubation time of 1 h.

4. Conclusion

The crystal structure of nAGE10 presents a different dimer organization to that previously published for AGE structures, while the monomeric unit is very similar. The dimerization interface involves the front faces of each AGE monomer, and analysis revealed that this assembly is the most probable dimer association for AGEs. The front-to-front dimer organization leaves open previous hypotheses regarding the location of the ATP-binding site and raises the question of whether ATP binding affects the dimer interactions at the interface. This opens new perspectives to analyze and understand the role of ATP in the regulation of AGE activity. Another difference from the previously published AGE structures was the presence of an ethylene glycol molecule and a chloride ion within the active site of nAGE10. Comparison with the active sites of other members of the AGE superfamily, cellobiose 2-epimerases and *yihS*-encoded isomerases, suggests that the chloride ion may be a conserved element of the active site and that ethylene glycol mimics the way that substrates of AGE may bind. In addition to this, it was possible to formulate a hypothesis for the binding of the *N*-acetyl group of hexosamines. nAGE10 can be used for the one-pot synthesis of sialic acid at 25°C when coupled with the NAL from the psychrophilic bacterium *A. salmonicida*. The insight gained from this structure offers new residues to consider for mutation studies. Together, these observations open new paths of investigation with new residues to consider, which may lead to a better understanding of AGEs.

Acknowledgements

The authors would like to thank Seila Pandur for the cloning of the nAGE10 construct into the cloning vector, Adele Williamson for help with the DSC experiments, and Man

Kumari Gurung and Tor Olav Berg for their help with the TBA assay.

Funding information

This work was funded by the Norwegian Research Council as part of the NorzymeD project (project ID 221568) and the University of Tromsø – The Arctic University of Norway.

References

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

Ahmed, Z., Timsah, Z., Suen, K. M., Cook, N. P., Lee, G. R., Lin, C.-C., Gagea, M., Marti, A. A. & Ladbury, J. E. (2015). *Nature Commun.* **6**, 7354.

Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L., Studholme, D. J., Yeats, C. & Eddy, S. R. (2004). *Nucleic Acids Res.* **32**, D138–D141.

Beitz, E. (2000). *Bioinformatics*, **16**, 135–139.

Carugo, O. (2014). *BMC Struct. Biol.* **14**, 19.

Datta, A. (1970). *Biochemistry*, **9**, 3363–3370.

Denger, K., Weiss, M., Felux, A. K., Schneider, A., Mayer, C., Spiteller, D., Huhn, T., Cook, A. M. & Schleheck, D. (2014). *Nature (London)*, **507**, 114–117.

Du, F., Zhang, M., Li, X., Yang, C., Meng, H., Wang, D., Chang, S., Xu, Y., Price, B. & Sun, Y. (2014). *Biochem. Biophys. Res. Commun.* **452**, 1034–1039.

Edgar, R. C. (2004*a*). *BMC Bioinformatics*, **5**, 113.

Edgar, R. C. (2004*b*). *Nucleic Acids Res.* **32**, 1792–1797.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L., Tate, J. & Punta, M. (2014). *Nucleic Acids Res.* **42**, D222–D230.

Fujiwara, T., Saburi, W., Inoue, S., Mori, H., Matsui, H., Tanaka, I. & Yao, M. (2013). *FEBS Lett.* **587**, 840–846.

Fujiwara, T., Saburi, W., Matsui, H., Mori, H. & Yao, M. (2014). *J. Biol. Chem.* **289**, 3405–3415.

Ghosh, S. & Roseman, S. (1965). *J. Biol. Chem.* **240**, 1531–1536.

Holm, L. & Laakso, L. M. (2016). *Nucleic Acids Res.* **44**, W351–W355.

Hu, S., Chen, J., Yang, Z., Shao, L., Bai, H., Luo, J., Jiang, W. & Yang, Y. (2010). *Appl. Microbiol. Biotechnol.* **85**, 1383–1391.

Itoh, T., Mikami, B., Hashimoto, W. & Murata, K. (2008). *J. Mol. Biol.* **377**, 1443–1459.

Itoh, T., Mikami, B., Maru, I., Ohta, Y., Hashimoto, W. & Murata, K. (2000). *J. Mol. Biol.* **303**, 733–744.

Kaneko, T., Nakamura, Y., Wolk, C. P., Kuritz, T., Sasamoto, S., Watanabe, A., Iriguchi, M., Ishikawa, A., Kawashima, K., Kimura, T., Kishida, Y., Kohara, M., Matsumoto, M., Matsuno, A., Muraki, A., Nakazaki, N., Shimpo, S., Sugimoto, M., Takazawa, M., Yamada, M., Yasuda, M. & Tabata, S. (2001). *DNA Res.* **8**, 205–213.

