Exploring the role of multiple *tonB* transporters in *Tannerella forsythia*

By

Ahmed Ali Almuntashiri

'180245639'

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

The University of Sheffield

School of Clinical Dentistry

Unit of Oral and Maxillofacial Pathology

10 /11 / 2022

Summary

Periodontitis, a chronic inflammatory disease, leads to the loss of gum, bone, and tooth. Periodontitis is characterised by a dysbiosis in which levels of certain anaerobes increase later during the biofilm development, and the most prominent of these anaerobes are the so-called red complex bacteria, namely, *Tannerella forsythia* (*T. forsythia*), *Treponema denticola* (*T. denticola*), and *Porphyromonas gingivalis* (*P. gingivalis*). *T. forsythia*, a periodontal pathobiont, utilises a dedicated transport and catabolic mechanism to utilise the host-derived sugar sialic acid, which is necessary for its biofilm growth and cell interactions. The best studied member of the sialic acids is 5-N-acetylneuraminic acid (Neu5Ac) which can be found in saliva, mucus, and gingival crevicular fluid. The *nan* operon of *T. forsythia* encodes for transport, scavenging, and catabolic proteins and enzymes for sialic acid utilisation, which in *Tannerella* is dependent on the TonB system (*tonB-exbB-exbD*) and is involved in molecular uptake in a range of bacterial species (e.g., ions, B_{12} etc.). Crucially, this also requires the involvement of TonB-Dependent Transport proteins (TBDT) specific for sialic acid, namely NanO and NanU.

In contrast to model species like *Escherichia coli* (*E. coli*), bioinformatics revealed that *T. forsythia* has four putative *tonB* genes that were identified largely based on identifying a conserved TonB signature domain. These four *tonB* genes are present in all the *T. forsythia* genomes sequenced so far. Here, the function of the multiple putative *tonB* genes to transport Neu5Ac was tested by a mutagenesis programme alongside cloning of *tonB* and sialic acid uptake genes *nanO* and *nanU* for testing of their transport role in a heterologous expression and transport system in an *E. coli* transport assay strain, which lacks it is own *tonB*, as well as the outer membrane sialic acid transporters. As a result, we found that *tonB BFO_0233* was preferentially required to utilise Neu5Ac or the heavily sialylated salivary glycoprotein mucin, indicating a similar situation to the oral cavity. In each observed phenotype, no significant difference was found between all deletion mutants in the presence of NAM.

During a bioinformatics study, we aimed to establish the extent of TonB dependent transporters (TBDTs) in 13 genomes of *T. forsythia*. Therefore, the TBDTs were identified and characterised bioinformatically, and their presence or absence as part of putative PUL, and which TBDT domains were present. In ATCC 43037, for instance, we found 62 TBDTs and predicted their domain structures, 42 of which were associated with (PUL complexes). In contrast, the BU0063, periodontal health-associate isolate, showed a presence of only 35 TBDTs. This study was also able to compare the TonB box of these TBDTs and of the NanO

sialic acid porin and assess the ability of this TBDT to be inhibited by targeting its conserved TonB box via peptides provided exogenously. The resulting data showed that the designed TonB box GASVVE affected not only the uptake of sialic acid, but also the biofilm formation when sialic acid or mucin was utilised as a growth substrate, which may provide a route for future therapeutic treatment for periodontitis.

Additional bioinformatic analysis was used to compare the presence of all virulence factors between 13 genomes in *Tannerella* after re-sequencing *T. forsythia* ATCC 43037 strain. Furthermore, it was recognised that *T. forsythia* ATCC 43037 has a unique outer membrane transporting Neu5Ac and NeuGc, but it was not known whether this porin of sialic acid was generic to nonulosonic acid (Sia-like sugars). We found during bioinformatical analysis that there were higher identities between the sialic acid outer membrane of ATCC 43037 strain (NanO) and other sialic acid outer membrane of 12 *Tannerella's* strains based on nucleotide acid and amino acid sequences.

Bacteria with potential *nan* operon feeding pathways for sialic acid were bioinformatically investigated. The distribution of these *nan* operons among 2500 bacterial species from NCBI and PULDB databases were examined. This analysis showed genomes from these bacterial species that represent 12 orders, eight classes, and four phyla. These bacteria phylum are *Bacteroidetes*, *Proteobacteria*, *Verrucomicrobia*, and *Gemmatimonadetes*. Of the four phyla, this study concluded the presence of *nan* operons in 222 bacterial species, which belonged to 31 bacterial families. The dissemination of 222 *nan* operons isolated from human microbes was 67 compared to 155 non-human microbes. This study elucidates the importance of TonB, *nan* operons, and TBDTs in Neu5A acquisition, which lays the foundation for future studies considering the survival mechanisms of *T. forsythia* in Neu5Ac-restricted environments.

After the role of four *tonB* and *nanO* genes have been investigated along with the potential peptide for their inhibitions, the change of the whole transcriptional profile remains undefined under Neu5Ac and mucin conditions. Further investigation was conducted to examine the whole transcriptional profile of *T. forsythia* and the expression of the genes associated with both Neu5Ac and mucin conditions, imitating their survival and influence within the oral biofilm. Transcriptomic data concluded upregulation of 379 and 200 genes in response to Neu5Ac and mucin conditions, respectively. The entire transcriptional profile of *T. forsythia* in response to Neu5Ac and mucin conditions was described for the first time, which could provide an important basis and experimental direction for additional research into the mechanisms of periodontitis.

Acknowledgments

I would like to thank Allah almighty, the most merciful and compassionate, for providing me guidance, help, and strength during this time. Then, I would like to thank Professor Graham Stafford for his time, knowledge, feedback, and supervision. My appreciation to Dr. Joey Shepherd for her help and supervision. To Mr. Jason Heath and Mrs. Brenka McCabe, thank you all for your help, guidance, and support. I have really benefited from their advice and support. I am so grateful for Dr. Elizabeth K Court, Dr. Ashley Gains, and Dr. Katherine Ansbro who provided me with laboratory help and support. I would like also to thank my lab mates for their laboratory help and support during this time. To the Saudi government and Qassim University, thank you for providing me with this opportunity and the funding to work with it. Honestly, without their help and support, this work would not have been possible. Thank you for believing in me and thank you for giving me this opportunity.

Most importantly, to my mother, father, and family, thank you all for your love, encouragement, prayers, and support. To my mother and father, I recognise and appreciate your huge contribution to my success, and I want you to know that I owe everything I have to you. I feel so blessed to have you both in my life.

To all my previous teachers, instructors, and professors, I learned a lot from you, and I appreciate your help, encouragement, and support during my education journey. Although I may not see many of you again, I will never forget your names. To those who have passed away from my family, friends, and previous professors, I still remember your names, and you are in my thoughts and prayers.

Contact info:

Email: a.monti181@gmail.com

aalmuntashiri@qu.edu.sa

Twitter: @AAAmuntashiri

Table of Contents

List of Figures

List of Tables

Chapter I Literature Review

Introduction

The human microbiome has an essential role in many physiological functions including human development, physiology, nutrition, and immunity.[1] Bacterial communities associating with the oral microbiome can cause periodontal disease.[1] Periodontal disease (Periodontitis) is a noncommunicable disease inducing inflammation and some chronic immunological diseases.[2] The World Health Organisation has classed periodontal disease as a "global burden."[3] In developed countries, the prevalence of overall periodontal diseases affecting the population ranges between 30% to and 70%; whereas severe periodontitis affects between 7% and 15% of the human population worldwide.[4] In The United States of America (USA) for instance, the prevalence of periodontitis in adults is approximately 47% including mild, moderate, and severe periodontitis.[5] In The United Kingdom (UK), the National Health Service spends over £1.6 billion on the treatment of periodontal disease annually.[3] Periodontitis is polymicrobial with a prominent role for the Red complex bacteria, *Tannerella forsythia* (*T. forsythia*), *Treponema denticola* (*T. denticola*), and *Porphyromonas gingivalis* (*P. gingivalis*).[6] While considerable attention has been paid to these red complex bacteria (reviewed in this chapter); there is still much to discover regarding the mechanisms and biology of *T. forsythia's* contribution to periodontitis. As such, treatment of periodontal diseases has remained largely unchanged for 30 years, with non-surgical root debridement still the treatment of choice. While in more serious cases, gingival surgery to reduce pocket depth or restore gum attachment is employed.[7] This highlights the need for improved treatment of periodontitis, with one idea being to target the key pathogens or a key virulence trait to try and restore health to the patient's oral cavity.

1.1. Periodontium and Periodontal Disease

The periodontium describes the structures surrounding and supporting the teeth in the maxillary and mandibular bones, including the gingival tissue and supporting bone.[2] Attacks by oral microbes cause (amongst other factors) periodontal disease, the inflammation of the gum and supporting structures of the tooth (Figure 1.1).[2] Different modifiable and unmodifiable risk factors are correlated with the evolution of periodontal disease.[8] Initiation of this disease is when bacteria accumulate around the tooth on the gingival or subgingival level leading to dental plaque. Bacterial accumulation becomes resistant to clearance by host defence systems and subsequently leads to an inflammatory response in the tissues surrounding the tooth/teeth (Gingivitis). Gingivitis (reversible inflammation) is the first form of periodontal disease and is characterised by red, swollen and bleeding gums.[2]
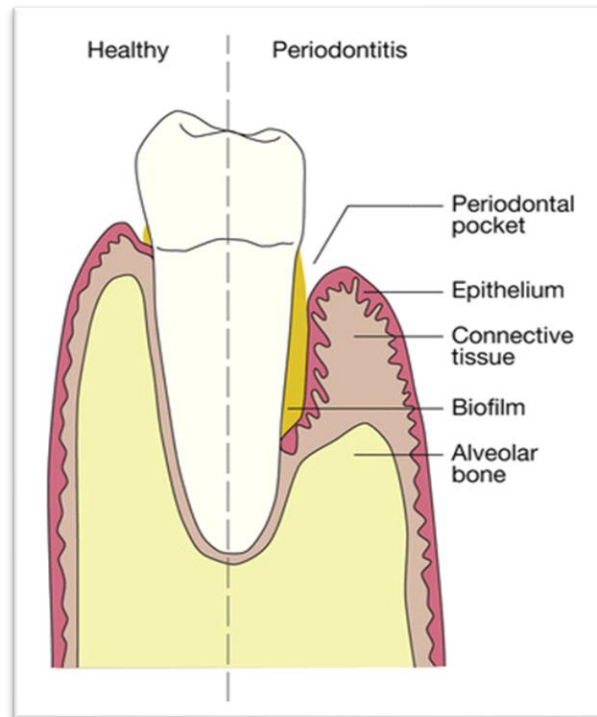
Figure 1.1 Periodontium and periodontal disease.
This figure shows the difference between healthy and unhealthy periodontium. "Reprinted with permission from (Cambridge University Press). Copyright (2013)"

The pathogenicity of these bacteria is enlarged with production of numerous toxins and enzymes. These enzymes and toxins increase pocket depth and create living conditions for red complex bacteria resulting in periodontitis that damage the periodontium and consequences result in many cases in resorption of the alveolar bone and tooth loss.[8]

1.2. Oral Microbiome

The oral cavity comprises discrete microbial habitats, including lip, tongue, gingival sulcus, teeth, cheek, hard and soft palate. The oral microbiota consists of more than 700 species of bacteria, viruses, and fungi that colonise the oral cavity.[9] With a steady transition of oxygen tension and nutrient availability, these microorganisms develop their communities via a dynamic process. This dynamic process enables the bacteria to attach to the oral surfaces, communicate and cohesive among constituent organisms, and adapt to the biofilm (See section 1.6.1.).[7]

1.3. Microorganisms Associated with Periodontal Disease

The first study identifying microbial complexes in subgingival plaque was completed in 1998.[6] Investigation of more than 13,000 samples using cluster analysis and community ordination techniques showed six closely associated bacterial groups, some of which were

observed in association with the healthy state in the mouth and the rest with periodontal disease.[6] Figure 1.2 details these complexes. Four of these complexes are associated mainly with healthy individuals: yellow, purple, blue, and green complexes. The others red and orange complexes are related to periodontal disease and are located within a community of bacteria in subgingival plaque. The orange complex includes *Prevotella* spp.*, Fusobacterium nucleatum* subspecies*, Fusobacterium periodonticum, Campylobacter gracilis, Campylobacter rectus, Streptococcus constellatus, Eubacterium nodatum,* and *Campylobacter showae*.[6] Socransky concluded that the orange complex is the bridge species for early colonisers and red complex bacteria, when the observation of the red complex bacteria was rarely observed in the absence of the orange complex.[6] Besides this, the orange complex can utilise and release nutrients from the dental plaque biofilm for other bacteria. The study suggested that the periodontal disease is a polymicrobial disease, demonstrating Gram-negative anaerobes as the disease-causing bacteria.[6]
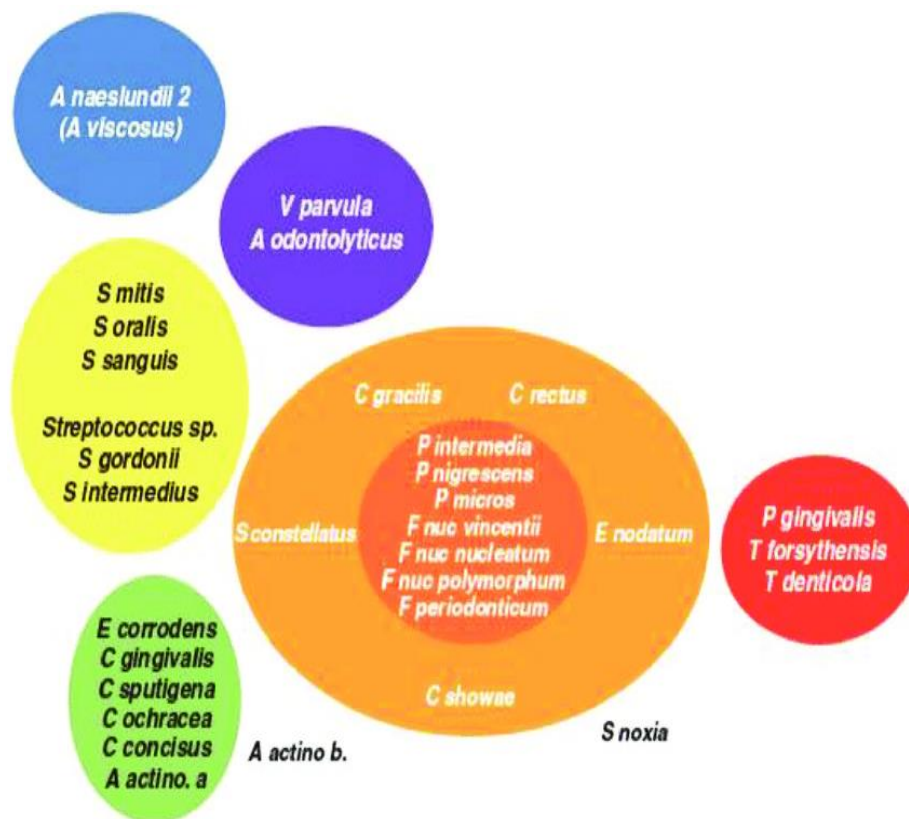
Figure 1.2 The complexes of bacteria initiated the periodontal pocket.
Four of these complexes are related with the oral health: yellow, blue, green, and purple. Two of these complexes are associated with periodontitis: orange and red. "Reprinted with permission from (Ref:5); Appendix VIII: 8.1. Copyright (2005) Journal of Clinical Periodontology. "

An additional study examined Socransky's conclusion and supported the idea of the multiple associated bacterial groups.[10] The study found that the first layer of supragingival plaque contains *Lactobacillus* spp., *Actinomyces* spp., *Streptococcus* spp., and *Candida* spp. The second layer of the subgingival plaque had *Streptococcus* spp., and the Gram-positive aerobes. The study suggested this transition might be due to the availability of nutrients.[10] *T. forsythia*, *Actinomyces* spp., *Spirochaetes*, *F. nucleatum*, and the new *Synergistes* spp were found in the deeper subgingival plaque layer.[10] Furthermore, a recent study used 16S pyrosequencing technology to investigate these complexes via the extraction of DNA. Gram negative bacteria dominance in this disease was confirmed, along with unknown additional microbiome species relating to this disease.[11]

1.4. Red Complex Bacteria

The red complex of bacteria consists of *P. gingivalis*, *T. denticola***,** and *T. forsythia*, and all are considered to be major contributors to the progression of periodontal disease.[12]

1.4.1. *Porphyromonas gingivalis*

The human microbiota consists of abundant members of Gram-negative *Bacteroidetes*, especially in the gut. *Bacteroidetes*, outside the gut, often can cause disease. The oral *Bacteroidete*s with the best-known examples are *P. gingivalis* and *T. forsythia* involved in periodontitis. *P. gingivalis* is a black-pigmented, Gram-negative obligate anaerobe, and is found in over 80% of subgingival plaque samples obtained from patients with severe cases of chronic peridontitis.[13] Different strains of *P. gingivalis* vary in pathogenicity as demonstrated by their differing abilities in causing abscesses in murine models, with some strains being more invasive. Many *P. gingivalis* virulence factors contribute to the progression of periodontal disease. One of these virulence factors is the capsular polysaccharide. This enables the adherence to the oral structures (teeth or mucosal surface), adherence to the other bacteria to resist the force of saliva flow, and aggregation with the orange complex 'bridging pathogens'.[14] Many (but not all) strains of *P. gingivalis* possess a capsule of one of the six serotypes (K1-K6), which have been shown to modulate cytokine suppressors and influence inhibition of the inflammatory response**.**[15] Lipopolysaccharide (LPS) is the major outer membrane component and consists of the core oligosaccharide, the O-antigen and the Lipid A molecules.[16] LPS can influence the host's innate immune response and suppression to the Toll-like receptors-mediated immunity.[17]

Besides this, the LPS assists the bacterium in causing tissue destruction and disrupting the bone remodelling process via inducing the pro-inflammatory cytokines IL-1β and TNF-α expression.[18,19] Following attachment and the subsequent signalling cascade, bacteria progress to invade the host cell as the invasion in general allows for protein phosphorylation, protein synthesis, and ion fluxes. Invasion results in facilitating the replication of this bacterium in a nutrient-rich environment and enabling the tissue destruction.[20] It is proposed that *P. gingivalis* pathogenesis is in its ability to following invasion of the host cell and enabling invasion of other Gram-negative species.[21] In addition, *P. gingivalis* has various molecular mechanisms of invasion, such as fimbriae, gingipains, haemagglutinins and outer membrane proteins. All these various molecular mechanisms of invasion have presented several evidences indicating the involvement of *P. gingivalis* for invasion determinations.[22]

P. gingivalis is unlike other human gut *Bacteroidetes* that specialize in degrading glycans, as *P. gingivalis* is a saccharolytic and feeds completely on peptides for growth. The generation of peptides is through multiple proteases, one of the best known is gingipains, which are abundant surface-anchored and secreted cysteine endoproteases containing trypsin-like activity. Both RagA (a TonB-dependent transporter) and RagB (a substrate-binding surface lipoprotein) are dynamic outer-membrane oligopeptide-acquisition machine that is essential for the efficient utilization of protein substrates in this bacterium as the structure of RagAB contains an ensemble of peptides.[22]

1.4.2. *Treponema denticola*

T. denticola is Gram-negative anaerobic spirochete and is a member of the *Treponema* genus species.[23] *T. denticola* possess a wide variety of virulence factors allowing for host tissue penetration, survival, and replication within the oral environment.[23] The majority of the virulence factors of *Treponema* have cytotoxic activities, such as chymotrypsin. These activities degrade collagen and fibronectin, to enable the interaction of *T. denticola* with epithelial and endothelial cells.[24] This interaction is facilitated via two outer membrane proteins: oligopeptide transporter unit (OppA) and major outer sheath protein (Msp). Msps are surface-exposed loops, such as MspA and MspTL. MspA and MspTL enables this bacterium to evade the immune system and initiate the production of pro-inflammatory cytokines IL-8, TNF-α, IL-1β, and IL-6. These cytokines are initiated via macrophages and epithelial cells to destroy the host tissues through an inflammatory response.[25,26] Of the major virulence factors, dentilisin factor that consists of three lipoproteins encoded by the prcB-prcA-prtP gene locus.[27] Dentilisin factor is a protease complex and possesses proteolytic activity allowing *T. denticola* to induce the permeability

of intercellular junctions and the shrinkage of cells. Furthermore, *T. denticola* has cystalysin, a cytoplasmic protein, that exhibits haemolytic activity and thereby contributes towards pathogenicity.[28] *T. denticola* has other various surface proteins enabling co-aggregation with other periodontal pathogens (*P. gingivalis*, *T. forsythia*, and *F. nucleatum*).[29] This bacterium also shows the capability of adherence to other bacteria and host proteins forming multispecies biofilms.

1.4.3. *Tannerella forsythia*

In 1986, Tanner and others isolated bacteria from patients with progressive periodontitis (deep periodontal pockets) and proposed *Bacteroides forsythus* as a novel Gram-negative species in the *Bacteroides* genus.[30] *T. forsythia* is a late colonizer within the oral biofilm and was described in the beginning as "fusiform (spindle-shaped)" morphology by Anne Tanner.[30] This bacterium was cultured on fastidious agar plates and requires and has a dependency on N-acetylmuramic acid (NAM) for its growth when it was noticed by Wyss that growth defects and morphological changes, such as fusiform morphology, are consequences of NAM depletion impairing cell wall biogenesis.[31] In 2002, a study used 16s rRNA sequencing to analyse *Bacteroides forsythus* to separate it from the other genera.[32] As a result, *B. forsythia* was reclassified as *T. forsythensis* and then renamed '*T. forsythia*' (Figure 1.3). *Tannerella* is found in all forms of periodontal disease, from early gingivitis to aggressive periodontitis.[33] *T. forsythia* requires *N*-acetylmuramic acid (NAM) for its growth, but cannot synthesise it alone.[31] *T. forsythia* was found to cause low alveolar bone density in rats with *P. gingivalis* itself.[34] This suggested that *T. forsythia* can act synergistically and it interacts with the other red complex bacteria for NAM. The shape of this bacterium is rod shaped and the cells have a regular short rod shape, with the Gram staining of light pink colour. Without NAM, *T. forsythia* grows poorly with a mixture of several morphotypes, such as a filamentous fusiform form and smaller rods with spheroids.[34]
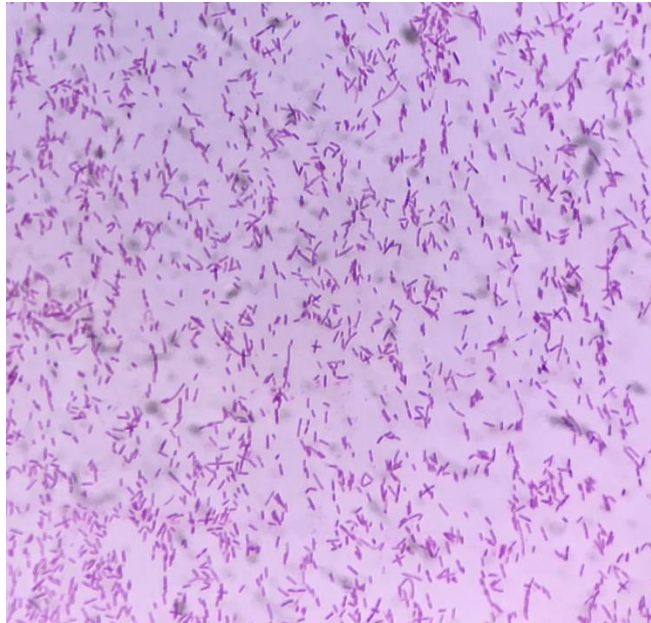
Figure 1.3 Gram-staining of *T. forsythia* grown with NAM.
Four-day-old plate of *T. forsythia* was harvested and placed on a microscope slide before staining.

1.5. Putative Virulence Factors of *T. forsythia*

Virulence factors are essential in pathogenicity of bacteria during periodontitis. While less is known about the virulence factors of *Tannerella*, several virulence factors have been identified. Although the Gram-negative bacteria have an outer membrane structure attaching a lipoprotein structure and linking to a thin peptidoglycan layer, some of these species have complex surface layers known as Surface layers (S-layers).[35] In general, S-layers regulate the cell envelope proteins, are composed of a single protein or glycoprotein (40-200 kDa), and can function as maintenance of cell shape, as protective barriers, and as surface recognition.[36] Of the red complex bacteria, *T. forsythia* is the only member that possesses an S-layer fully covering the bacterial cells that is ~22-nm-thick facilitating adhesion and invasion, suggesting a significant role in the early stage of oral biofilm and periodontal disease, as adhesion to oral surfaces is the initial step in the bacterial colonization. Unlike other bacteria, the S-layer of *T. forsythia* is comprised of water-insoluble proteins, attached to nine monosaccharide units, and is encoded by two large proteins, both of which showed identical glycan mass profiles.[37] The first glycoprotein (230kDa) encoded by the *TfsA* (3.5Kb) gene, whereas the other glycoprotein (270kDa) encoded by the *TfsB* (4.1Kb) gene.[38] This layer is synthesized by a protein *O*-glycosylation system and bound to multiple threonine or serine

residues within an amino acid target motif: D(S/T)(A/I/L/M/T/V). The S-layer is completed in the bacterial periplasm and exported via a type IX glycoprotein secretion system.[39] Glycosylation of this layer allows a change in the physicochemical properties of *T. forsythia* by introducing charged sugar residues and modulating cell hydrophobicity. This makes *T. forsythia* an ideal candidate for influencing bacterium's interaction with other bacteria and host interactions in oral biofilm.[37] S-layers of *T. forsythia* change the expression of IL-1β, and TNF-α that are released by macrophages along with mediating the section levels IL-8 of the human gingival fibroblasts (HGFs), which are the primary and most abundant cells in the periodontium.[39] Deficient S-layer reduces monospecies biofilm formation in a salivary glycoprotein mucin and varies *TF's* localization within the multispecies consortium.[37] Another virulence factor is the outer membrane protein A (OmpA), which is assumed to be antigenic with certain part exposed to the cell surface.[40] *Tannerella* also encodes a protein called BspA (98kDa) containing Leucine Rich Repeat (LRR) motifs and two other bacterial proteins and host receptors.[41] BspA proteins assist in colonisation via binding to fibronectin and fibrinogen of the extracellular matrix.[42] Other protease virulence factors such trypsin-like, *prtH*, karilysin, and mirolase that might damage host tissue and exacerbate infection.[43–45] *T. forsythia* additionally encodes several stress proteins Dps, AhpC, SodF, and OxyR; some of which are involved in oxidative stress resistance.[46] Additionally, *T. forsythia* expresses eight known glycosidases, such as sialidase, cellobiosidase, α-glucosidase, β-glucosidase, arabinosidase, fucosidase, glucosaminidase, and galactosidase.[47]

1.6. Aetiology of Periodontitis

The subgingival environment is challenging to oral bacteria due to the presence of rich immune and inflammatory mediators.[1] These mediators can maintain host–microbe homeostasis in the periodontium resulting in periodontal health. Major environmental factors, such as poor oral hygiene, dietary habits, and smoking can cause alteration in the composition of the microbiome, which leads to several chronic diseases of the mouth.[1] Besides that, non-modifiable risk factors (genetic factors) including the host response is suggested to cause periodontal disease.[1] Several hypotheses were proposed to explain the cause of periodontal disease. In 1970, it was proposed that the presence of a few pathogen species were involved in causing the disease from the large collection of subgingival plaque bacteria.[48] This hypothesis led to the idea that targeting these species might produce prevention treatment. However, Theilade hypothesised that the cause of periodontal disease is a result of overall action of the entire plaque microflora.[49] This indicates that bacterial accumulation, with no specific bacterium, has the ability to cause disease. This hypothesis was supported when

isolation of *P. gingivalis* was found in the periodontium of several healthy patients.[50] Although both hypotheses are seemingly opposing, both hypotheses complement each other as plaque is a polymicrobial infection. In response to both hypotheses, the "ecological plaque hypothesis" described the key elements between these hypotheses.[51] This theory states that an enrichment of oral pathogens causes an imbalance in the resident microflora leading to plaque.[51] To clarify, predominantly Gram-positive aerobic bacteria shift to the more disease-associated Gram-negative anaerobic bacteria. In 2012, the Keystone-Pathogen hypothesis stated that some species have effects on their community leading to their abundance.[21] This theory of microbial dysbiosis expanded where the keystone pathogen population decreases the number of beneficial bacteria alongside an increase in pathogens.[52] The keystone pathogen of periodontal disease is considered to be *P. gingivalis*. This is because this bacterium is capable of instigating inflammation at relatively low numbers and altering the community organisation.[53] Demonstration of this bacterium at relatively low numbers on a mouse model instigated inflammation with alterations to the organisation of the oral microbiota.[54] Overall, two primary factors can contribute to periodontal disease: a microbial factor in subgingival plaque and an immunological factor.

1.6.1. Subgingival Plaque and Biofilm Formation

Biofilms are relevant to many human bacterial infections. It is suggested that biofilm bacteria have differing phenotypes to their planktonic counterparts. One reason is because biofilm supports bacterial tolerance to different environmental conditions, such as pH, temperature, and oxygen levels.[55] In the oral cavity, the biofilm initiates with colonisers of the sub- and supra-gingival surfaces following by community of microorganisms (Figure 1.4).[51] Planktonic bacteria recognise different sites, such as α-amylase from where the pellicle (protein that forms on the surface of enamel) has once been formed. The recognition of these planktonic bacteria in biofilms is influenced by the physical properties of their cell surfaces, such as surface charge, hydrophobicity, or electrostatic interactions.[56,57] For instance, hydrophobic effect interactions assist the adhesion of oral bacterium *Streptococcus sanguis* to the salivary pellicle. Additionally, the formation of a stable biofilm is facilitated by the surface components including fimbriae, pili, or flagella to bind reversibly to the pellicle.[56] The presence of fimbriae is a requirement of *Streptococcus parasanguinis* for stable biofilms whereas loss of flagellum glycosylation affects *Campylobacter spp'* ability to form microcolonies and biofilm.[58] Initial attachment becomes permanent once the biofilm matures, followed by synthesising the outer membrane components to direct binding incapable bacteria.[59]
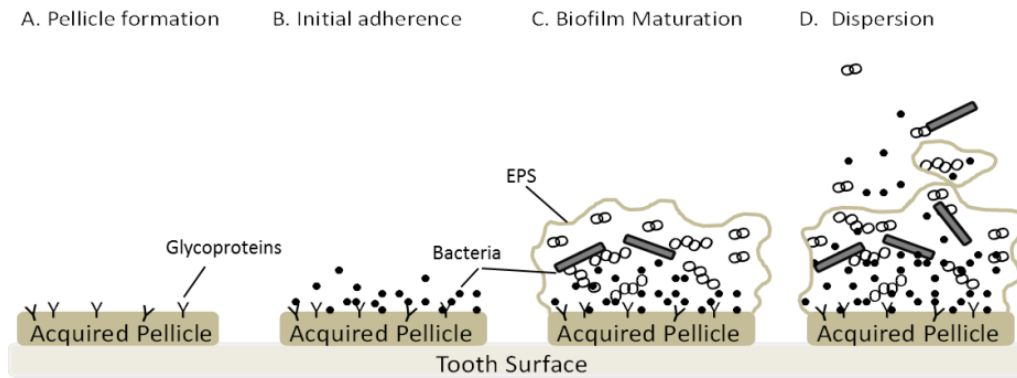
Figure 1.4 Oral biofilm development. A. Pellicle formation comes from salivary glycoproteins attached to the tooth surface. B. This adhesion binds to the glycoproteins. C. Irreversible form of bacterial adherence after it was reversible. D. Spread to colonise a new site. "Reprinted with permission from (Ref:22). Copyright (2016) While Rose eTheses Online."

With the development of early pioneering bacteria, later colonising bacteria recognise the cell surface receptors or polysaccharides leading to the bacterial aggregation and a maturation of the biofilm where water channels and permeable layers permit necessary nutrients to pass in the biofilm.[55] To clarify, co-adhesion is the further process allowing other microbes adherence to already attached organisms on a surface through food chains, gene transfer, and cell–cell signalling. Host proteins, glycoproteins, saliva, and gingival crevicular fluid are the primary nutrients for oral microorganism.[60] These microorganisms break down proteins with mixtures of proteases and peptidases while glycosidases are necessary for the complete degradation of host glycoproteins. Food chains develop with primary organism feeder becoming the main source of nutrients for secondary feeder.[55] This includes catabolism of complex host molecules to metabolism products (e.g. CO_2, CH_4, H_2S). To determine the order of colonization in a microbial community, a study quantified the overlap in metabolic functions of eleven species, whose genomes were sequenced and whose annotations are available.[61] The specific metabolic distance was used to measure the level of variation between the sets of enzymes in the two organisms (Figure 1.5 for more details). These eleven species were found to have smaller average of pairwise metabolic distances to make this build-up favourable.[61] Among biofilm mature, species with array of novel host factors are capable in developing nutritional interdependencies and physical interactions. One of these is *T. denticola* which produces succinic acid to support the growth of *P. gingivalis* and in return, *P. gingivalis* generates isobutyric acid to stimulate the growth of *T. denticola*.[62] Metabolism pathways in different layers of the biofilm are a major factor driving the order of colonization and structured

community as the metabolic activity establishes the phase for transitions between health and disease states.[61]

In addition, it is ideal in biofilm for close proximity of cells to provide conditions for horizontal gene transfer (HGT). HGT increases bacteria adaptive to changes in the oral environment as HGT allows oral bacteria to benefit from metagenome.[63] Extracellular DNA (eDNA) is part of the biofilm matrix and assists in adhesion and in nutrient storage. Of such, eDNA release was found between *P. gingivalis* and *F. nucleatum*.[64]

Organisms are able to communicate with and respond to other microbial cells in biofilm.[62] Responses of organisms depend on the concentration of the signalling molecules. Cell-cell signalling can enable cells to adapt to various environmental stresses and regulate genes influencing the ability of pathogens to cause disease. Autoinducer-2 (AI-2) for example, may be "universal language" for inter-species and inter-kingdom communication.[62] *F. nucleatum* plays a role in periodontal disease, produces AI-2 and showed that this bacterium had a differential effect on biofilm formation. Cultured *F. nucleatum* with *Streptococcus gordonii* had enhanced the biofilm formation whereas cultured *F. nucleatum* with *S. oralis* had reduced the biofilm formation.[65] *F. nucleatum*, *S. sanguinis* and *Streptococcus salivarius* were observed supporting the bridging of early and late colonising species in biofilm like *Tannerella*,[59,62] whereas the *T. forsythia* involvement enhances the attachment and invasion of *P. gingivalis* outer membrane vesicles.[37] In contrast, deficient-gingipain-Lys of *P. gingivalis* reduced growth of *T. forsythia* in multispecies biofilms.[66] Oral biofilm is not only exposed to temperature, pH, salt, but also to oxidative stress surviving environmental challenges and improve the overall fitness of the oral biofilm.[67] In the presence of oxygen, *Aggregatibacter actinomycetemcomitans* feeds on lactate produced by *S. gordonii* while *S. gordonii* is protected by *A. actinomycetemcomitans* from oxidative stress as it does not make its own catalase.[62] *S. gordonii* is suggested to reduce oxygen tension to levels encouraging growth of an obligate anaerobe like *P. gingivalis*, as both interact with their long and short fimbriae.[58] *T. forsythia* was found significantly to upregulate oxidative proteins, including Dps (TF0105), AhpC (TF0383) and Hsp20 (TF2697), which form part of the defence mechanism in *T. forsythia* and promotes persistence in biofilm between periods of periodontal tissue destruction. Likewise, Dps and AhpC of *Prevotella intermedia* were found to be upregulated in biofilm of oral bacteria.[67] The interaction between these pathogens may induce a severe pro-inflammatory response for monocytes and macrophages and lead to periodontal disease.[59]
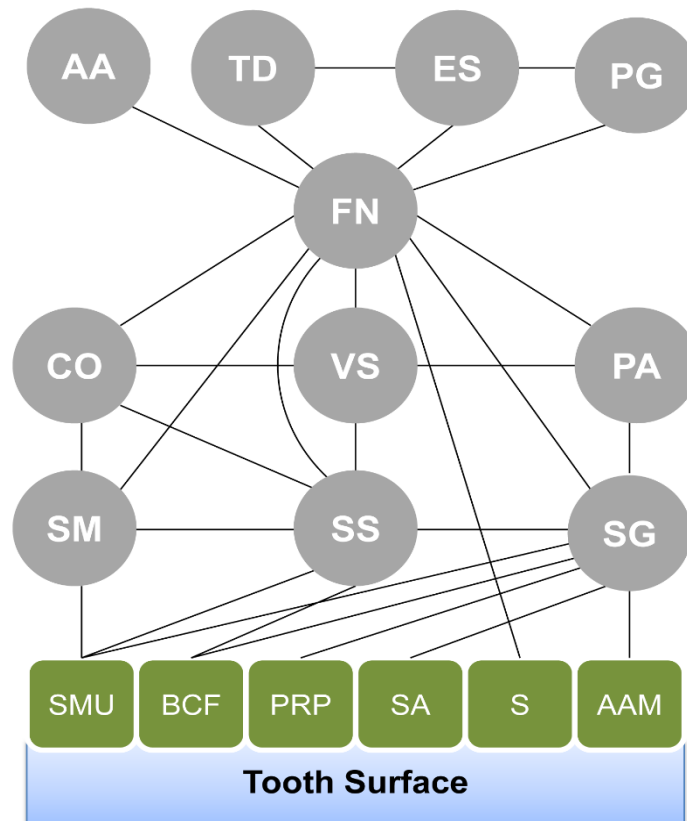
Figure 1.5 Model of dental biofilm.
This figure represents dental biofilm based on the metabolic activity, starting with components of the salivary pellicle (rectangular nodes) and followed by organisms (circular nodes). The lines often illustrate interactions by adhesin molecules. Abbreviations of salivary receptors from right to left are: Alpha amylase (AAM), Statherin (S), Salivary agglutinin (SA), Proline rich protein (PRP), Bacterial cell fragment (BCF), and Sialylated mucins (SMU). Organisms from right to left, Layer 1: *S. gordonii* (SG), *S. sanguinis* (SS), and *Streptococcus mitis* (SM). Layer 2: *Propionibacterium acnes* (PA), *Veionella* (represented by *Veillonella parvula*) (VS), and *Capnocytophaga ochracea* (CO). Layer 3: *F. nucleatum* (FN). Layer 4: *P. gingivalis* (PG), *Eubacterium* (represented by *Eubacterium eligens*) (ES), *T. denticola* (TD), and *A. actinomycetemcomitans* (AA). In this figure, the initial colonizer stage is related to metabolism enzymes utilizing carbohydrate and proline in saliva for energy. The next layer of the biofilm is enriched of fumarase and succinate dehydrogenase to provide a pathway for proline metabolism. Besides that, this layer has a citrate cycle and propionate, all of which require lactate to metabolize carbohydrate and convert it into cytotoxic by-products. This layer also has butyrate affecting gingival epithelial cells to induce apoptosis for providing an enriched food source. The third layer consists of porphyrin metabolism and is essential to *P. gingivalis*. The fourth layer contains nitrogen-related and citrate acid cycle genes (TCA-cycle), which shunts amino acids from tissue degradation into cellular metabolism within the TCA cycle.[61] Copyright: © 2013 Mazumdar et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. https://doi.org/10.1371/journal.pone.0077617.g001

1.6.2. Periodontitis and an Immunological Factor

The variation of oral microbiota compositions depends on different life events: dietary diversification, age, antibiotics, and hormonal change.[68] The oral microbiota can have a beneficial effect for the host including protection against pure pathogens and development of the systemic immune system. These functions are possible due to a qualitative and quantitative balance between the diverse bacteria within the microbiota. However, microbial imbalance due to inflammation altering competitiveness of multiple interactions among the microbes results in some pathogens in the oral cavity increasing in number and decreasing the beneficial symbionts in the community. This disruption of the oral microbiota is dominated by low-abundance microbial pathogens causing inflammatory disease.[68] Presence of biofilm alone is not sufficient to advance into periodontitis where complex interaction is recognised between immune response mediators and the biofilm.[69] Biofilm accumulation results in deprivation of oxygen site and an increased flow of gingival crevicular fluid (GCF). GCF delivers host molecules (e.g. haemin) acting as substrates for proteolytic bacteria along with host defences components (e.g. cytokines, neutrophils, immunoglobulins).[62] For example, *T. forsythia* and *P. gingivalis* in biofilm lifestyle were found to downregulate multiple genes in their butyrate production pathways indicating knock-on effects on epithelial cell behaviour.[67,70] This ideal environment for the growth of obligatory anaerobic and protein dependent bacteria leads to a shift from a symbiotic microbiome to dysbiosis.[69] Inflammation by orange complex for instance, induced micro-ulceration of the sulcular epithelium leading to the leakage of blood and thereby iron into the gingival crevice.[6,69] This resulting environment is conducive for periodontitis-associated species such as *P. gingivalis* and *A. actinomycetemcomitans,* where they induce the destruction of the periodontal tissue and provide new tissue breakdown-derived nutrients for the bacteria by a dysregulated host inflammatory immune response.[69] Adhesion and invasion of *T. forsythia* and *T. denticola* induce the proinflammatory mediators IL-1β, TNF-α, and IL-8(15)[25,26] with a reduction in host-cell cytotoxicity to contribute to the continued periodontitis by *T. forsythia*.[67] Mediating the expression of released macrophage IL-1β, and TNF-α can stimulate bone resorption *in vitro* and *in vivo*,[71] whereas mediating IL-8 attacks neutrophils and develops acute inflammation.[72] S-layer of *T. forsythia* also was suggested to shield the lipopolysaccharide (cell envelope architecture of *TF*) from activating Toll-like receptors along with bacterial DNA of *T. forsythia*.[73] *P. gingivalis* and *A. actinomycetemcomitans* subvert the host response along with other bacteria.[69] Recent studies proposed the action of *P. gingivalis* in three different ways. First, the increase of *P. gingivalis* lipopolysaccharide (LPS) expression to

facilitate survival and multiplication can reduce the TLR4 response.[74] Second, *P. gingivalis* acts on the gingival epithelial cells to inhibit the production of IL-8 through secreting a serine phosphatase.[75] Third, the ability of *P. gingivalis* to produce gingipains (membrane bound and soluble arginine-specific cysteine proteinases) helps to avoid complement-mediated detection. The role of gingipains is to cleave two factors (C3 and C5) and turn them into cell activator (C5a) and phagocytosis enhancer (C3a).[76] The pathogenicity of oral bacteria produce several enzymes and toxins leading to the collapse of the oral health of the host until periodontal disease happens.