Klermund, L., Groher, A. & Castiglione, K. (2013). *J. Biotechnol.* **168**, 256–263.

Klermund, L., Riederer, A., Groher, A. & Castiglione, K. (2015). *Protein Expr. Purif.* **111**, 36–41.

Kragl, U., Gygax, D., Ghisalba, O. & Wandrey, C. (1991). *Angew. Chem. Int. Ed. Engl.* **30**, 827–828.

Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.

Lee, J.-O., Yi, J.-K., Lee, S.-G., Takahashi, S. & Kim, B.-G. (2004). *Enzyme Microb. Technol.* **35**, 121–125.

Lee, Y.-C., Chien, H.-C. R. & Hsu, W.-H. (2007). *J. Biotechnol.* **129**, 453–460.

Lee, Y.-C., Wu, H.-M., Chang, Y.-N., Wang, W.-C. & Hsu, W.-H. (2007). *J. Mol. Biol.* **367**, 895–908.

Liaimer, A., Jensen, J. B. & Dittmann, E. (2016). *Front. Microbiol.* **7**, 1693.

Liao, H.-F., Kao, C.-H., Lin, W.-D., Hsiao, N.-W., Hsu, W.-H. & Lee, Y.-C. (2012). *Process Biochem.* **47**, 948–952.

Luchansky, S. J., Yarema, K. J., Takahashi, S. & Bertozzi, C. R. (2003). *J. Biol. Chem.* **278**, 8035–8042.

Maru, I., Ohta, Y., Murata, K. & Tsukada, Y. (1996). *J. Biol. Chem.* **271**, 16294–16299.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Nayar, S. & Bhattacharyya, D. (1997). *FEBS Lett.* **409**, 449–451.

Sola-Carvajal, A., Sánchez-Carrón, G., García-García, M. I., García-Carmona, F. & Sánchez-Ferrer, Á. (2012). *Biochimie*, **94**, 222–230.

Sparta, K. M., Krug, M., Heinemann, U., Mueller, U. & Weiss, M. S. (2016). *J. Appl. Cryst.* **49**, 1085–1092.

Svenning, M. M., Eriksson, T. & Rasmussen, U. (2005). *Arch. Microbiol.* **183**, 19–26.

Tabata, K., Koizumi, S., Endo, T. & Ozaki, A. (2002). *Enzyme Microb. Technol.* **30**, 327–333.

Takahashi, S., Hori, K., Takahashi, K., Ogasawara, H., Tomatsu, M. & Saito, K. (2001). *J. Biochem.* **130**, 815–821.

Takahashi, S., Miura, R. & Miyake, Y. (1988). *Biochem. Int.* **16**, 1053–1060.

Takahashi, S., Ogasawara, H., Hiwatashi, K., Hata, K., Hori, K., Koizumi, Y. & Sugiyama, T. (2005). *Biomed. Res.* **26**, 117–121.

Takahashi, S., Ogasawara, H., Takahashi, K., Hori, K., Saito, K. & Mori, K. (2002). *J. Biochem.* **131**, 605–610.

Takahashi, S., Takahashi, K., Kaneko, T., Ogasawara, H., Shindo, S. & Kobayashi, M. (1999). *J. Biochem.* **125**, 348–353.

Takahashi, S., Takahashi, K., Kaneko, T., Ogasawara, H., Shindo, S., Saito, K. & Kawamura, Y. (2001). *J. Biochem.* **129**, 529–535.

Takahashi, S., Takahashi, K., Kaneko, T., Ogasawara, H., Shindo, S., Saito, K. & Kobayashi, M. (1999). *J. Biochem.* **126**, 639–642.

Tyler, T. R. & Leatherwood, J. M. (1967). *Arch. Biochem. Biophys.* **119**, 363–367.

Wang, S.-Y., Laborda, P., Lu, A.-M., Duan, X.-C., Ma, H.-Y., Liu, L. & Voglmeir, J. (2016). *Catalysts*, **6**, 212.

Wang, T.-H., Chen, Y.-Y., Pan, H.-H., Wang, F.-P., Cheng, C.-H. & Lee, W.-C. (2009). *BMC Biotechnol.* **9**, 63.

Warren, L. (1959). *J. Biol. Chem.* **234**, 1971–1975.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). *Acta Cryst.* D**67**, 235–242.