1.6.3. Oral Microbiome Variations in Health and Disease

The oral cavity is a gateway for microorganisms entering into the human body.[2] In 2007, The World Health Assembly emphasised the link between oral health and general health, as the oral microbiome causes several oral infectious diseases that lead to heavy monetary healthcare burden.[77] Although the links that may exist between systemic disease and periodontal disease remain speculative due to the difficulty of these periodontal-systemic associations, periodontal disease shows association with numerous systemic diseases and conditions.[2] These systemic diseases include diabetes, pulmonary diseases, cancers, rheumatoid arthritis, and coronary artery disease. The invasion process of oral bacteria and other bacteria products was proposed in two main ways: bloodstream invasion and the digestive tract invasion (Figure 1.6). Bloodstream invasion is possible due to close proximity of vasculature to the periodontal pockets, where the diffusion of periodontal bacteria, bacterial products, immunocomplexes, and mediators of inflammation reach different sites of the human body after invading the epithelium and the connective tissue, then the bloodstream.[78] The second possibility of invasion process by oral bacteria is via the digestive tract. Oral bacteria and other bacterial products and inflammatory molecule disseminate during digestion. Bacteria that survive the acidic pH of the stomach can survive and multiply in the gastrointestinal tract.[79]
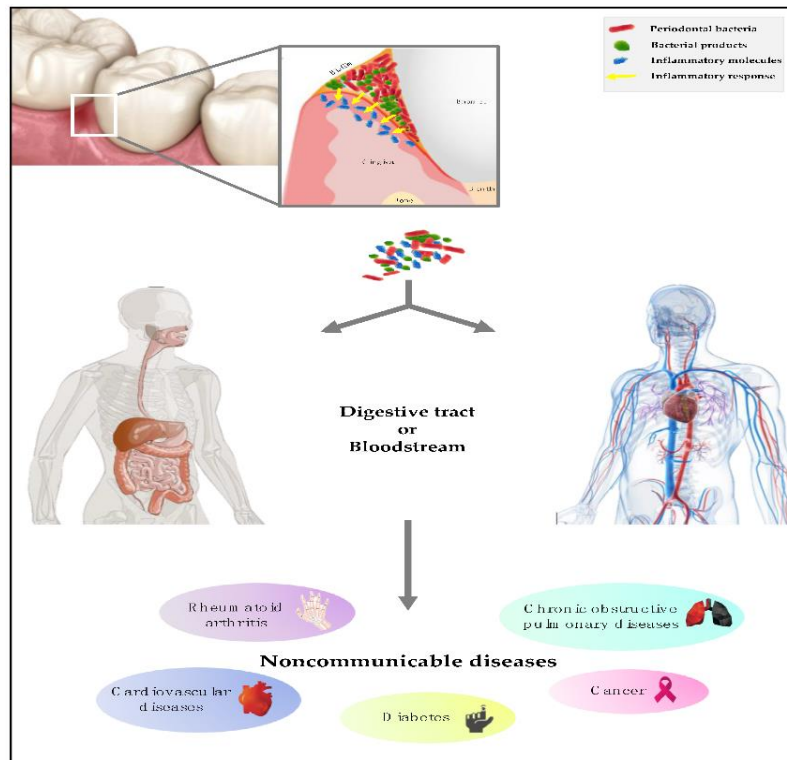
Figure 1.6 Invasion process of oral bacteria.
Explanation of periodontitis effects on human body via digestive tract or bloodstream.
"Reprinted with permission from (Ref:2). Copyright (2019) Journal of Microorganism
http://creativecommons.org/licenses/by/4.0/"

1.7. Sialic Acid Metabolism in Bacteria

The outer membrane structure of Gram-negative bacteria is covalently attached to a lipoprotein structure linking to a thin peptidoglycan layer. Peptidoglycan is a polymer forming a mesh-like layer outside the plasma membrane of several bacteria to construct the cell wall. This polymer consists of amino acids and two other sugars: N-acetylmuramic acid (NAM, also known as MurNAc) and N-acetylglucosamine (GlcNAc).[80] N-acetylmuramic acid contains the ethers (an oxygen atom connected to two alkyl or aryl groups) of lactic acid and N-acetylglucosamine.[81] The chemical formula of NAM is $C_{11}H_{19}NO_8$.[81] While these two sugars are missing from the human bodies, microorganisms have a wide range of glycosidase activities to acquire carbohydrate moieties directly from host glycoproteins, not only from their environment. Thus, microorganisms can substitute for their growth using sialic acids/neuraminic acid (5-amino3,5-dideoxy-D-glycero-D-galacto-2-nonulosonic acid, abbreviated as Neu).[82] In addition, microorganisms can use sialic acids to decorate their cell surfaces, as sialic acids feature at terminal positions of their surfaces.[83] Sialoglycoconjugates are another indication of sialic acids as sialic acid is linked to sugars, such as

oligosaccharides, some polysaccharides, and lipids (glycolipids) or proteins (glycoproteins).[84] For instance, several human glycoproteins are terminally glycosylated with sialic acids, which act as a point of adhesion for bacteria and viruses, as well as have roles in human cell biology such as labelling its self.[85]

The structure of sialic acid (Neu) is represented in N-acetylneuraminic acid (Neu5Ac) (Figure 1.7A), which is an acidic sugar with a group of nine-Carbon sugars, and it is expressed in some bacteria. Neu5Ac is the most prominent member of sialic acid in humans. It is found in the terminal moiety of a variety of human glycoproteins. Such terminal moiety of sugar is present in integrins, TLRs (Toll-like receptors) and the blood group antigens (sialyl-Lewis A/sialyl-Lewis X). Other common sialic acids are derivatives of N-acetylneuraminic acid such as N-glycolylneuraminic acid (Neu5Gc) (Figure 1.7B), which is absent from humans, but it is found in mammals and can be transferred to human cell.[83,85] N-glycolylneuraminic acid is a derivative from CMP-Neu5Ac when an extra oxygen atom is added to the N-acetyl group and catalyzed by CMP-Neu5Ac hydroxylase.[83] The difference between N-acetylneuraminic acid and N-glycolylneuraminic acid is only at position five (C-5) of the carbon rings.[84]
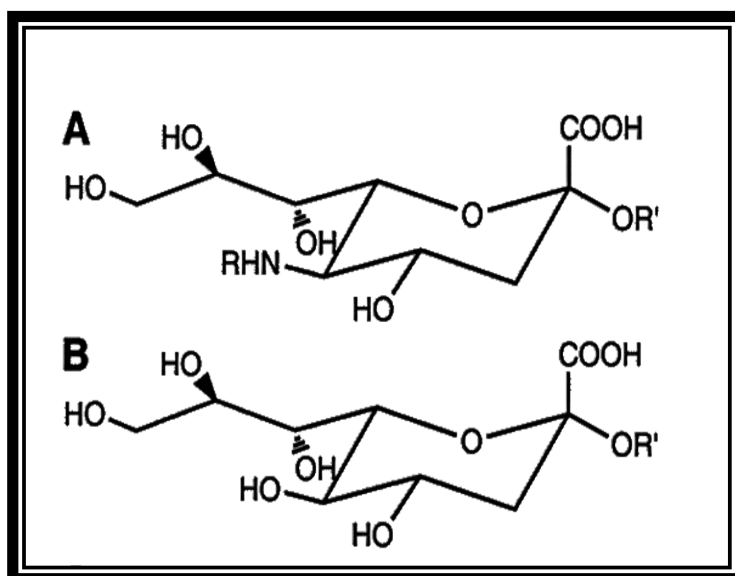
Figure 1.7 Structure of: A. N-acetylneuraminic acid. B. N-glycolylneuraminic acid.
"Reprinted with permission from (Ref:86); Appendix VIII: 8.2 Copyright (2002) American Chemical Society."

Furthermore, diversity of sialic acid structures is generated via a combination of C5 position with multiple modifications of hydroxyl groups at either C4, C7, C8, and C9, as modifications

are less usual than acetylation. These modifications combine different chemical processes including either acetate, lactate, sulfate, or phosphate esters.[84]

In eukaryotic organisms, sialic acid acts (i) as chemical messengers to mediate cell-cell regulation and to stabilize glycoconjugates with different charge repulsion, (ii) control membrane transport and regulate the functions of the transmembrane receptor, (iii) control perm selectivity of the glomerular endothelium and slit diaphragm.[83] Microorganisms use four routes to obtain sialic acid for their cell surface decorations: *de novo* synthesis, donor scavenging, trans-sialidase, and precursor scavenging. *De novo* synthesis is via the biosynthetic genes *neuDBACES* of N-acetylneuraminic acid transporters. Examples of bacterial species which use *de novo* synthesis are *E. coli* K1 and *Mannheimia*. Another route is in microorganisms like *Neisseria gonorrhoeae* that lack *neuABC*. This bacterium uses donor scavenging of extracellular sialyltransferase and CMP-Sia. In addition, some organisms possess a trans- sialidase that does not activate sialic acid or sialyltransferase; instead, it transfers sialic acid from its acceptors to a terminal galactosyl or N-acetylgalactosamine residue of lactose acceptor in species such as a eukaryotic parasite *Trypanosoma cruzi*. Precursor scavenging is an additional mechanism of cell surface sialylation found in microorganisms that lack *neuBC*, but encode a *neuA* orthologue as well as several sialyltransferases, as in *Haemophilus influenzae*. [86,87]

T. forsythia has fastidious growth requirements with one of its unique characteristics being the need for N-acetylmuramic acid (NAM) as a growth factor. The reason for this was later shown to be due to a lack of the second sugar N-acetylglucosamine (NAG).[88–90] *T. forsythia*, unlike other bacteria, depends on exogenous N-acetylmuramic acid (MurNAc or NAM) for growth. A combination of NAM and N-acetylglucosamine (GlcNAc) in bacteria forms the peptidoglycan amino sugar backbone, but *T. forsythia* cannot synthesize its own peptidoglycan amino sugars since it lacks *de novo* synthesis enzymes (MurA and MurB) involved with the formation of the PGN precursor uridine diphosphate-N-acetylmuramic acid (UDP-MurNAc). While *T. forsythia* has an inability to *de novo* synthesize NAM and NAM is not known to be synthesized by the human host and is not available once *T. forsythia* penetrates the epithelial cell, *T. forsythia* has to attain or substitute this compound from external sources for viability.[91] With absence of NAM in the oral cavity, there was a need to understand *Tannerella* survival, as earlier studies highlighted the use of sialic acids (Neu5Ac) as a growth factor for several bacteria.[89,90] Thus, sialic acids were assessed to confirm whether or not they were capable of stimulating the growth of *T. forsythia* biofilms in the place of NAM. Using various concentration of sialic acids (1.5mM to 6mM) as the sole carbon source, it was

demonstrated that *T. forsythia* utilized sialic acids as a growth factor in biofilm, and the growth of *T. forsythia* was increased with sialic acids at higher concentration (6mM). While *T. forsythia* has the ability to use sialic acids (Neu5Ac) as a growth factor, it was essential to understand the ability of *T. forsythia* to utilize alternative forms of Neu5Ac.[88] Sialyllactose is the sugar head group on the GM3 ganglioside and is abundant on mammalian plasma cell membranes. A mixture of 2,3- and 2,6-sialyllactose were used to assess the ability of *T. forsythia* to use it as a growth stimulant. Data show that *T. forsythia* can use both sialic acid sources.[88] Another form of Neu5Ac is Neu5Gc (N-glycolylneuraminic acid) that is produced on the surface of cells in higher (non-human) mammals and present in the human gut or oral cavity from usually dietary sources. It is recognized by the human body as a 'foreign' antigen and incorporated into human glycoproteins.[92] Assessment of *T. forsythia* grown with Neu5Gc in the absence of Neu5Ac and NAM showed its ability to utilize and most likely to transport NeuGc into the cell.[88] As *T. forsythia* can use sialic acid to replace NAM and can attach and colonize epithelial cells, this bacterium can obtain sialic acid for biofilm growth and survival in epithelial cells. *T. forsythia* has been shown to be an essential pillar for biofilm development in the existence of sialylated glycoproteins. Therefore, *T. forsythia* has a novel sialic acid utilisation and uptake system suggesting its survival in the oral community.[88]

1.7.1. Sialidases (Neuraminidases)

Some bacteria can synthesize sialic acids while others can express the sialidase enzyme needed to capture sialic acid.[84] Sialidases are group of enzymes categorised as GH33 Glycosyl Hydrolase Carbohydrate Active Enzymes and have conserved catalytic domains with a 5-6-blade β-propeller structure.[93,94] Sialidase enzymes are encoded through the accessory genes in the *nan* system and can play a nutritional role or contribute to the virulence of some pathogenic organisms such as *P. gingivalis*.[95] Sialidase enzymes share multiple copies of the conserved Asp-box repeats of (Ser-X-Asp-X-Gly-X-Thr-Trp).[96] The action of sialidase enzymes includes release of chemokines from epithelial cells,[97] exposure of sialic acid-masked epitopes for adhesion,[98,99] promotion of biofilm formation,[27] and degradation of host glycoproteins to obtain nutrient.[100] Sialidase structure has two sizes based on the molecular mass and a differential calcium requirement: small sialidase or large sialidase. For instance, small sialidase can be found in *Clostridium perfringens* (40 kDa), whereas large sialidase can be found in *Clostridium tertium* (80 kDa).[101]

Sialidase activity is important in periodontitis and is linked to a poor response to standard treatment. The sialidase in *Tannerella* is *nanH* with the role to access the glycoconjugates: carbohydrates linked with proteins, saccharides, and lipids on buccal and

gingival epithelial cells.[102] In addition, human salivary glycoproteins contain multiple complex sugars including fetuin and mucin.[103] Fetuin has both α2,3- and α 2,6-sialic acid linkages and is found in saliva and gingival crevicular fluid.[103] Mucin also has a bond to N-acetylglucosamine containing linked sialic acid by α 2,6-glycosidic.[104] The role of *T. forsythia nanH* gene was examined in adhesion, and obtaining energy during the formation of dental biofilm. As a result, the *nanH* (part of the *nan* gene cluster-Figure 1.8) was found to use human glycoproteins as a source of sialic acid and to support biofilm growth, as the sialidase gene of *T. forsythia*. The *nanH* also can cleave α-2,6- and α-2,3-linked sialic acid; however, fetuin showed a release of sialic acid at higher concentration compared to mucin.[105] One possible reason for *T. forsythia* having lower affinity for mucin is that mucins has a high level of diacetylation. Diacetylation affects sialidases, which indicates that nutritional source for oral bacteria like *T. forsythia* might be wasted. The human body has a good level of spread diacetylation, which in theory can affect sialidase functions through either inhibition or slow cleavage rates,[106,107] or can modulate sialic acid to assist in the biological roles in development and autoimmunity among others.[108,109]
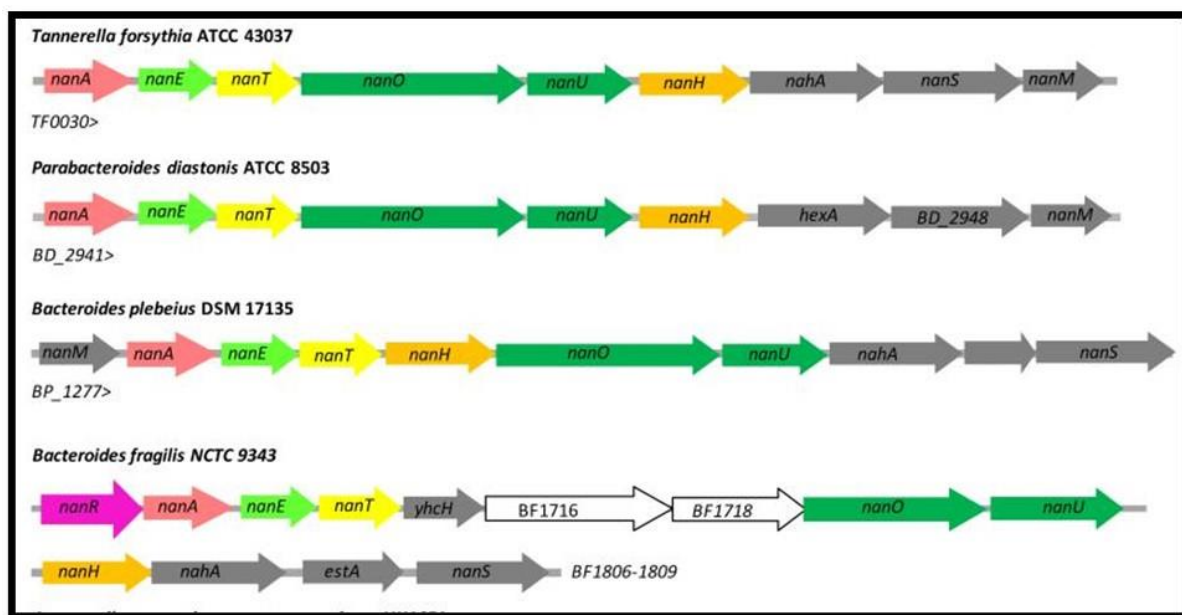
Figure 1.8 Sialic acid scavenge, transport, and catabolism clusters.
This figure illustrates some confirmed or predicated sialic acid gene clusters in the *nan* gene operons of the annotated organisms. Key: inner membrane transporter: *nanT* (major facilitator superfamily permease); TonB-dependent sialic acid transport (TBDTs): *nanO, nanU, nanC*; Catabolic genes: *nanE* (N-acetylmannosamine-6P epimerase), *nanA* (neuraminate lyase); Scavenger gene: *nanH*; other accessory genes: *nanS* (sialic acid 9-O-acetylesterase), *nanM* (sialic acid mutarotase), *nahA*/*hexA* (beta hexosaminidase), *estA* (sialyl

transferase), *yhcH* putative (glycolyl sialic acid processing enzyme). "Reprinted with permission from (Ref:203); Appendix VIII: 8.3. Copyright (2011) Molecular Oral Microbiology."

In other bacteria however, *nanS* produce sialate-O-acetylesterases to help several Gram-negative bacteria survive in mucin-rich environments.[110] The *nanS* assists the survival of bacteria via encoding an esterase enzyme with specificity for acetyl groups to remove the second acetyl group (Diacetylation) and aiding in the efficient harvest of sialic acid.[110] Bioinformatics analysis showed two distinct domains of *nanS* amino acid sequence from residues 1 to 564. First, residues 1-180 are mature proteins in *T. forsythia'* nanS and have homology with SGNH serine hydrolase family. Second, residues 260–564 seem to be related to the DUF303 domain. As the sialic acid in humans contains a second acetyl group attached via an *O*-group, *nanS* acts to remove 9-O-acetyl groups (maybe those at 7-O and 8- O positions) from diacetylated sialic acids in glycoproteins and mucin, and transform them into Neu5Ac.[110] In addition, *nanS* can cleave acetyl groups from the surface of oral epithelial cells (human N-glycans) and can act upon the prevalent sugar in the mammalian hosts (Neu5Gc,9Ac).[110] Involvement of *nanS* with *nanH* increases the access of *nanH* to sialic acid and assists in the efficient harvest of sialic acid from human glycoproteins. Besides *nanS*, two other accessory genes that involve with *nanH* scavenging for sialic acid are *nahA* (a beta-hexosaminidase) and *nanM* (interconversions of isomers of sialic acid). As the *T. forsythia* only possess *nanS* in the red complex bacteria to produce a sialate-O-acetylesterase, this increases the suspicion of providing the sialic acids to other oral microbes.[110]

  Tannerella sialidase has conserved primary sequence motifs like the five Asp-boxes, the F/YRIP motif, and a catalytic arginine triad. Despite other bacteria sialidase revealed function with (pH~5-6) an acidic pH,[48] sialidase activity of *Tannerella* can likely function under a wide variety of pH conditions to enable efficient colonisation of *T. forsythia* in the sub- and supra-gingival plaque biofilms. The optimal activity of *nanH* was in a range of a pH between 5.2 to 5.6, with a range of *nanH* sialidase activity of pH between 4.0 to 7.6 values, where pH 7.6 showed 30% of maximal activity. One possible clarification of this phenomenon is that *nanH* can associate with attached surface to change and shape the local environment of the protein and then optimise the pH.[47] Investigation of *nanH* under using 3-SL and 6-SL trisaccharides showed that *nanH* could cleave both 3-and 6-linked sialic acid under several physiological-mimicking conditions.[47] The physiological salt concentrations in the body can be mimicked at 200 mM of NaCl; however, the presence of NaCl did not significantly affect

affinity or the catalytic efficiency of the sialidase enzyme (*nanH*). Since N-terminal domain of *Tannerella' nanH* is conserved among other sialidase, *nanH* has a broad putative to carbohydrate-binding module (CBM) with a clear affinity for oligosaccharide glycoconjugates.[111]

As *nanH* sialidase role can release α2,3- and α 2,6-sialic acid linkages from the surface of gingival epithelial cells, additional investigation had explored the role of *nanH* on the surface of oral epithelial cells using Lewis antigens from the biochemical prospective. Sialyl Lewis antigens (SLe^A and SLe^X) were used to probe the ability of *nanH* acting on oral epithelial cell surfaces, as Sialyl Lewis antigens are presented in secreted glycoproteins, including salivary mucins. The study used treated cell with either 50 nM *nanH* in PBS, or untreated incubated cell with only PBS. These cells then were stained by either anti-SLe^A or anti-SLe^X Ig. By using fluorescence spectroscopy, *nanH* sialidase reduced the level of SLe^A (from 3.51 to 1.17 MFI/cel) and the level of SLe^X (from 3.26 to 1.03 MFI/cell). This confirmed that *nanH* with broad CBM putative acted on the complex human cell surface glycan.[111] Further investigation indicated that *T. forsythia* survives during biofilm inside the epithelial cells with critical indication of sialic acid uptake.[91]

1.8. TonB-Dependent Transport Systems

Gram-negative bacteria require chemical elements for their growth and supports. The outer membranes of Gram-negative bacteria have specific receptors that bind to different elements with synthesis transporters.[112] Many of these transporters fall into the category of TonB-dependent transporters (TBDTs) representing different ligands from individual zinc ions to polysaccharides. Such different ligands include iron, vitamin B$_{12}$, chemical homo-logues, nickel chelates, and carbohydrates.[112] These elements initially transit from the outer membrane and into the periplasmic space before other proteins facilitate inner membrane translocation.

Study of the sophisticated mechanism of TonB-dependent regulatory system (receptors) and other model organism systems in microbial genetics started in the strain *E. coli* K12. Initially, mutation in *E. coli* resulted in resistance to infection of bacteriophage T1 and the identification of two resistant genes *tonA* and *tonB* (T-one). Ferrichrome transporter (FhuA) was renamed after *tonA* gene (the outer membrane protein) and serves as the transporter for the hydroxamate-type siderophore ferrochrome, as this strain has a mutation to iron chelating agent (enterochelin).[113] Several homologs of Fhu were also noticed in associating with setting of environmental signals, pathogenicity, and uptake of macromolecules. Other roles inlude the ferric enterobactin transporter (FepA), which encodes the outer membrane receptor for enterobactin siderophore. Enterochelin is a cyclic trimer of

2,3-dihydroxy-N-benzoyl-L-serine to make iron solubilize from the insoluble ferric polymers. Once the ferric-enterochelin complex arrives to the cell, ligand moiety secretes hydrolysis enzymes with the assistance of 2,3-dihdroxybenzoylserine fragments to release iron to cellular metabolism. These TonB-dependent receptors (FhuA and FepA) were used as a reference to identify other TonB-dependent transporters/transducers.[113] The features of TonB-dependent transporters (TBDTs) are involved in transport and trans-envelope signal transduction from the outer and periplasmic membranes into the cytoplasm.[113] To achieve this, the TBDT uses a 22-stranded β-barrel to span the outer membrane with a globular plug domain that acts as a receptor for specific substance/metal chelates.[112] The plug domain is folded into the barrel interior with a conserved N-terminal globular domain (Figure 1.9).[114] In other words, the structure can be defined with two key domains, a 22 β-strand barrel which spans the outer membrane and a plug domain held within that barrel.
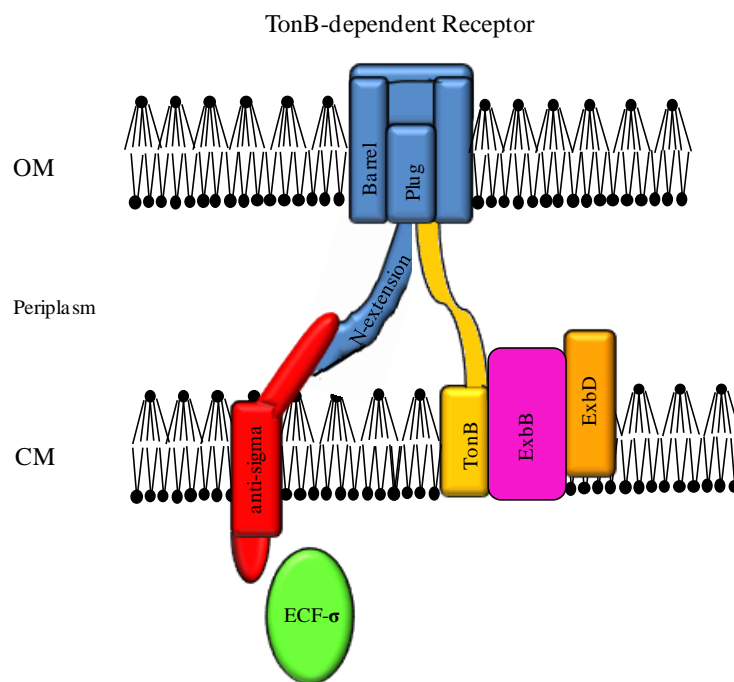
Figure 1.9 Schematic of TonB-dependent regulatory systems.
TonB-dependent transport regulatory systems contain an outer membrane TonB-dependent transducer (Plug, blue) attached to its energizing TonB system (TonB-ExbB-ExbD proteins). A cytoplasmic membrane-localised an ECF-subfamily sigma factor. Abbreviations: OM, outer membrane and CM (cytoplasmic membrane). "Reprinted with permission from (Ref:111); Appendix VIII: 8.4. Copyright (2004), Elsevier."

Several genes encoding the TBDTs in *E. coli* are distributed throughout the chromosome. Collectively these genes include *fhuA*, *fecA*, *BtuB*, *fhuE*, *cirA*, *fepA*, and *fiu*.[113] Two operons *fecABCDE* and *fhuACDB* are customised to bind diferric dicitrate and bind ferrichrome respectively. Downstream genes in both operons encode an ABC transporter for siderophores transit across the cytoplasmic membrane.[112] Transcription of the *fecABCDE* operon is dependent on the extra-cytoplasmic function sigma factor (ECF).[115] Interaction of *fecABCDE* operon with the ECF factors derived from the N-terminal extension on TBDT for the transport and regulation of siderophores. For a complete signal transduction after *fecA* transduces a signal across the outer membrane to *fecR* and *fecI*, the interaction of TonB system (see section: 1:9) at the periplasm via the TonB box (see section: 1:10) is required (Figure 1.10).[112] Although iron is necessary for *E. coli*, it can be toxic as it may lead to the production of reactive radicals.[116] The Ferric Uptake Regulator (Fur) can control iron levels and regulates the genes involved in iron homeostasis via small RNA (sRNA) and RyhB genes.[116] In the absence of iron or iron limiting, Fur cannot bind DNA, which leads to iron derepression via encoded iron transporters genes, and siderophore biosynthesis and iron metabolism proteins.[117,118] In the presence of iron, Fur binds DNA using ($Fe2^+$ cofactor) and blocking the expression of dozens of genes.[119] Fur can represses transcription also of the TonB protein (see section: 1.9.1), and the other two proteins ExbB and ExbD operon by binding upstream of ExbB (see section: 1.9.2), which they will be discuused later in this stduy.[112]
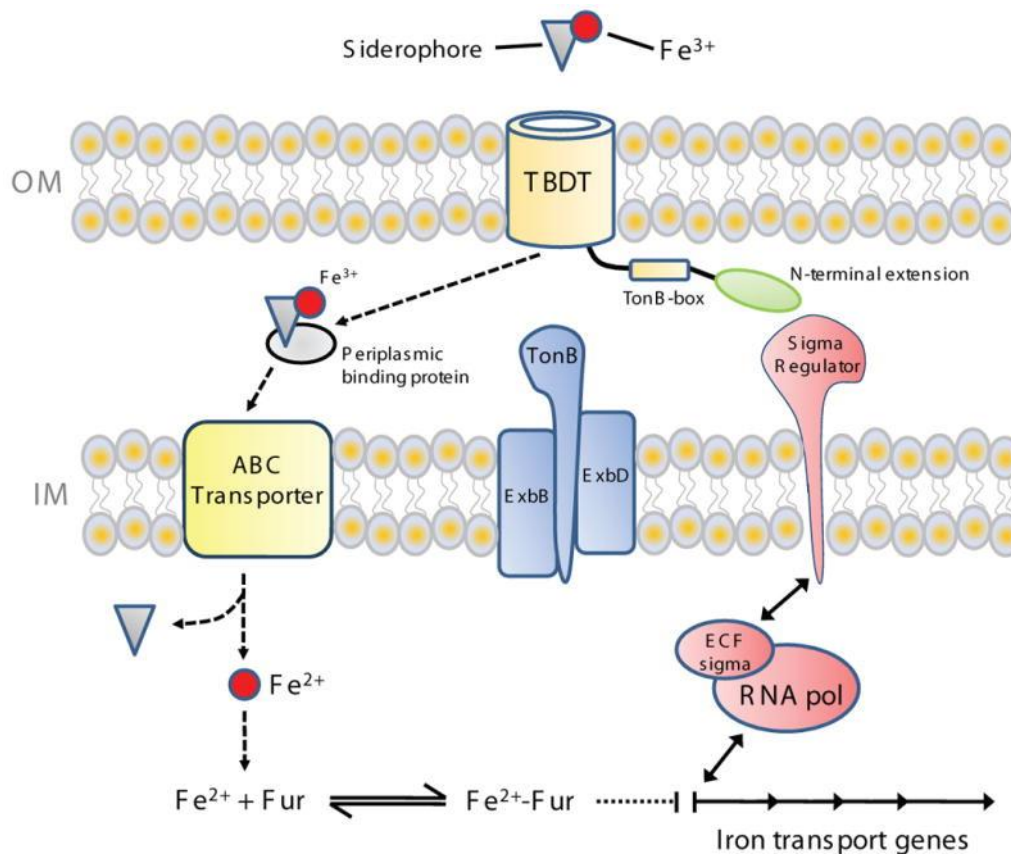
Figure 1.10 Transport and regulation of siderophores.
The outer membrane receptors bind to the ferric siderophores and dependent on the TonB-dependent receptors achieving energy from proton motive force (PMF) in the inner membrane. This energy transduces via TonB system (TonB, ExbB, and ExbD proteins). TonB box motif moved the TonB protein forward at its location to recognise and interact with ferric siderophores. ExbB and ExbD proteins obtain energy from the PMF process to recycle back the TonB for more energy. Transit of ferric siderophores into the inner membrane also requires both the periplasmic binding protein and the ABC transporter. Upon arrival of ferric ion (Fe3+) to the cytoplasm, reduction of this iron into ferrous ion (Fe2+) is destined for either storage or combination into enzymes. Ferrous ion (Fe2+) binds to Fur protein and then transports by the *fecABCDE* genes, which they are regulated by FecA receptor. This initiates the involvement of inner membrane regulator (FecR) and cytoplasmic sigma factor (FecI). "Reprinted with permission from (Ref:114). Appendix VIII: 8.5. Copyright (2010) Annual review of microbiology."

The general structure of these domains is various and suggested to be more than 50 structures based on different bacteria with and without ligand. Blanvillain and others surveyed TBDTs in bacteria showing that 15.5% of bacteria have more than 30 TBDTs, 71% of bacteria have 14 TBDTs or less, and 21% of bacteria do not have TBDTs proteins.[120] They stated also that bacteria with five or more TBDTs/Mpb ratios have an overrepresentation of

TBDTs, which is the case in *T. forsythia* as it has very high TBDT/Mbp ratio of >15.[120] Unlike the well-defined TBDT for iron and other substances uptake, the uptake of iron through TBDT was different in *Neisseria*. It was noticed that the capture of transferrin iron and its uptake required both the TBDT TbpA and the surface-exposed lipoprotein TbpB.[121]

1.8.1. Polysaccharide Utilisation Loci

The human microbiota illustrates a valuable community of microorganisms that impacts host health. This community includes fungi, viruses, protozoa, and bacteriophage where the most well-studied members are the bacteria. There are many bacterial species within the mammal body, the *Bacteroidetes* is one of the dominant phyla of bacteria and contain the largest number of Gram-negative bacteria. Several bacteria phyla including the *Bacteroidetes* and *Firmicutes* can metabolize a variety of complex carbohydrates and are characterised as glycan metabolism. Glycans include starches, fructans, cellulose, hemicelluloses and pectins are component of different plants and humans' diet, which some microbial species cannot digest in the absence of dietary input. Abigail Salyers' laboratory work on *Bacteroides thetaiotaomicron*, a prominent human gut's *Bacteroidete*, provided a template to understand glycans metabolism in these bacteria, which is the cell envelope-associated multi protein system. This system was named Sus (starch utilisation system), which enables the bacterium to bind and degrade this carbohydrate. Later on, microbial genome sequencing showed that many saccharolytic *Bacteroidetes* have derivatives of this prototypic system "Sus-like systems". While this system is capable of substrate binding and degradation, this system highlighted concerted activities of multiple different gene products.[122]

1.8.2. Starch Utilisation System

The prototypic Starch Utilisation System (Sus) contains eight adjacent genes: s*usRABCDEFG* (Figure 1.11).[122] These genes encode different proteins comprising the cell envelope-associated apparatus. Five of these genes (*SusCDEFG*) localize to the outer membrane, whereas the rest localize in the periplasm and cytoplasm. *SusC* is one member of the TonB-dependent receptor family and is a member of the outer membrane-spanning β-barrel proteins that required energy from the proton motive force, and the TonB-ExbBD system for solutes and macromolecules transport. Other four genes (*SusDEFG*) are lipoproteins with a bacterial signal peptidase II recognition motif.[122] Suggestion about these four Sus lipoproteins indicated their exposure and interaction with the external environment, and their location at the

outer leaflet of the plasma membrane.[123] One study mutated both *susC* and *susD* in the wild-type *B. thetaiotaomicron* assessing their interaction on the cell surface. Mutation resulted in inhibition of 60% of the binding and interaction with starch. *SusG* is responsible for hydrolysis, whereas *susEFG* assists in the utilisation. Inclusion of *susE* only with *susCD* increased over 80% of the affinity for starch in the wild type.[123] Both periplasm *susAB* and cytoplasm *susR* have the signal peptidase I recognition motifs. Periplasm *susAB* are more responsible about additional glycoside hydrolases.[122] Disruption of either *susA* or *susB* does not eliminate growth on starch, as one of these genes is sufficient to depolymerize oligosaccharides.[52] One study suggested a preference for *susA* to target the larger oligosaccharides and a preference for *susB* to target shorter substrates like maltotriose.[124] SusR protein spans the cytoplasmic membrane, extends domains into periplasm and cytoplasm, and activates transcription in all seven genes.[122]

Consequently, the Sus-like system *in B. thetaiotaomicron* had been investigated.[122] Searching of Carbohydrate-Active Enzymes (CAZy) Database annotated that the Genome of *B. thetaiotaomicron* has 261 glycoside hydrolases and polysaccharide lyases.[125] Besides that, the genome of this bacterium contains 208 homologs of *susC* and *susD* for targeting other nutrients.[126] These *susC* and *susD* are components of larger gene clusters known as Polysaccharide utilisation loci (PULs). Both *susC* and *susD* of PULs have reminiscent functions (61%) in degrading enzymes and regulators to prototypic Sus.[127] Besides *B. thetaiotaomicron*, polysaccharide utilisation loci are observed in *Bacteroides ovatus* (found in human gut and can grow on plant hemicelluloses) has a trait of PULs (*susD* and *susE*) and *Flavobacterium johnsoniae* (environmental *Bacteroidete*).[122]
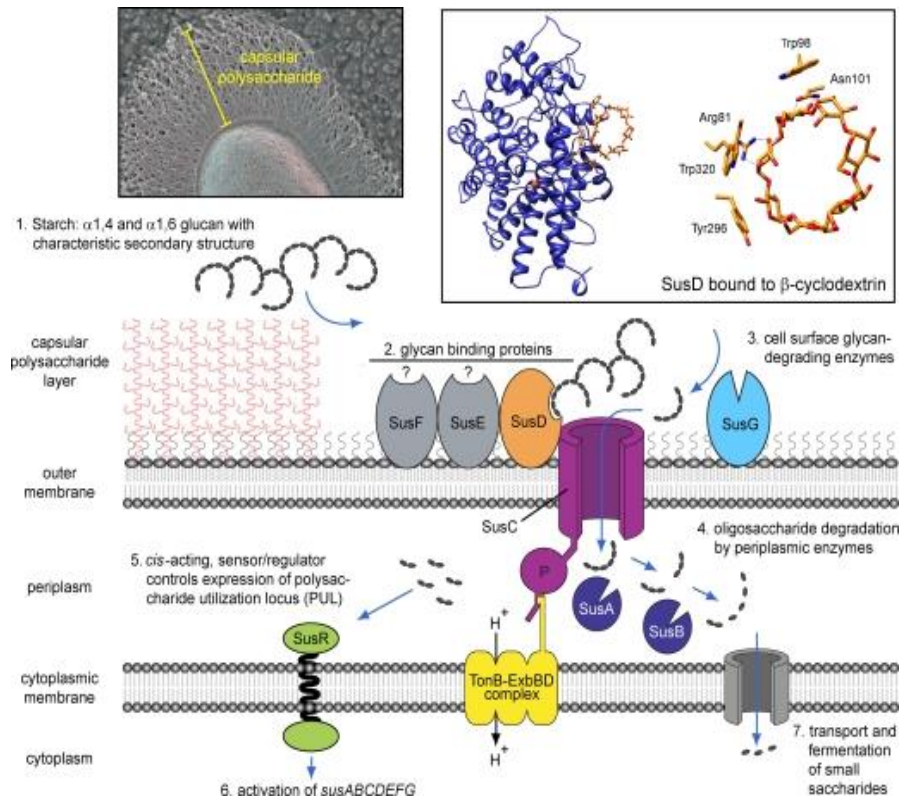
Figure 1.11 Functional model of glycan processing.
The process of starch metabolism starts when the surface capsular polysaccharide layer transit glycans via the intensive efforts of SusCDEF. SusE and SusF interact with higher affinity to starch and its oligosaccharides to sequester it at the cell surface. SusD with its oligosaccharide-binding pocket (composed of an arc of aromatic residues) can bind up to three individual starch, as the starch has the three-dimensional structure of its helices. Then SusG, an outer membrane α-amylase, degrades starch via glycoside hydrolases to generate smaller oligosaccharides, which transported across the outer membrane by SusC proteins into the periplasmic compartment. Periplasmic SusA and SusB degrade oligosaccharides into their component mono- or disaccharides prior to final transport to the cytosol. SusR sensor/regulator activates expression for the free saccharides and then imports depolymerized sugar across the cytoplasmic membrane. "Reprinted with permission from (Ref:127). Copyright (2009), The Journal of Biological Chemistry."

Pairing *susC* and *susD* of PULs can assist in defining the functions of many unknown genes and the discovery of new proteins involved in glycan metabolism.[128] Sequence analysis of *nanO* gene of *T. forsythia* pointed out that it is one member of the TonB-dependent receptor protein (TBDT), and it has similarity to the amino acid sequence of *susC*, the starch utilisation system.[88] The *nanU* gene of *T. forsythia* has similar homology with the *susD* which had never previously been found to have a role in sialic acid transport.[88] Testing the function of *nanO* and *nanU* genes from *T. forsythia* via cloning them into the *E. coli* cloning vector (pCRTOPO-2.1), both genes were able to grow using Neu5Ac or NeuGc as a sole carbon and energy source.

This indicates that the outer membrane transporters of *T. forsythia* encoded by the genes *nanO* and *nanU*.[88] Additional study has investigated more about the outer membrane of *T. forsythia*.[128] As the outer membrane of *T. forsythia* has *nanOU* putative sialic acid-transport genes, similar putative transport genes are available in other several genomes of species in the *Bacteroidetes* group, available in many human pathogens, and available in commensal species.[128] One of which is *B. fragilis* NCTC-9343 that has similar homology putative genes (*BF1719* and *BF1720*) with 69.0% and 63.1% amino acid identity respectively. One way to further investigate the function of *T. forsythia nanOU*, the same putative genes from *B. fragilis* needed investigation in a similar manner to the novel family of sialic acid transporters. Cloned *BF-BF1719* and *BF1720* genes into the *E. coli* system were grown in a similar manner to the *T. forsythia-nanOU* genes, indicating the function of sialic acid transport system (*BF1719* and *BF1720*) in *B. fragilis* with confirming the use of sialic acid in this group of important human-associated bacteria. The role of *nanU* as monomeric high-affinity sialic acid-binding protein (only interacts with free sialic acid and indirect sialylated glycoprotein-derived conjugates) was tested and confirmed with expressing its C-terminal His6 tag in *E. coli BL21λ*(DE3) cells, as well as confirmed by tryptophan fluorescence spectroscopy binding to monomeric only compared to the model sialo-conjugate sugars $3^{'}$- and $6^{'}$-siallyl-lactose- targets for the NanH sialidase from *T. forsythia*. Investigation of *nanU* function with *nanO* function indicated that *nanU* enhanced *nanO* function, while *nanO* and *nanU* are requirement for maximal activity in providing sialic acid to internal catabolic pathways.[128] Therefore, free sialic acids from glycoprotein are bound to the extracellular neuraminate uptake protein (*nanU*), which then pass it to the neuraminate outer membrane permease (*nanO*) (Figure 1.12-B).[128]
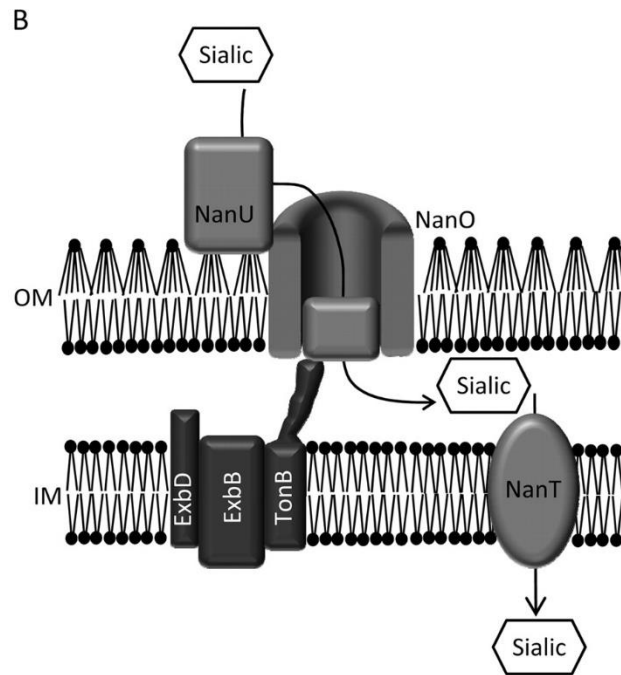
Figure 1.12 Diagram of assumed sialic acid utilisation.
(B). Proposed sialic acid uptake pathway. Sialic acids from glycoprotein are bound to the extracellular neuraminate uptake protein (NanU) before kicking it to the neuraminate outer membrane permease (NanO). The journey of sialic acid starts from the outer membrane into the inner membrane (NanT) before the catabolism process. The TonB system is located in the inner membrane and consists of TonB, ExbB and ExbD proteins. "Reprinted with permission from (Ref:88); Appendix VIII: 8.6. Copyright (2010), American Society for Microbiology."

While there are several studies about the bacteria sialic acid transport inner membrane, outer membrane, and TonB protein systems, the function of TonB system in *T. forsythia* was previously mentioned.[128] As the outer membrane of *T. forsythia* consists of NanOU proteins for sialic acid transport, their functional dependence on TonB system (TonB–exbB–exbD proteins complex) was not clear. A designed system in *E. coli* used to assess the function of TonB system of *T. forsythia* as the *E. coli* involves with sialic acid transport. The constructed system had missing *nanC* and *nanR* as they were sialic acid and its regulator genes, respectively (*E. coli* Δ*nanCnanR*[amber] Δ*ompR*::Tn10[tet]). The *tonB* of this system was deleted and replaced by inserting a kanamycin-resistance cassette (Δ*tonB*::FRT-Km-FRT). Both *nanO* and *nanU* inserted into this system: Δ*nanCnanR*(amber) Δ*ompR*::Tn10 (tet) Δ*tonB*::FRT-Km-FRT. Incubation with either glucose or sialic acid showed that an unaffected growth of *T. forsythia-nanOU* strain on glycose, but in contrast, no growth was observed for *T. forsythia-nanOU* strain on sialic acid. This is an indication that *nanOU* activity requires the function of TonB system (TonB–exbB–exbD proteins).[128]

1.9. TonB System

In Gram-negative bacteria, the outer membrane consists of water-filled channel proteins called porins. Those porins can be specialised porins, and they may have limited sizes for diffusing nutrients into the periplasmic space. However, some molecules cannot diffuse from the exterior of the cell into the periplasmic membrane. Three reasons can explain the incapability transport of these outer membrane porins. First, the outer membrane cannot endure the necessary nutrients as the capability of the outer membrane to pass substances only with sizes less than 600 Da.[129] Second, these nutrients are in low concentrations in the extracellular medium hindering the uptake of such nutrients. Third, the outer membrane and the periplasmic membranes are depleted in chemical energy sources such as nucleotide triphosphates and nucleotide hydrolysis, which they provide energy for transportation.[130,131] Thus, the transport of those elements and nutrients in bacteria requires active transport or energy for their uptake from other sources. Proton Motive Force (PMF), much like the AP synthase, is an energy source that can process the transport by initiating the connection between the outer membrane and the cytoplasmic membrane.[119,132] Once nutrient binds to the TBDT receptor through extracellular loops and apical regions of the plug domain, this plug domain protects the nutrient from diffusion. To activate this energy source, TonB-dependent transporters or outer membrane receptors must interact with the TonB system. This interaction starts with large surface loop motion rearrangements which disseminate via the protein interior and reposition the TonB box to the centre of the β-barrel (Figure 1.13-A). The TonB system is unique system in Gram-negative bacteria, located in inner membrane complex, and consists of either TonB protein or TonB, ExbB and ExbD proteins (Figure 1.12-B). Both ExbB and ExbD partner together to produce the proton translocation part of the motor system while TonB protein can span the periplasmic space and physically contact the outer membrane receptor to energise the transport of nutrients or element into the periplasmic space by transducing the protein motif force across the inner membrane (PMF). This interaction results in the opening barrel domain gate of the TBDT or receptor and then the entry of the bound nutrient into the periplasm.[133]

Interestingly, the reported characterisations of biochemistry and biophysics present the structure of ExbB and ExbD with a range of possible stoichiometries. The possible stoichiometries for TonB system components showed the presence of 3 ExbD and 6 ExbB,[134] 2 ExbD and 5 ExbB,[135] and 2 ExbD, 7 ExbB1, and 1 TonB.[136] This suggests a considerable role of uncertainty about the stoichiometry of the TonB system components that need further explanation.

Nutrient acquisition in the host by Gram-negative bacteria is assembled to oxygenated and non-oxygenated niches. The mechanism function of the TonB system is found to be regardless of oxygenation conditions. In *E. coli* K12, the TonB system is needed to energise the uptake of seven TBDTs whereas the clinical isolates of *E. coli* have between 8 to 22 TBDTs.[137,138] The energy transduction mechanisms of TonB system (TonB/ExbB/ExbD) were explored in aerobic and facultatively anaerobic bacteria. In oxygenated environments, the TonB system is generally for the dominant iron valence acquisition like insoluble ferric iron. In contrast, free ferrous iron, Fe^{3+}-siderophore chelates, siderophore enterobactin, at low pH and/or under anaerobic conditions, is highly rich with multiple reported pathways for iron utilisation dependent on the TonB system.[139]
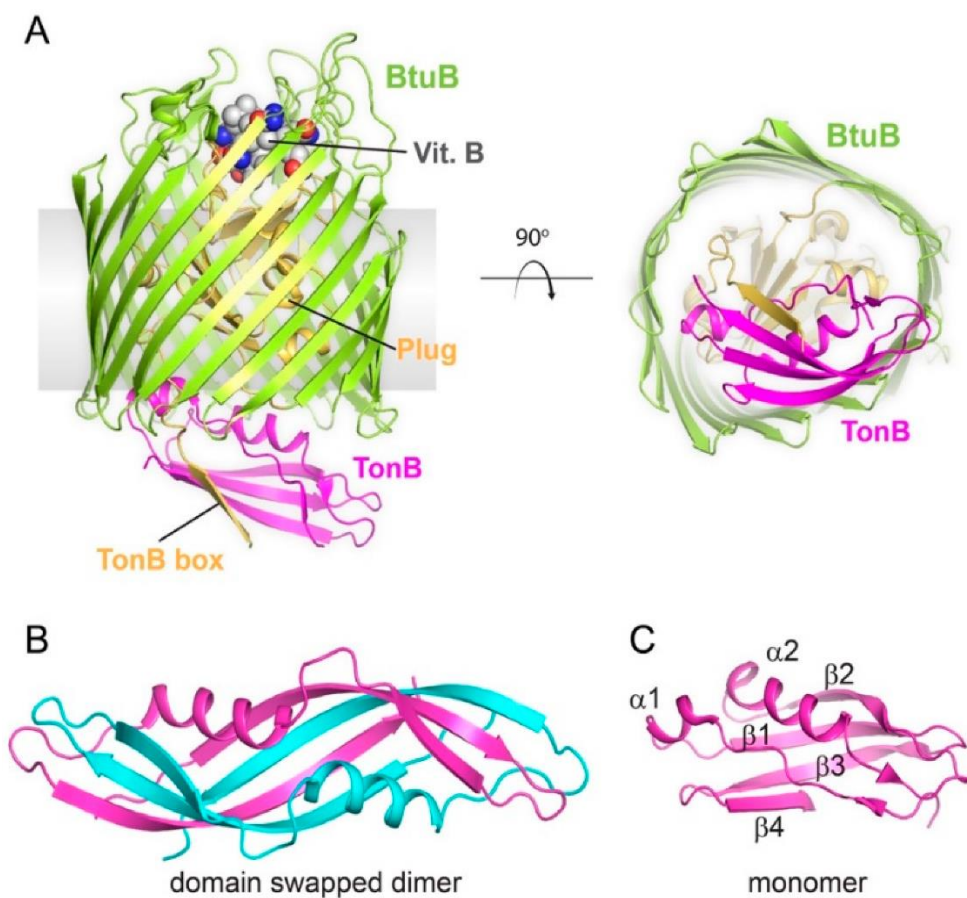
Figure 1.13. Structures of TonB-dependent transporter (TBDT) and TonB protein.
A. The panel using the X-ray structure shows the TonB-CTD (magenta colour) interacts with the BtuB TBDT (green colour). The BtuB TBDT consists of a barrel domain (green colour) and a plug domain (gold colour). On the right panel, the TBDT is shown from the periplasmic space. B. TonB protein is shown in dimer structure with each TonB monomer in different colour using the X-ray structure. C. This panel shows the TonB monomer using the NMR structure. "Reprinted with permission from (Ref.133); Appendix VIII: 8.7. Copyright (2020) MDPI."

1.9.1. TonB protein

In 1970, the TonB system was dicovered, but it took decades before isolation and purification. The role of TonB protein was initially characterised in resistance to *Ti* and *phi80* phages, and to the colicins *B*, *I*, and *V*. This characterisation appeared later as an impairment in these strains' ability to accumulate iron, although in the presence of distinct systems in these strains that were found in accumulation of iron (enterochelin, siderochromes, and hydroxamate siderochromes systems).[140][141] Frost and Rosenberg recognised that iron and other ligand-dependent responses had classical energy-dependent (membrane-associated permeases), as in the absence of specific chelating agents for instance, these cells could still grow slowly and accumulate iron. They examined the relationship between the TonB protein and iron transport (2,3-dihydroxybenzoate and ferrichrome).[142] In *E. coli*, the preparation of *tonB* mutants was via transduction techniques with the generalised transducing phage Plkc. These mutants were tested in liquid media containing either citrate and various hydroxamate siderochromes, or iron in the presence of citrate. To isolate a blocked *tonB* strain in enterochelin biosynthesis, but it has normal iron transport abilities, AN92 strain [with a mutation the *trp* operon and unable to synthesize enterochelin (*aroB-*)] was used for the preparation of transductions. The study transduced *tonB* mutant into strain AN92 from a trp$^+$ donor strain (resistance strain to phage phi80) to result in AN408 (*aroB$^-$*, *tonB$^-$*) and AN409 (*aroB$^-$*, *tonB$^-$*). A mutation in *tonB* gene affected all the known active transport systems for iron.[142] Additional examination of TonB protein role was recognized in other ligand complexes.[131] This understanding of TonB protein came out of studying cyanocobalamin (CN-Cbl) transport across the cell envelope of *E. coli*. Reynolds and others tested *BtuC* gene (one of TBDTs) involvement in B$_{12}$ transport and its role with different transporter genes. By obtaining different *E. coli* mutagenesis strains, the tested *BtuC* indicated reasonability about taking up the B$_{12}$ into the periplasmic space. However, the release and transport of B$_{12}$ required participation of different transporter genes when B$_{12}$ is bound to BtuA domain, as the cobalamin is too large to transport via pores in the outer membrane.[131] Likewise, most of the elements transport in Gram-negative bacteria require the proton motive force (PFM) generated across the inner membrane. The stage and type of this energy for the active uptake examined in *E. coli* with mutants lacking the membrane-bound ATP synthase. In a minimal medium supplemented with vitamin B$_{12}$, all the mutants lacking the membrane-bound ATP synthase retained the transport phenotypes of their parents. In a similar medium, strains with ATP and *btuC* (responsible about B$_{12}$ uptake) mutant cells showed limited uptake to the level of B$_{12}$ binding to the BtuB protein in the outer membrane. This is a strong evidence that the proton motive force is required for B$_{12}$ transport across the outer

membrane.[100] Additional study by Postle and her colleagues has confirmed the need of PMF to iron transport delivering by the TonB system.[143] Besides essential nutrients for bacterial growth, the TonB system can act negatively on a bacterium. Bacteriocin is produced to inhibit the growth of closely or similar bacteria strain and can hijack the TonB system to facilitate binding at the surface through linking siderophore and delivering energy for bacteriocin uptake through the outer membrane. During respiration, TonB system pumps proton motive force into the periplasmic space between the cytoplasmic and outer membranes. This resulted in the generation of a proton gradient across the cytoplasmic membrane with other protons in the periplasm to generate the energy.[100]

1.9.1.1. Location and Crystal Structure of TonB Protein

Attention around TonB protein captured scientists to investigate its location. Initial cellular location of TonB protein was presented via amino and carboxyl termini. Postle and Skare suggested that the amino- terminus hydrophobic (N-terminal domain) can function as an anchor to tether the TonB protein to the cytoplasmic membrane and as an export signal. In addition, the TonB N-terminal domain can facilitate translocation of TonB to the periplasm.[144] Their study's result was consistent with other studies such as chemotaxis chemo-transducers study[145] and leader peptidase study,[146] which they indicated the location of N- terminus in the cytoplasmic membrane. Postle and Skare have suggested also the hydrophobic carboxyl-terminus is on the other side of hydrophobic amino-terminus, but it may be free in the periplasm, anchored to the cytoplasmic, or anchored in the outer membrane.[144] Hannavy and others investigated the membrane topology of *tonB* using alkaline phosphatase (PhoA) fusions to multiple *tonB* amino acids and confirmed that TonB protein is anchored in the cytoplasmic membrane via the amino-terminus.[147] They also determined the location of *tonB* via exploring the carboxy terminus. Their study used *tonB* (239)-PhoA fusion to be adsorbed with *Phi80* phage in order to measure the irreversible *Phi80* adsorption, which would measure *tonB* function; after previous studies had shown adsorption between bacteriophage Phi80 and FhuA receptor. Testing the strain in M9 media, *tonB* (239)-PhoA fusion strain grew and retained some function, which confirmed periplasmic localization of the PhoA for hydrophobic carboxyl terminus.[147] The location of the C-terminus of *tonB* in the periplasm is to localise the outer membrane and contact the TonB box via its N-terminal periplasmic region.[143,148] A later description of TonB location using Nuclear magnetic resonance (NMR) about its central rigid portion along with bioinformatic predictions about TonB domains indicated that the presence of TonB is an inner membrane-anchored protein spanning the periplasm to interact with TBDT. The TonB protein is described as a protein domain containing multiple sheets (i.e. protein

domains) anchored in the cytoplasm membrane via its an amino terminal (putative) transmembrane helix with two domains in the periplasm: proline-rich and a carboxy-terminus. In *E. coli* for instance, the topology studies showed three domains constructing the TonB protein. The first domain of the TonB protein has a short cytoplasmic membrane of 32 residues, 20 residues of which is a single α-helical transmembrane domain. The second domain is a linker domain, to span the periplasm, between 33-150 residues including Pro-rich and flexible regions. The third domain is 90 residues of the structurally well-defined C-terminal domain (CTD).[149]

Similar to other limitations on many protein structures in the protein data bank, there is limited reports for the structures of the TonB system components. Both X-ray crystallography and NMR were used to define the structures of soluble periplasmic domains of TonB. With several reported distinct conformations of C-terminal domain (CTD) of TonB protein, the crystal structures of this domain concluded two conformations.

The structure of TonB-CTD, containing a construct of 85 or 75 residues, was found to be intertwined dimers with different dimerization interfaces (Figure 1.13-B). This dimerization of TonB-CTD shows one α-helix and three β-strands.[150] The TonB dimerization also was investigated *in vivo* to suggest the presence of this TonB structure.[151] Wiener's study mentioned their observation of the second coincident TonB that cannot use the TonB box in binding leading to much lower affinity. They justified that the TonB box is the recognition section for TonB protein and after binding the substrate to the transporter, the transporter can induce the disorder transition in the TonB box.[152] Dyson's study added that the TonB protein interaction with TBDT is temporary as the TonB protein may interact with multiple TBDTs to operate multiple uptake cycles. This suggests that the TonB box helps obtain high specificity with low affinity and sweep out a larger volume within the periplasm. The presence of salt bridge between the TonB and TonB box may be essential for longer range electrostatic attraction and nucleation of β-strand development of the TonB box.[153] A few years later, Postle and others tested the existence of the dimeric TonB *in vivo* as this dimers structure was reported *in vitro* with several similar results bearing the last 85 or 77 residues. Residues of 150 to 239 of *E. coli* TonB-CDT was scanned with Cys substitutions to confirm the presence of dimerization of TonB I232C and N233C. The Cys substitutions positioned TonB I232C and N233C at large distances from one another, suggesting the non-existence of dimeric crystal structures of TonB. Further scanning of TonB-CFT revealed irreplaceable residues.[154] In contrast to the dimers structure, the structure of TonB-CTD in the presence of a longer construct involving 92 residues was found to be monomer (Figure 1.13-B and C).[133,150,155] The crystal structure was

investigated in *E. coli* for TBDTs BtuB and FhuA of B_{12} and ferrichrome ligands, respectively. The TonB protein was monomeric when coupling the BtuB and FhuA TBDTs.[156,157] The monomeric TonB-CTD of BtuB and FhuA confirmed the previous results from Peacock's study of the solution structure of the CTD, which included the first-time residues from 103 to 239. The Peacock's study aimed to investigate the crystal structure of TonB-CDT as the majority of the crystal structure did not include the Gln160 residue of TonB protein enabling interaction with TonB box.[155] Peacock's study claimed that the presence of intertwined dimer is not needed for the TonB protein to initiate an interaction with the TonB box. They justified their claim with residues 165 to 170 of TonB protein which are in an α-helical conformation in the monomeric structure compared to β-sheet in the dimeric structure. This indicated the dimeric structure generated a non-presence of a totally different secondary structure element in the intertwined dimer structure while monomeric structure is known to swap proteins and exchange secondary structure. Furthermore, the environment of β3-strand was buried in the intertwined dimer with the absence of backbone amide or carbonyl groups to another strand generating hydrogen bonds in contrast to the monomeric structure. From residues 197 to 199, the backbone dihedral angles in the monomeric structure enable β2-strand to return to α2-helix, whereas the backbone dihedral angles in the intertwined dimer is extended.[155]

In 2013, Freed and others conducted a further study examining the structure of the TonB protein and TonB Box. They aimed to yield essential insight into the nature of protein coupling between the TBDTs and inner membrane. Previous quantitative binding studies examined and characterised the TonB in solution, but these results were not inconsistent. These studies used different methods with various affinities, such as the choice of detergent, the choice of different soluble TonB fragment due to difficulties in investigating the full TonB, and the choice of a single TBDT concluding different binding stoichiometries. Thus, the fluorescence anisotropy was used to evaluate the binding affinity of a soluble TonB fragment using the following TBDTs FecA, BtuB, and FhuA. Freed's study concluded that the structure of the TonB-CDT in solution is an intertwined dimer to position the TonB protein close to the TBDT via interacting with the peptidoglycan layer whereas changing to monomer upon binding to the TBDT.[158] A few years later, another study examined both extended and flexible C-terminal residues of the TonB protein in *Pseudomonas aeruginosa*. The confirmation of NMR structure and molecular dynamics simulation showed a monomeric structure of the TonB-CTD upon binding the TBDT.[159]

The C-terminal domain of TonB protein was reported with multiple structures based on a common of conserved residues between TonB proteins. Of which, there is two α-helices

bundled with 3-stranded antiparallel β-sheet specifying as $\alpha_1\beta_1\beta_2\alpha_2\beta_3$ topology (Figure 1.13-C).[160] In a monomeric TonB, the TonB helices α1 and α2 are oriented toward the TBDT whereas the β_3 can make a parallel β-strand contact with the TonB box of the TBDT.[160,161] In FhuA TBDT for instance, four residues (Val225, Val226, Leu229, and Lys231) were found to interact with these residues on the FhuA complex (Ile9, Thr10, Val11, and Ala13), forming an interprotein β-sheet with TonB β1 to β3.[156] Cysteine-scanning crosslinking studies identified similar pairwise interactions of TBDT (BtuB and FecA) and TonB proteins. In both TonB:FecA and TonB:BtuB, the TonB Gln160 highly interacts with TBDT residues Leu8 and Val10, the TonB Gln162 highly interacts with TBDT residues Leu8 and -Ala12, and the TonB Tyr163 highly interacts with BtuB-Ala12.[157] In addition, the NMR structure reveals a short β-strand at the C-terminus (β_4) of the monomeric TonB. This additional C-terminus (β_4) is reported to interact with β_3 to structure a 4-stranded antiparallel β-sheet or to facilitate dimerization in the crystal structures. However, this additional short β-strand (β_4) at the C-terminus was reported to be a disorder in *E. coli* once TonB interacts with the following TBDTs, FhuA:TonB, FoxA-TonB, and BtuB:TonB.[160,161]

The crystal structure showed few changes upon interaction of FhuA complex and TonB protein. These changes are occurred in the C-terminal of the plug domain and in the periplasmic of BtuB, FecA, and FhuA barrels. The TonB protein dominates roughly one-part of the periplasm-exposed surface area of this each TBDT complex and closes the barrel lumen from the edge, suggesting the possibility of more than one TonB molecules binding to the TBDT during the uptake cycle.[156] In FhuA for instance, these changes were noticed on several interface residue pairs of FhuA complex and TonB protein where these residues were seen to have a well-defined electron density and bound by hydrophilic interactions. Within the interface of the C-terminal of the plug domain, a single electrostatic interaction was positioned in the switch helix region of TonB protein to form a hydrogen bond between carbonyl oxygen atom of FhuA Ala26 and TonB Arg166. It was previously noticed that ferrichrome binding to FhuA receptor can translocate the FhuA N-terminus and unwind the switch helix. Thus, it appears the interaction and formation of a hydrogen bond between FhuA Ala26 and TonB Arg166 can stabilise FhuA and TonB interaction, as well as unwind switch helix of the FhuA upon ligand binding. Furthermore, and in the periplasmic of FhuA barrel, double electrostatic interactions were noticed between Ala591 and Asn594 of FhuA and Arg166 of TonB forming a hydrogen bond.[156]

1.9.2. ExbB and ExbD Proteins

Suppression of receptor genes by *tonB* mutations pointed out the interaction between the TonB protein and additional receptor proteins as the function of *tonB* involved in transport processes across the outer membrane. The Exb locus has proteins involved in the TonB-dependent transporters and their locus recognised with less sensitivity to colicin B. Mapping of Exb locus identified two genes, termed *exbB* and *exbD* with encoded polypeptides of 244 and 141 amino acids respectively. Their cloning into plasmid could restore growth on ferrichrome as the sole iron source and showed two molecular masses of 26 kDa (ExbB) and 15.5 kDa (ExbD).[162]

The role and process of transcriptional regulation of the ExbBD proteins were not clear. Ahmer and others discussed that if ExbB and ExbD were not transcribed as an operon, then the mutations in each proteins could not distinguish the relative contributions to the energy transduction process.[163] Their result confirmed that both ExbB and ExbD are signal transducers, expressed as an operon, part of the TonB system, interdependent, and divided to opposite side of the cytoplasmic membrane. ExbD protein has a N-terminus with a single transmembrane domain (43-63 residues) similar to the TonB protein with the majority of the soluble domain occupying the periplasm including a flexible linker (46-133 residues), and a C-terminal periplasmic domains (134-141 residues). ExbD forms a homodimer with ExbB and is required for TonB activity, so ExbD accompanies TonB protein into the periplasmic space once TonB pass through the periplasmic membrane, however, it is unknown how the force and movement are propagated to TonB.[164] ExbD forms formaldehyde-linked complexes with the other TonB and ExbB proteins and plays a central role in the proton translocation due to highly conserved Asp residue in its TM domain.[164] This movement during the proton translocation is predicted to be through either or both a piston mechanism or/and a rotation. ExbB consists of three transmembrane domains and has several proteins with the majority of them in the cytoplasm like its carboxy-terminal domain.[143] The mission of ExbB is to stabilize TonB protein in the cytoplasmic membrane when TonB protein interacts with the TonB box region on the periplasmic domain of the TBDT because it does contain conserved residues (Thr148 and Thr181) that may be involved directly in proton translocation.[143] Besides that, ExbB can form formaldehyde cross-linked tetramer and a tetramer cross-linked to ExbD, TonB, and other different proteins.[143] Both ExbB and ExbD appear to harvest the PMF and transduce it into the TonB protein, allowing the TonB carboxy terminus to converse TonB protein from high-affinity outer membrane association to a high-affinity cytoplasmic membrane association.[143]

The absence of either or both TonB and ExbD proteins shows that they are proteolytically unstable. ExbB protein can express by itself and appears to be the scaffold on which ExbD and TonB assemble.[143] However, neither TonB nor ExbB can form ExbD dimer.[164] This led to the quantification of these two proteins in *E. coli* strains. ExbB exists in the cells at the ration of about 3.5-fold compared to one-fold of ExbD, although they are encoded on an operon.[136] Roles of either or both ExbB and ExbD proteins can determine the distribution of TonB protein between the outer membrane and cytoplasmic membrane. With the absence of either or both ExbB and ExbD proteins, the TonB protein associated only with the outer membrane.[165] Further investigations were performed on substitutions of ExbB carboxy terminus (residues 195 to 244). Scanning the carboxy-terminal cytoplasmically localized tail of ExbB resulted in four substitutions at residues: N196A, D211A, A228C, and G244C. All these residues were dominant except G244C. All four residues decreased or prevented the PMF cross-link between *tonB* and *exbD* when they mutated via extra-long PCR. Mutant *exbB* also prevented the proteinase K-resistant conformation of *tonB*. This indicates the signal transduction from cytoplasm to periplasm via *exbB*.[143] Similarly, inactivation of *exbD* prevented efficient formaldehyde cross-linking between *exbD* and *tonB*.[164] Taken together, presence of *exbB* and *exbD* is essential for a *tonB* gene; all are needed to complete the cycle of energy and uptake as a complete set of TonB system.

1.9.2.1. Crystal Structure of ExbD Protein

The ExbD protein is described as three domains where the first domain is an extended N-terminus (~ residue 40 to 65), the second domain is an additional periplasmic structure (~ residue 60 to 135), and the third domain is a short flexible C-terminus (~ residue 30 to 145). The structure of ExbD was shown by NMR to be a dimer and to have consisted of two α-helices assembled against 5-stranded β-sheet.[166] Although there was a prediction of monomer ExbD due to observation of the α-helical transmembrane pore created by the pentamer of ExbB (Figure 1.14-B), the recent Crosslinking and EPR experiments explored more about the complex of ExbB-ExbD concluding a dimer of ExbD (Figure 1.14-C). This observation of dimer ExbD was in the structured C-terminal periplasmic space with the dimerization domain. The presence of dimer ExbD was challenged by the cryoEM using complete ExbB-ExbD subcomplex embedded in lipid nanodiscs.[135] Two ExbD transmembrane I domains were clearly shown by cryoEM in the ExbB pentameric pore. The presence of these two ExbD transmembrane domains (α-helices) are similar to each other and the two important Asp25 residues facing opposite directions to the preserved Thr148 and Thr181 residues, which they create a polar ring within the ExbB pentameric TM pore (Figure 1.14-F).[134] The second

periplasmic domains of ExbD are highly mobile and were not visible in the cryoEM map.[167] Further exploration by cryoEM showed the dimer interface of the ExbD TM α-helices starts from Val20 to Ala41 residues, whereas from Asn12 to Val20 cytoplasmic residues are in an extended conformation.[167] In contrast, Maki-Yonekura and others have a different recent conclusion about the presence ExbD dimer. They stated in their X-ray crystallography and single particle cryo-EM an increase of ExbD transmembrane helices from two to three within the central channel with an increase of the pH (Figure 1.14-B).[134]
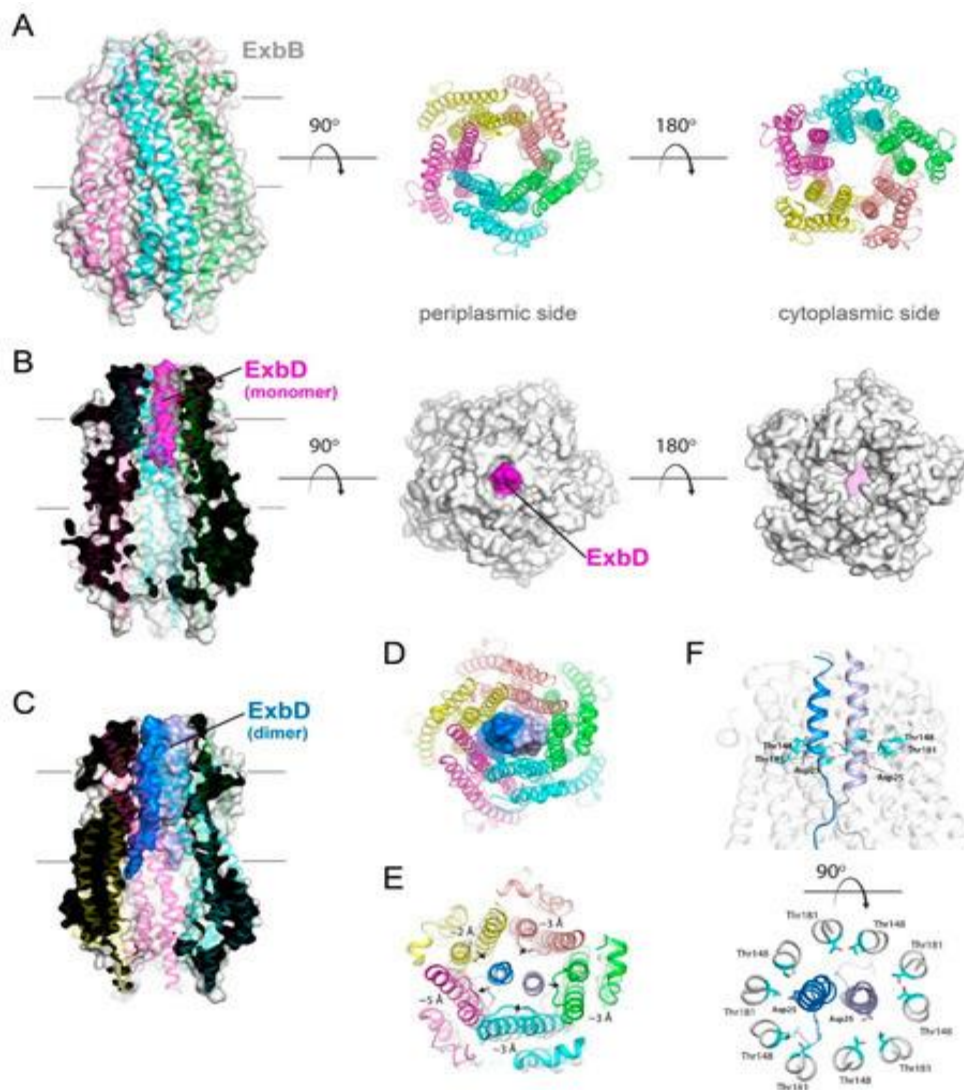
Figure 1.14. Structures of the dimer ExbD and pentamer ExbB subcomplex.
(A) This panel shows the ExbD dimer where the ExbB is a pentamer consisting of 5-fold symmetry on the cytoplasmic membrane and other predicted 5-fold symmetry on the periplasmic membrane. This structure was solved at neutral pH, meaning some structures of ExbD might be disordered in the electron density. (B) At low pH, the X-ray structure was taken for the ExbB/ExbD subcomplex. On the right panel, the magenta colour represents the transmembrane helix for ExbD inside the pore of the ExbB pentamer. On the left panel, the plane of the membrane shows cutaway representation of the ExbB/ExbD subcomplex. (C) A dimer structure of ExbD is shown by the cryoEM structure within the pentameric pore of ExbB. (D) The shade blue colour represents ExbD dimer and shows ExbD dimer is in the surface with each chain of ExbB. This view was down the pore from the periplasmic side using the cryoEM structure. (E) As in panel D, but this view shows above the pore from the cytoplasmic membrane. (F) The conserved residues in ExbB (Thr148 and Thr181) and ExbD (Asp25) are labelled, and these residues are suggested to play a role in proton transport during energy production. "Reprinted with permission from (Ref:133) Appendix VIII: 8.7. Copyright (2020) MDPI."

1.9.2.2. Crystal Structure of ExbB Protein

The stoichiometry of ExbB within the TonB system is still not fully understood during energy transduction accommodating different conformations with the ExbD dimer. The use of NMR reveals the structure of ExbB where ExbB can be copurified with TonB and ExbD or only ExbD. The structure of ExbB was revealed to form stable oligomers on its own.[168] The subcomplex of ExbB-ExbD reveals four copies of ExbB and two copies of ExbD using electron microscopy and biophysical characterisation.[169] Different stoichiometries *in vivo* and *in vitro* can explain inconsistencies in these studies about the ExbB-ExbD complexes.[170] Cottreau used negative stain electron microscopy to claim a stoichiometry of five and six copies of ExbB once ExbB co-expressed with ExbD and in the absence of ExbD, respectively.[170] The conclusion of Cottreau's study was restricted by a limitation of the physiological technique. With native membranes of *E. coli* cells, a study used mass spectroscopy to examine the crystal structure of ExbB, concluding that a pentameric form of ExbB was found with no overexpression of *E. coli*. This provides an overall structure for populated physiological state of ExbB.[171] Two recent additional studies exploring the TonB system using atomic resolution crystal structures showed that the form of ExbB is either a pentamer (Figure 1.14-B) or a hexamer (Figure 1.15).[134,135] These two studies used a full length or truncated construct of ExbD to purify the complex. Both of these studies concluded a monomeric ExbB that consists of short loops linking seven α-helices with different length. This suggests that ExbB can oligomerize as either a pentamer or hexamer supporting Cottreau's conclusion about the presence or absence of ExbD. Pramanik and others suggested the sixth copy is a mixture of ExbB-only oligomers and ExbB-ExbD complex.[168] A large central cavity is created inside both pentamer or hexamer form to direct link between cytoplasm and periplasm holding TM domains of ExbD (Figure 1.15-D).[134,135] All these proposed experimentally structures of ExbB indicate that ExbB has the ability to modulate its lateral contacts forming pentamer and hexamer structures. Ability of ExbB formations can encourage transmembrane domains of ExbD to undertake conformational changes while the TonB system is active and transducing energy.[133]
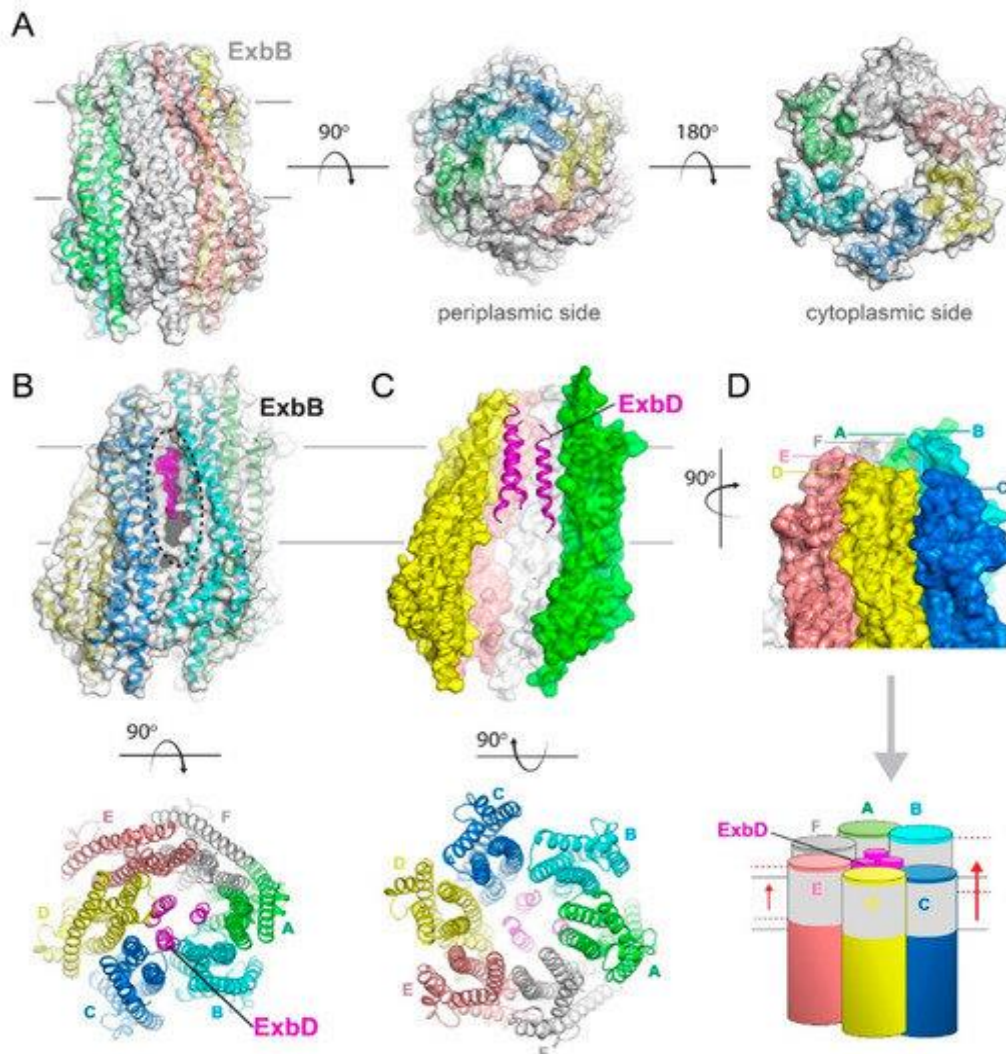
Figure 1.15. Structures of the trimer ExbD and hexameric ExbB subcomplex.
(A) The ExbB is shown in a hexamer structure consisting of 6-fold symmetry on both the cytoplasmic and periplasmic sides. (B and C) These two panels show a trimer of ExbD (magenta) inside a hexamer of ExbB by using the cryoEM structure. Both panels were adjusted in the structures using the cryoEM where each ExbB is monomer, but it generates an unusual large portal that was highlighted by the black dashed oval as seen in panel B, an angle in the complex as seen in panel C, and the ExbB hexamer in both panels is no longer symmetric. (D) This panel presents each monomer of ExbB with trimer ExbD. "Reprinted with permission from (Ref: 133; Appendix VIII: 8.7. Copyright (2020) MDPI."

1.10. TonB Box

The 'TonB box' is a semi-conserved stretch of five amino acids and has a β-strand conformation pairing with the existing β-sheet of TonB protein.[112] It interacts physically with TonB protein to facilitate the transit of TBDTs family. The monomeric form of *E. coli* TonB-CTD was predicted to give a 1:1 complex in binding with various TonB box regions of TBDTs.[155] Later studies with the TBDTs FhuA and BtuB confirmed a 1:1 complex in binding of the periplasmic domain of TonB with TonB box regions.[156,157] The facilitated transition and interaction with TonB protein begins once the Ton-box changes its relative position after the substrate binds to the TBDT. The TonB box, a facilitator of the elements transport, is located based on the type of transported element, such as at the N terminus of the plug domain, into the periplasm, inside the plug domain within the barrel or not visible.[112] In the siderophore binding for example, a study tested its location and mobility in apo- and siderophore transporters.[172] To determine the location of TonB box, the study used spin labelling and electron paramagnetic resonance spectroscopy (EPR). Unfolded TonB box was seen upon siderophore bound TBDTs while the apo structure shows a folded (or ordered) TonB box. This indicates that siderophore binding transduces signals from the outer membrane, which results in unfolding (or increased mobility) of the TonB box. These signals arrive to the TonB system (TonB-ExbB-ExbD) to provide energy, as the TBDT is ligand-loaded.[173] As the TonB system is essential for energy of elements transit, energy in the form of proton motive force is the actual target of transport facilitator. Equally, the hypothesis of all TBDTs is that dependence on the PMF only was examined in the presence of iron through FhuA and FecA. Howard and others mutate the TonB box to assess whether FhuA and FecA could alternatively obtain the PMF without the role of TonB box. A damaged TonB box shows a blockage of TBDT FhuA and FecA indicating that the binding of TonB protein to the TBDTs is not dependent only on an energising step from the PMF, as this conclusion was separately confirmed later with the fluorescence anisotropy.[158,174]

1.10.1. Crystal and Mechanical Structures of TonB Box

Further studies explored the direct interaction between the TonB box and the TonB protein during the uptake of iron and B_{12}. The crystal structure described the contact of *E. coli* TonB where it adapts α1 and α2 helices to contact plug domain through its central β-sheet of the TonB box as it is recruited by the BtuB complex and to make a parallel β-sheet interaction with three-stranded β-sheet of the TonB protein.[157] The hydration of β-sheet of the plug domain renders it susceptible to disruption by TonB protein via transmission of a relatively small force. Further exploration of the interface between BtuB complex and TonB

protein shows an area of 1481 A^2 of this interaction, whereas around 37% of this interface is TonB box:TonB (542 A^2).[157] The conclusion of the TonB box role was further tested by creating a hypothetical model between the *E. coli* TonB-96 and *P. aeruginosa* TonB-96 through superimposing residues into the CTD. This replacement of residues retained the function of the TonB Box interaction with the *E. coli* TonB and BtuB complex. This not only indicated similarity of the TonB box function, but also indicated the interaction between TonB protein and TonB box are mostly hydrophobic due to the highly conserved Valine (Val) and Phenylalanine (Phe).[159] This is consistent with the previous alignment of TonB boxes of BtuB, FecA, and FepA TBDTs as the next residue of Threonine (Thr12) is a position of conserved hydrophobicity occupied by Valine residue. The residue Thr12 in the FhuA TonB box consists of multiple atoms; carbonyl oxygen atom of this residue is the main chain with a hydrogen bond binding the FhuA TonB box.[156]

The crystal structures also showed a direct interaction between TonB box and TonB protein in their β-sheets in the C-terminal domains.[175] Based on the chemical shift using isothermal titration calorimetry measurements during uptake in *E. coli*, it was shown that the TonB box binds specifically to the β3-strand on the TonB-CTD extending the β-sheet.[155] Pawelek and others added to the β-sheets that the presence of a negative charge of Glu56 in the TonB box of FhuA receptor and the presence of positive charge of Arg166 at α1of the TonB can disrupt the plug domain and lower the energy barrier for pore opening, both Glu56 and Arg166 are located in the TEE motif.[156] Likewise, the presence of a positive charge of Arg271 was noticed in *P. aeruginosa* TonB, however, a further exploration of the positive and negative interaction between TonB and TonB box resulted in transferring Lys32 from *P. aeruginosa* TonB as a positive charge to the *E. coli* TonB. This complementation showed the positive charge (Lys32) was able to disrupt the negative charge (Asp6) making a salt bridge in the TonB box of BtuB to facilitate the uptake of B$_{12}$.[159] In addition, Hickman et al. study proposed a model of mechanical pulling of the interaction between TonB box and TonB protein to change the structure of the plug domain and remain stable during unfolding of the plug domain.[175] This mechanical pulling was played by the protein only creating a channel through the TBDT.[175]

Mapping the FecA TonB box to BtuB shows that the TonB protein Gln160 and Gln162 residues crosslinking occurs strongly with TonB box Asp6, Leu8, and Val10 residues, and Asp6, Val9, and -Val10 residues, respectively. Peacock and others characterised the chemical shift changes occurring in TonB protein upon interaction with the TonB box of FecA to conclude the large residues changing in TonB protein (Gln160, Ala167, Gly174,

Ile232, and Asn233).[155] A surface description of the interaction region of the structure between the TonB protein and TonB box illustrates extra restrictions on the TonB box sequence. In the parallel β-sheet of TonB protein and TonB box, this β-sheet formation occurs in TonB box residues 6 to 12 (i.e. BtuB) and TonB residues 226 to 232. This β-sheet shows unknown or no participation of the odd-numbered residues of the TonB box compared to the even-numbered residues contribution in hydrogen bonding. Cadieux and others hypothesized that Proline (Pro) and Glycine (Gly) cannot form inter-strand hydrogen bonds and replacement of the even-numbered residues in the TonB box may damage this β-sheet formation compared to the odd-numbered residues.[176] In the BtuB TBDT, two residues' mutations were introduced in to two *E. coli* strains by replacing Leu8 by Pro and Val10 by Pro in first strain and replacing Thr7 by Pro, Val9 by Pro, and Thr11 by Pro in the second strain. A deficient B_{12} uptake was reported in the first strain, whereas a retained function of B_{12} uptake was seen in the second strain compared to the wild-type.[176] Gudmundsdottir and others replaced each residue from 6 to 12 in the TonB box of BtuB by either Pro or Gly to examine the role of TonB box. A complete uptake inactivity was seen when Leu8 was replaced by Pro and Val10 was replaced by Gly.[177]