STRUCTURAL BIOLOGY

Volume 74 (2018)

Supporting information for article:

The crystal structure of the *N*-acetylglucosamine 2-epimerase from *Nostoc* sp. KVJ10 reveals the true dimer

Marie-Josee Haglund Halsør, Ulli Rothweiler, Bjørn Altermark and Inger Lin Uttakleiv Ræder

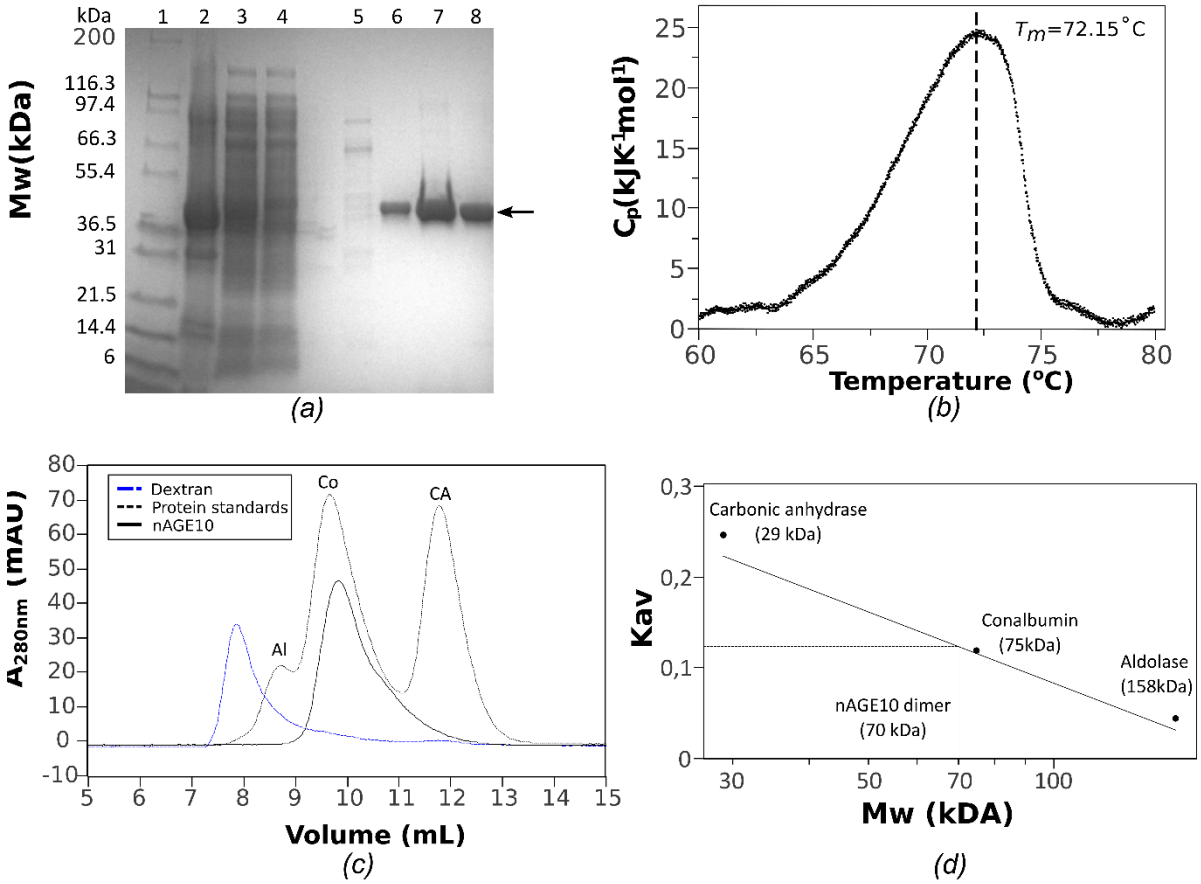


Figure S1 Purification and determination of melting temperature of nAGE10. *(a)* One-step purification of nAGE10. Lane 1: insoluble fraction. Lane 2: soluble fraction. Lane 3: Flow through. Lane 4: Washing at 5% elution buffer. Lanes 5-8: Elution fractions. The first elution fraction (lane 5) represents a peak at 25% elution buffer, while the others (lanes 6-8) encompass a peak at 60% elution buffer. A total of 37,8 mg purified nAGE10 was harvested from those. *(b)* Differential scanning calorimetry (DSC) experiment. The thermal denaturation of nAGE10 was followed from 5 to 95°C at a rate of 1°C/min. The thermogram shows the transition phase between 60 and 80°C. The melting temperature T_m is indicated by a vertical dashed line crossing the transition peak at T=72.15°C. *(c)* Size exclusion chromatography (SEC) of native nAGE10. *(d)* Standard curve for the SEC experiment.