Taken together, the structure of plug domain of TBDTs is a four-strand β-sheet and has the characteristic of single-molecule mechanical unfolding experiments.[178] Two previous studies showed a marked anisotropy in the quantity of force required to unfold the plug domain β-sheet. A lower force was suggested in perpendicular to the β-sheet of the plug domain and a higher force was suggested to parallel β-sheet of the TonB protein.[179,180] The interaction of the TonB box and TonB protein forms a four-strand β-sheet; however, the orientation of this interaction was shown to be rotated approximately 90° regardless both structures. This interaction suggested a much-needed force to disrupt its inter-strand interaction before the disengagement of TonB protein from the TBDT.[181] This results in a mechanical pulling force to drive the active transport. This pulling force leads to delivering conformational rearrangement opening the N-terminal of the plug domain of the TBDT and allowing the transport of the substrate via the lumen of the β-barrel. The degree of unfolding plug domain may depend on the size of the substrate which ranges from a few hundred to extra thousand Daltons. The size and shape of the plug domain where the substrate binds may stay within the β-barrel and expose a permeation pathway. Otherwise, the plug domain may move out the β-barrel to unfold uptake.[182] This unfolding hypothesis was tested on disulphide linking of plug domain and β-barrel domain. The disulphide formation was placed at the β-barrel (residue 533) and at the N terminus (residue 27) generating mutants to reveal that the

model of β-sheet was completely folding in the hydroxamate transporter FhuA lacking transport activity.[183] In contrast, in the hydroxamate transporter FhuA, Eisenhauer et al. study disulphide between 109 and 356 residues and between 112 and 383 residues to conclude that the model of β-sheet was partially folding enabling the uptake.[160] Shultis and others mentioned the residues between plug domain structure (i.e. β-sheet, N terminus, and a linker between them) and the attachment of the plug domain on the barrel can separate and modify the interaction between TonB protein and TBDT.[157] They also favour the partial folding of the plug domain as the full folding of the plug domain may prevent a conformational change of interaction with the TonB protein while the TonB box is located on the N-terminus, which was recently proven using mechanical gating of interaction between proteins of outer and inner membranes.[157,175]

1.11. Theories of Peptidoglycan (PG) Architecture

The periplasm space is an overcrowded aqueous compartment including enzymes involved in amino acid and sugar transport, proteases and chaperones for protein quality control, and nucleases and chemotaxis inhibiting uptake of foreign DNA. In the periplasm space, there is a structural polymer tied to the outer membrane, which is known as the PG layer. This layer supports mechanical strength to determine and maintain the cellular shape, and protect the bacterium not only from antibiotics, but also for the recognition of the immune system. The structure of PG consists of conserved repeating units of MurNAc and GlcNAc residues, which they are joined by β-(1,4)-glycosidic linkages. These chains of residues can vary in length from one species to another and shape multiple unique chemotypes of PG.[184]

There are few relevant theories describing the PG architecture and this certainly emphasize various structures of the PG in Gram-negative bacteria. One of which is a theory introduced by Merouech and others (planar network [PN]) where they used the NMR to structurally solve the synthesized N-acetylglucosamine and N-acetylmuramic acid (NAG-NAM) oligosaccharide. They synthesised oligopeptide chains attached to the NAM lactyl group from the NAG-NAM oligosaccharide. This synthesis showed 3-fold symmetry of the pentapeptides on NAM and their cross-linked generated a hexagonal honeycomb, which was similar to the PG matrix. These hexagonal cell, with the glycan chains vertical to the cell surface, were found to organise and support the outer membrane proteins.[185] Another finding theory about the PG architecture supports the vertical scaffold (VS) model. This finding was based on the crystallographic data concluding similar diameter of outer membrane cells within PG predicted matrix and TonB-CTD.[170,186,187] The VS model suggested a regular

framework for the outer membrane proteins indicating the interaction with the TonB protein. These two theories raised a question about the physical impairments to protein-protein interfaces in the periplasm in the PN model. In the periplasm membrane, protein-protein interaction can occur between TBDT and TonB, between inner membrane exporters Pal and TolAQR,[188] and between AcrAB and MdtABC and TolC.[189] This suggests the movement of several interactions including PG and TonB protein. In the VS model using *E. coli*. creation of PG with extra amenable to cross movement of trans-cell envelope proteins, such as TonB was noticed with a lower extent of cross-linking.[190] This model enables TonB protein to be less restrictive and to bind easily to the TBDT.[191] In the PN model, the horizontal array of glycan strands and the hexagonal cells of cross-linked peptides can cause the outer membrane to create physical and conceptual barriers to important interaction, suggesting the possible role of TonB. Yet, the individual residues responsible for binding between PG and TonB protein are not yet established.[184]

1.11.1. Peptidoglycan (PG) and TonB

Progress in the understanding of the TonB protein through its biochemical properties clarified more about the architecture of the Gram-negative bacterial cell envelope. The TonB protein is driven by PMF and has an unexpected role in overall affinities for PG and outer membrane proteins. Prediction of TonB interaction with PG was investigated in a treated TonB with monomeric formaldehyde. A disruption of TonB localisation by high salt to the outer membrane was seen and analysed by Western blotting, suggesting an interaction between TonB and PG through electrostatic interactions.[192] In addition, the role of *E. coli* TonB binding PG was predicted based on similar structures of N-terminal, pro-rich, and C-terminal domains between the TonB protein and the proline-rich (LdtC/formerly YcsF). The proline rich LdtC has affinity for PG as LdtC is transpeptidase to eliminate $_D$-Ala from the PG tetrapeptide stem and links Braun's lipoprotein.[160] An additional role of LdtC is to characterise a family of putative periplasmic proteins (YnhG, YbiS, and ErfK).[193] YnhG, YbiS, and ErfK proteins can confer affinity for PFG and has a hydrophobic possible inner membrane anchor, proline-rich sequence, and a lysin (LysM) motif. LysM domains (27 kDa) are found in PG-degrading enzymes of bacteria, in bacteriophage baseplate assemblies, in PG-binding proteins, in innate immunity proteins of both plants and humans, and in many other protein classes.[194,195] In the presence of a dimeric TonB, analysis showed the LysM motif shared four sites of structural relatedness, and this homologies between both proteins are occurred in a dimeric TonB because PG-binding surfaces include multiple monomeric

elements. This suggests that dimeric TonB expresses affinity for PG while the monomeric TonB attaches and recruits the TonB box.[196]

Furthermore, there is a significant affinity in sequences and structural homologies between both ExbB and ExbD and LysM domain of the PG-binding. The interaction of ExbB and ExbD with the PG in a transient binding leads to frictional resistance, which reduces the amount of its rotational motion and anchors this motion to the rotation of TonB. Furthermore, it is predicted that the LysM motif can confer transient binding to the PG and energised turning of the ExbBD complex, promoting another movement for the ExbBD complex via the cell envelope. This produces a mechanism for catalysis of the substance uptake and catalysis of intermittent adsorption of rotating ExbBD to PG as it was seen in the possibility of protein motion among glycan strands arranged in the VS model.[184]

1.12. Mechanism for Energy Transduction by the TonB System

Several models have suggested energy transduction from the inner membrane to the TBDTs at the outer membrane by the TonB system. In recent decades, Chang and others proposed the propeller model for energy transduction based on their prediction of the TonB dimer. The propeller model suggests that the dimer of TonB attaches to the TonB box of the TBDT and alters the conformation of the plug domain by rotation. This model was challenged by other crystal structure studies that predicted the monomer of TonB upon binding to the TBDT.[155,197] The shuttle model is another suggested hypothesis in which the electrochemical PMF triggered TonB to relocate to the outer membrane from the inner membrane promoting the uptake of substance/metal through the TBDT.[198] However, the hypothesised removal of the TonB N-terminus from the inner membrane bilayer in *E. coli* and reinsertion of this terminal domain to the outer membrane suggests difficulty in reuniting with membrane thermodynamics and biochemistry. Two studies investigated this idea with the physiology of green fluorescent protein (GFP)-TonB fusion proteins to disprove this hypothesis for relocation between membrane bilayers.[154,193] Instead, proton gradient across the inner membrane can generate electrochemical force from the electrochemical proton gradient which the TonB system may harvest. This introduced the third hypothesis of the rotational surveillance and energy transfer (ROSET) model. In the ROSET model, Jordan et al. evaluated this theory of energy transduction of the TonB system from the inner membrane to the outer membrane.[199] They used green fluorescent protein (GFP) to generate a fluorescence system for the TonB analysis concluding the presence of energised motion of TonB in the inner membrane combining with the electrochemical proton gradient by ExbBD.[199] Furthermore, the duration of GFP was analysed to find the ability of TonB protein

to reorient the GFP and emit this light with a different directional vector compared to the polarised excitatory light. It was noticed during GFP analysis that unused PMF with the bacterium exposure to an energy inhibitor, such as 2,4-dinitrophenol (DNP) and proton ionophores carbonyl cyanide *m*-chlorophenyl hydrazone (CCCP) can cause a slowdown in the motion of the TonB protein. Deletion of ExbB and ExbD led to insensitive motion TonB for the same energy inhibitors, indicating the movements of TonB are driven by electrochemical force.[200,201] Additional consideration was gained with anisotropy data for GFP that fluorescence change is dipole in a time frame, with excluding the possibility of conformational change in the membrane regions of TonB system, as well as excluding the possibility of N-terminal hydrophobic helix removal. This led to the most reasonable interpretation of the reorientation of GFP-TonB due you to the rotation of TonB protein.[200,201] Thus, the ROSET model is hypothesised to describe the TonB as a rotor and describe ExbBD as a cylinder with three scenarios (Figure 1.16). First, TonB can turn with proton transmission generates an equal force that increases the rotor in one direction and drives the stator in the other direction. Second, in the flagellar system the stator is attached within the inner membrane suggesting the same in TonB system where only the rotor (TonB) will turn. Third, similar in a fluid bilayer, the rotor and the cylinder have the ability to move with their angular velocities in the other directions, depending on their masses and inertias. This model is based on dimer TonB with the affinity of localising the PG. The rotational motion of N-terminal domain in the inner membrane with the movement of ExbBD complex can cause the TonB C-terminal dimer to turn and wend through the PG polymer. This enables the TonB C-terminal dimer to survey the basement of the outer membrane, until the movement of TonB box for the uptake substance/metal, the dimer of TonB is shifted to a monomer with TBDT and a monomer with the peptidoglycan layer.[194]
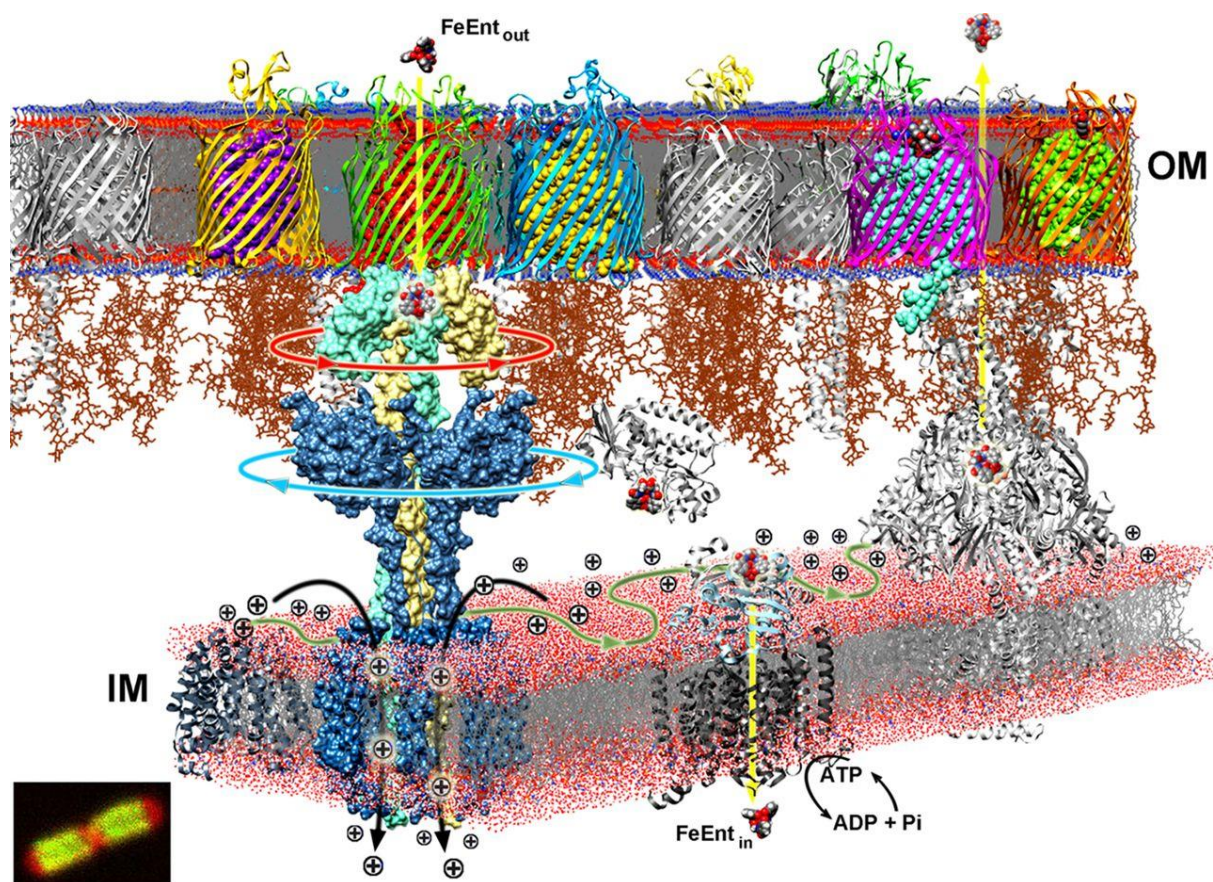
Figure 1.16. ROSET model of TonB action.
Ribbon or space-filling formats is used in the TBDT and the TonB system. Other TBDTs are shown in shades of gray. Peptidoglycan (PG) is shown in brown wireframe following the VS architecture. The ferric enterobactin (FeEnt) is presented in elemental colours and space-filling format. The TonB-CTD dimer (light blue and pale yellow) at the periplasmic space is for localising and binding the PG with low-affinity. The inner membrane PMF-driven rotation of TonB and ExbB (oligomeric-blue) in other directions can move the TonB CTD underside of the OM and move TonB-NTD and ExbBD complex via the inner membrane (green path) and the PG. In the presence of TonB box (ligand-bound), the motion of monomeric TonB-CDT enrols the polypeptide into its β-sheet transferring kinetic energy to the TBDT and allowing the uptake of iron. The yellow arrows in the outer and inner membranes show the uptake pathway of FeEnt through FepA and FepCDG. Proton pathways via TonB system are illustrated with black arrows. "Reprinted with permission from (Ref: 199); Appendix VIII: 8.8. Copyright (2016) PMC."

In contrast to the ROSET model, is the pulling model where a pulling force binds the TonB-CDT to the TBDT delivering energy, unfolding the plug domain, and allowing the ligand to enter the periplasmic space (Figure 1.17). Gumbart et al. investigated how energy is transferred from TonB protein to the TBDT and how energy transduction begins uptake. They examined TonB in complex with BtuB using a simulation of Molecular dynamics (MD).[160] These simulations concluded a mechanical mode of coupling where the interaction between TonB and BtuB is intense to attach the TonB to the TonB box and apply force regularly unfolding the plug domain. This bridge is based mainly on two stable hydrogen-bonded β-strands along with 11 salt bridges between the TonB and BtuB. Namely, these two stable hydrogen-bond were Arg166 and Arg204 from TonB and Glu426 and Asp471 from BtuB.[160] The role of pulling model between the TonB and TBDT led to the protein–protein interactions in membrane protein gating. The durable connection under force was measured using the single-molecule force spectroscopy. An atomic force microscopy (AFM) probe was used to attach it to the TonB-CTD via its N-terminus before attaching it to an immobilised BtuB. The interaction of the probed TonB results in a powerful connection that leads to the mechanical unfolding of a portion of the plug domain before TonB ultimately disconnects.[175] This suggested the possibility of mechanical force and postulated TonB rotation transducing the energy.
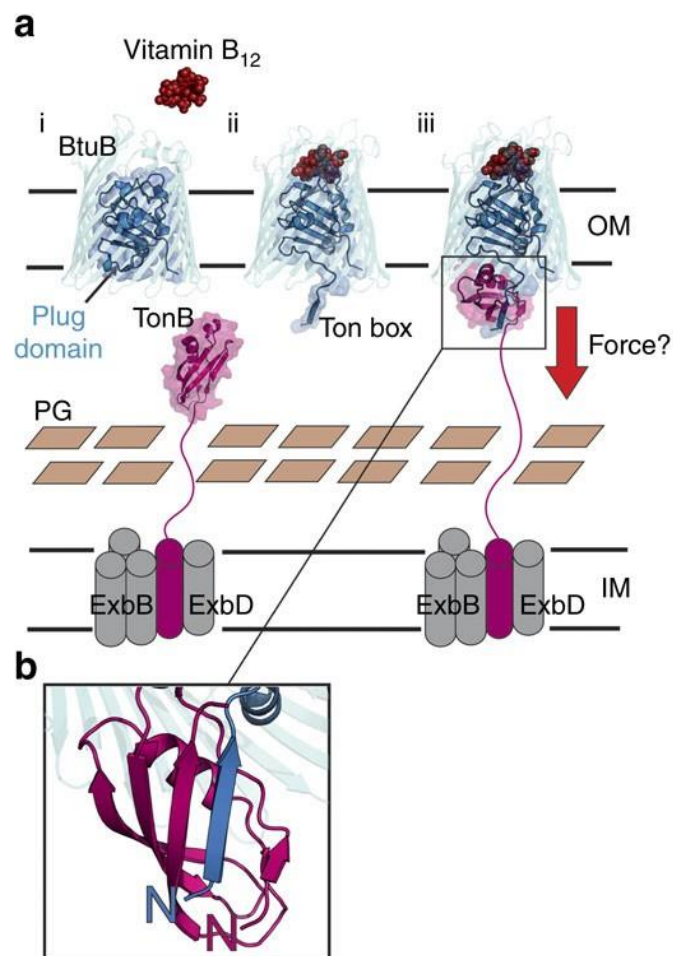
Figure 1.17. Pulling model of energy transduction by the TonB system.
(a) Schematic of TBDT transport in *E. coli* showing vitamin B_{12} transit. (i) The barrel is
closed by a plug domain (arrow), preventing the transport of vitamin B_{12}. (ii) The binding of
vitamin B_{12} generates a rearrangement of the plug domain and releases the TonB box into the
periplasmic space for interaction with the TonB. (iii) Connection of the inner membrane and
outer membrane through the TonB protein is to cause the full or partial unfolding of the plug
domain by application of force (red arrow), allowing the passage of vitamin B_{12}. (**b**) This
shows the interaction of the TonB-CTD (pink) and BtuB (blue) at the TonB box where the
parallel orientation of the β-strand expansion of this interaction. "Reprinted with permission
from (Ref:160) Appendix VIII: 8.9. Copyright (2020) creativecommons."

1.13. Sialic Acid Catabolism and Transportation

Once the sialic acid is cleaved via *nanH*, *T. forsythia* has mechanisms for sialic acid transport from the extracellular to intracellular environments. *E. coli* is the best example to compare the transport of sialic acid in *Tannerella* to (Figure 1.18).[87] Sialic acid outer membrane transporters are well-known.[202] Initially in *E. coli*, it was shown that the NanC porin permits sialic acid to traverse the outer membrane, and under the regulation of NanR regulator. Two explanations of why *E. coli* possess NanC as a specific channel for Neu5Ac while the molecular mass of Neu5Ac is 309 Da and non-specific porins exclusion size in *E. coli* is 600 Da. First, free sialic acid concentration in most animal tissue is low, which diffuses with other general porin. Consequently, NanC channel allows efficient uptake of this carbon and nitrogen source. Second, Neu5Ac induces NanC synthesis and is required the entry of Neu5Ac in the absence of expression of the general porins outer membrane protein C (OmpC) and outer membrane protein F (OmpF) that allow the diffusion of hydrophilic molecules.[202] In contrast, *T. forsythia* has a novel outer membrane TonB dependent sialic acid uptake system, part of the *nan* gene cluster that located via 16-kb section (Figure 1.8). This system consists of a putative outer membrane transport encoded by the genes *nanO* (*TF0033*) and *nanU* (*TF0034*).

Catabolism of sialic acids in *E. coli* enabled identification of the *nan* systems. The *nan* system defines as orthologues of encoded genes and is regulated in the presence of sialic acid.[84] Upon completion of *E. coli* K-12 genomic sequencing for example, the *nan* operon consisted of multiple genes: *nanAETKR* and multiple open reading frames: *yhcHIJ*, which encode a protein homologous to NagC, Mlc, and ROK family (repressor, ORF, and kinase). Sialic acid then is transported via a secondary transporter (*nanT*) and degraded intracellularly by *nanA* to the amino sugar *N*-acetylmannosamine (ManNAc) and pyruvate (Figure 1.18-black arrows). *NanK* is an ATP-dependent kinase specific for ManNAc, whereas the *nanE* is a reversible 2-epimerase to phosphorylate ManNAc to ManNAc-6p. A mutation in the *nan* operon (*nanATEK-yhcH*) identified the evidence of the *nan* operon that can be expressed in the presence of sialic acid. Once *nan* operon (*nanATEK-yhcH*) repressed transcription, *nanR* was expressed indicating that the *nanR* regulate *nan* expression. Next, N-Acetyl-D-Glucosamine 6-Phosphate (GlcNAc-6-P), a result of *nanE* epimerization, converted to fructose-6-P and glucosamine-6-P by deacetylase *nagA* and deaminase *nagB* to use it as a carbon source or amino acid source for cell wall biosynthesis.[84] The process to peptidoglycan conversion likely starts with GlmS (l-glutamine:d-fructose-6-P amidotransferase), GlmM (autophosphorylation), and GlmU (GlcNAc-1-P uridyltransferase) proteins, which they are synthase GlcN-6-P to UDP-GlcNAc, once Neu5Ac passes nanT (Figure 1.18-green arrows). Next, the process of

converting Uridine diphosphate N-acetylglucosamine (UDP-GlcNAc) to N-acetylmuramic acid (NAM) and then to the structure of peptidoglycan is not known.[84]
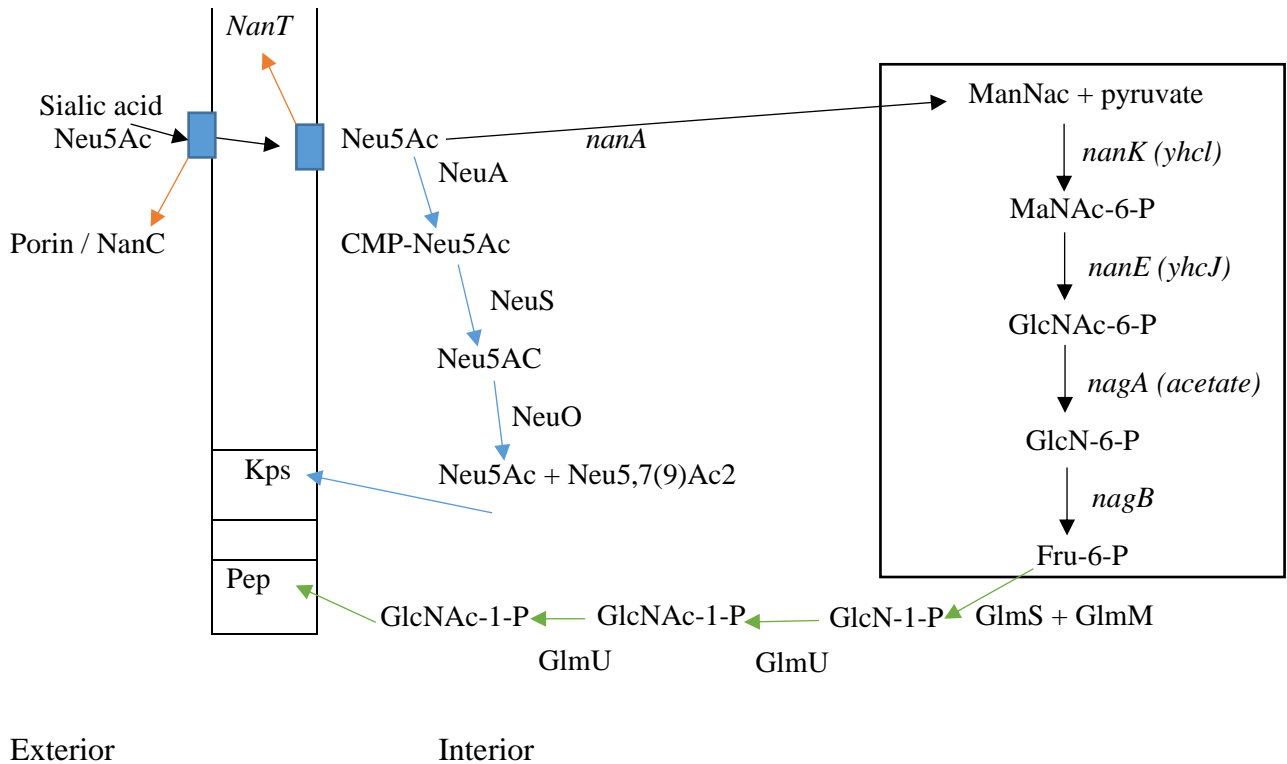
Figure 1.18 Overview of sialic acid utilisation in bacterial pathogens. Black arows show the process of *nan* system and catbolism of Neu5Ac. Blue arrows shows the process of capsular polysialic acid (PSA) biosynthesis and exports. Green arows explain the process of coverting sialic aicds to Peptidoglycan (Pep). The figure was drawn by the author, but the idea came from the Sialic acid utilisation by bacterial pathogens. "Adapted and edited from (Ref:74); Appendix VIII: 8.9. Copyright (2004) American Society for Microbiology."

In *T. forsythia*, sialic acids transportation across the inner membrane is achieved via the *nanT* type system (Figure 1.19-B). This was studied initially by a heterologous expression method where the inner membrane transport gene of *T. forsythia* was expressed in an *E. coli nanT* mutant to confirm its role with sialic acid transport, while later Honma et al., showed a growth defect of a *nanT* mutant of *T. forsythia*.[88,91] After the sialic acid is transported via the inner membrane (*nanT*), the catabolism pathway includes *nanA* gene (*TF0030*), the requirement of a neuraminate lyase, converts sialic acid into ManNAc (N-acetyl mannosamine) and pyruvate (Figure 1.19-A). *NanE* (*TF0031*) is N-acetylmannosamine epimerase, which converts ManNac to N-acetylglucosamine (GlcNAc). *T. forsythia* has a predicated enzyme (RokA [TF1994]) that phosphorylates hexoses and is proposed to phosphorylate N-acetylglucosamine to become GlcNAc-6-Phosphate and maybe then to peptidoglycan (NAM). Both *nagA* and *nagB* convert GlcNAc-6-Phosphate to fructose-6-P.[87,203] However, *T. forsythia* lacks the *nagE* (N-acetylglucosamine PTS permease) and *murP* (phosphotransferase system) homologues and the NAM synthesis pathway steps (MurA and MurB). Those steps include UDP-N-acetylglucosamine-enolpyruvate transferase (MurA) and UDP-enolpyruvate reductase (MurB).[203] Nevertheless, *E. coli* shows ability using glycolyl sialic acid (NeuGc) as a growth substrate especially in humans.[204] This is an explanation that while humans are unable to synthesize NeuGc,[205] *E. coli* can process into human glycoproteins on the surface of gut and probably oral epithelial cells.[92,206] *T. forsythia* might use NeuGc to convert it into NAM to stimulate its growth and build block in cell wall biosynthesis, acting in a manner similar to Neu5Ac.[13] The *nan* system in *T. forsythia* is not only similar to *E. coli*, but it is related to other gastrointestinal anaerobes including *Bacteroides fragilis* and *Parabacteroides distasonis* (Figure 1.8).[203]
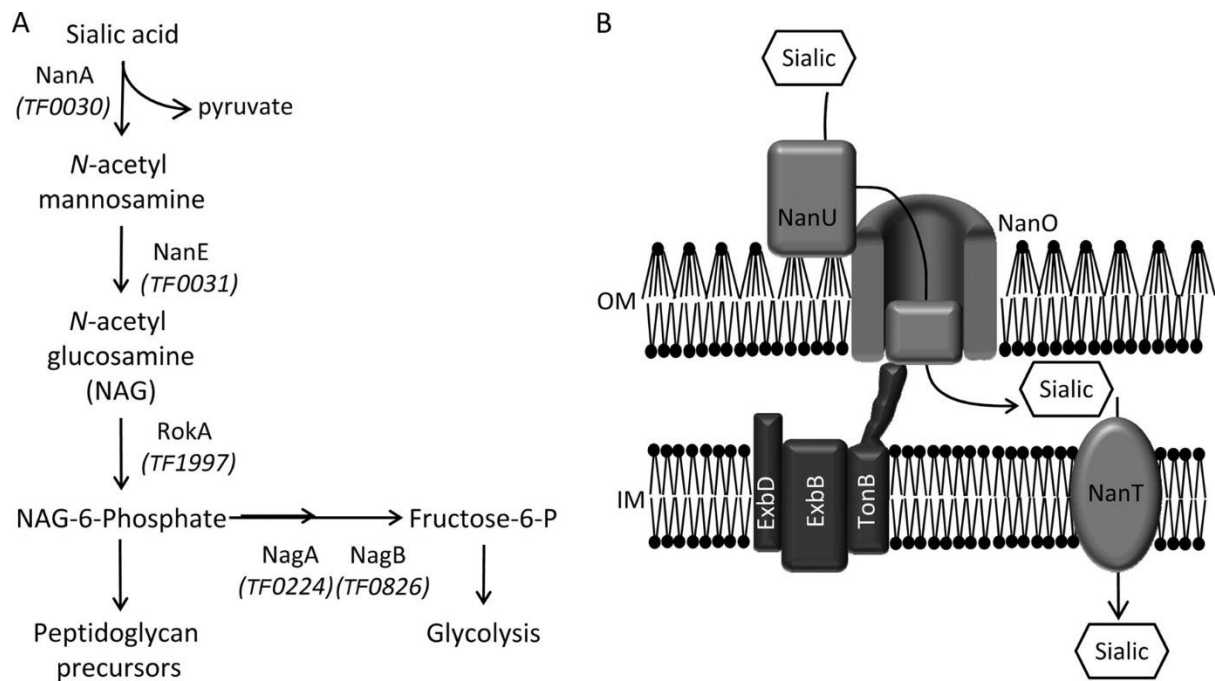
Figure 1.19 Diagram of assumed sialic acid utilisation in *T. forsythia*.
(A) Putative sialic acid utilisation pathway (B). Proposed sialic acid uptake pathway. "Reprinted with permission from (Ref:88; Appendix VIII: 8.6. Copyright (2010), American Society for Microbiology."

1.14. Aims of the Study

Several human-colonising bacteria are adapted via a wide array of metabolic adaptations to their gut or oral environments. In this context, the oral-dwelling pathogen *T. forsythia* associated with the periodontitis processes has intriguing complete novel pathways for utilisation and harvest of sialic acids, which seems to be key for its survival and virulence. In addition, the genome of *T. forsythia* possesses an extensive array of potential glycan-harvesting systems. In order to utilise and harvest sialic acid, this organism possesses a specific outer membrane based sialic acid uptake system that relies on the ability of the energising protein TonB to drive transport into the periplasm. As outlined above, TonB sits in the inner membrane with a domain that protrudes into the periplasm and contacts the outer membrane uptake system proteins (NanOU). In *T. forsythia*, as in several other organisms, there exists multiple TonB proteins that service a multitude of outer membrane transport and signalling systems (>60).

This work tested the hypothesis that "multiple TonB proteins have differing target transport substrates and partners, interaction of which can be blocked using peptides." The overriding aim of my project was to investigate the function of these multiple TonB proteins as well as, how they interact with their TBDTs and whether this is a targetable weak spot for the design of novel antimicrobials. A further aim was to establish how gene expression changes in *T. forsythia* in response to growth on sialic acid and mucin in order to establish whether other genes might be key to this process and give some insight into regulation of the *nan* operon. In addition, we conducted a bioinformatics survey of TBDTs in *T. forsythia* and of the distribution of *nan* operons in microbes in order to gain an insight into conserved function, domains and residues.

The main aims of this study are thus summarised for each result chapter as:

Chapter III. To characterise the role of multiple *tonB* genes with Neu5Ac: by mutagenesis in *T. forsythia* and using an *E. coli* based heterologous testing system.

Chapter IV. To bioinformatically identify structures of TBDTs in all *Tannerella* available genomes before generating a potential peptide-based inhibitor targeting the TonB box.

Chapter V. To investigate the distribution of sialic acid *nan* operons in microbes to gain an insight into conserved function, domains, and residues before predicting their feeding pathways through *nan* operons.

Chapter VI. To investigate global gene expression in *T. forsythia* in response to Neu5Ac and mucin by RNA-sequencing.

Chapter II Materials and Methods

2.1. Bacterial Strains and Plasmids

In this study, the wild-type *Tannerella forsythia* ATCC 43037, *Tannerella forsythia* 92A.2, *E. coli* MG1655, and derivatives were used for all experiments (Table 2.1). For complementation experiments, this study used *E. coli* MG1655 quadruple mutant (Table 2.1). Plasmid manipulations with PCR products are listed in table (Table 2.2).

Table 2.1. Bacterial strains used in the study

Bacterial strain	Genotype	Reference
T. forsythia ATCC 43037	Wild-type	Professor William Wade, King's College London, London, UK
T. forsythia 92A.2	Wild-type	Professor Graham Stafford University of Sheffield S10 2TA Sheffield, UK
ATCC 43037-Δ*nanO*	Δ*nanO*::*ermF*; knock-out strain of *nanO*	This Study; Dr. Ahmed Ali Almuntashiri
ATCC 43037-Δ*BFO_0333*	Δ*BFO_0333*::*ermF*; knock-out strain of *BFO_0333*	This Study; Dr. Ahmed Ali Almuntashiri
ATCC 43037-Δ*BFO_0233*	Δ*BFO_0233*::*ermF*; knock-out strain of *BFO_0233*	This Study; Dr. Ahmed Ali Almuntashiri
ATCC 43037-Δ*BFO_0953*	ΔΔ*BFO_0953*::*ermF*; knock-out strain of *BFO_0953*	This Study; Dr. Ahmed Ali Almuntashiri
ATCC 43037-Δ*BFO_3116*	Δ*BFO_3116*::*ermF*; knock-out strain of *BFO_3116*	This Study; Dr. Ahmed Ali Almuntashiri
E. coli MG1655	Wild-type (F^- λ^- *ilvG- rfb-50 rph-*)	Professor Barry Wanner, Purdue University, West Lafayette, IN, U.S.A
E. coli MG1655 Δ*nan*Δ*tonB*	F^- λ^- *ilvG$^-$ rfb*-50 *rph*-1 Δ*nanCnanR*(*amber*) Δ*ompR*::*Tn*10(*tet*) Δ*tonB*(*FRT-Km-FRT*)	Reference:128
E. coli DH5α	Δ(*lacZ*)*M*15 *gyrA*96 *recA*1 *relA*1 *endA*1 *thi$^-$*1 *hsdR*17	New England Biolabs UK

Table 2.2. Plasmids used in this study

Plasmid	Description	Phenotype	Reference
pBAD18	High-level expression vector containing the arabinose pBAD promoter (colE1).	Amp^R	Ref: [207]
pAcTrc99a	High-level expression vector containing all elements of pTRC99a except terminators downstream of M.C.S. (p15A).	Cm^R	Ref: [208]
pBAD18-*nanOU*-ATCC 43037	*nanO* and n*anU* cloned into pBAD18 vector.	Amp^R	This Study
pBAD18-*nanO*-ATCC 43037	*nanO* cloned into pBAD18 vector.	Amp^R	This Study
pBAD18-*nanU*-ATCC 43037	n*anU* cloned into pBAD18 vector.	Amp^R	This Study
pAcTrc99a- *BFO_0333*-ATCC 43037	*BFO_0333 tonB* cloned into *NdeI* and *XbaI* sites of pAcTrc99a vector.	Cm^R	This Study
pAcTrc99a- *BFO_0233*-ATCC 43037	*BFO_0233 tonB* cloned into *NdeI* and *BamHI* sites of pAcTrc99a vector.	Cm^R	This Study
pAcTrc99a- *BFO_0953*-ATCC 43037	*BFO_0953 tonB* cloned into *NdeI* and *BamHI* sites of pAcTrc99a vector.	Cm^R	This study
pAcTrc99a- *BFO_3116*- ATCC 43037	*BFO_3116 tonB* cloned into *NdeI* and *BamHI* sites of pAcTrc99a vector.	Cm^R	This study
pBAD18-*TF0033,34*-TF92A.2	*TF0033* and *TF0034* cloned into pBAD18 vector.	Amp^R	This Study
pBAD18-*TF0033*-TF92A.2	*TF0033* cloned into pBAD18 vector.	Amp^R	This Study

pBAD18-*TF0034*-TF92A.2	*TF0034* cloned into pBAD18 vector.	Amp^R	This Study
pAcTrc99a-*TF1354*-TF92A.2	*TF1354 tonB* cloned into *NdeI* and *XbaI* sites of pAcTrc99a vector.	Cm^R	This Study
pAcTrc99a-*TF1255*-TF92A.2	*TF1255 tonB* cloned into *NdeI* and *BamHI* sites of pAcTrc99a vector.	Cm^R	This Study
pAcTrc99a- *E. coli tonB*	*E. coli tonB* cloned into *NdeI* and *BamHI* sites of pAcTrc99a vector	Cm^R	This study
pJET (Clone Jet)	Blunt end cloning storage vector	Amp^R	Thermo Fisher Scientific
pMK- Δ*nanO*	Mutant *nanO* gene replaced with EM cassette	Km^R	This study
pJET-Δ*BFO_0333*-ATCC 43037	Mutant *tonB BFO_0333* gene replaced with EM cassette	Amp^R	This study
pMK- Δ*BFO_0233*-ATCC 43037	Mutant *tonB BFO_0233* gene replaced with EM cassette	Km^R	This study
pMA- Δ*BFO_0953*-ATCC 43037	Mutant *tonB BFO_0953* gene replaced with EM cassette	Amp^R	This study
pMS- Δ*BFO_3116*-ATCC 43037	Mutant *tonB BFO_3116* gene replaced with EM cassette	Spc^R	This study

Ap^R –Ampicillin resistance; Cm^R –Chloramphenicol resistance; Km^R –Kanamycin resistance; Spc – Spectinomycin resistance

2.2. Growth Media and Conditions

2.2.1. *Tannerella forsythia*

T. forsythia was grown either on Fastidious Anaerobe agar plates (FA) or liquid cultured in the Tryptic soy broth (TSB) medium supplemented with NAM. For FA agar preparation, the media was prepared by weighing FA powder (4.6 mg/ml), adding distilled water, and sterilising by autoclaving at 15 pounds psi (121°C) for 15 minutes on a liquid cycle. The FA was supplemented with 5% horse blood, 10 µg/ml NAM (filter sterilisation), and 50 µg/ml Gentamicin (filter sterilisation). Preparation of liquid culture TSB includes autoclaving 3 mg/ml of trypticase soya broth (Melford), 4 mg/ml of yeast extract, 5 µl/ml haemin, and 1 µl/ml menadione and

distilled water; followed by adding 4 µl/ml of NAM (10 µg/ml). Both FA-NAM plates and broth culture were incubated at 37 °C in an anaerobic atmosphere (10% CO_2, 10%H_2 and 80%N_2) for four to five days. For mutagenesis FA plates, 5µg/ml of Erythromycin was added with Gentamicin in the FA agar.

2.2.2. *Escherichia coli*

The wild-type *E. coli* and its derivatives were grown on Luria-Bertani (LB) agar or in LB liquid following the manufacturer's instruction with appropriate different concentrations of antibiotics. Preparation of LB agar includes autoclaving LB powered (Sigma Aldrich) and distilled water (40 g/L). Preparation of LB broth includes LB powder (Fisher Scientific) and distilled water (37 g/L). The LB plates were incubated at 37 ºC and the LB broth medium was shaken at 200 rpm.

2.2.3. Antibiotics preparations

Antibiotics were prepared in distilled water or distilled water and ethanol, followed by filter sterilisation. Antibiotics were added to the autoclaved liquid and solid media at the concentrations detailed in Table 2.3. All antibiotics were stored at -20 ºC.

Table 2.3. Antibiotics

Antibiotic	Dissolved	Final Concentration
Ampicillin	Water	50 µg/ml
Chloramphenicol	70% ethanol	25 µg/ml
Tetracycline	70% ethanol	10 µg/ml
Kanamycin	Water	50 µg/ml
Streptomycin	Water	50 µg/ml
Gentamicin	Water	50 µg/ml
Erythromycin	90% ethanol	5 µg/ml

2.3. DNA Preparations and Manipulation

2.3.1. Extraction of Genomic DNA from *T. forsythia*

Tannerella forsythia ATCC 43037 and *Tannerella forsythia* 92A.2 strains were grown for three days on FA agar plates and then for two days in TSB broth. Extraction of genomic DNA for each strain followed the protocol provided by the Promega Wizard® kit. The extracted DNA was estimated by NanoDrop (Thermo Scientific, UK) before it was used immediately for further analysis or stored at -20ºC.

2.3.2. Extraction of Genomic DNA from *E. coli*

E. coli (MG1655) strain was grown for one day on LB agar plate before its inoculation into LB broth and grown overnight. Genomic DNA of *E. coli* was extracted following the protocol provided by the Promega Wizard® kit. The extracted DNA was estimated by NanoDrop (Thermo Scientific, UK) before it was used immediately for further analysis or stored at -20ºC.

2.3.3. Plasmid Extraction

The DNA plasmid was transferred into 5 ml of LB broth medium with appropriate antibiotics and shaken overnight, after being grown for one day on LB agar plate. The plasmid DNA isolation followed the protocol of Bioline ISOLATE II Plasmid Mini Kit. The extracted plasmid DNA isolation was estimated by NanoDrop (Thermo Scientific, UK) before it was used immediately for further analysis or stored at -20ºC.

2.3.4. Polymerase Chain Reaction (PCR)

The PCR in this study was performed in either a total volume of 20 or 25µl using *Phusion®* High-Fidelity DNA polymerase or DreamTaq Polymerase. Template optimisation of the PCR samples or colony PCR was tested with several dilutions of template stock and annealing temperature (Table 2.4). Storage of the PCR samples was at -4ºC for the next day's work or -20ºC for an extended period.

Table 2.4. PCR reaction for PCR samples or colony PCR

Action	Temperature	Time
Initiation	95-98°C	5 minutes
Denature (30cycle)	95-98°C	30-45 seconds
Annealing (30cycle)	50-62°C	30-45 seconds
Extension (30cycle)	72°C	1 minute
Final extension	72°C	5-9 minutes
Final hold	4°C	-

2.3.4.1. PCR Amplification with Phusion Polymerase

This study used *Phusion®* High-Fidelity DNA Polymerase (Thermo Scientific, UK) because it has proof-reading capability and a lower error rate. The *Phusion®* High-Fidelity DNA Polymerase was used in this study to amplify ORF of *tonB* and TBDT genes from *T. forsythia* and one *tonB* gene from the *E. coli* as a control. It was also essential for the study cloning and mutation experiments. The *Phusion* polymerase PCR reaction was carried out as in the table 2.5.

Table 2.5. *Phusion* **polymerase PCR reaction**

Component	Final concentration	20µl reaction
5X Phusion HF or GC Buffer	1X	4 µl
10 µM Forward Primer	0.5 µM	1 µl
10 µM Reverse Primer	0.5 µM	1 µl
10 mM dNTPs	200 µM	0.4 µl
Phusion DNA Polymerase	1.0 units/50µl PCR	0.2 µl
Template DNA	< 250 ng	variable
Nuclease-free water	-	to 20 µl

2.3.4.2. PCR Amplification with DreamTaq DNA Polymerase

DreamTaq DNA Polymerase (Thermo Scientific, UK) was used to amplify and check colonies from the ligation process or mutation. The DreamTaq DNA polymerase for all standard PCR reactions was carried out as in the table 2.6.

Table 2.6. DreamTaq DNA polymerase PCR reaction

Component	20µl reaction
Green PCR Master Mix (2X)	12µl
Reverse primer	1µl
Forward primer	1µl
Template DNA	variable
Nuclease-free water	to 20µl

2.3.5. Primers

This study constructed several primers for *tonB* genes from both strains of *T. forsythia* (ATCC 43037 and 92A.2) and *E. coli tonB* are shown in table 2.7. Likewise, table 2.7 shows that the primers for TBDTs from both ATCC 43037 and 92A.2 strains (*nanOU*, *TF0033,34*, *nanO*, *TF0033*, *nanU*, and *TF0034*). Table 2.8 shows the constructed primers for the *T. forsythia* mutagenesis. Furthermore, this study examined *E. coli ΔnanΔtonB* strain before transforming *tonB* and TBDT ligations using table 2.9.

Table 2.7. Primers used in this study. The restriction enzyme site is underlined, and the ribosome binding site is highlighted.

Primer	DNA oligonucleotide sequence
nanOU-43037-KpnI-F	5'AA<u>GGTACC</u>TAATTTTGTTTAACTTT**AAGAAGGA**GATATACAT-ATGAAAAGAATTTTAAAAAATG 3'
TF0033,34-92A.2-KpnI-F	5'AA<u>GGTACC</u>TAATTTTGTTTAACTTT**AAGAAGGA**GATATACAT-ATGAAAGGAATTTTAAAAAATG 3'
nanO-43037-KpnI-F	5'AA<u>GGTACC</u>TAATTTTGTTTAACTTT**AAGAAGGA**GATATACAT-ATGAAAAGAATTTTAAAAAATG 3'
nanO-43037- Hind III-R	5' CCC<u>AAGCTT</u>GGG-TCAAAAGTTACATTAATCCCG 3'
TF0033-92A.2-KpnI-F	5'AA<u>GGTACC</u>TAATTTTGTTTAACTTT**AAGAAGGA**GATATACAT-ATGAAAGGAATTTTAAAAAATG 3'
TF0033-92A.2-BamHI-R	5' CGC<u>GGATCC</u>-TTAAAATGTTACGGATAAGCC 3'
nanU-43037-KpnI-F	5'AA<u>GGTACC</u>TAATTTTGTTTAACTTT**AAGAAGGA**GATATACAT-ATGAAAAGATAGCAAATCATATTG 3'
nanU-43037- XbaI-R	5' GC<u>TCTAGA</u>GC-TTATTCATACCCCGGAGTCTGC 3'
TF0034-92A.2-EcoRl-F	5' AA<u>GAATTC</u>TAATTTTGTTTAACTTT**AAGAAGGA**GATATACAT-ATGAAAAGTTTATCATATATTTG 3'
TF0034-92A.2–Xbal-R	5' GC<u>TCTAGA</u>GC-TTATTCATACCCCGGAGTCTGC 3'
BFO_0333-43037- Ndel-F	5' GAATTC<u>CATATG</u>-GAAACATGGACTTATTTAC 3'
BFO_0333-43037- XbaI-R	5' GC<u>TCTAGA</u>GC-CTAATTAAACTTCTTCTTAAT 3'
TF1354-92A.2-Ndel-F	5' GAATTC<u>CATATG</u>-GAAACATGGACTTATTTAC 3'
TF1354-92A.2-Xbal-R	5' GC<u>TCTAGA</u>GC-CTAATTAAACTTCTTCTTAAT 3'
BFO_0233-43037-NdeI-F	5' GGAATTC<u>CATATG</u>-GGAGCATTCTTTTCTTATTCG 3'
BFO_0233-43037-BamHI-R	5' CGC<u>GGATCC</u>-TTTCGTTTGAAGCGAGTC 3'
TF1255-92A.2- NdeI-F	5' GGAATTCCATATG-GGAGCATTCTTTTCTTATTCG 3'
TF1255-92A.2- BamHI-R	5' CGCG<u>GATCC</u>-TTATTTCGTTTGAAGCGAGTC 3'
BFO_0953-43037-NdeI-F	5' GGAATTC<u>CATATG</u>GAAATCAAGAAATCGCG 3'
BFO_0953-43037-BamHI-R	5' CGC<u>GGATCC</u>TTACTGATTTAAACGGAA 3'

TF1960-92A.2-NdeI-F	5' GGAATTC<u>CATATG</u>GAAATCAAGAAATCGCG 3'
TF1960-92A.2- BamHI-R	5' CGC<u>GGATCC</u>TTACTGATTTAAACGGAA 3'
BFO_3116-43037-NdeI-F	5' GGAATTC<u>CATATG</u>AAGTTTGAAAAGAATGA 3'
BFO_3116-43037-BamHI-R	5' CGC<u>GGATCC</u>TTATCTTAATGAGTAGCG 3'
TF0783-92A.2-NdeI-F	5' GGAATTC<u>CATATG</u>AAGTTTGAAAAGAATGA 3'
TF0783-92A.2- BamHI-R	5'CGC<u>GGATCC</u>TTATCTTAATGAGTAGCG 3'
Ec-tonB-NdeI-F	5' GGGAAA<u>CATATG</u> CGAGGTAAATCAAGGGTCAT 3'
Ec-tonB-BamHI-R	5' CGC<u>GGATCC</u>TTACTGAATTTCGGTGGTGCC 3'

Table 2.8. Primers used for amplifying *T. forsythia* ATCC 43037 flanking genes.

Primer	DNA oligonucleotide sequence
nanO-43037-F	5' TGATGGGCTTCTGTACGAC 3'
nanO-43037-R	5' TCAGCGTGTCGTTCTCG 3'
BFO_0333-43037-F	5' AACGCTCGAAAGACTTTTTG 3'
BFO_0333-43037-R	5' GGTGTACACATTCGAGTT 3'
BFO_3116-43037-F	5' ATGGGGGTTCTCCTCATC 3'
BFO_3116-43037-R	5' GAGCATGATTATAAAAGGC 3'
BFO_0953-43037-F	5' GAACAAAACTATCCTCGAG 3'
BFO_0953-43037-R	5' GCGCTAATTCATCCATCGC 3'
BFO_0233-43037-F	5' TGATACCGGCTGTCAGTTC 3'
BFO_0233-43037-R	5' TCTGTGATCCGTTTGTCTG 3'
Erythromycin-F	5' ATGACAAAAAGAAATTGC 3'
Erythromycin-R	5' CTACGAAGGATGAAATTTTTC 3'
pJET-F	5' CGACTCACTATAGGGAGAGCGGC 3'
pJET-R	5' AAGAACATCGATTTTCCATGGCAG 3'

Table 2.9. Primers used for amplifying *nanC*, *ompR*, and *tonB* of *E. coli* Δ*nan*Δ*tonB* strain.

Primer	DNA oligonucleotide sequence
Before the start codon of *nanC*-F	5' ATCCTACAAATGAGTGAAATTT 3'
After the end codon of *nanC*-R	5' ATTCATTGTGACTGTCTCCTG 3'
150 bp before the start codon of *nanC*-F	5' CACAGGAGACCTTATACAAA 3'
20 bp before the start codon of *ompR*-F	5' AACCTTTGGGAGTACAAACA 3'
20 bp after the start codon of *ompR*-R	5' CGCAATCGCCTCATGCTTTA 3'
120 bp before the start codon of deleted *tonB*-F	5' AAGGGTAATTACGCCAAAAATG 3'
K2-R	5' GCAGTTCATTCAGGGCACCG 3'
Kt-F	5' CGGCCACAGTCGATGAATCC 3'

2.3.6. Agarose Gel Electrophoresis

Agarose gel electrophoresis (1%) was carried out to analyse DNA in TAE buffer (40mM Tris-acetate, 1mM EDTA, pH 8.0). Before casting, ethidium bromide was added to a final concentration of 1 µg/ml. The gel was placed in Bio-rad Wide/Mini-Sub Cell GT Cell system and covered with 1x TAE buffer. The PCR products were stained with 1 µl of 6x loading buffer (BioLabs, UK) containing: 2.5% Ficoll®-400, 10 mM EDTA, 3.3 mM Tris-HCI, 0.08% SDS, 0.02% Dye 1, 0.001% Dye 2. The HyperLadder™ 1 kb from Bioline used in this study as a size standard. The gel was run at variable voltage between 80 to 120 V for 30 to 60 minutes. DNA was visualised under ultraviolet (UV) light using Syngene G:BOX transilluminator and photographed using the GeneSnap system (SynGene).

2.3.7. Purification

The PCR products were cleaned up from salts, primers, enzymes, and other contaminants following the Bioline purification kit according to the manufacturer's protocol. Similarly, the same kit was used to clean-up extracted DNA and vectors from agarose gel after restriction digestion (section 2.3.8.). Examination of clean up extracted DNA and vector was performed by running 2 µl to 5 µl of the DNA on a gel.

2.3.8. Digestion of DNA and Vector Products with Restriction Enzymes

DNA fragment of interest directly from cleaned PCR products or extracted from agarose gel electrophoresis was double digested via two restriction enzymes as in table 2.10. Similarly, two restriction enzymes were used to double digest the plasmid of interest (pBAD18 or pAcTrc99a) following BioLabs's protocol, as in table 2.11. Antarctic Phosphatase enzyme was used to remove terminal phosphate molecules to prevent the vector from self-ligation. Digestion was carried out at 37ºC for around four hours in a heating block or water bath. Heat inactivation or PCR clean-up protocol (section 2.3.7) was performed for further cloning process.

Table 2.10. Double-digest reaction protocol for DNA

Component	50 µl reaction
Restriction enzyme 1	2 µl
Restriction enzyme 2	2 µl
Buffer	5 µl
DNA	variable
Nuclease-free water	to 20 µl

Table 2.11. Double-digest reaction protocol for vectors

Component	50 µl reaction
Restriction enzyme 1	2 µl
Restriction enzyme 2	2 µl
Buffer	5 µl
Antarctic Phosphatase enzyme	2.5 µl
Antarctic phosphate buffer	5 µl
Plasmid DNA	variable
Nuclease-free water	to 50 µl

2.3.9. Extraction of DNA and Vector of Interest from Agarose Gel

After the digestion with restriction enzymes, DNA fragments of appropriate size were visualised via UV transilluminator to be excised from the gel-using a scalpel and placed in 1.5 ml microfuge tubes. The plasmid of interest was extracted from agarose gel using the gel extraction kit (Bioline, UK) to recover DNA and plasmid from the gel following the manufacturer's protocol.

2.3.10. Ligation of PCR Products into Vectors

The digested DNA inserts were ligated into plasmid vectors with compatible ends (pBAD18 or pAcTrc99a). Before the ligation process was performed, the concentration of both insert DNA and vectors was estimated by Thermo Scientific NanoDrop 1000 spectrophotometer. Different ratios of cloning a fragment into a plasmid vector can be carried out depending on the volume and concentration. In general, a 3:1 ratio can be applied for sticky ends ligation. The calculated amount of insert DNA and vector followed this formula:

$$\frac{\text{ng of vector x kb size of insert}}{\text{Kb size of vector}} \quad X \quad \frac{\text{molar ratio of insert}}{\text{molar ratio of vector}}$$

Based on the previous formula, an appropriate amount of insert DNA and vector was added to make the ligation reactions in a total volume of 10 or 12µl (Table 2.12). The ligation reactions were incubated at 16°C overnight, followed by heat inactivation at 65°C for 12 minutes.

Table 2.12. Ligation reaction mixture

Component	Amount
Insert DNA	~5-7 µl
Vector	~1-3 µl
Ligase Buffer	1 µl
T4 DNA ligase	1 µl

2.3.10.1. Cloning of PCR Products into Cloning Vector

The cloneJET^{TM} PCR cloning kit was used in some parts of this study to clone synthetic DNA segments provided as blunt end fragments from GeneArt. This kit has an advanced positive system for the highest efficiency cloning of PCR products or synthetic DNA. (Table 2.13).

Table 2.13. Blunt-end cloning protocol

Component	Volume
2X Reaction Buffer	10 µl
PCR product	1 µl
pJET1.2	1 µl
T4 DNA Ligase	1 µl
Water, nuclease free	Up to 20

2.3.11. Presence and Confirmation of Insert

2.3.11.1. PCR Colony

The presence of insert was screened after the ligation of *tonB* or TBDT genes with the vector using PCR colony technique. Similarly, the presence of insert was screened after the transformation of constructed mutant *tonB* gene in *T. forsythia*.

PCR colony was performed by randomly choosing between 10 to 15 colonies from a selective antibiotic agar plate. Each colony was picked up and suspended in 50 µl of RNase free water to be boiled at 95ºC for 10 minutes before spinning down at 11,000 xg. From the supernatant, 5 µl of each colony suspension was used for PCR amplification using DreamTaq polymerase (section 2.3.4.2). Colonies with presence of plasmid showed the right size on

agarose gel electrophoresis (Section 2.3.6.), followed by additional restriction digestion confirming the size of DNA insert and vector on agarose gel.

PCR colony was also used for screening the constructed mutant *tonB* and TBDT genes in *T. forsythia*. An overlap PCR was carried out to amplify one side of the flanking gene with the other side Erythromycin cassette. Four constructed primers (Table 2.8) were used to test each TonB mutant as in table 2.14.

Table 2.14. TonB mutant test technique

Primer A	Primer B
Flanking gene forward-primer	Flanking gene reverse-primer
Flanking gene forward-primer	Erythromycin gene reverse-primer
Erythromycin gene forward-primer	Flanking gene reverse-primer
Erythromycin gene forward-primer	Erythromycin gene reverse-primer

2.3.11.2. DNA Sequencing Technique

All tonB and TBDT genes ligations to either pBAD18 or pAcTrc99a were sequenced to confirm successful ligations. Likewise, sequencing samples was essential as a second technique following the PCR colony confirming all counteracted mutants in *T. forsythia*. Sequencing samples were prepared in a total volume of 7 µl containing the DNA template with a volume of 5 µl and either forward or reverse primer with a volume of 2 µl before sending them to Eurofins genomics. After receiving the sequencing samples, FinchTv programme was used for analysing and confirming each sequence.

2.3.12. Generation of *T. forsythia* Mutagenesis

Mutant strains were generated to assess their individual roles in sialic acid transit. Each gene was knocked out and replaced with an Erythromycin resistance cassette (ery^R), using a construct that contained 1000 bp flanking DNA on either side of the ery^R cassette. In all cases, gene synthesis constructs were designed in silico and ordered from GeneArt™, Thermo Fisher Scientific. Briefly, the DNA sequence is optimised for maximum production using the company GeneOptimizer software to include/exclude restriction enzyme sites. Genome constructs are supplied as linear DNA and cloned into pJET vector with 100% sequence verified. Alternatively, genes were pre-cloned into the GeneArt pMA and pMK plasmids and upon receiving the products, they were suspended in 50 µl of RNase free water

and transferred into competent DH5α cells (section 2.4.1.) Each product was plated on a selective agar plate before transformation into *T. forsythia*.

2.4. Transformation of Bacteria

2.4.1. Transformation of DNA into Commercial Competent DH5α Cells

DH5α competent *E. coli* was used for general cloning and sub-cloning applications. The competent cells were thawed on ice for 10 minutes before ligation reactions or plasmid transformation and gently mixed. Heat shock was given to the cells in a 42ºC water bath for one minute and then placed back on ice for two minutes. LB liquid (1 ml) was added to the mixture and left shaking at 37ºC at 200 rpm for 30 minutes. Before overnight incubation at 37ºC, the transformation of these cells was spread on antibiotic plates as appropriate.

2.4.2. Preparation of Calcium Chloride Competent Cells

E. coli Δ*nan*Δ*tonB* strain was taken from the overnight culture before inoculation into 100 ml of fresh LB and monitoring the growth to OD_{600} of 0.5-0.6 (2 to 4 hours). This mixture was then centrifuged at 10,000 xg (4°C) and resuspended in 10 ml cold 0.1 M $CaCl_2$, followed by incubation on ice for 20 minutes and centrifugation again at 4°C (Repeated twice). Cells were then gently resuspended in 5 ml of cold $CaCl_2$ (0.1 M) with 15% glycerol and then dispensed in microtubes. The cells were frozen at –80°C until needed.

2.4.3. Transformation of Calcium Chloride Competent Cells by Heat Shock

This study added the extracted plasmid of each *tonB* gene and one of TBDT to the calcium chloride competent cells (section 2.4.2). This mixture of cells was incubated on ice for 30 minutes and then heat shocked at 42°C for a minute. This mixture was resuspended in 1 ml of LB and left shaking at 37ºC at 200 rpm for 30 minutes. The cells were then plated on an agar plate having four antibiotics and incubated at 37ºC.

2.4.4. Transformation of Mutant *T. forsythia* Using Natural Competency

The wild-type *T. forsythia* was grown on FA agar for 72 hours before a colony was removed and inoculated into 10 ml TSB broth with supplements as described in section 2.2.1. The bacteria suspension was incubated for around 36 hours to reach an OD_{600} ~0.5-0.7. From the suspension, 1-2 ml was centrifuged for 2 minutes at 11,0000 xg and resuspended in 1 ml fresh TSB medium containing 1 µg/ml haemin, 1 µg/ml menadione, and 10 µl/ml of NAM. Extracted plasmid of constructed mutant TonB gene (≥200 ng/µl) was added to the cell suspension before being incubated for 24h. The incubation mixture was then centrifuged and

resuspensded in 200 µl fresh TSB medium and plated onto Erythromycin-containing blood agar plates.

2.4.5. Transformation of Mutant *T. forsythia* Using Electroporation

An inoculum loop of wild-type *T. forsythia* was suspended into 10 ml TSB medium supplemented with 5% fetal calf serum to reach an OD_{600} ~0.2-0.4 (early exponential). The bacterium suspension was then centrifuged at 5000 xg for 10 minutes and resuspended in 5ml of ice-cold sterile distilled water. Two additional centrifugations with ice-cold sterile distilled water were performed before a final resuspension in 200 µl ice-cold dH_2o. Approximately, 100 µl of the cells were placed in a pre-chilled 0.1 mm cuvette on ice before adding up to 50 ng/µl of *tonB* mutant plasmid to the same cuvette. The Bio-Rad Micropulser system was used for electroporation (Ec1, 1.8kV, 200ohm, 125uFaraday, 4-5 msec to ensure voltage delivery). Cells were placed on ice to recover for a minute before transferring cells to 10 ml of fresh TSB medium with no Erythromycin. The cells were placed in an anaerobic incubator for 18h, followed by centrifugation and resuspension into 1 ml of TSB medium, and plated onto Erythromycin-containing FA-blood agar plates.

2.5. Protein Profile Analysis

The protein samples were analysed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and stained with Instant Blue. Briefly, an inoculum loop of bacteria was removed and resuspended in 2 ml of sterile PBS. After taking the OD_{600}, SDS liquid was added to denature proteins in preparation for electrophoresis (section 2.5.1.3), followed by incubation at $98^{o}C$ for 10 minutes. The resolving gel was prepared as in section 2.5.2.2 and poured into BioRad glass plates to ~1 cm from the top of the lower plate. Isopropanol (100 µl) was added immediately to remove air bubbles and prevent the top of the gel from drying out. Once the resolving gel set, the isopropanol was removed and washed with dH_2O. The stacking gel was prepared as in (section 2.5.2.3.) and poured on top of the resolving gel, placing the comb to form wells. The gel was positioned into a BioRad protein gel tank (Tetra System) and the tank was filled with SDS running buffer (section 2.5.1.2.). The comb was then removed and 5 µl of pre-stained protein ladder (EZ-RunTM) was added to the first lane, and the rest lanes included 10 µl of each sample. At 160 mA, the gel was run until the loading dye reached the bottom of the gel (around 2h). The gel was removed, stained with InstantBlue™ (Expedeon), and left shaking for two hours at room temperature.

2.5.1. SDS-PAGE Buffers

2.5.1.1. SDS-PAGE Upper and Lower Tris

SDS-PAGE upper and lower Tris buffer was prepared with the following components.

Table 2.15. Upper and lower tris preparation

Component	Upper Tris buffer	Lower Tris buffer
Tris Base	6.06 g	18.17 g
Sodium dodecyl sulphate (SDS)	0.4 g	0.4 g
Sterile distilled water (dH$_2$O)	Up to 100 ml	Up to 100 ml

2.5.1.2. SDS-AGE Running Buffer

SDS-PAGE running buffer was made with four components (Table 2.16).

Table 2.16. Running buffer preparation

Component	Volume
Tris Base	12 g
Sodium dodecyl sulphate (SDS)	4 g
Glycine	57.5 g
Sterile distilled water (dH$_2$O)	Up to 1000 ml

2.5.1.3. 2X SDS Loading Buffer

Before running protein samples, loading buffer is prepared to be added to the protein samples using component as in table 2.17.

Table 2.17. Loading buffer preparation

Component	Volume up to 50ml
100mM Tris-HCl solution	40 ml
20% Glycerol	10 ml
Sodium dodecyl sulphate (SDS)	1 g
Bromophenol Blue	0.1 g
200 mM dithiothreitol	Added before using the loading buffer.

2.5.2. SDS-PAGE Gel Analysis

This section has reagents and their quantities for preparing 12% trisglycine SDS-polyacrylamide gel.

2.5.2.1. Ammonium Persulphate Preparation

In a 5 ml tube, the Ammonium persulphate preparation was prepared as in table 2.18

Table 2.18. Ammonium persulphate preparation

Component	Volumes
Ammonium persulphate	0.25g
Sterile distilled water (dH$_2$O)	5ml
This salt is stored in fridge	

2.5.2.2. Resolving Gel

In a 50 ml tube, all the reagents (table 2.19.) are added in order to prepare the resolving gel (Table 2.19).

Table 2.19. Resolving gel preparation

Reagents	12% Gel
Sterile distilled water (dH$_2$O)	4.3 ml
Acrylamide	3 ml
TEMED	5 µl
Lower Tris	2.5 ml
Ammonium persulphate	350µl

2.5.2.3. Stacking Gel

Stacking gel is made in a 50 ml tube before added to the resolving gel.

Table 2.20. Stacking gel preparation

Reagents	Volume
Sterile distilled water (dH$_2$O)	4.725 ml
Acrylamide	2.1 ml
Upper Tris	0.975 ml
Ammonium persulphate	100 µl
TEMED	17 µl

2.6. Disruption of *E. coli* Cell Wall

This study followed the protocol of the BugBuster protein Extraction to liberate soluble and insoluble membrane proteins. An overnight culture of the wild-type and Δ*nan*Δ*tonB E. coli* strains was centrifuged at 14,000 xg for 10 minutes before resuspension of the pellets into 300 µl of BugBuster. Next, the cell suspension was incubated on a shaking platform for 20 minutes at room temperature before applying a slow spin and removing unlysed cell at 3000 xg for 5 minutes. A faster spin was then applied at 7,000 xg for 5 minutes to isolate the soluble fraction proteins. Both the supernatant and pellets were resuspended in SDS-PAGE buffer and boiled at 98°C before loading directly into the SDS gel or frozen at –20°C until needed.

2.7. Minimal Media (M9)

The M9 salt was autoclaved for 15mins at 15psi on a liquid cycle (Table 2.21), whereas these two chemicals, $MgSO_4$ and $CaCl_2$ were separately prepared, sterilised by autoclaving, and added to the M9 salts. Trace elements and iron chloride solutions were prepared separately, autoclaved, and added to M9 to improve the growth. Filter sterilisation was required for Thiamine (vitamin B1) and Cobalamin (function in amino acid synthesis and breakdown and importantly in *tonB* transportation). Sialic acid (carbon source) was filter sterilised and frozen at –20°C before being added to M9 salts, whereas autoclaved 50% glycerol was used separately as an alternative carbon source.

For M9 minimal agar plate, this study added 1.5% of Agar Bacteriological (Agar No.1 [Oxoid]) to the autoclaved M9 minimal medium.

Table 2.21. Preparation of 5X M9 salts

Component	Volume
$Na_2HPO_4.7H_2O$	64 g
KH_2PO_4	15 g
NH_4Cl	5.0 g
NaCl	2.5 g
Sterile dH_2O	Up to 1 litre

Table 2.22. Preparation of MgSo$_4$ and CaCl$_2$

Component	Volume
100 mM CaCl$_2$	1.74 g per 100 ml
1M MgSo4	24.6 g per 100 ml

Table 2.23. Trace elements solutions

Component	100x per litre
MnCl$_2$	100 mg
ZnCl$_2$	170 mg
CuCl$_2$	43 mg
CoCl$_2$	60 mg
Na$_2$MoO$_4$	60 mg

Table 2.24. Iron chloride solutions

Component	Volume up to 100 ml
50 mM FeCl$_3$	1.35 g

2.7.1. Minimal Medium Growth Assay

The *E. coli* Δ*nan*Δ*tonB* strain with each *tonB* and a member of TBDT ligations was grown in M9 for an overnight pre-culture in a shaker (37ºC, 200 rpm), with only glycerol as a sole carbon and energy source. The M9 minimal medium contained trace elements and iron chloride solution (Tables 2.22, 2.23, 2.24, and 2.25). The next day, the strains were measured at OD600 using a spectrophotometer (Bio-Rad Laboratories) before being centrifuged and washed three times in sterile phosphate-buffered saline (PBS). These cultures were diluted to an optical density at 600 nm (OD_{600}) of 1 in M9. In a 50 ml falcon tube, we added all components as illustrated in table 2.25. The final volume in each well of a 96-well microplate was 200 µl, including 10 µl of the bacterial suspension and 3 mM, 6 mM, or 15 mM of individually added sialic acid. The 96-well microplate was then placed into a plate reader (Tecan Sunrise Plate Reader) for 12 hours, with readings every 30 minutes.

Table 2.25. Minimal media (M9)

Component - final concentration	Volume
5X M9 salts	10 ml
$CaCl_2$ (100 mM)	50 µl
Thiamine (4 µg)	5 µl
$MgSO_4$ (1 M)	50 µl
Trace elemnts (100x)	5 ml
FeCl3 (50 mM)	5 µl
Cobalamin (0.005 mM)	5 µl
Antibiotics if needed	5 µl
Sialic acid (3 mM, 6 mM, and 15 mM) or glycerol (50%)	Sialic acid can be dissolved in PBS up to 10 mg/ml or higher in dH_2O. It is preferred to be added directly into the 96-well microplate.
Sample	10 µl at OD_{600}=1
Sterile dH_2O	Up to 50 ml

2.8. Assessment of the wild-type and mutants' *T. forsythia* strains

2.8.1. Planktonic Growth of *T. forsythia*:

The wild-type *T. forsythia* ATCC 43037 and mutant's *T. forsythia* strains were maintained at 37°C in an anaerobic cabinet (containing CO_2, H_2, and N_2) for four-days on Fastidious Anaerobe agar supplemented with NAM. The planktonic growth of either the parental strain or the mutant's strains was prepared by harvesting and washing these stains three times in sterile PBS before being diluted to an optical density at OD_{600}=1 in the TSB medium. The TSB medium was supplemented with autoclaved 3 mg/ml TSB, 4 mg/ml yeast, 5 µl/ml haemin, and 1 µl/ml menadione (vitamin K_3) and this medium was incubated overnight before use. The final volume in each well of a 96-well microplate was 200 µl, including 20 µl of the bacterial suspension to give an OD_{600} of 0.1 before individually adding to each well either NAM (0.34 mM-4 µl/ml) or Neu5Ac (6 mM). The 96-well microplate then is placed into an anaerobic cabinet and the growth was monitored for seven days by taking out the plate every day from the anaerobic cabinet at a specific time, covering it with 'SealPlate', measuring growth at OD of 575 nm, and returning it as quick as possible to the anaerobic cabinet. The reading data was transferred into the excel sheet and GraphPad for analysis.

2.8.2. Biofilm Growth of *T. forsythia*

The biofilm growth of *T. forsythia* was monitored and assessed in two assays. First, Four-day-old wild-type and mutant *T. forsythia* ATCC 43037 strains were washed three times in PBS before their inoculations at a final optical density at OD_{600} of 1 into the TSB medium including either NAM or sialic acid (Neu5Ac) as indicated in the relevant figures and legends (0.34 mM NAM, and 1.5 mM, 3 mM, or 6 mM of Neu5Ac). The TSB medium was supplemented with autoclaved 3 mg/ml TSB, 4 mg/ml yeast, 5 µl/ml haemin, and 1 µl/ml menadione (vitamin K_3) and this medium was incubated overnight before use. The bacterial suspension, NAM, and Neu5Ac were individually added into each well of the 96-well microplate before their incubation for seven days (200 µl total volume included 10 µl of the bacterial suspension to give an OD_{600} of 0.05). In the presence of mucin (purchased from sigma-Aldrich / 84195-52-8 – MW 5000 g/mol), 15 µM (0.75 mg/ml), 30 µM (1.5 mg/ml), and 60 µM (3 mg/ml) were mixed with sterile PBS and left shaking for 3 hours before filter sterilisation with 'Nalgene MF75 PES 75 mm x 0.45 um x 250 ml Filter Unit'. After filtration, each well was coated with mucin (200 µl) overnight at 4°C before washing out these wells again two times with PBS on the following day. In the presence of peptide inhibitor, 2 µl (100 µg/ml) was added only to the well of wild-type *T. forsythia* ATCC 43037. The biofilm growth was counted using the cell numbers for mature biofilms after removing the TSB medium and gently washing the wells two times in sterile PBS. Then pipette tips and PBS were used to scrape each biofilm well and dilute it 1:10 before counting cells using a Helber counting chamber (Hawksley) and microscope with a phase-contrast lens (magnification, ×1,000).

Second, the biofilm formation was evaluated by crystal violet staining after seven days of incubation. The wild-type and mutant *T. forsythia* strains were adjusted to an OD_{600} of 1 in the TSB medium after washing these strains three times in sterile PBS. The TSB medium was supplemented with autoclaved 3 mg/ml TSB, 4 mg/ml yeast, 5 µl/ml haemin, and 1 µl/ml menadione (vitamin K_3), and this medium was incubated overnight before use. Aliquots (200 µL) of bacterial suspensions and either 0.34 mM of NAM, or 3 mM, 6 mM, and 10 mM of Neu5Ac were individually dispensed into the 96-well microplate including 20 µl of the bacterial suspension to give an OD_{600} of 0.1, followed by incubation at 37 °C for seven days into an anaerobic cabinet. In the presence of the peptide inhibitor, 2 µl (200 µg/ml) was added only to the well of wild-type *T. forsythia* ATCC 43037. The medium and the contents of the wells were then removed and poured off before washing out the plate

only one time with sterile PBS. After washing, each well was stained with 200 μL of 0.1%
crystal violet solution and incubated for 30 minutes at room temperature. The crystal violet
solution was then removed, and the wells were washed with sterile PBS as much as possible
to remove any unbound dye. The plate was dried for 20 minutes before the formed biofilms
of the wild-type and mutant *T. forsythia* strains were dissolved with 100 μL of ethanol-
acetone, and the optical density was measured at OD_{575}.

2.8.3 Thiobarbituric Acid Assay

Free sialic acid can be quantified following Aminoff's Thiobarbituric Acid (TBA)
assay.[209] This assay includes thiol group to label free sialic acid in each sample and allowed
free sialic acid to be arbitrarily quantified by spectrophotometry. This assay has the
advantage of generating a standard curve on the control sialic acid concentrations, which they
went under the TBA assay and spectrophotometry to conclude the uptake of sialic acid in
each sample. The wild-type and mutant *T. forsythia* strains were washed three times in PBS
and adjusted to OD_{600} of 2 in PBS. Both concentrations of Neu5Ac (500 μM and 1000 μM)
were tested where each Neu5Ac concentration was added to a bacterial suspension (total 100
μl of bacterium in PBS and Neu5Ac in a 500 μl Eppendorf tube by making a pin hole in the
head). Each experiment included triplicates of each strain, and cells were then incubated for 3
hours at 37°C in an anaerobic cabinet. To estimate free Neu5Ac, a supernatant of 50 μl was
collected and placed in a PCR tube from each 500 μl Eppendorf tube after centrifuging the
incubated sample bacteria at 13, 000 *g* for 3 minutes. Next, the 25 μl of periodate solution (20
mM periodate in 60mM H_2SO_4) was added to the 50 μl supernatant before incubating these
PCR tubes for 30 minutes at 37°C. The reaction of each sample was halted at a given time
point by oxidation with 20 μl of 2% (w/v) sodium arsenite solution in 500 mM HCI and
incubated for two minutes at room temperature before adding 100 μl of Thiobarbituric Acid
to label oxidised Neu5Ac with a thiol group (100 mM / pH 9.0). All reactions were then
heated at 95 °C for 7.5 minutes and cooled for two minutes at room temperature. Next, the
whole reaction for all samples was added to a clear 96 well plate, and the absorbance was
measured at OD 549 nm (Tables 2.26, 2.27, and 2.28).

Table 2.26. Preparation of Sodium periodate

Component	Volume
20 mM Sodium periodate (NalO$_4$)	0.043 g
2 M Sulfuric acid (H$_2$SO$_4$)	300 µl
Sterile dH$_2$O	Up to 10 ml

- Dissolve NalO$_4$ in dH$_2$O before the addition H$_2$SO$_4$. The reaction is wrapped in foil and stored in a spark-free fridge.

Table 2.27. Preparation of Sodium meta-arsenite

Component	Volume
Sodium meta-arsenite (NaAsO$_2$)	0.4 g
500 mM Hydrochloric acid (HCI)	2 ml
Sterile dH$_2$O	Up to 20 ml

- The reaction is stored in a spark-free fridge.

Table 2.28. Preparation of Thiobarbituric acid

Component	Volume
2-thiobarbituric acid	0.144 g
Adjusted to pH9 with NaOH	-
Sterile dH$_2$O	Up to 10 ml

- The reaction is stored in a spark-free fridge.

2.9. Statistical Test

The data generated from this study were statistically analysed using GraphPad Prism v9.3.1 software. Data are produced in triplicate from at least three independent experiments using the mean and standard deviation (SD). A student's t test was used for differences between two groups, whereas one-way analysis of variance (ANOVA) followed by Tukey's post-hoc multiple comparison test was applied for multiple comparisons between three or more group data. The statistical significance of this study was considered at $p < 0.05$.

2.10. Total RNA-Purification

The wild-type *T. forsythia* ATCC 43037 strain was maintained at 37°C in an anaerobic cabinet (containing CO_2, H_2, and N_2) for four-days on Fastidious Anaerobe agar supplemented with NAM. The wild-type *T. forsythia* ATCC 43037 strain was washed three times in PBS before their inoculations at a final optical density at 600 nm (OD_{600}) of 1 into an overnight incubated TSB medium including 4 mg/ml yeast, 5 µl/ml haemin, and 1 µl/ml menadione (vitamin K_3). In nine individual 12-well plates, each well has 2 ml total volume, including 50 µl of the bacterial suspension to give an OD_{600} of 0.05. Of the nine individual 12-well plates, we directly added NAM (0.34 mM) to each well (total three individual 12-well plate). Likewise, we added Neu5Ac (6 mM) to each well (total three individual 12-well plate) before their incubation for seven days in an anaerobic cabinet. In the presence of mucin, 30 µM (3 mg/ml) was mixed with sterile PBS and left shaking for 3 hours before filter sterilisation with 'Nalgene MF75 PES 75 mm x 0.45 um x 250 ml Filter Unit'. After filtration, each well was coated with 2 ml mucin (total three individual 12-well plate) overnight at 4°C before washing out these wells again two times with PBS on the following day and adding 2 ml total volume of TSB including 50 µl of the bacterial suspension to give an OD_{600} of 0.05. Three biological replicates were generated for NAM, Neu5Ac, and mucin. Next, the TSB medium and the contents of the wells were removed and poured off before washing out each plate two times with 2 ml sterile PBS. The cells were collected in a 1.5 Eppendorf tube by scraping each well using pipette tips and PBS, and concentrated by centrifugation at 15,000 xg for 1 minute. Total RNA isolation was performed using the QIAGEN bacteria kit (Mini Kit) and treated with the RNase-Free DNase set to remove and digest DNA during RNA purification. The RNA purity was assessed in two ways. First, all RNA isolations were checked using a NanoDrop ND-1000 Spectrophotometer, checking the ratio of absorbance at 260/280 and 260/230, indicating purity of contamination and purity of unwanted organic compounds, respectively. Second, we used 1% agarose gel electrophoresis,

confirming the process of the common precursor (47S rRNA) through a ratio of intensities of 28S, 18S, and 5.8S/5S rRNA bands. Total RNA purifications were stored at -80 °C before sequencing.

2.10.1. RNA-Sequence Analysis

The RNA samples from NAM, Neu5Ac, and mucin were sent to the 'GENEWIZ Germany GmbH' to perform Illumina NovaSeq, 2x150bp configuration. The service includes the removal of the adapter sequences and low-quality reads before generating clean reads that were accurately compared with the reference genome of *T. forsythia*. The normalisation of aligned reads to each gene was applied before taking the average of three biological replicates. Once the quantification of gene expression was completed, differential gene expression analysis of RNA-seq data (DESeq2) was used to perform statistical analysis on the expression data and screen for significant gene expression. Integrated Differential Expression and Pathway analysis (iDEP) database were used to conduct Gene Ontology (GO) and Pathway significant enrichment analyses and conduct hypergeometric testing to identify significant enrichment in differentially expressed genes between NAM versus Neu5Ac and NAM versus mucin.[210] Furthermore, Gene Ontology (GO) enrichment analysis was performed using 'Comparative Go' webserver.[211] KEGG and KOBAS databases were used to summarise significant enrichment in differentially expressed genes.[212,213]

Chapter III

Role of TonB transporters in *T. forsythia*

3.1. Introduction

Once the ligand binds the TBDT, conformational changes lead the TonB box with high affinity to periplasmic exposure interacting with the C-terminal domain of TonB. It is predicted then, but largely unknown, that the TonB propagates the energy obtained from the PMF at the inner membrane to force and alter the confirmation of the TBDT plug domain. This PMF energy is transduced from the inner membrane through the ExbB and ExbD, which transmit this energy to the TonB protein. This alteration helps the bound ligand to travel from the outer membrane into the periplasmic space. When the bound ligand is imported to periplasmic space, the TBDT plug domain returns to a normal state and a dissociation of TonB protein.[133] The main aim of this study was to characterise the role of multiple *tonB* genes with Neu5Ac since Neu5Ac uptake is not only important for *T. forsythia* growth in biofilm but also for intracellular survival, suggesting a role for Neu5Ac uptake in virulence. To examine and confirm the function of individual *tonB* genes with Neu5Ac transport, this study constructed gene deletion for these genes from the parent organism *T. forsythia* as well as cloned all the *tonBs* for heterologous complementation in *E. coli* system.

3.2. Results

3.2.1. Bioinformatics Analysis and Identification of *tonB* Genes

As outlined in the introduction, transport of sialic acid across the cell envelop of *T. forsythia* is energised by TonB dependent transport (NanO).[214] As a first point of investigation, the *tonB* complement of *T. forsythia* was examined bioinformatically as these *tonB* genes are located apart from one another. The type strain and most well studied is ATCC 43037, however, due to historical error in annotation,[215] the closely related strain 92A.2 was the first sequenced and remains the best annotated. Therefore, the *tonB* genes in both genomes were investigated.

On inspection of the annotated genomes of two *Tannerella* strains and using a combination of BLAST and Pfam databases, the genomes were found to contain four putative *tonB* genes in each strain (Table 3.1.).

Table 3.1. Genes identified as putative TonB proteins and chosen for investigation in this study

Gene Name of ATCC 43037 Strain	Length			Gene Name of 92A.2 Strain	Length		
	bp	aa	kDa		bp	aa	kDa
BFO_0333/(Tanf-11055)	1867	621	69.07	TF1354	1867	621	69.07
BFO_0233/(Tanf-08975)	1440	480	53.28	TF1255	1440	480	53.28
BFO_0953/(Tanf-12470)	699	232	25.86	TF1960	699	232	25.86
BFO_3116/(Tanf-07665)	837	278	30.09	TF0783	837	278	30.09

*Gene names inside parenthesis are alternative name based on different ATCC 43037 sequences.

With the presence of multiple *tonB* genes in both strains, the gene sequences were compared with one another (Table 3.2.). Alignment of the DNA of the *tonB* genes between the 43037 and 92A.2 strains showed 99.71 % identity between BFO_3116 *tonB* (ATCC 43037) and TF0783 *tonB* (92A.2). Likewise, there is 99.88 % DNA sequence identity between BFO_0953 *tonB* (ATCC 43037) and TF1960 *tonB* (92A.2). Other *tonB* genes however, showed some differences between their DNA sequences. The similarity between BFO_0233 and TF1255 and between BFO_0333 and TF1354 was 96.75 % and 97.54 %, respectively. Thus, this study decided to investigate all four *tonB* genes of the ATCC 43037 strain and two *tonB* genes (TF1354 and TF1255) of 92A.2 strain with less identity to identify *tonB* involvement with sialic acid (Sequences are shown in the Appendix VIII: 8.10).

Table 3.2. Similarities between *tonB* gene DNA from ATCC 43037 and 92A.2 strains.

Light green highlights similarities between *tonB* genes of the ATCC 43037 whereas dark green highlights similarities between *tonB* genes of the 92A.2 strain.

Genes	Similarity % BFO_0333	Similarity % BFO_0233	Similarity % BFO_0953	Similarity % BFO_3116	Similarity % TF1354	Similarity % TF1255	Similarity % TF1960	Similarity % TF0783	Similarity % E. coli *tonB*
BFO_0333		48.32	26.29	32.19	**97.54**	48.39	26.17	32.04	28.59
BFO_0233	48.32		33.01	44.22	48.43	**96.75**	33.79	40.96	33.92
BFO_0953	26.29	33.01		49.34	26.22	33.01	**99.88**	100	50.90
BFO_3116	32.19	42.22	49.34		32.19	42.22	100	**99.71**	53.54
TF1355	**97.54**	48.39	26.17	32.04		48.32	26.17	32.04	28.56
TF1255	48.43	**96.75**	33.79	40.96	48.32		33.79	40.96	35.73
TF1960	26.22	33.01	**99.88**	100	26.17	33.79		49.34	50.90
TF0783	32.19	42.22	100	**99.71**	32.04	40.96	49.34		53.54
E. coli *tonB*	28.59	33.92	50.90	53.54	28.59	35.73	50.90	53.54	

3.2.1.1. Genetic Context of *tonB* in *T. forsythia*

Next, the protein domain composition and genetic context of each TonB was examined, with the focus on the ATCC 43037 genome sequence. Figure 3.1 shows the size and predicted protein domains present within the four tonB gene/ protein in *T. forsythia*. These TonB predicted proteins all contain the signature TonB C-terminal domain (Pfam: PF03544) but have different compositions. Two of the TonB proteins (BFO_0333/TF1354 and BFO_0233/TF1255) have the peptidase family M56 domain, which is responsible for resistance to beta-lactam antibiotics. The TonB BFO_0333 has a CarboxypepD-reg domain to control signal and analyse individual C-terminal amino acids from polypeptide chains.[216] The TonB BFO_0233 has the ThiJ/PfpI family, and it is known as Glutamine amidotransferase. This enzyme is to transfer the ammonia group from glutamine forming a new carbon-nitrogen group.[217] CedA/B Type II is suggested with a role to repress the DnaAcos that is responsible for the inhibition of cell division. While these domains are often seen in TonB proteins, these role in TonB function is not known.[216–218]
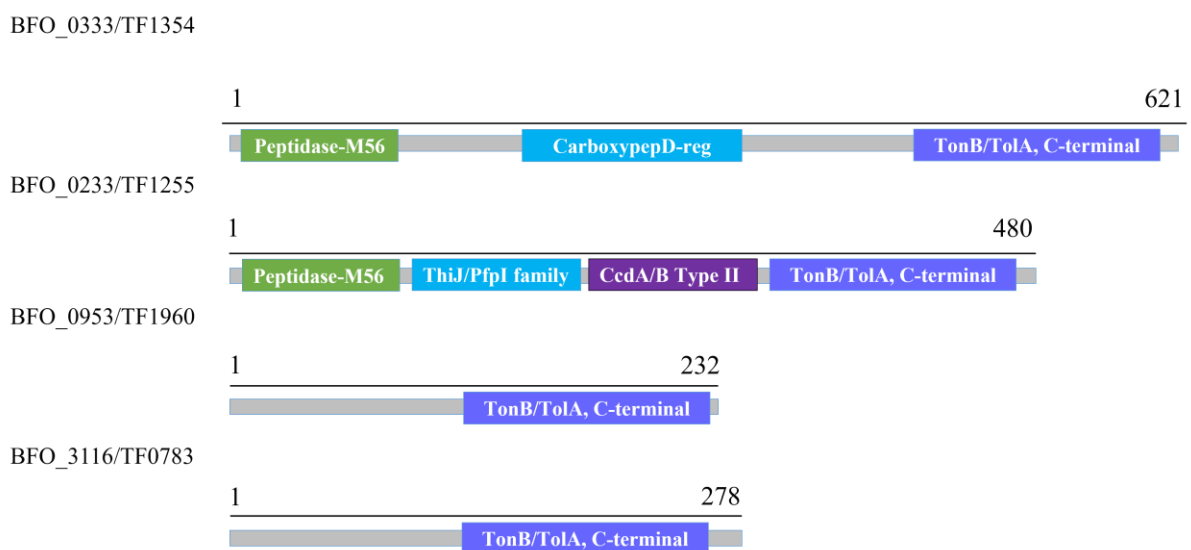
Figure 3.1 Analysis of amino acid for families and domains in TonBs.
This figure shows different length of amino acid sequence of TonB proteins of *T. forsythia*. Pfam and InterPro websites were used to analyse and identify these proteins families and domains. All these TonBs have conserved C-terminal domains.

A second point of investigation was to identify and investigate the genetic context around each TonB, including other components of the TonB system, namely ExbB and ExbD (Figure 3.2). The first TonB system in the genome contains TonB BFO_0233/TF1255 which is positioned at 260,239 bp. This TonB system is a monocistronic *tonB* locus, and it does not contain a copy of *exbB* or *exbD*. It is flanked by an upstream open reading frame (ORF) transcribed in the same direction, coding for a hypothetical protein and a downstream ORF in the same direction, coding for a predicted penicillinase repressor. Likewise, the second TonB system in the genome BFO_0333/TF1354 is located at 357,116 bp and is a monocistronic *tonB* locus without a copy of *exbB* or *exbD*. It is flanked by an upstream ORF coding for hypothetical protein and transcribed in the opposite direction, and a downstream ORF coding for a transcriptional regulator and in the opposite direction. The third TonB system (*BFO_0953/TF1960*) is organised into an operon with its corresponding *exbB* and *exbD* and an additional *exbD* gene copy (located at 1,033,408 bp). This system is flanked by an upstream ORF transcribed in the same direction, coding for a Lrp/AsnC ligand binding domain and a downstream ORF transcribed in the opposite direction, coding for a type IX secretion system protein (PorQ involved with other components in the attachment of T9SS substrates to the cell surface). The fourth system is organised into an operon consisting of a *tonB* gene (*BFO_3116/TF0783*) corresponding *exbB* and *exbD* genes, and the *tonB* gene is located at 3,165,40. This TonB system is flanked by an upstream ORF coding for pyridoxal phosphate biosynthesis protein and transcribed in same direction, and a downstream ORF coding for a DJ-1/PfpI protein and transcribed in the same direction. It is worth noting that downstream *tonB BFO_3116* by three genes, there are three related genes transcribed in the same direction that code for *susC*, *susD*, and an *ABC* transporter. Genes in the vicinities to the *tonB* genes are different, all of which are in the cytoplasm with multiple functional classes.
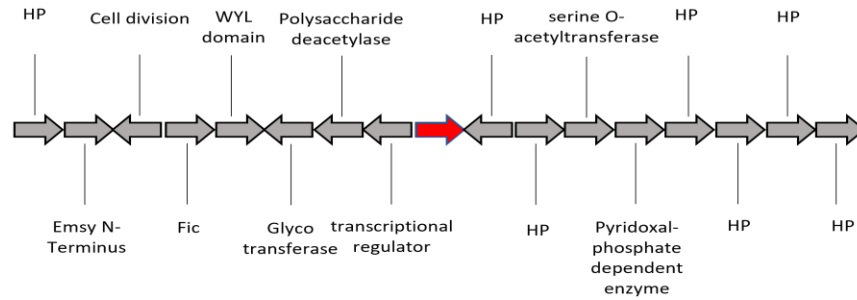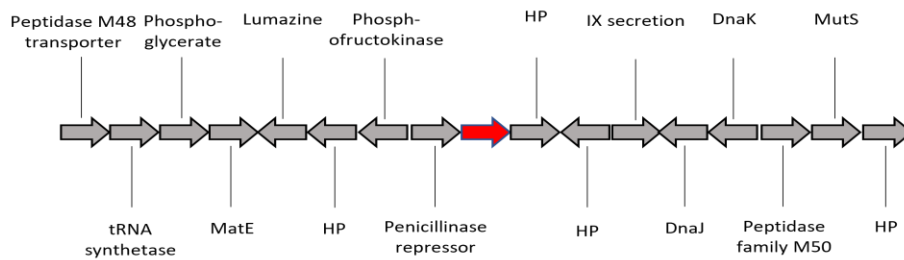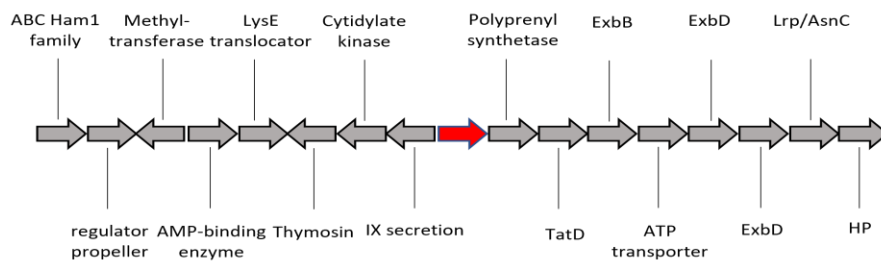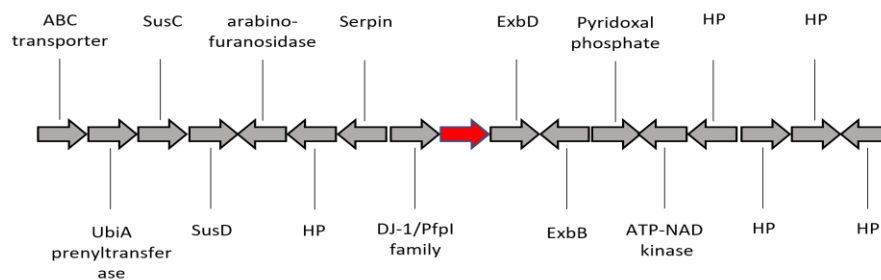
Figure 3.2 Identification of ExbB and ExbD location

This figure shows genetic context of each TonB system including ExbB and ExbD locations. TonB protein is coloured in red with eight genes on each side.

3.2.1.2. The Carboxy-Terminal Domain (CTD) of TonB protein in *T. forsythia*

The TonB C-terminal domain (CTD) is the best conserved domain of the TonB protein and the most well studied. CTD can deliver the interaction via TonB-box with the TonB-dependent transporters. In *E. coli*, the CTD is around 90 amino acids of the protein. Previously, Byron and others generated a phylogenetic tree of nine clusters of 28 TonB proteins representing 263 TonB sequences based on the CTD, where these clusters do not depend on the taxonomy of the organism. All four *Tannerella* TonB proteins were aligned within the representative clusters of phylogenetic analysis of the C-terminal domains. Based on the neighbour-joining bootstrap tree, generated alignment showed that all CTD of *Tannerella* fall into two different phylogenetic clusters. The TonB BFO_0953/TF1960 falls into the 1C cluster whereas the rest TonB proteins fall into the 3A cluster (Figure 3.3). This result indicates each *tonB* gene found within the *T. forsythia* genome is a distinct genetic unit.
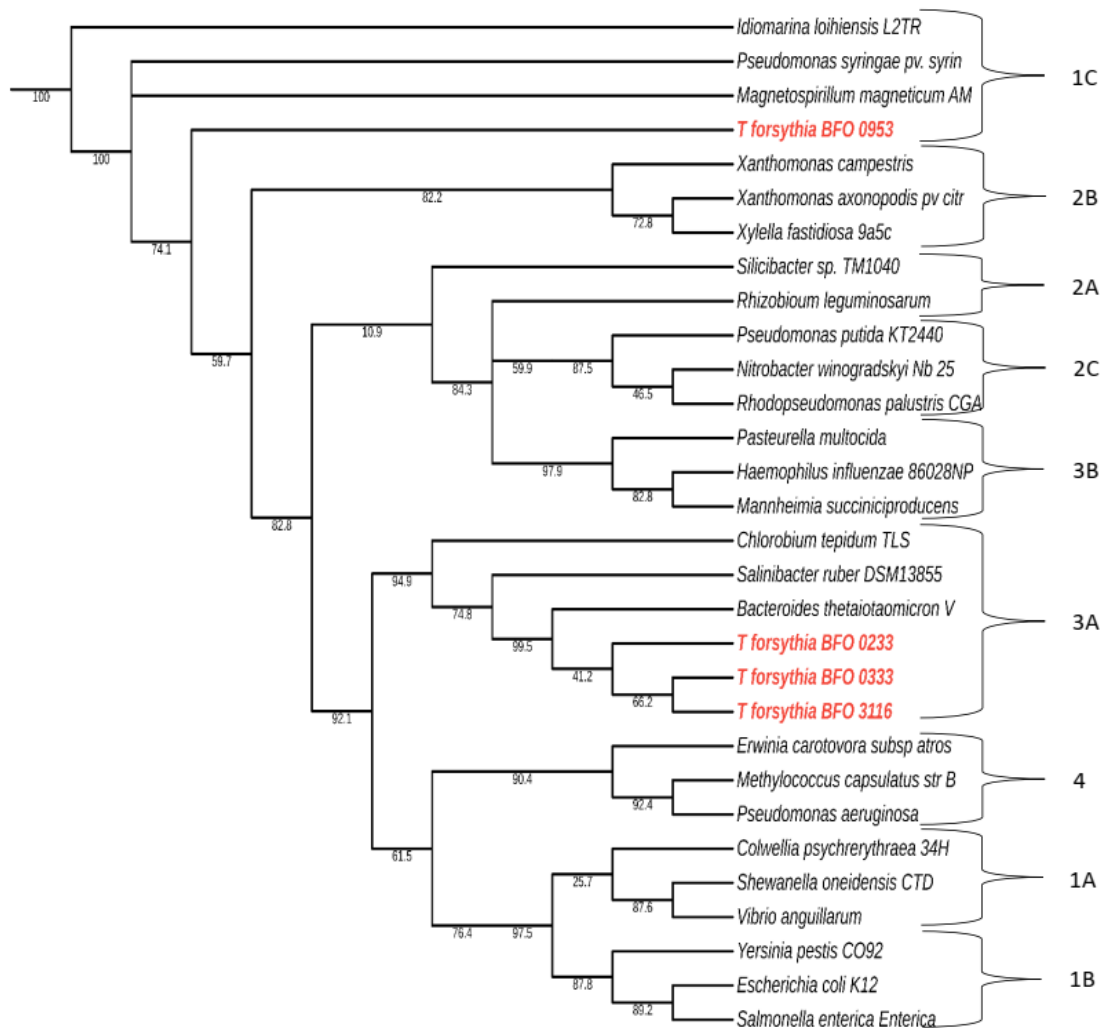
Figure 3.3 Phylogenetic tree of TonB C-terminal domains.
Neighbour-joining Bootstrap tree of the CLUSTAL W was designed to represent previous
alignment of the C-terminal domains of TonB proteins. This tree was edited to align four *T.
forsythia* CDT TonB proteins, which they are highlighted with red colour. Nine clusters are
shown representing 27 Gram-negative bacteria. The tree was drawn and viewed with
interactive tree of life (iTOL).

In general, the structure of TonB consists of three domains: a C-terminal domain, a periplasmic linker domain with a Proline-rich domain, and a N-terminal domain with single or multiple transmembrane helices. Due to the variability of domains and linker regions between TonB proteins, a subdivision for TonB domains was used in this study.[149] This study used the *E. coli* TonB protein as an established guide to subdivide and analyse domains of *T. forsythia* TonB proteins as it is well characterised with conserved residues well defined.

Figure 3.4 shows amino acid level alignment between the four TonB proteins of ATCC 43037 strain to the *E. coli* TonB. This alignment marks the presence of the conserved TonB C-terminal domain in all TonB proteins as well as the signature *E. coli* TonB "YP" motifs in all cases, although BFO_3116 had the variant LP motif in it is place.[149] Additional examination of the C-terminal domain of *E. coli* TonB determined secondary structures of 2 α-helices and 3 β-strands. As a result, the level of conservation is seen for all four TonBs sharing the consensus of the β1 and β2 regions (EGQVKV and GRVDNVQILSA). The level of less conservation is also seen for all four TonB proteins of *Tannerella* in α2 and β3 regions (REVKNAMRR and VVNILFK) (Figure 3.4).

```
                 α1
E.coli_TonB   ------------------------------------------------------------
BFO_0333      KEILAHEQIHVRQHHSFDILLAELSCAVCWFNPMAWMLKKEIHSNLEFLVDHRVVKDGAD
BFO_0953      ------------------------------------------------------------
BFO_0233      AAIIAHETAHIRLHHSWDLMAAHVFVILQWFNPFAWLLYRDLRNTHEYEADKSVLDEGFE
BFO_3116      --------------------------------------------------

                                                         α2
E.coli_TonB   ----------------------------------------------TLLSVCIHGAVVA
BFO_0333      ARSYQYHLLRMAHQPVQAALANQFNTS-PLKKRIRMLNAKQSPKVKLVAYTLVLPLALLL
BFO_0953      ------------KKSRKADLEGGKGLS--------------------ILMGIVVGLAVLF
BFO_0233      AQHYQLLLIKKAVGTRLYSMANGYNHSNNLKKRITMMLQKKSNAWARLKYALVLPLAATT
BFO_3116      -------------------------------------EKNDIIALIGSIIFHLSILL

                                           GGQxxLMxxIA
E.coli_TonB   QEQPK-------------RDVKPVESRPASPFENTAPARLTSSTATAATS----KPVTSV
BFO_0333      -------------------IFTVVEEMPQFPGGQAALMQYIARNI----------------
BFO_0953      -------------------IFMVVEEMPEFPGGQAALMSFIAKSI----------------
BFO_0233      ---PEDT-----------VVFMVVEQMPEFPGGQGELMKYIAQNI----------------
BFO_3116      TEQRKVEAERKKQTDEQRRKEQAIRDRVAGAFGSTGGSENNNQGDAASGSGNQGSPFGNS

                            α1          β1               β2
E.coli_TonB   ASGP------------RALSRNQPQYPARAQALRIEGQVKVKFDVTPDGRVDNVQILSAK
BFO_0333      -----------------------KYPVIAQENGIQGRVVCQFIIDERGNTADVQVVQGV
BFO_0953      -----------------------KYPVVAQENGIQGRVTCSFVVNKDGSIVDAEVIRGI
BFO_0233      -----------------------KYPVKAIEKKMEGRVVVSFVVSCDGSLRDFQVVRSV
BFO_3116      DSGANEGVGGIGSFNLNGRSIGAGGLPKPAYTIQEEGRIVINITVDPRGNVIMAEIGKGT

                        α2                    β3
E.coli_TonB   P---ANMFEREVKNAMRRWRYEP----GKP-GSGIVVNILFKIN--------------GT
BFO_0333      D---PLLDKEAVRLIEEMPRWKP----GKQRGKPVSVEYTVPVNFRLTKGAVCSIQKDTT
BFO_0953      D---PSLDKEALRVINTMPKWKP----GKQRGKPVRVKFTVPINFRLNQ-----------
BFO_0233      D---PDLDAEAIRVLKLMPNWKP----GKQRGKNVSVKYTLPINFSLSA------KKDS-
BFO_3116      NIDHATMRKSALE-AARKAKFNSIQGTNNQSGT-ITYRYSLR-----------------
E.coli_TonB   TEIQ--------------------------------------------------------
BFO_0333      RQTQSRQAPIRYYLNGKEIFSERMKSIPPAEIENIHVLNTENRQMEHKDKTASKIIIITT
BFO_0953      ------------------------------------------------------------
BFO_0233      LQTK--------------------------------------------------------
BFO_3116      ------------------------------------------------------------
```

Figure 3.4 Amino acid alignment of *tonB* genes.
Multiple sequence alignment was performed for the *E. coli tonB* and the four *tonBs* of *Tannerella* ATCC 43037. TonB N- and C-terminal domains in all five *tonBs* were highlighted. Green colour highlights low and high consensus of the reported β-strand regions from *E. coli tonB*. Purple colour highlights all reported α-helices on both N- and C-terminal domains that were reported in *E. coli tonB*. In addition, we highlighted the new predicted helix on the Periplasmic linker domain (GGQxxLMxxIA) (section 3.2.1.3). This alignment was prepared using MAFFT server.

The second domain of TonB is the periplasmic linker domain including proline-rich region. The previous investigation of this domain on *E. coli* TonB showed no presence of conserved transmembrane helix or strand domain compared to other Gram-negative TonB sequences.[149]

The third domain of TonB is the N-terminal region. In *E. coli*, the TonB N-terminal domain was reported to have a conserved (SXXXH) motif, which is in the transmembrane α-helix. This motif is essential for connecting TonB to the cytoplasmic membrane electrochemical gradient.[219] Again this *E. coli* conserved motif was present, although with less similarity among the four TonB proteins. In addition, the TonB protein was reported with variable sizes of N-terminal domain. Unlike *E. coli* TonB protein, the longer N-terminal domains were found to have a highly conserved loop region that codes for an M56-Zn2+ peptidase. A conserved putative zinc-binding motif of an M56-Zn2+ peptidase (HEXXH) was found in BFO_0333 and BFO_0233 of *Tannerella* TonB proteins.

3.2.1.3. Secondary structure of four *Tannerella* TonB proteins

Besides the similar motifs between *Tannerella* and *E. coli* TonB proteins, the secondary structure of the four TonB proteins was additionally explored using several different online servers, I-TASSER, Phobius, SignalP, and LipoP -1.0.[220–223] On the N-terminal domain, the presence of additional transmembrane helix and strand motifs were investigated in each TonB protein. The TonB BFO_0333 consists of a long N-terminal domain containing four transmembrane and three strand motifs. The BFO_0233 TonB has six transmembrane and two strand motifs on the N-terminal domain and it is the only TonB with a predicted export signal peptide for a secretory pathway. Only one transmembrane helix motif was found in both TonB BFO_0953 and TonB BFO_3116. The presence of N-terminal stand motif was found in TonB BFO_3116, but none was found in TonB BFO_0953. We explored more about the presence of signal peptide and cleavage sites, predictions of lipoproteins, and export signal peptide on N-terminal domain. The significance of the lack of these domains to the function of *Tannerella* TonB homologs resulted in no further analysis.

Unlike previous TonB protein studies,[149] this study noticed a consensus sequence of a transmembrane helix (GGQxxLMxxIA) on the periplasmic linker domain of the *Tannerella* four TonB proteins. However, the lack of conserved helix and strand motifs on this part of TonB made it question its importance. This transmembrane helix is highlighted in figure 3.4.

On the C-terminal domain, a bioinformatic analysis revealed that two transmembrane helix and six strand motifs in the C-terminal of BFO_0333, two transmembrane helix and three strand motifs in the C-terminal of BFO_0233, two transmembrane helix and three strand motifs in the C-terminal of BFO_0953, and one transmembrane helix and three strand motifs in the C-terminal of BFO_3116 (Table 3.3). The presence of six strand motifs in the C-terminal of BFO_0333 contrasted with the well-studied *E. coli* TonB. These transmembrane helix and strand motifs are highlighted in each amino acid sequence of *Tannerella* TonB (Appendix VIII: 8.10).

Table 3.3. Number of predicted transmembrane helix and strand motifs in *T. forsythia* TonB proteins.

Domain	BFO_0333/TF1354	BFO_0233/TF1255	BFO_0953/TF1960	BFO_3116/TF0783
N-terminal helix	4	6	1	1
N-terminal stand	3	2	0	1
Signal peptide in N-terminal	0	0	0	0
Signal peptide of lipid attachment in N-terminal	0	0	0	0
Export signal peptide in N-terminal	0	1	0	0
Pro-rich helix	1	1	1	1
Pro-rich stand	0	0	0	0
C-terminal helix	2	2	2	1
C-terminal stand	6	3	3	3

Taken together, these results show that each *T. forsythia* ATCC 43037 and 92A.2 strain codes for four different TonB proteins. These putative TonB proteins are predicted to have several roles including the necessary energy to TonB-dependent transporters. However, it is not known whether a TonB protein in this bacterium works as independent TonB protein, complete unit of TonB with ExbB and ExbD, or even as hybrid complex due to potential functional cross talk between proteins. With different protein varieties, the role of TonB to transport energy for substance acquisition may change to allow the expression of needed function. Several questions could be raised as this bacterium codes for monocistronic TonB, as well as TonB associated with ExbB and ExbD. With this in mind, the next step in this project is to explore more about the role of TonB systems first in delivering energy for the sialic acid uptake.

3.2.1.4. Annotations and Multiple Sequence Alignments of sialic acid specific TBDTs NanO and NanU

In terms of sialic acid transport, the TonB proteins are known to work in concert with the TonB-dependent transporters (TBDT)/Polysaccharide utilisation locus (PUL)-like gene products (neuraminate outer membrane porin) *nanO* and *nanU* (extracellular neuraminate uptake protein) (ATCC 43037 strain), and *TF0033* (similar to *nanO*) and *TF0034* (similar to *nanU*) (92A.2 strain) (homologues of *susC* and *susD*) (Table 3.4.).[88] Surprisingly, despite having functional equivalence, alignment of the amino acid gene sequences (NanO vs TF0033) from the two strains 43037 and 92A.2 showed fairly conservation (95.26 % identity) (Table 3.5). The alignment showed high level of sequence identity between both homologues at the N- and C-termini (Figure 3.5). In particular, the conserved N-terminal region encompasses the 'TonB-Box' and the 'plug' domains; whereas conservation in the β-Barrel receptor domain is lower, suggesting some plasticity in this region that still allows function.

Table 3.4. Genes identified as TBDTs and chosen for investigation in the study

Gene Name of ATCC 43037 Strain	Length			Gene Name of 92A.2 Strain	Length		
	bp	aa	kDa		bp	aa	kDa
BFO_2205/NanO	3384	1127	125.208	TF0033	3336	1111	123.43
BFO_2206/NanU	1569	522	58.053	TF0034	1572	523	58.164
Combined NanOU-equivalent to SusCD	4953	1649	183.26	TF0033,34 equivalent to SusCD	4908	1634	181.59

Table 3.5. Identities and similarities of TBDTs from ATCC 43037 and 92A.2 strains

	DNA Identity %	DNA Identity %	Amnio acid Identity %	Amnio acid Identity %
Genes	BFO_2205-NanO	BFO_2206-NanU	BFO_2205-NanO	BFO_2206-NanU
TF0033	94.32%		95.26%	
TF0034		100%		32.34%

```
                              N-terminal Domain
WT_NanO      MKRILKNVVLGGVFLFISLHPNGVNAKNNLGISETNISQQNKNVSGTVTDQSGEPVIGAS
WT_TF0033    MKGILKNVVLGGVFLFISLHPNGVNAKNNLGISETNISQQNKNVSGTVTDQSGEPVIGAS
             ** **********************************************************

             TonB Box
WT_NanO      VVEKGTRNGTVTDIDGNYSLSTASNAVLVFSYIGYASQEVSVAGKTKIDVVLKEDVAQLS
WT_TF0033    VVEKGTRNGTVTDIDGNYSLSTASNAVLVFSYIGYASQEISVSGKTKIDVVLKEDVAQLS
             ***************************************:**.******************

                          Plug Domain
WT_NanO      EVVVTGYGGRQLRSKVTNSISKVKEETLSQGLFSNPAQALSGAVAGLSVSQTSGNPGATP
WT_TF0033    EVVVTGYGGSQKRATLTTAISKMDDKVLQTAAYSNVGQALQGSVTGLRVVNTIGQPGSNP
             ********* * *:..*.:***.::..*.  . :** .***.*:*:** * :* *:**:.*

WT_NanO      TLVLRGGTNFDG-SGSPLILIDGQVRSSLSDINPDDIESMEVLKDAGATAIYGARANDGV
WT_TF0033    NIVLRGGATISGNNNSALIIVDGIVRESMQGLSPDDIESMQILKDAASTAIYGARANGGV
             .:*****:..*  ..*.**:**** **.*:..:.*****:****.:*****.***.**.**

WT_NanO      ILVTTKRGKSGRA-EVNLKAKFGMNYFHDSYQFLDAGQYLYWMRSAYKNAYVDELKNPDG
WT_TF0033    ILIETKKGTSGGASSVNYKFKVGVNYAREGYGFLNAQDYIHYNRLGFKRTGRKNVDTQMG
             **: **:*.** * .** * *.*:** ::.* **:* :*:::: *  :*.:  .::. *

WT_NanO      TKVKSWSSLGSLTGATPYGTGNAYFGKDGVTPLDGNKHASAVWSPMKY-NDKLAFLLDQG
WT_TF0033    ----------------YGIGNSLFD-------------------IRYLSDETAHLQKEG
                             ** **:.*.                    ::* .*: *.* .:*

WT_NanO      WQTMTDPVYGDK-IIYKNTF--MDEFNIQSPAYSQDYNLSVNGGNNKGSYYAGLGYNKSD
WT_TF0033    WLSMPDPFYEDRTILFKDYYGTLDDAVFDNSAMTHDHHLSFNGGNDKATYAASLGYYSED
             * :*.**.* *: *::*: .  :*:  :*..:**.****:*.:* *.*** ..*

                                   TonB-dependent Receptor Domain
WT_NanO      GTAYGNWYKRITFTLNADYKLKEWLTSNSSFTADATWYGLPPSQADEANYFSRVFSVPP
WT_TF0033    GQVKGTGLQRFTGSLNASYKVLPFLAVKGGTTY---SWQRQPSLWIGTYEFFYRTRSQRP
             *  . *.  :*:* :***.**: :*: :.. *:   :* *.  . ::* *. * *

WT_NanO      TFRGYNAKGEMLLGPNSGDGNQQYNFKQFIRDNNTDKFTINQSFTLDIM-KGLSLKLGTI
WT_TF0033    TWNPWLADGSPASGFGTGDGNPLYYKDRLTRKNGTRRQTYTMGFTLDIIPQKLVLNGNAS
             *:. : *.*.    * .:**** *   .:: *.*.* : *  .****:: : * *: .:

WT_NanO      WYFSEGKYEAFNKDYLSSPGNM-NTARHTSASYKRTLDQTYNAILNYKRQINKDHYLDAM
WT_TF0033    LYHYDHQEEEFNKKYQTQNSNTPNTTRQARAKYLKRNQQQVSVTLSYTDTFADRHNLDVM
             *. : * ***.* :. :* **:**:: .* *.*.*. : *.*. .:. *.*

WT_NanO      AGFEYYDSYEKGFNAAGSGAPTDDFMDLGLTSKKENQRSIDSWHSRQRIMSFFGRVNYDY
WT_TF0033    FGGEYFNHHEFLFDATTEGSPLDDIATLNVGATRTKTTSEKTGY---RILSAFGRANYNY
             * **:: :*  *:*: .*:* **:  *.: :.: *  *.:    **:* ***.**:*

WT_NanO      QSKYLLSIVMRKDGYSKLTEENRWGIFPGVSAGWVFNKEEFMAGT--TDILSFGKLRASF
WT_TF0033    DMKYLLTLTARYDGISRLV-DNRWGFFPGVSAGWNVMEEEFFKESPISDYVSNLKPRISY
             : ****::. * ** *:*. ;****;******** . ;****: : :* :* * * *:

WT_NanO      GLNGNVNKNFVGNYTVQGSYGSLT-YNGSTGFLLG---------AIPNPYLLWEKSRTFEV
WT_TF0033    GVNGNVSA--IGNFEAYGIYGQTKDYAGTTGFQSGYESNGAFIGLMNSKLRWEQSQTFEV
             *:****.   :**:. * **.  * * *:****  .: *.  * * ** *:**:*****

WT_NanO      GLDLGFLENRINANLTYYNRLTTDKYANITIPSTSGVNSVVSNNGKFQNQGLEFELAFRI
WT_TF0033    GLDFSFFNNRLSFIADYYDRRTKDLLTSLALPGYTGYNSITTNLGILRNYGFELEVKANI
             ***:.*::**:.   **:* *.*  :.:::*. :* **::* * * ::* *:*:*:  .*

WT_NanO      IDNKD-WKWNISWNGAMNRNKVIQLPDNGLERNRQGAFQVYTGNGNEKMWVGGYQEGQTP
WT_TF0033    LQLKNGFTWDMSANVSTVANKVVKLPYNGNERNRQGGVQVWNPEKGALAWLGGTEEGGKL
             :: *: :.*::* * :    ***:;** ** ******,.**:. : .   *:** :** .

WT_NanO      NDLYAFVAEGIYRSQDEI---PKGLIDITSGGFGSTGRPLYGGAEGYHKLSESQKKNALP
WT_TF0033    GVLYGYKQEHIFKDWDDVKAKANKRLDHVASLYGP------GLADEY-----KDKKGWQP
             . **:. *.*::.:. .*..   .: :*. :.*.        *  .:     . :**. *

WT_NanO      IQPGDIKWKDVNGDGVIDNFDRVKIGNTVPKWTGGISTSLSWKSLTLSARMDYALGFKAV
WT_TF0033    IEPGDVCWADLNNDGIIDGKDRVEIGNIFPNVTGGFSTTLGYKGLSLYARFDYALGHTIY
             *:***. * *:*.**.**:.** *** .*: *.* ***:**:*.:*.*:* **:****..

WT_NanO      DWRTPWFMGAAQGTFNTIAETKDTWTPDNPNAAYPKYVWADQLGKRNYSRT---------
WT_TF0033    NDLAARSLGQYQGSFNIIDMVKDTWSETNSSSDLPKFYYADQLSKKNITRSNNANTAANN
             :  :.  :*  **;** *  .****. *..:  **: ;****.*:* ;*:

WT_NanO      -TSMFIYKGDYLALRELTLAYKLPALWIEKARLNAVELSVTGQNLGYLTQAKHLFSPEKA
WT_TF0033    NSSRFYEKGDYMALRELTLSYALPSRWIRKASMQEATVSITGQNLFYITGYTGV-SPEPA
             :* * ****;*******;* **; **.** ::  . ;*:***** *;*   . ; *** *

WT_NanO      N--------NNGGYPLPRTVIFGINVTFFSPEKANNNGGYPLPRTVIFGLNVTF
WT_TF0033    VSNYYGRGIDNGRYPTPRTVLFGLSVTF-------------------------
                      :** ** ****;**:.***
```

Figure 3.5 Amino acid alignment of NanO and TF0033.
Alignments using T-coffee alignment programme showed high consensus sequence between both sequences. Conserved residues between *nanO* and *TF0033* were indicated by stars below the alignment. The plug domain in the outer membrane highlighted with red colour whereas the blue colour represents the TonB-dependent β-barrel receptor domain in both strains.

In addition, the second member of PUL is NanU/TF0034 (homologous SusD). Although the ATCC 43037 NanU and the 92A.2 TF0034 are analogous in predicted function and their DNA sequences are 100 % identical, the global alignment of their amino acid sequences showed relatively low levels of identity (32.34 % identity) (Table 3.5). This may suggest that the conserved part between both proteins is for sialic acid binding.[214] Amino acid alignment was performed for NanU and TF0034 showing conservation between both homologues, with particularly high homology at the C-terminal domains in both amino acid sequences (Figure 3.6).

```
WT_NanU      MKKIANHIVLWIAFMVVLMSCDSLNLAPEDYFGSGNFWKNEAQVNGALIGLHADLRRSYF
WT_TF0034    MKKFIIYL-MAITSPALLLQCNSLEMEPESSITDANYWKTEAHFSAFNTGLHSLLRKYSY
             ***:  :: : *:  .:*:.*:**:: **. : ..*:**.**:...   ***: **:  :

WT_NanU      MFYHLGEIR----------GGTQRVGSSSLNTSLHYANERSNLFDKDRTGIRNWYGLYDN
WT_TF0034    HYFLLGESRANLYCDDNPFGGEATQGIERLPY---------NTLNKENPVVSNFGGLYVP
              :: *** *        **    * . *      * ::*:.. : *: ***

WT_NanU      LFQVNHFIDQVENKCTFLSDESRKNLLAQAYGFRALYYFMLYKTYGGVPIVTETEILNGK
WT_TF0034    IERLNMMIRRT-SETNLLTDADKNYYLGEAYGMRAYLYFHLLRSWGDV--VLSTDPTDGA
             : ::* :* :. .: .:*:* .:: *.:***:** ** * ::*.* * .*: :*

WT_NanU      ATAEKFYVERATPQA-TMEFIKKDILKSEEYYAGKTVSNNYPKMMWSKAATLMLKAEIYL
WT_TF0034    SLDVFNLAKAASPAADVMKQIKADIEASEKAF-GDNHSFKYGRHYWSLPATWMLKGEVYL
             :       .: *:* * .*: ** **  **: : *.. * :* :  ** .** ***.*:**

WT_NanU      WAAKVSIQGYTATGTDDLKVAKTALNEVVGKFE--LLKDFGSIFSTSNRNNKELVFTLHF
WT_TF0034    WSGR-----QMGGGNADYQVARMAL-EQVGKADVALSEKFESVFAFSNKKNKEIIFTIHN
             *:.:      . *. * :**: ** * *** :  * :.* *:*: **::***::**:*

WT_NanU      ADREAT---NNASSFLYQQALFVGQVYGRNGKKIENDTLNLKGTGGVFRSEYTYDFW-KT
WT_TF0034    QKDEFSLWDDRYRMNMLPQDAYMTLYYDEEGKPYQKEPA-LEKLRGLIRLQINYDMYQKL
              . * :   :.    : * ::   *..:** :::.  *:   *::* : .**:: *

WT_NanU      YDAADQRRDATFLEYYG----DSLLTASKFGCVMKKGVGSINENNNRVYDA-DIIIYRYA
WT_TF0034    FRDEDSRKRATLKAVYSKDKDDNLKYVAPYANKFK---GTLVEGNSSCSMLDDYPIYRYA
             :    *.*: **:   *.     *.* .: :. :*   *:: *.*.      * *****

WT_NanU      DALLMMAEVSNGLGEPCADYINQVRKRAYGKH------------------YEEHSYKES
WT_TF0034    DCLLMLAEVKVLLGEDPAEEINKVRERAYGANYFKENKDKVAYPNDTDPAVYTNNRYVGS
             *.***:***. *** *: **:**:**** :          * :: *  *

WT_NanU      DFATNELAILHERDKEFVWEGKRWFDVVRMRDASKNSLAFSASANYPAGSPLISASEK--
WT_TF0034    DADALE-AILKERLREFMFEGKRWYDLRLL------------------GEPYLSKYSKAN
             *  : * ***:** :**::*****:*:  :                   *.* :*  .*

WT_NanU      -HKLLWPLDIGTMNINPLLKQTPGYE
WT_TF0034    ASRLLWPINEDVLTNNPLLKQTPGYE
              :****:: ..:. ***********
```

Figure 3.6 Amino acid alignment of NanU and TF0034.
Alignments using T-coffee alignment programme showed somewhat consensus between both sequences. Red colour indicates conserved residues between NanU and TF0034 sequences.

3.2.2. Establishing Function of Individual *tonB* and *nanOU* Genes in Sialic Acid Transport

In this study, we planned to use two methods to establish the function of each *tonB* using heterocomplementation in *E. coli* MG1655 (section 3.2.2.1.) and then to examine the question around the function of the individual *tonB* in *T. forsythia* using knockout mutagenesis (section 3.2.2.2).

3.2.2.1. Heterologous Complementation in *E. coli* System

To establish function of each *tonB* for sialic acid transport, a heterologous complementation system that was previously developed in the lab was employed. Lack of a genetic complementation system in *T. forsythia* and the ease of manipulation of *E. coli* created the opportunity for heterocomplementation in *E. coli* MG1655. This system potentially was used to transform each *tonB* with either *nanOU* or *TF033,34* and establish the function of each *tonB* with sialic acid transit in unison when expressed on compatible plasmids and modified to allow testing function of the different *tonB* genes from *T. forsythia*. This system works by utilising strains that lack the *E. coli* outer membrane sialic acid transporter *nanC* and ompCF (via OmpR control), as well as having a mutation in the sialic acid catabolism operon transcriptional repressor NanR to allow maximal expression of the *nan* operon sialic acid uptake genes; this is known here as the Δ*nan* strain (MG1655 Δ*nanCnanR*(*amber*) Δ*ompR*::*Tn10*(*tet*)). For testing the function of *tonB* genes, this gene is also missing, making the Δ*nan*Δ*tonB* strain (MG1655 Δ*nanCnanR*(*amber*) Δ*ompR*::*Tn10*(*tet*) Δ*tonB*::*FRT-Km-FRT*).[128] This *E. coli* system was tested and used in our laboratory previously in the experiments proving function of *TF0033-34* and of *tonB* function by Roy et al.,[88] this strain cannot grow in minimal media where sialic acid is the only carbon source, whereas complementation allows growth. This enables a simple test for function which can be complemented by use of a sialic acid measurement assay to assess uptake this sugar by the cells that is known as the ThioBarbiturate (TBA) assay,[209] and was developed to test function of the *nanT* gene in *T. forstyhia* in our lab.[91]

3.2.2.2. Inspection of ΔNanΔTonB Strain Before Gene Complementation

The Δ*nan* strain was gifted to our school before previous work at our laboratory introduced a deletion to the *tonB* gene (Δ*nan*Δ*tonB*). As this strain was stored for a long time, it was necessary for this project to check and confirm the presence of deleted genes and confirm their activities. Thus, this study designed several Polymerase Chain Reaction (PCR) primers to check the inactivity of sialic acid outer membrane (*nanC*), *ompR* regulator gene of the osmoregulatory expression of the *ompF* and *ompC* genes, and *tonB* gene, all they are explained in three stages.

For the sialic acid outer membrane (*nanC*), this study does not know how *nanC* was deleted and where the mutation was inserted. Since the size of *nanC* gene is 717 bp, we extracted the sequence of this gene from NCBI and designed three sets of PCR primers to test both the WT *E. coli* MG1655 strain and Δ*nan*Δ*tonB* MG1655 strain. The first set of primers was a 150 bp forward primer from the start codon of *nanC* and a reverse primer after the end codon (887 bp). The expected 887 bp was confirmed and showed at lane '1' for the WT *E. coli* and at lane '2' for the altered strain (Figure 3.7). The second set of primers was a forward primer at the start codon of *nanC* and a reverse primer at the end codon of *nanC* (717 bp). The size of 717 was estimated at lane '3' for the WT *E. coli* and at lane '4' for the altered strain. The third set of primers was a forward primer of 20 bp before the start codon and a reverse primer of 20 bp after the end codon (757 bp). The size of this product was confirmed at lanes '5' and '6' for both the WT *E. coli* and altered strains, respectively. Surprisingly, there was no difference in size products between the WT *E. coli* and altered strains, indicating the needed further steps of testing for *nanC* gene.
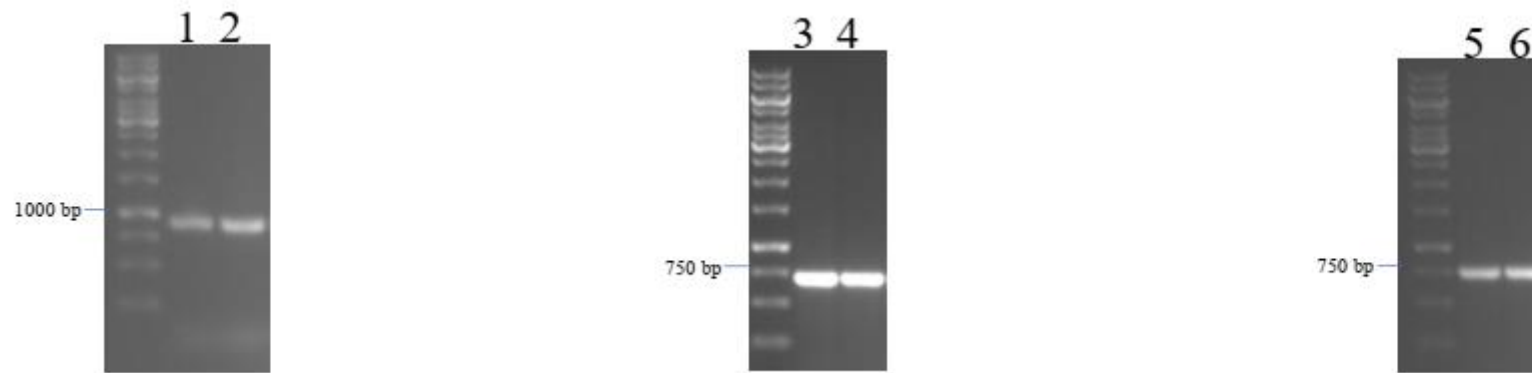
Figure 3.7 Agarose gel confirming the sialic acid outer membrane in the WT _E. coli_ and the altered Δ_nan_Δ_tonB_ strains.
The nanC was amplified with three sets of primers on the PCR using Phusion HF™ polymerase enzymes. A designed 150 bp forward primer from the start codon and a reverse primer after the end codon results in a product of 887 bp for the WT and the altered strains, lanes 1 and 2 (First set of primers). Lanes 3 and 4 shows a similar product of 717 bp for the WT _E. coli_ and the altered stains using the second set of primers. A product of 757 bp was showen for the third set of primers of testing both the WT _E. coli_ and the altered stains (lanes 5 and 6).

Next, we sought additional clarification of *nanC* gene by DNA sequencing for both the WT *E. coli* and altered strains. As a result, the sequence of *nanC* from the altered strain shows a similar start codon compared to the WT *E. coli*. It shows that both the WT *E. coli* and altered strains have the same initiation codon, but they are different on the second amino acid. The WT NanC translation is Met-Lys, whereas the alter strain is Met-Ser, although in *E. coli* it states that the most common second amino acid following the Met (AUG) is lysine (Lys), mostly with its AAA codon.[224] Thus, such a difference in the second amino acid may affect the level of expression of the protein. Also, the difference comes between the WT and the alter strain on the third amino acid as the alter strain has a stop codon (TAG=UGA), indicating a change to the expressed protein. However, the rest of the DNA sequence between the WT and the altered strain are the same. Therefore, this altered strain is mutated with a site-directed mutagenesis affecting its protein production and structure (Figure 3.8).
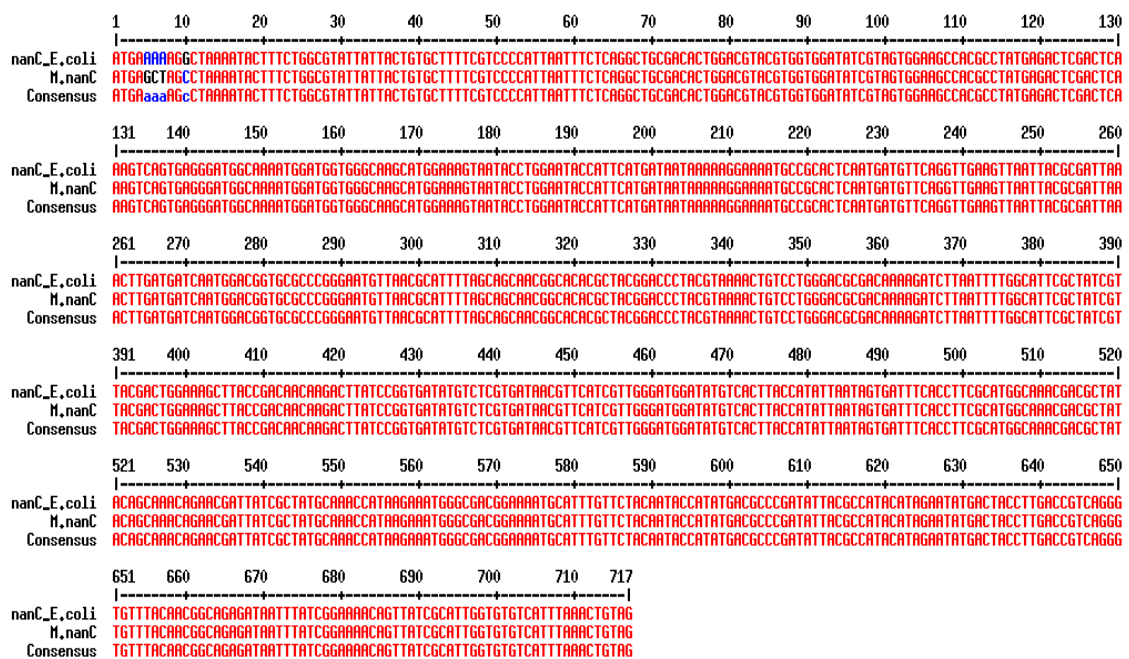
Figure 3.8 Alignments of *E. coli* MG1655 *nanC* compared to altered strain.
Alignments between two *nanC* genes from two *E. coli* strains showed high consensus homology between both sequences. The difference between the WT *E. coli* strain and Δ*nan*Δ*tonB* MG1655 strain was presence on the second and third codons.

OmpR (27.355 kDa) is a cytoplasmic protein and is one of the regulatory proteins that can activate the major outer membrane proteins OmpF (40.25 kDa) and OmpC (40.48 kDa) of *E. coli* in response to medium osmolarity. OmpF and OmpC form diffusion channels in the membrane to allow access for small hydrophilic molecules (650 Da). OmpR can control a role in the regulation of a wide variety of genes. These genes are part of the expression of the *csgDEFG* operon, flagellar master regulon *flhDC*, enterochelin synthesis, metabolism, cell envelope, and signal transduction and transport.[225] In addition, OmpR can regulate *omrA* and *omrB* which regulate the outer membrane protease *ompT* along with iron transport genes *cirA*, *fecA*, *fepA*.[226] The global regulator OmpR of *E. coli* MG1655 strain was deleted previously using Tetracycline cassette. Since this strain was gifted to us many years ago, the type of cassette and how it inserts were not known as the OmpR consists of N- and C-terminal domains. Thus, this study obtained the sequence of *ompR* gene from NCBI (710 bp) and designed a 20 bp forward and a 20 bp reverse PCR primers to compare this gene between both the WT *E. coli* MG1655 strain and the altered version strain (Δ*nan*Δ*tonB*). As a result, the PCR gel shows a product size of 756 bp for the WT *E. coli* (lane 1) as expected compared to a different size for the altered version strain (Δ*nan*Δ*tonB*) (lane 2) (Figure 3.9).

A further assessment for the altered version strain (Δ*nan*Δ*tonB*) was continued to check the deletion of *tonB* gene. The mutagenesis of the *tonB* was introduced in our laboratory previously using the pKD13 plasmid of Kanamycin cassette. The pKD13 plasmid has several locations of PCR primers (Kt, K2, K1) to confirm the presence of Kanamycin cassette. Here, we designed a 120 bp forward primer before the cassette and used K2 as a reverse primer to obtain a product of 883 bp (lane 3, figure 3.8). We also used the cassette forward primer (Kt) and the cassette reverse primer (K2) to obtain a product of 471 bp (lane 4, figure 3.9). Both set of primers confirm the deletion of *tonB* gene.
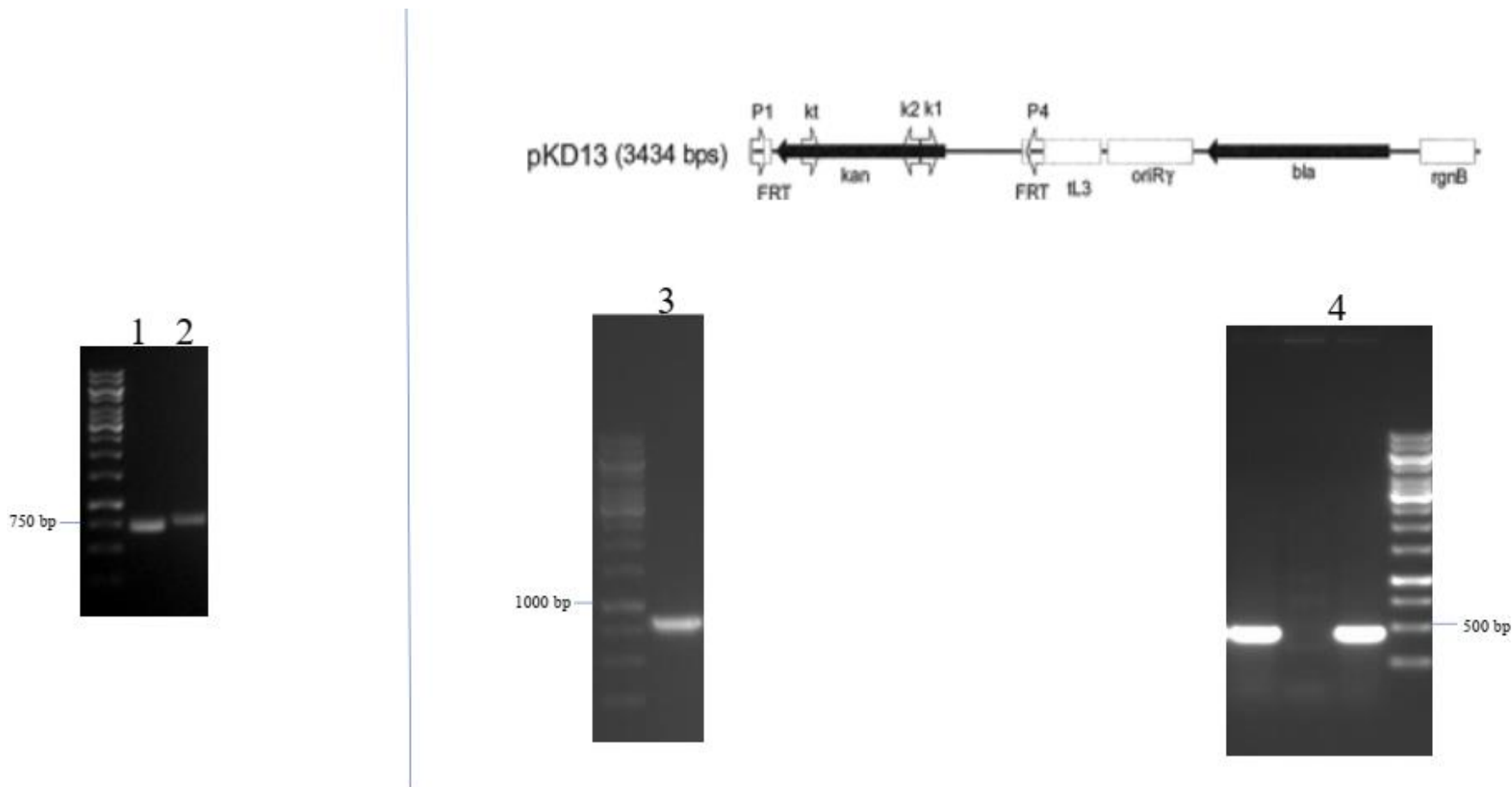
Figure 3.9 Agarose gel confirming the *ompR* and *tonB* deletions in the WT *E. coli* and the altered (∆*nan*∆*tonB*) strains.
The *ompR* was amplified with two primers on the PCR using Phusion HF^{TM} polymerase enzymes. A designed 20 bp forward primer from the start codon of *ompR* and a reverse primer 20 bp after the end codon results in a product of 756 bp for the WT *E. coli* strain and a different size for the altered strain (>750 bp), lanes 1 and 2. On the top right side of the figure, the locations of PCR primers (Kt, K1, K2) on pKD13 plasmid are shown. To confirm deletion of *tonB* gene, lane 3 shows a product of 883 bp from a combination of 120 bp forward primer and K2. Furthermore, lane 4 shows a product of 471 bp from the cassette forward primer (Kt) and the cassette reverse primer (K2).

A further membrane protein profile was used to confirm the absence of NanC (27.88 kDa), OmpC (40.48 kDa), and OmpF (40.25 kDa) from soluble and insoluble based on disruption of the cell wall of *E. coli*. Both soluble and insoluble membranes were electrophoresed on a sodium dodecyl sulfate 12% polyacrylamide gel. As a result, the absence of NanC was noticed and confirmed from the insoluble proteins. Likewise, we were able to detect the absence of OmpC, but not the OmpF (Figure 3.10).

Although the OmpR can bind three tandem sites from the transcriptional start site at both the OmpF and OmpC promoters, the OmpR can differently regulate the OmpC and OmpF.[225] Both OmpF and OmpC have a protective role as a barrier for antibiotics and other toxic agents entering inside the cell. However, it was found that the OmpC was significantly deceased compared to OmpF under concentration gradient imipenem and meropenem stress.[227] Furthermore, the affinity model suggested that the OmpR at low osmolality can activate the expression of the OmpF, whereas at high osmolality, OmpR can activate expression of the OmpC. Later by Mattison and others showed that an *ompR* mutant can alternate model for differential porin gene expression.[228] Chakraborty and others tested the *ompR* mutant *in vivo* using a strain with over-expression showing binding at *ompF* and altered binding at *ompC* on the DNase I footprinting assays. As a result, the *ompR* mutant did not activate transcription of both *ompF* and *ompC*, which indicates the role of OmpR with different conformations when bound to both OmpC versus OmpF. Binding at *ompF* but not at *ompC* suggests the possibility of an affinity model where the *ompR* mutant is locked into a low osmolality conformation.[229] A further possibility is that the *ompR* is bound to *ompC* DNA, but a conformational change to facilitate RNA polymerase (RNAP), oligomerisation, or other events may be required for transcriptional activation.[230] This may favour an old conclusion of Sheridan and others where they predicate the integration host factor (IHF) that protects approximately 40 bp upstream of the OmpR mediating duplex destabilisation favouring RNAP binding.[231]
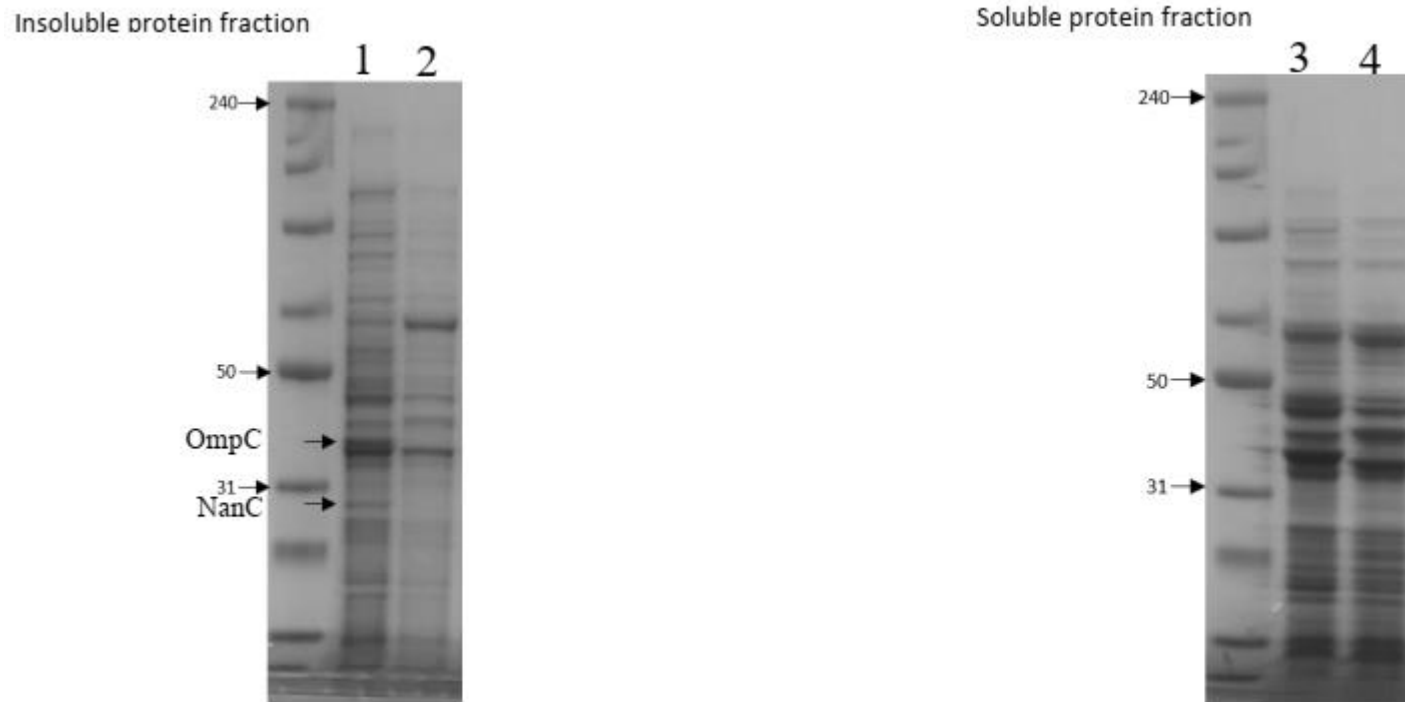
Figure 3.10 Identification of OmpC and NanC absence.
Soluble and insoluble membranes fraction were prepared from overnight culture of *E. coli* MG1655 and Δ*nan*Δ*tonB* strains. Soluble and insoluble proteins were electrophoresed on a sodium dodecyl sulfate 12% polyacrylamide gel. Insoluble protein fraction shows the presence of OmpC and NanC in *E. coli* MG1655 strain (lane 1) and absence of OmpC and NanC in Δ*nan*Δ*tonB* strain (lane 2). Soluble protein fraction shows no missing band between *E. coli* MG1655 strain (lane 3) and Δ*nan*Δ*tonB* strain (lane 4).

3.2.2.3. Amplification and Ligation of *tonB* genes

After identifying the *tonB* and TBDT genes in both strains of *Tannerella* and testing the Δ*nan*Δ*tonB* strain to check it has the correct mutations, PCR primers were designed to amplify these genes from *T. forsythia*. As a positive control for the complementation experiments, the *tonB* gene of *E. coli* was used to confirm reliability of this complementation. These primers included restriction sites that are present in the cloning sites of the plasmids chosen for these experiments, while the 3' primers also contain stop codons (Table 2.7). Due to the high identities between *tonB BFO_0953* (ATCC 43037 strain) and *tonB TF1960* (92A.2 strain), and *tonB BFO_3116* (ATCC 43037 strain) and *tonB TF0783* (92A.2 strain), this study excluded *tonB TF1960* and *tonB TF0783* from 92A.2 strain. Polymerase Chain Reaction (PCR) was used to amplify these genes after DNA extraction from *T. forsythia* and *E. coli*, with agarose gel electrophoresis to confirm each gene correct size (sections 2.3.4. and 2.3.6.). Gel analysis showed successful amplification of *nanOU* (lane 1), *TF0033,34* (lane 2), *BFO_0333* (lane 3), *E. coli tonB* (lane 4), *nanO* (lane 5), *TF1354* (lane 6), *BFO_0233* (lane 7), *TF1255* (lane 8), *TF0033* (lane 9), *BFO_0953* (lane 10), and *BFO_3116* (lane 11) (Figure 3.11). This process was then followed by PCR purification to remove primers before restriction and cloning (section 2.3.7.).
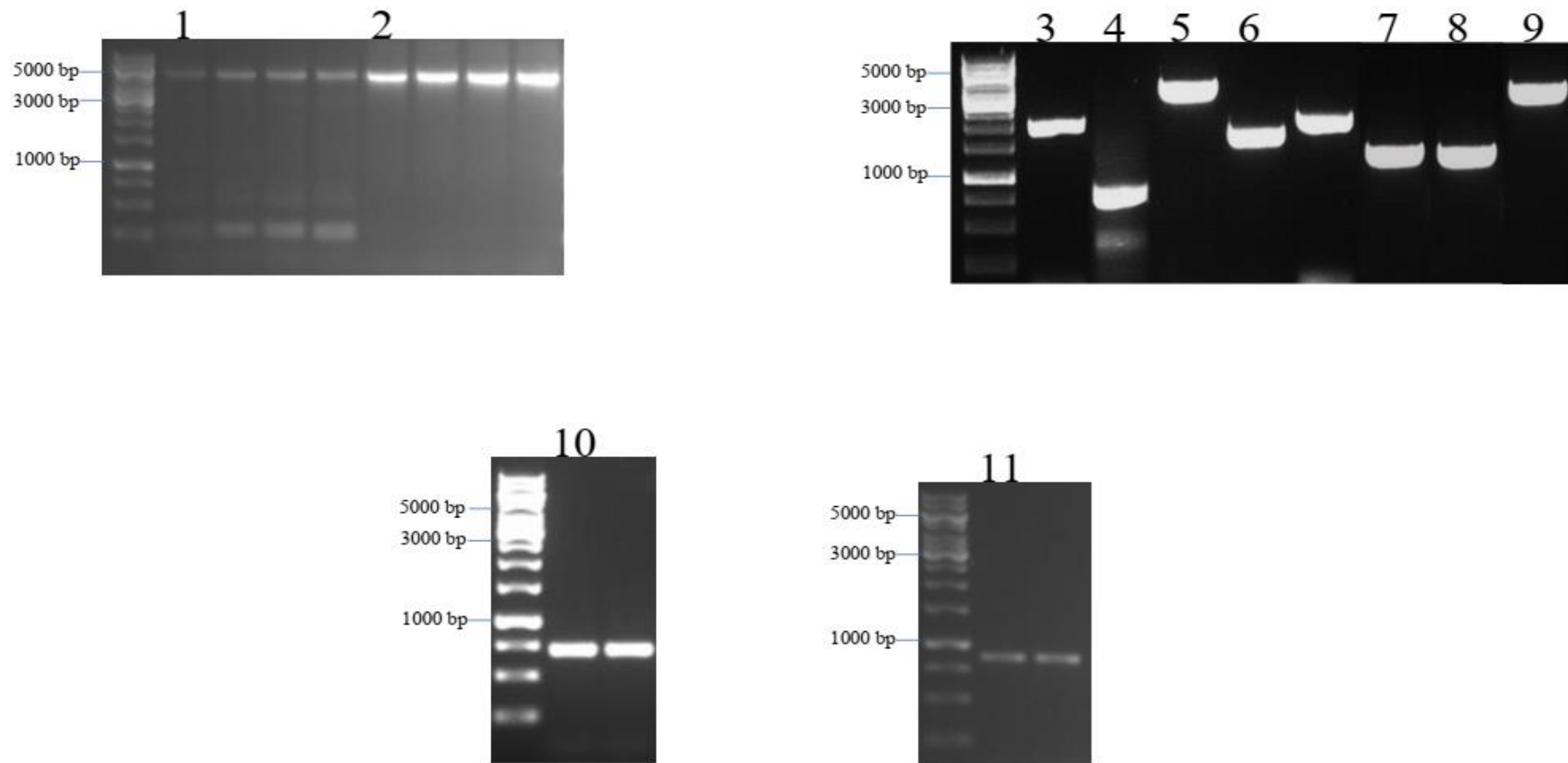
Figure 3.11 Amplification of TBDTs and putative *tonB* genes in *T. forsythia* and *E. coli* strains.
The PCR products were analysed by gel electrophoresis and confirmed by estimating every gene size using known sizes of 1Kb DNA Ladder. This figure presents *nanOU* (lane 1 – 4953 bp), *TF0033,34* (lane 2 – 4908 bp), *BFO_0333* (lane 3 – 1867 bp), *E. coli tonB* (lane 4 – 720 bp), *nanO* (lane 5 – 3384 bp), *TF1354* (lane 6 – 1867 bp), *BFO_0233* (lane 7 – 1440 bp), *TF1255* (lane 8 – 1440 bp), *TF0033* (lane 9 – 3336 bp), *BFO_0953* (lane 10 – 699 bp), and *BFO_3116* (lane 11 - 837).

The next stage was to prepare these PCR products for ligation into the pBAD18 or pAcTrc99a plasmids. These plasmids were chosen for several reasons, firstly, both contain inducible promoters for control of expression of the genes inserted 3' of these promoters. In the case of pBAD18 (Figure 3.12), the inducer is arabinose, driven by the AraB promoter, contains a useful Multiple Cloning Site (MCS), and the pBR322 origin of replication with ampicillin resistance.[207] Of note, pBAD18 has been used in the lab previously for expression of the *nanOU* genes in complementation experiments, but it was lost when I started this project. In case of pAcTrc99a (Figure 3.13), it contains the pTrc promoter, and is inducible with IPTG, as well as having an MCS. PAcTrc99a contains a different origin of replication (deriving from pACYC184) and antibiotic resistance cassette (Cm).[232] Overall, we have two plasmids that can only be co-transformed into *E. coli* based on their alternative replication origins and antibiotic resistances, but also that have separate inducer profiles to allow fine control of each genes expression in experiments if needed (NB, in previous experiments no induction was needed for complementation). In these experiments, the *nanOU*, *nanO*, *TF033,34*, and *TF033* genes were to be cloned into pBAD18, while *tonB* genes into pAcTrc99a.
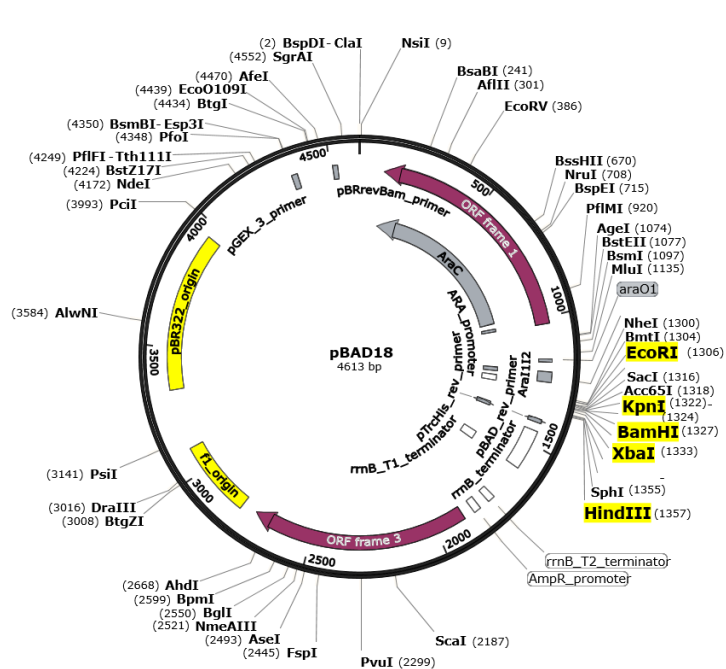
Figure 3.12 pBAD18 plasmid map.
This plasmid was used to carry *nanOU*, *nanO*, *TF0033,34*, and *TF033* of ATCC 43037 and 92A.2 respectively. Restriction sites used in this study are highlighted in yellow. This map is displayed by SnapGene software.
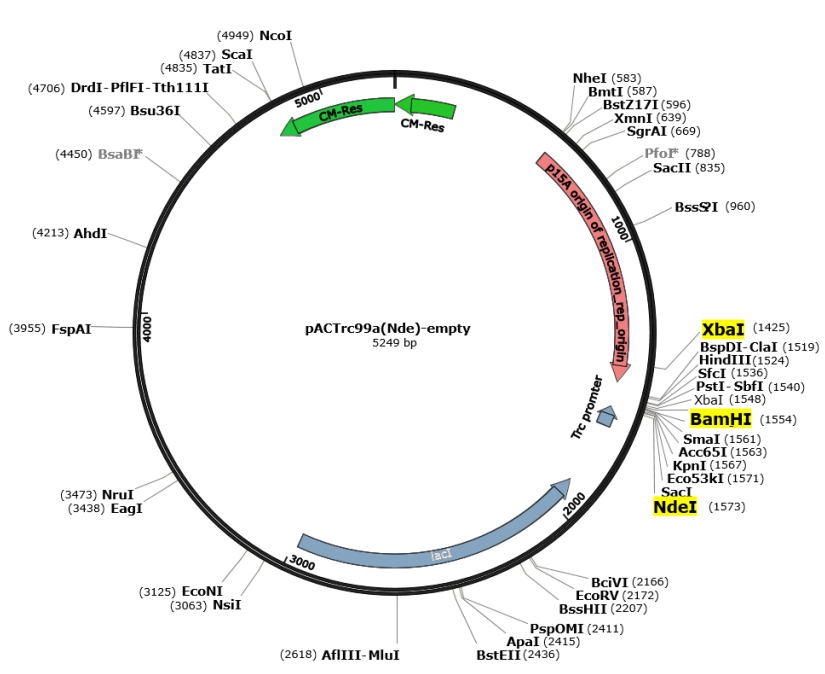
Figure 3.13 pAcTrc99a plasmid map.

This plasmid was used to carry all *tonB* genes of *T. forsythia* and *E. coli*. Restriction sites used in this study are highlighted in yellow. This map is displayed by SnapGene software.

Before cloning the PCR products into pBAD18 or pAcTrc99a plasmid (Figures 3.12 and 3.13), a double digestion of vectors and amplified PCR products were performed using the appropriate restriction enzymes as in the table 3.6. The process of this digestion explained in section 2.3.8, followed by purification (section 2.3.7.), and running on an agarose gel electrophoresis to confirm the products presence (data not shown). Once size of the product confirmed, DNA and vectors were extracted from agarose gel (section 2.3.9.), followed by purification (section 2.3.7.), and prepared for the ligation process.

Table 3.6. List of the restriction enzymes and buffers used

Strain/vector	Restriction enzyme 1	Restriction enzyme 2	Buffer
*nanOU/*pBAD18	KpnI-HF	XbaI	CutSmart Buffer
nanO/pBAD18	KpnI-HF	HindIII-HF	CutSmart Buffer
*BFO_0333/*pAcTrc99a	NdeI	XbaI	NEBuffer2.1
*BFO_0233/*pAcTrc99a	NdeI	BamHI	NEBuffer3.1
*BFO_0953/*pAcTrc99a	NdeI	BamHI	NEBuffer3.1
*BFO_3116/*pAcTrc99a	NdeI	BamHI	NEBuffer3.1
*TF0033,34/*pBAD18	KpnI-HF	Xbal	CutSmart Buffer
*TF0033/*pBAD18	KpnI-HF	BamHI	CutSmart Buffer
*TF1354/*pAcTrc99a	NdeI	XbaI	NEBuffer2.1
*TF1255/*pAcTrc99a	NdeI	BamHI	NEBuffer3.1
*E. coli/*pAcTrc99a	NdeI	BamHI	NEBuffer3.1

After the digested PCR products and plasmids were extracted from the gel, the DNA of (*nanOU*, *nanO*, *TF003334, and TF0033*) was each ligated separately into the pBAD18 plasmid (section 2.3.10.). Similarly, the *tonB* genes from *Tannerella* and *E. coli* strains were ligated into pAcTrc99a plasmid. After overnight incubation at 16ºC, the ligation products were transformed into DH5α cells. Colonies that appeared on a selective agar medium were randomly chosen and confirmed by colony PCR method using the original cloning primers to identify positive colonies (section 2.3.4.2.). The presence of the band indicated initial successful ligation of insert into the plasmid of interest. Size of the bands 4953 bp, 4908 bp, 3384 bp, and 3336 bp confirmed the ligation of *nanOU* (lane 1), *TF0033,34* (lane 2), *nanO* (lane 3), and *TF0033* (lane 4), respectively (Figure 3.14). We also confirmed all *tonB* genes ligation to pAcTrc99a: *BFO_0233* (1440 bp – lane 5), *TF1255* (1440 bp – lane 6), *BFO_0333* (1867 bp – lane 7), *TF1354* (1863 bp – lane 8), and *E. coli* (834 bp – lane 9). The ligation of *BFO_0953* (699 bp – lane 10) and *BFO_3116* (837 bp – lane 11) were lost during the pandemic of COVID-19 for unknown reason, however, we failed to re-ligate these genes again for several times, leading us to order them from a company.
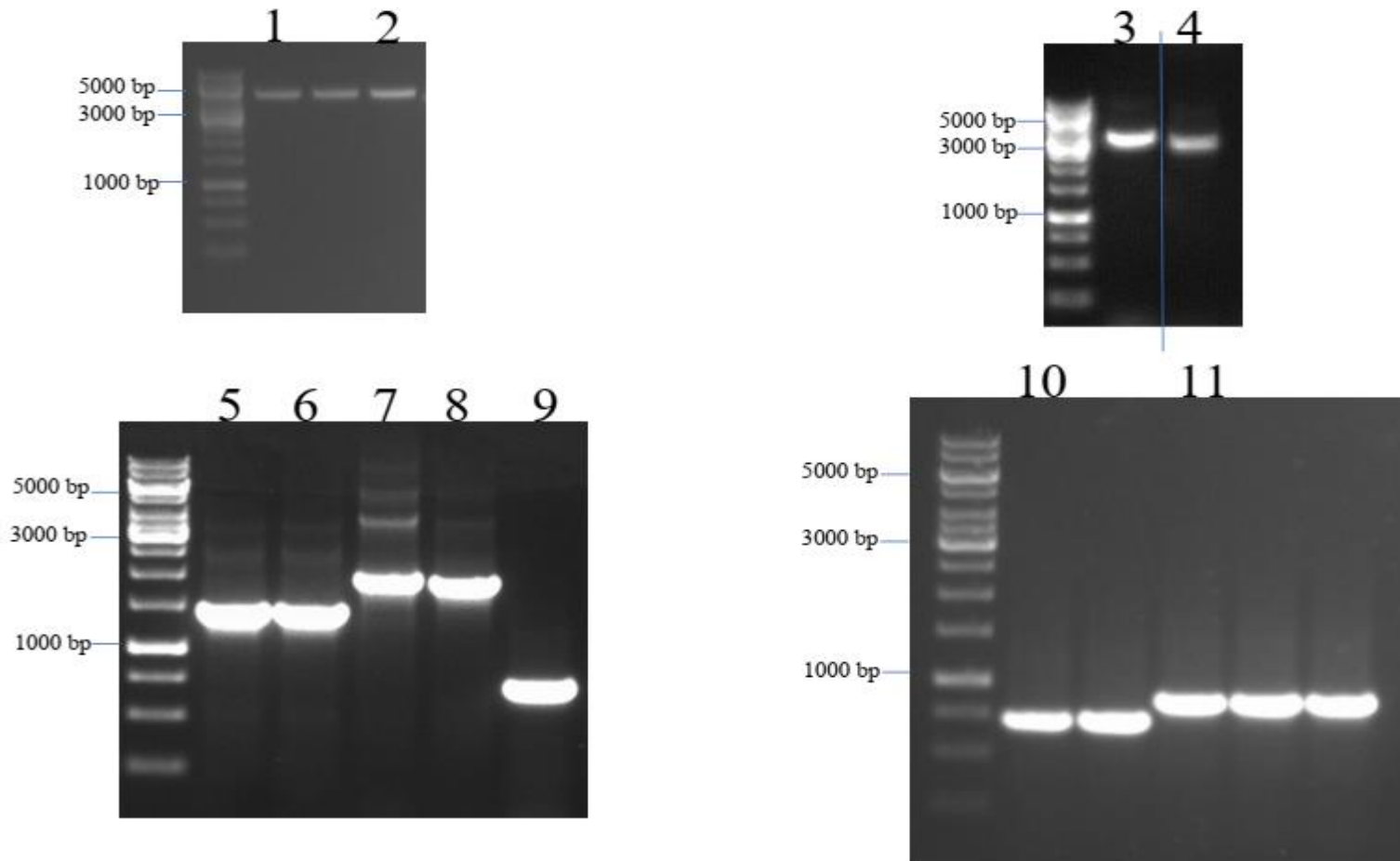
Figure 3.14 Agarose gel confirms ligation of TBDT and *tonB* genes.
Each gene was amplified with two primers on the PCR using Phusion HF™ polymerase enzymes. Each primer has a restriction enzyme site to be ligated into a vector of interest. This figure shows ligations of *nanOU* (4953 bp – lane 1), *TF0033,34* (4908 bp – lane 2), *nanO* (3384 bp – lane 3) and *TF0033* (3336 bp – lane 4) into pBAD18 vector. This figure also demonstrates ligation of *tonB* genes into pAcTrc99a vector: *BFO_0233* (1440 bp – lane 5), *TF1255* (1440 bp – lane 6), *BFO_0333* (1867 bp – lane 7), *TF1354* (1863 bp – lane 8), *E. coli* (834 bp – lane 9), *BFO_0953* (699 bp – lane 10), and *BFO_3116* (837 bp – lane 11).

118

Once colonies of ligations were analysed on agarose gels, they were confirmed using two methods. First, each colony was cleaned up and digested again with the same restriction enzymes resulting in a specific banding pattern of plasmid and insert (Table 3.3.). Figure 3.15 showed two bands of *nanO* (lane 1), *TF0033* (lane 2), *BOF_0333* (lane 3), *TF1354* (lane 4), *BFO_0233* (lane 5), *TF1255* (lane 6), *BFO_0953* (lane 7), and *BFO_3116* (lane 8), and their cloned vectors (pBAD18 is 4613 bp and pAcTrc99a is 5249 bp), indicating successful ligation (Figure 3.15). Second, 5 µl of insert was checked by DNA sequencing (section 2.3.11.2.).
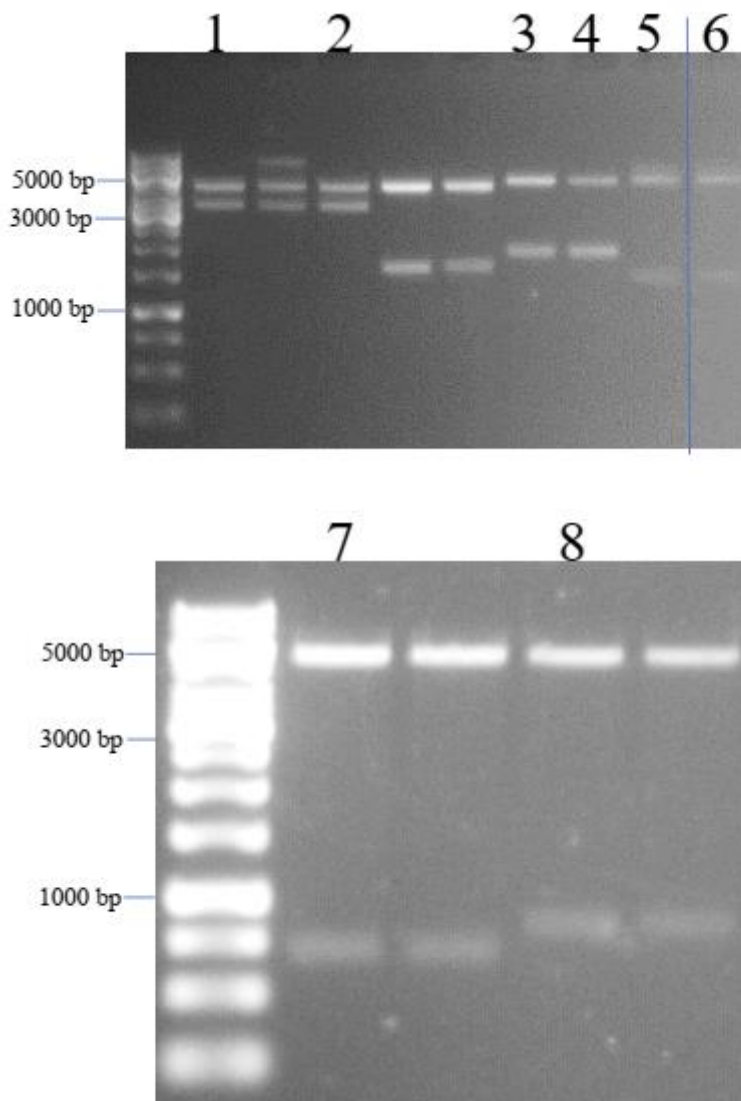
Figure 3.15 Agarose gel showing restriction digestion of ligation products.
Each cloned product was digested by two restriction enzymes to confirm the success of ligation. With 5 µl of samples, agarose gel-electrophoresis showed DNA fragments of *nanO* (lane 1), *TF0033* (lane 2), *BOF_0333* (lane 3), *TF1354* (lane 4), *BFO_0233* (lane 5), *TF1255* (lane 6), *BFO_0953* (lane 7), and *BFO_3116* (lane 8) and their pBAD18/ pAcTrc99a vectors.

3.2.2.4. Complementation Test in Minimal Media (M9)

To investigate the ability of *Tannerella's tonBs* to utilise different sialic acid as a carbon and energy source, this study used an established strain ([Δ*nanCnanR*(amber) Δ*ompR*::Tn10(tet) Δ*tonB*::FRT-Km-FRT] / or referred to as Δ*nan*Δ*tonB* MG1665). This system is devoid of its sialic acid outer membrane receptor (*nanC*) and inner membrane *tonB*, which cannot grow in the presence of sialic acid as the sole carbon and energy source. The growth of the mutant strain (Δ*nan*Δ*tonB* MG1665) was investigated in the presence of both glycerol and Neu5Ac. This experiment included a positive control of the wild type *E. coli* MG1665 whereas a negative control for this experiment was the *T. forsythia nanT*. As shown in the Figure 3.16, the Δ*nan*Δ*tonB* MG1665 strain in the presence of Neu5Ac (6 mM) was unable to utilise the sialic acid for growth, indicating similar phenotype to the negative control *T. forsythia nanT*. In the presence of glycerol, the Δ*nan*Δ*tonB* MG1665 deficient strain and the *nanT* were able to retain growth like the wild type *E. coli* MG1665.

Next, it was needed to check the growth of the mutant strain (Δ*nan*Δ*tonB* MG1665) with the chosen ligation vectors in the presence of both glycerol (1 %) and Neu5Ac (6 mM). As published previously, the sialic acid outer membrane *nanO* and *nanOU* were able to restore the sialic acid growth defect of ([Δ*nanCnanR*(*amber*) Δ*ompR*::*Tn10*(*tet*)] / or Δ*nan* MG1665 strain) in the presence of the *E. coli tonB*. Thus, the transformation of the empty chosen ligation vectors must not impact the growth of Δ*nan*Δ*tonB* MG1665 strain in the presence of Neu5Ac as both outer and inner membrane essential genes were removed. However, the complementation to Δ*nan*Δ*tonB* MG1665 strain with empty vectors containing either or both pBAD18 and pAcTrc99a was found to influence its growth in the presence of Neu5Ac. Likewise, in the presence of Neu5Ac, the transformation of empty pAcTrc99a to the Δ*nan*Δ*tonB* MG1665 strain with either sialic acid outer membrane porin from *Tannerella* ATCC 43037 (*nanO*) and for the first time from *Tannerella* 92A.2 strain (*TF0033* or *TF0033,34*) showed unexpected retain of growth to the wild type *E. coli* MG1665. In addition, no variations were observed in the growth yield of the complement Δ*nan*Δ*tonB* MG1665 with vectors compared to the wild type *E. coli* MG1665 in the presence of glycerol source (Figure 3.16).
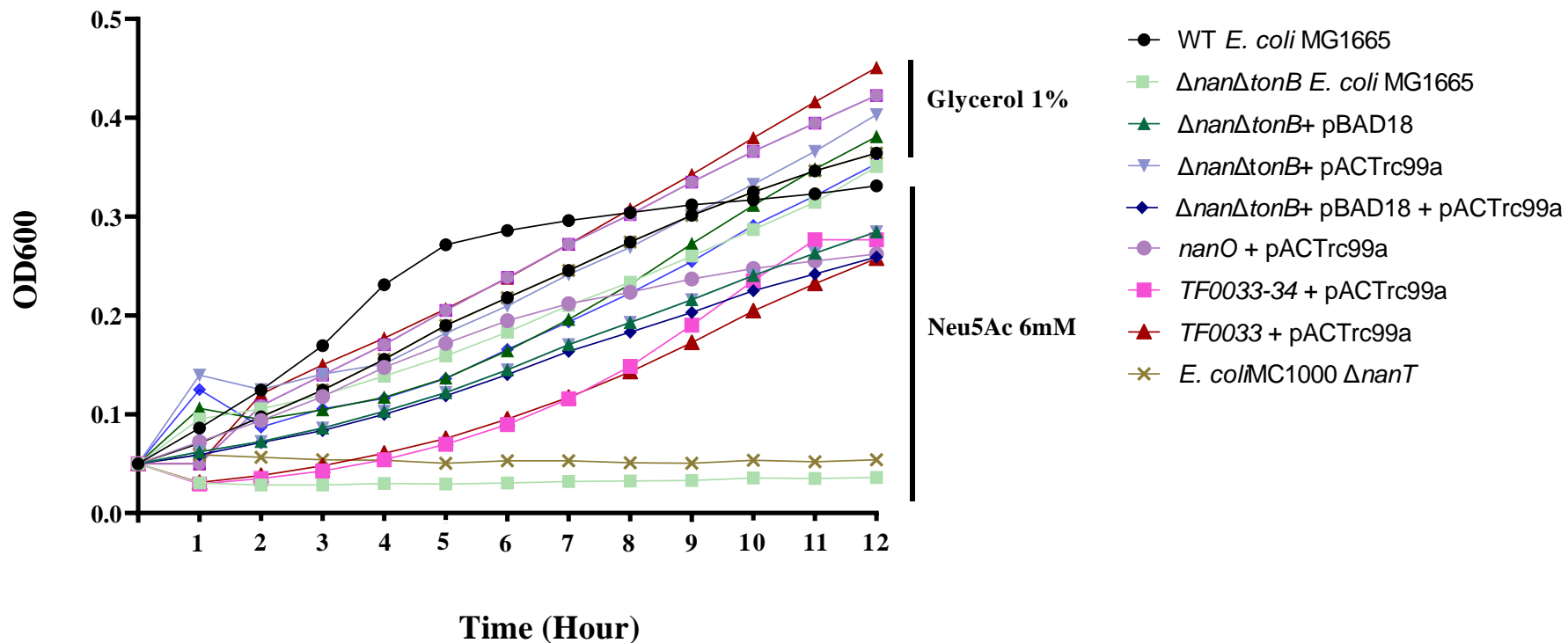
Figure 3. 16 Transformation of empty vectors to ΔnanΔtonB MG1665 strain.
Defined M9 medium with either 1% Glycerol (as a control) or 6 mM Neu5Ac (as a sole carbon and energy source) was used to test the phenotype of ΔnanΔtonB MG1665 strain before complementation. Both WT *E. coli* and Δ*nanT* were used as a positive and negative controls, respectively. Growth of ΔnanΔtonB strain was complemented with either empty pBAD18, empty pACTrc99a, or empty pBAD18 and pACTrc99a. Further growth complementation included either *nanO* + empty pACTrc99a, *TF0034,34* + empty pACTrc99a, or *TF0033* + empty pACTrc99a. These complementations were monitored at (A_{600}) and 37°C during culture for 12 hours.

Although this study established the genotype of the sialic acid transport and *tonB*-deletion mutant strain (Δ*nan*Δ*tonB* MG1665), we thought to continue the process of complementation, seeking additional clarification. The deleted *nanC* and *tonB* from the Δ*nan*Δ*tonB* MG1665 strain were complemented with each ATCC 43037 *Tannerella tonB* along with *Tannerella* sialic acid outer membrane porin (ATCC43037-*nanO*). These complementation strains were tested with the wild type *E. coli* MG1665 and the complement *tonB* of *E. coli* as positive controls. The *T. forsythia nanT* was used as a negative control in response to the available Neu5Ac (6 mM) in the established M9 medium. As expected, unfortunately, all growth of complementation strains with ATCC 43037 *Tannerella tonBs* and with the established sialic acid outer membrane porin (*nanO*) showed no difference compared with those positive controls in the M9 medium in the presence of sialic acid as a sole carbon and energy source (Figure 3.17). In the presence of glycerol (1 %), no variations were observed in the growth yield of the complement Δ*nan*Δ*tonB* MG1665 with *Tannerella tonB* genes compared to the wild type *E. coli* MG1665 and *T. forsythia nanT*.
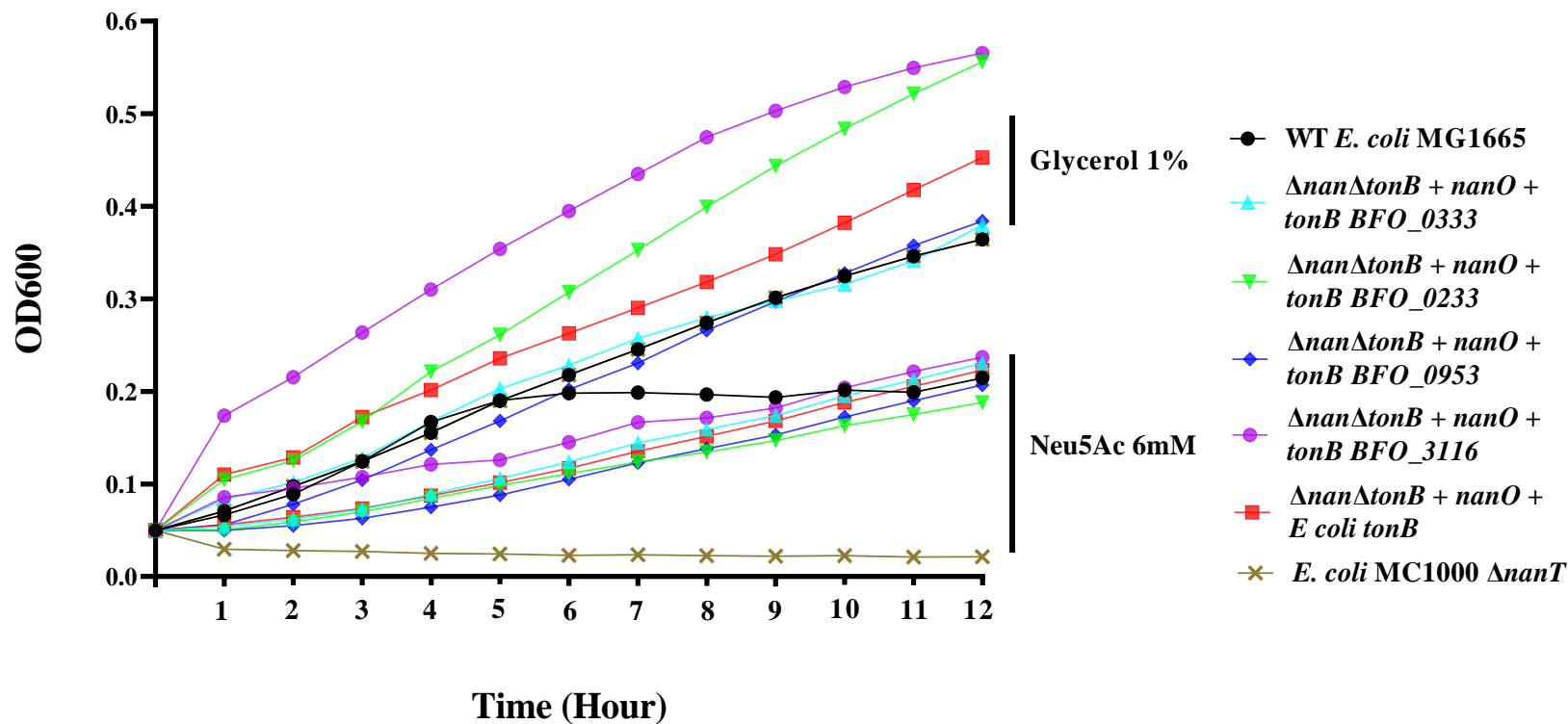
Figure 3. 17 Complementation of Δ*nan*Δ*tonB* MG1665 with ATCC 43037 *nanO* and *tonBs*.
Defined M9 medium with either 1% Glycerol (as a control) or 6 mM Neu5Ac (as a sole carbon and energy source) was used to test the phenotype of Δ*nan*Δ*tonB* MG1665 strain complemented with ATCC 43037 sialic acid outer membrane (*nanO*) along with one *tonB* of the inner membrane (*BFO_0333*, *BFO_0233*, *BFO_0953*, or *BFO_3116*). The WT *E. coli* and *E. coli tonB* were used as positive controls, and the Δ*nanT* was used as a negative control. These complementations were monitored at (A_{600}) and 37°C during culture for 12 hours.

Likewise, and for the first time of ligating sialic acid outer membrane (*TF0033*, or *TF0033,34*) from the 92A.2 strain, the Δ*nan*Δ*tonB* MG1665 strain was complemented with each 92A.2 *Tannerella tonB* along with the *Tannerella* sialic acid outer membrane porin (*TF0033*, or *TF0033-34*). As a result, there was no difference in growth between four *tonBs* of this strain with the established positive controls (Figures 3.18 and 3.19). Also, we examined these complementation strains growth levels in response to available glycerol (1 %) in the medium. No variations were observed in the growth yield of the complement Δ*nan*Δ*tonB* MG1665 with *Tannerella tonB* genes compared to the wild type *E. coli* MG1665 and *T. forsythia nanT*. In conclusion, we are not sure what exact reason affected these complementations, as well as time and laboratory restrictions hindering us from further investigations.
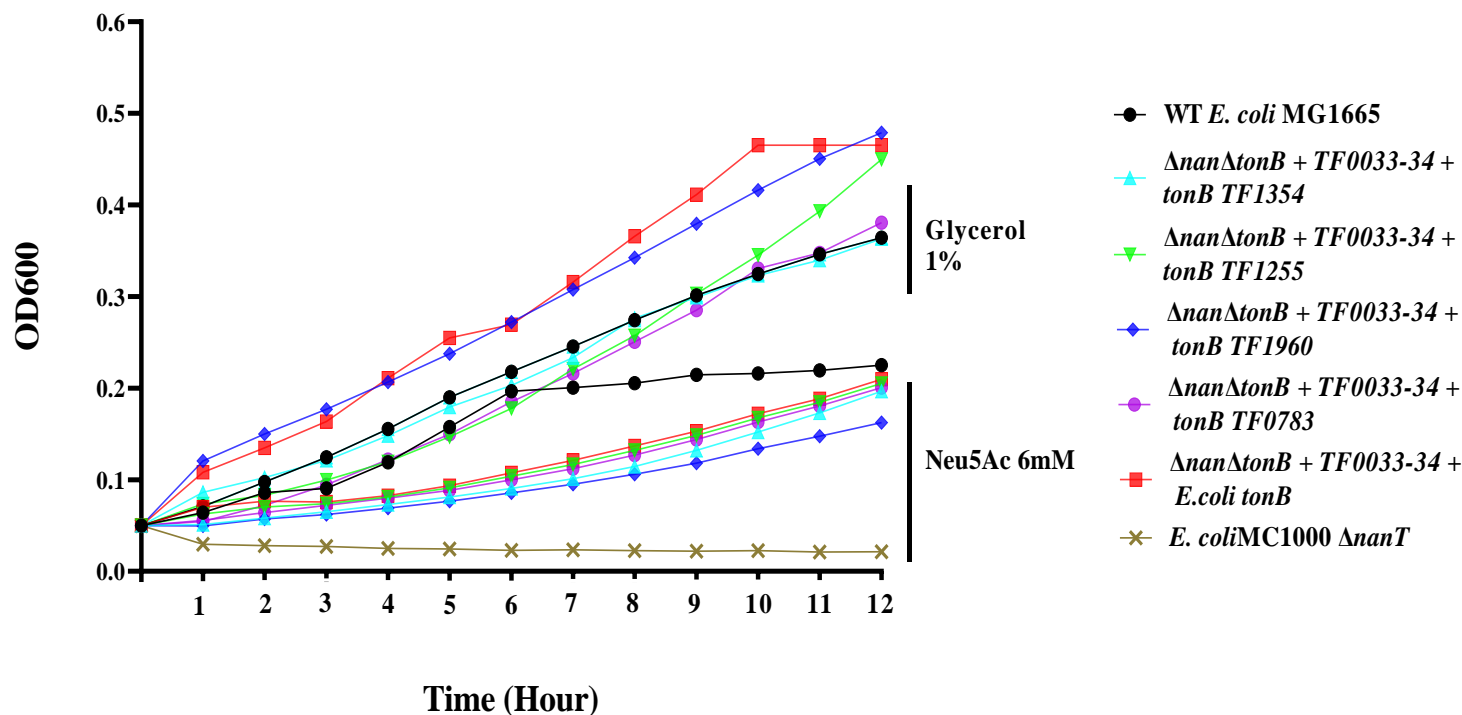
Figure 3.18 Complementation of ΔnanΔtonB MG1665 with 92A.2 *TF0033,34* and *tonBs*.
Defined M9 medium with either 1% Glycerol (as a control) or 6 mM Neu5Ac (as a sole carbon and energy source) was used to test the phenotype of ΔnanΔtonB MG1665 strain complemented with 92A.2 sialic acid outer membrane (TF0033,34) along with one *tonB* of the inner membrane (*TF1354*, *TF1255*, *TF1960*, or *TF0783*). The WT *E. coli* and *E. coli tonB* were used as positive controls, and the Δ*nanT* was used as a negative control. These complementations were monitored at (A$_{600}$) and 37°C during culture for 12 hours.
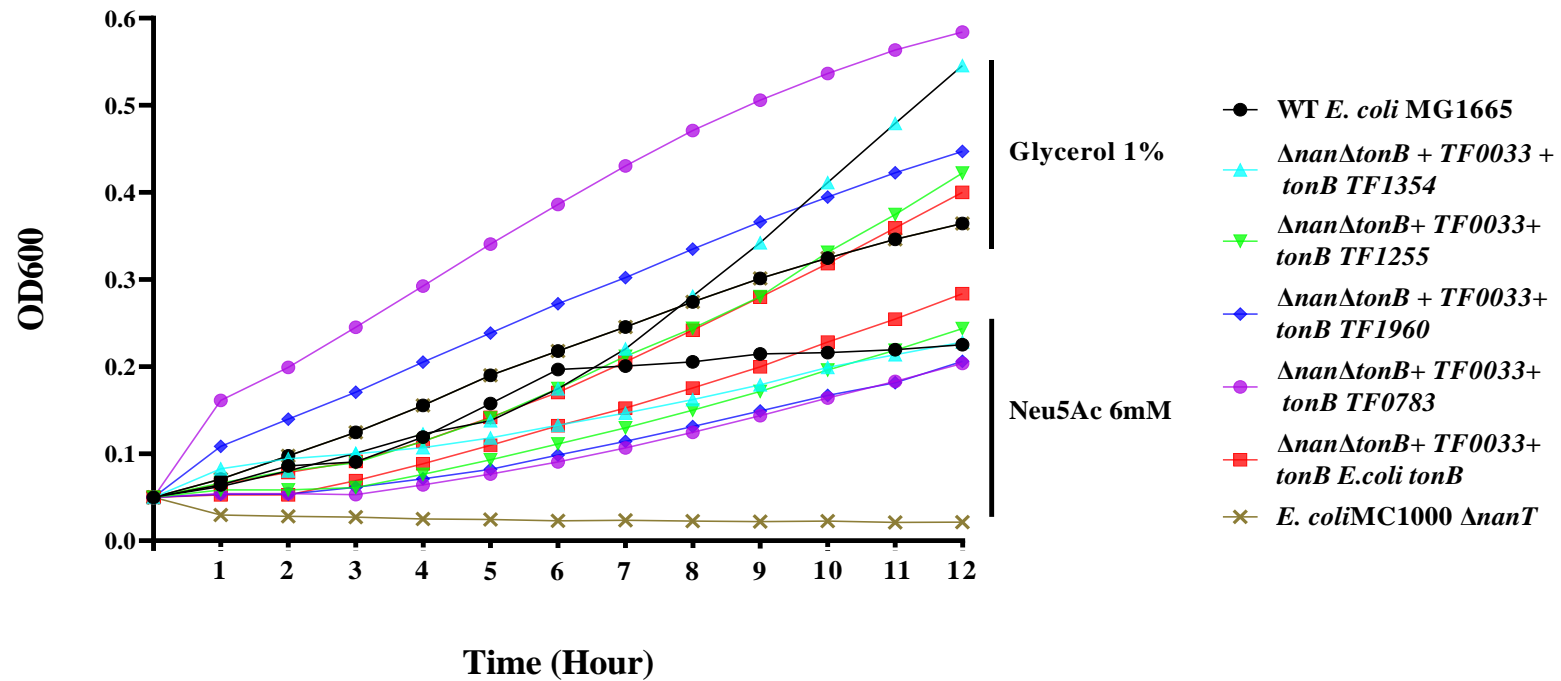
Figure 3.19 Complementation of Δ*nan*Δ*tonB* MG1665 with 92A.2 *TF0033* and *tonBs*.
Defined M9 medium with either 1% Glycerol (as a control) or 6 mM Neu5Ac (as a sole carbon and energy source) was used to test the phenotype of Δ*nan*Δ*tonB* MG1665 strain complemented with 92A.2 sialic acid outer membrane (*TF0033*) along with one *tonB* of the inner membrane (*TF1354*, *TF1255*, *TF1960*, or *TF0783*). The WT *E. coli* and *E. coli tonB* were used as positive controls, and the Δ*nanT* was used as a negative control. These complementations were monitored at (A_{600}) and 37°C during culture for 12 hours.

3.2.3. Generation of *T. forsythia* Mutants

3.2.3.1. Construction and Transformation of *tonB* Mutagenesis in *T. forsythia*

The second method used in this study to examine and confirm the function of individual *tonB* genes was a deletion of genes in *T. forsythia* directly (section 2.3.12.). Here we focussed on the ATCC 43037 strain to construct deletions via replacement of genes of interest with the Erythromycin resistance gene (*ermF*) for the following *tonB* genes: *BFO_0333*, *BFO_0233*, *BFO_0953*, and *BFO_3116* (Table 3.1) (Appendix VIII: 8.11). In addition, this study sought to fully confirm the role of *nanO*, but not *nanU* as our previous work showed it not to be essential.[128]

Previous other laboratory groups making genetic deletions in *T. forsythia* have focussed chiefly on electroporation to introduce knockout constructs into the organism, contained within suicide plasmids, or as naked PCR products. However, Tribble et al., as well as Naylor et al., used natural competence as a means to introduce DNA in the closely related organism *P. gingivalis*,[22,233] while Nishikawa et al., reported its use in *T. forsythia* for mutagenesis.[234] This work attempted both methods as part of mutagenesis strategy, these are outlined below (Figure 3.20). This study worked to improve efficiency of these methods based on several personal modifications. Briefly, in the case of electroporation, wild-type cells were grown in TSB broth before electroporation, incubated in TSB liquid broth for 24h after electroporation, and cells were then plated onto Erythromycin-containing blood agar plates (section 2.4.5). For natural competency, wild-type cells were grown in supplemented TSB medium before the DNA was added and transferred to Erythromycin-containing plates (section 2.4.4).

Electroporation

(Starter culture on BA-plate)

↓

TSB Broth (*T. forsythia* + NAM + haemin + menadione) for 2 days

↓

Add DNA plasmid (50 ng/µl)

↓

Electroporation

↓

Cells incubation in anaerobic incubator (24h)

↓

Ery- BA plate

Natural Competency

(Starter culture on BA-plate)

↓

TSB Broth (*T. forsythia* + NAM + haemin + menadione) for 2 days

↓

Add DNA plasmid (200 ng/µl or more)

↓

Cells incubation in anaerobic incubator (18h)

↓

Ery- BA plate

Figure 3.20 Schematic representation transformation of mutagenesis in *T. forsythia*.
The transformation of DNA plasmid cloning mutant genes into wild-type *T. forsythia* is
performed using two methods: electroporation and natural competency.

Nevertheless, before the method of transformation of the cells, the basic methodology for design of knockout constructs is to produce the gene of interest with two flanking regions of 1000 bp upstream and 1000 bp downstream of the gene of interest to eliminate the risk of polar effects in the mutant, such that the coding gene sequence is replaced by the ermF gene, between the natural ATG start codon and TAA (stop) codon (Figure 3.21). These constructs were synthesised without the codon optimisation via GeneArt™ and provided as cloned fragment into the GeneArt™ proprietary plasmid, which were then stored via transformation into *E. coli* DH5α before DNA plasmid was prepared for transformation into *T. forsythia*.

Step1: Identify the gene of interest manually from NCBI database

1000bp Flanking gene	*tonB* or *nanO* gene	1000bp Flanking gene
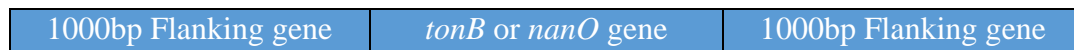
Step 2: Replace the *tonB* gene or the gene of interest with Erythromycin cassette (EM) *in silico*

1000bp Flanking gene	EM cassette (801bp)	1000bp Flanking gene

Step3. Send the edited gene into the company for mutant gene construction

Step4. Clone the mutant gene into a vector of interest

Step4. Transform the cloning vector into competent *E. coli* DH5α cells and placed on an antibiotic plate for selection
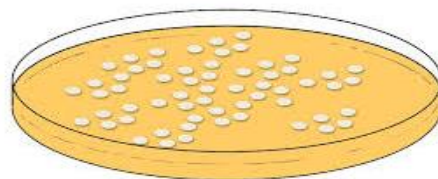
Step5. Extract the plasmid and transform it into wild-type *T. forsythia*

Figure 3.21 Generation of *T. forsythia* Mutants.
This scheme illustrates steps of mutagenesis in *T. forsythia*. Once this study replaces the gene of interest with EM *in silico* (Step 2), the gene was sent to the 'GeneArt™, Thermo Fisher Scientific' to construct this gene. This study then resumed the process of mutagenesis from step 4 to step 6.

At this point in the process, a set of primers was also designed to enable verification of gene insertion in the Chromosome of *T. forsythia* if Erythromycin resistant colonies were obtained, these are shown in figure 3.22.
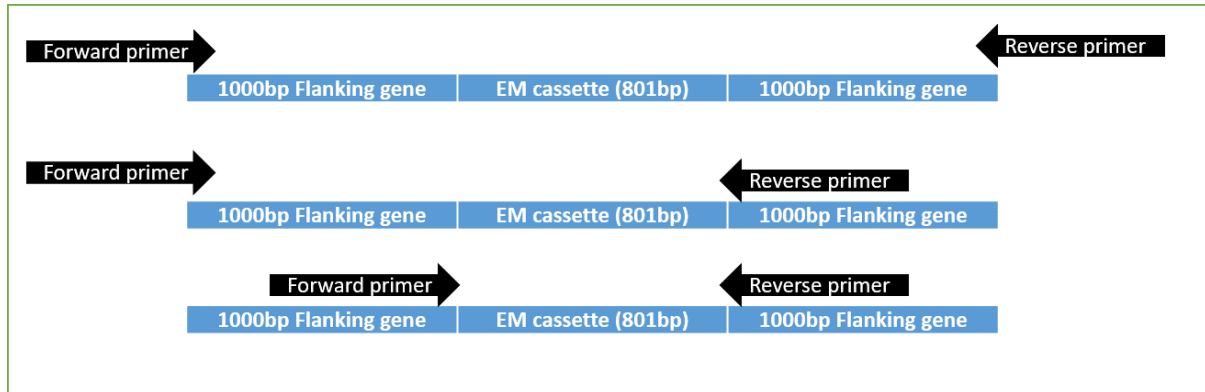
Figure 3.22 Primers method confirmation.
This figure explains three methods of using primers to identify and confirm the insert. The first method, to confirm the presence of EM insertion with 1000 bp on both sides, forward and reverse primers of the flanking genes were used to obtain a product of 2800 bp. The second method, a forward primer of the flanking gene and a reverse primer of the Erythromycin cassette were used to confirm the end of EM cassette (1801 bp). In contrast to the second method, this study obtained the same product of 1801 bp using a forward primer of the Erythromycin cassette and a reverse primer of the flanking gene. The third method, this study confirms the presence of Erythromycin cassette using forward and reverse primers of the Erythromycin cassette (801bp).

Once we received the order of mutagenesis as blunt-end PCR product, we cloned these products into cloning vectors. Transformation of cloning vectors containing the replacement of the gene of interest was verified with colony PCR. We used two sets of previous primers (Figure 3.22) to verify that the synthetic genes contained the right gene constructs (Figures 3.23). In addition, the five constructed genes were all sequenced for confirmation.
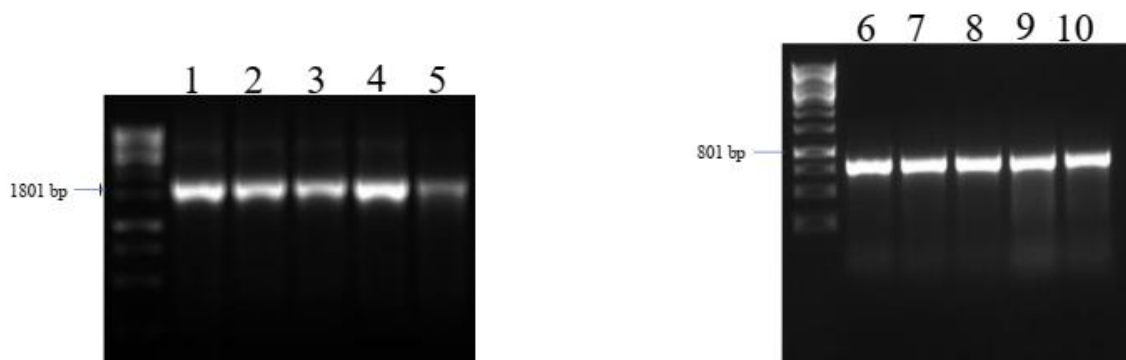
Figure 3.23 PCR analysis for mutant constructs.
Colony PCR was performed to confirm a product of 1801 bp using a forward primer of flanking gene and a reverse primer of Erythromycin cassette for *nanO* (lane 1), *BFO_0333* (lane 2), *BFO_0233* (lane 3), *BFO_0953* (lane 4), and *BFO_3116* (lane 5). Next, colony PCR was performed using Erythromycin specific primers to obtain a product of 801 bp for *nanO* (lane 6), *BFO_0333* (lane 7), *BFO_0233* (lane 8), *BFO_0953* (lane 9), and *BFO_3116* (lane 10).

3.2.3.2. Confirmation of *tonB* Mutagenesis in *T. forsythia*

DNA plasmids carrying replacement of gene of interest were then introduced into *T. forsythia* via the electroporation and natural competence outlined above. Erythromycin colonies were sub-cultured and screened for the presence of the correct deletion via colony PCR method using primers from table 2.8 (section 2.3.11.1.). Multiple combinations of primers were used to confirm the deletion of each gene (Figure 3.22). This study was able to generate five mutants in *T. forsythia*, four of which were *tonB* genes and one is the outer membrane TBDT (*nanO*) (Figure 3.24). Briefly, two primers were used to generate the upstream and downstream of both flanking regions including the size gene of each Δ*nanO* (5384 bp), Δ*BFO_0333* (3867 bp), Δ*BFO_0233* (3440 bp), Δ*BFO_0953* (2699 bp), and Δ*BFO_3116* (2837 bp) (lane A). Also, two primers were used to generate both flanking regions upstream and downstream, including the Erythromycin cassette (2801 bp – 'lane B'). Two primers were used to generate the flanking region upstream and the downstream of Erythromycin cassette (1801bp – 'lane C'). In contrast, two primers were used to generate the upstream of Erythromycin cassette and the downstream of flanking region (1801bp – 'lane D'). Finally, two primers were used to confirm upstream and downstream of the Erythromycin cassette (801bp – 'lane E'). This method assisted in the determination of the successful homologous recombination of the Erythromycin resistance cassette into the *T. forsythia* genome, thereby knocking out genes of interest in this study. Furthermore, all PCR products of these mutagenesis genes were confirmed via DNA sequencing (section 2.3.11.2). Generation of *T. forsythia* mutants sets the basis for analysis of the phenotypes and characterisation of the role of *tonBs* in the future.
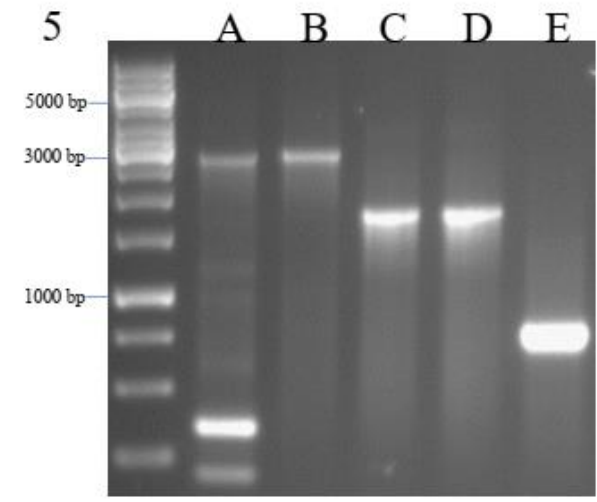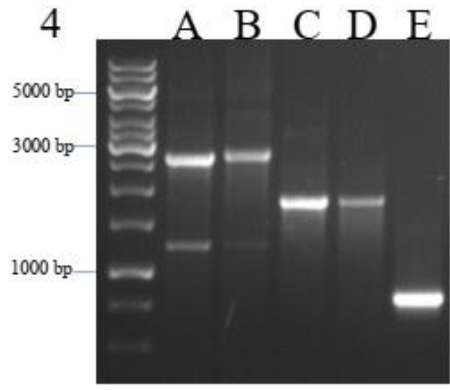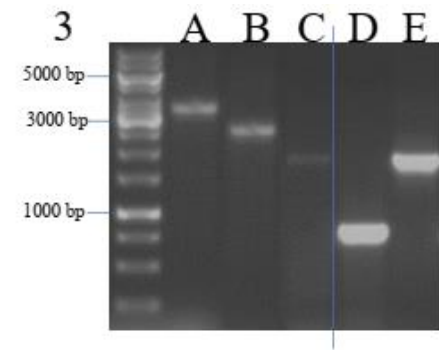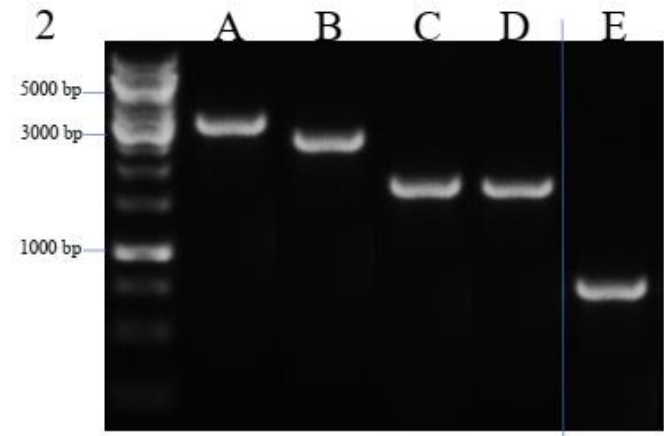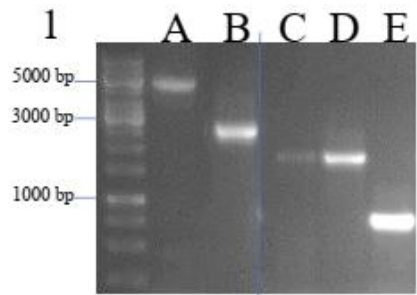
Figure 3.24 PCR analysis of the generation of *nanO* and *tonB* mutants.
This figure shows all combinations of primers to confirm the mutagenesis process in the wild type. Figure 1 shows mutant *nanO* where lane A presents the size of both flanking genes, including the size of WT-*nanO* (5384 bp). Figure 2 shows mutant *BFO_0333* where lane A shows the size of both flanking genes including the size of WT-*BFO_0333* (3867 bp). Figure 3 shows mutant *BFO_0233* where lane A shows the size of both flanking genes including the size of WT-*BFO_0233* (3440 bp). Figure 4 shows mutant *BFO_0953* where lane A shows the size of both flanking genes including the size of WT-*BFO_0953* (2699 bp). Figure 5 shows mutant *BFO_3116* where lane A shows the size of both flanking genes including the size of WT-*BFO_3116* (2837 bp). To confirm the deletion process, multiple types of primers combinations were used: forward and reverse primer of flanking gene including the Erythromycin cassette (2801 bp – 'lane B'), primers of the forward flanking gene and the reverse of Erythromycin cassette (1801bp – 'lane C'), two primers of forward Erythromycin cassette and the reverse of flanking gene, except figure 3 (1801bp – 'lane D'), and two primers of upstream and downstream of the Erythromycin cassette, except figure 3 (801bp – 'lane E').

3.2.3.3. Protein Profiling of Mutagenesis

Aside from PCR confirmation, it was essential to check whether these mutants altered the whole protein complement of wild-type *T. forsythia*. This would also serve to establish their identity as *T. forsythia* given that cultures are prone to contamination. To achieve this, a sample of plate-grown cells for WT- *T. forsythia* (lane 1), Δ*nanO* (lane 2), Δ*BFO_0333* (lane 3), Δ*BFO_0233* (lane 4), Δ*BFO_0953* (lane 5), and Δ*BFO_3116* was lysed using SDS-PAGE lysis buffer before analysis by SDS-PAGE gel. Figure 3.25 shows that a general observation of the whole cell protein profile between the wild-type *T. forsythia* and other mutants is an identical protein profile, with the prominent *T. forsythia* **S**-layer proteins (labelled with *) present.
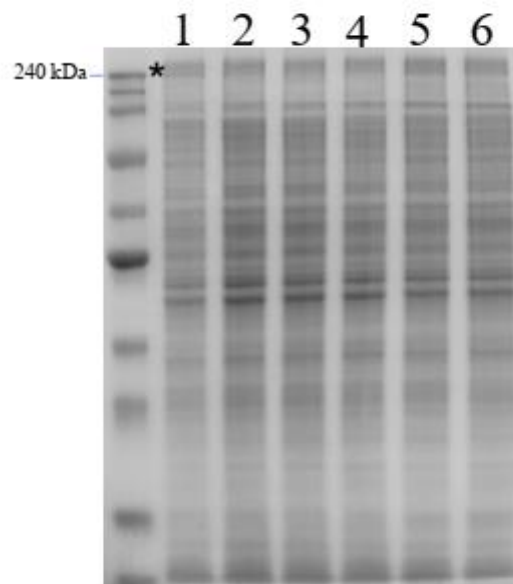
Figure 3.25 SDS-PAGE analysis of purified whole protein of WT-*T. forsythia* and all generated mutants.
An inoculum loop of each strain resuspended in 2 ml of TSB broth before taking the OD$_{600}$. The pellet from each suspended strain with TSB was mixed with SDS liquid before boiling and then applied to a 10 % SDS-PAGE gel. This figure presents WT- *T. forsythia* (lane 1), *nanO* (lane 2), *BFO_0333* (lane 3), *BFO_0233* (lane 4), *BFO_0953* (lane 5), and *BFO_3116* (lane 6).

3.2.4. Role of *tonB* Mutagenesis in *T. forsythia*

3.2.4.1. Optimisation of *Tannerella* Biofilm Growth

As a prelude to testing the effects of mutations in the *tonB* and *nanO* genes, a range of biofilm growth periods was tested to see which was most reliable. Here, seven days gave the most effective and reliable growth, on Neu5Ac, NAM and mucin substrates (Figure 3.26).
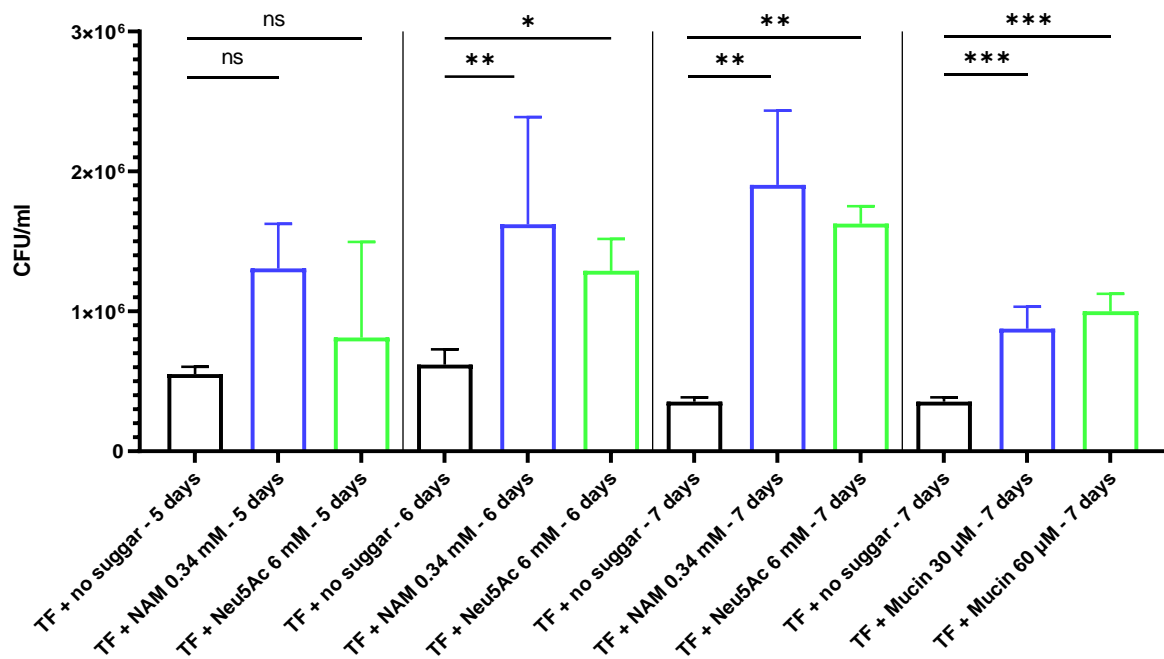
Figure 3.26 Time period of biofilm formation by *T. forsythia*.
This figure shows time period of biofilm formation on N-acetyl muramic acid (NAM) or sialic acid (Neu5Ac). *T. forsythia* was inoculated in a defined TSB medium with either NAM or Neu5Ac. Mucin was coated on the 96-well plate overnight at 4 °C before these wells were washed off with PBS two times on the following day. At the time point, the 96-well plate was washed two times with PBS and the biofilm was harvested and assessed by cell counting. To eliminate bias, this study counted all 16 squares of Helber counting chamber (Hawksley).

3.2.4.2. Planktonic Growth of *tonB* Mutagenesis

In TSB medium, the growth phenotype of ATCC 43037 isogenic *nanO*, *BFO_0333*, *BFO_0233*, *BFO_3116*, and *BFO_0953* mutants were compared to the parental strain and *nanT* for seven days.[91] The growth rate was monitored every day by taking out the plate from the anaerobic cabinet at specific time, covering it with 'SealPlate', measuring growth at OD of 575 nm, and returning it as quick as possible to the anaerobic cabinet. When the bacterial strains were grown in TSB medium containing NAM, the isogenic *nanO* and *tonB* genes mutants showed growth comparable to that of the parental strain and Δ*nanT*. However, in a TSB medium containing 6mM of Neu5Ac, the growth of Δ*nanO* and Δ*BFO_0233 tonB* showed a similar growth phenotype to Δ*nanT*, indicating they are required for *T. forsythia* to grow on Neu5Ac. The other *tonB* genes (Δ*BFO_0333*, Δ*BFO_3116*, and Δ*BFO_0953*) showed no growth defect compared to parent strain. The planktonic growth of all strains in TSB medium containing no carbon source was used as control (Figure 3.27).
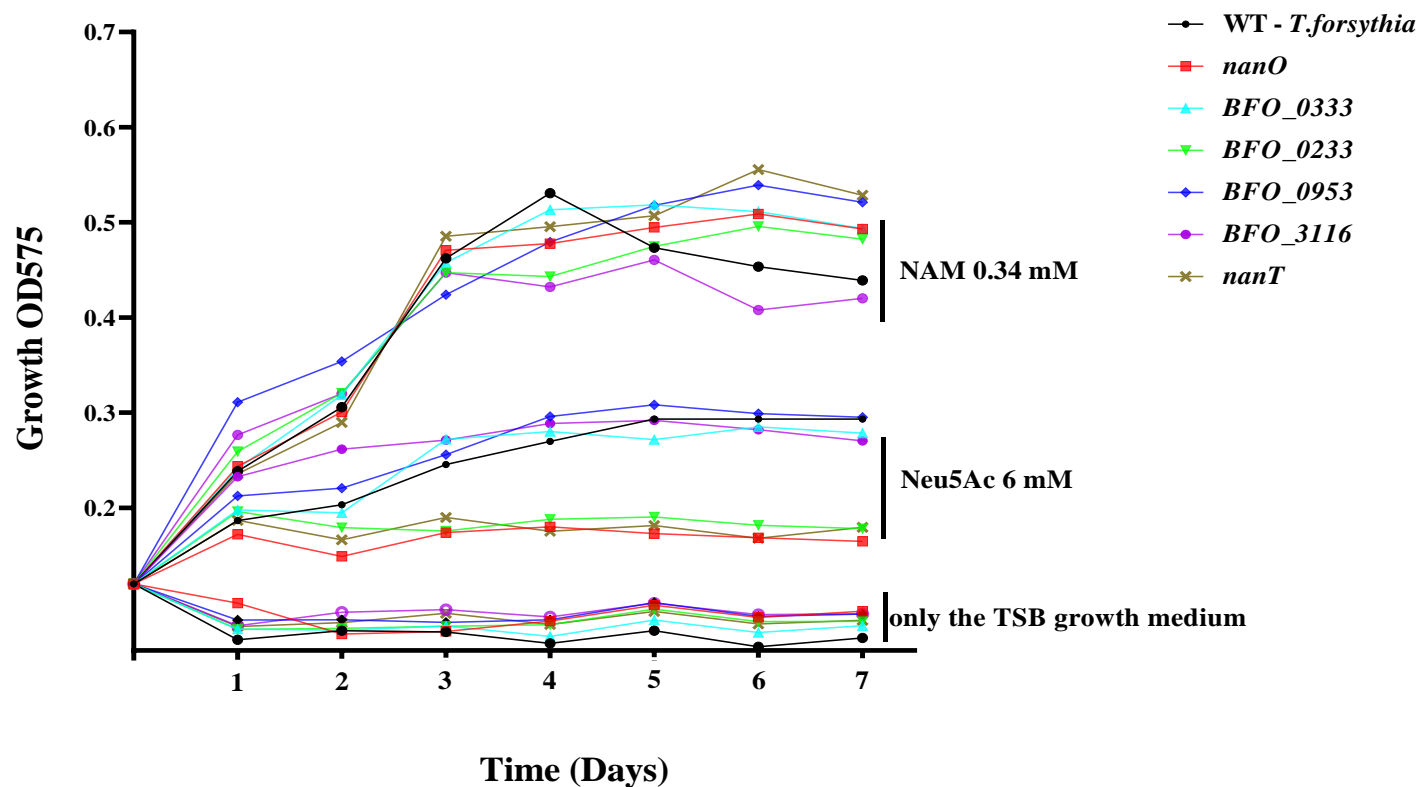
Figure 3.27 Growth of *T. forsythia* and deletion mutant strains on different substrates.
Growth rates of *T. forsythia* and the constructed mutants (*nanO*, *BFO_0333*, *BFO_0233*, *BFO_3116*, and *BFO_0953*) in TSB medium with N-acetyl muramic acid (NAM) or sialic acid (Neu5Ac). Four days old plate of each strain was harvested, washed three times in PBS, and adjusted to a cell density at OD_{600} of 0.1 before inoculating wells of a 96-well plate. Growth was monitored every day by measuring absorbance at 575 nm for seven days. The data points are the average of three independent experiments in triplicate wells.

3.2.4.3. Biofilm Formation of *tonB* Mutagenesis

The role of each *tonB* gene in the sialic acid uptake was examined by biofilm cell counting. On polystyrene wells and in the presence of sialic acid (6 mM), the wild type of strain ATCC 43037 and the negative control Δ*nanT* were compared to the isogenic *nanO* and four *tonBs* insertion mutants. Biofilm of these bacterial strains for 7 days resulted in biofilm formation of the wild type and *tonB* Δ*BFO_0333*, Δ*BFO_3116*, and Δ*BFO_0953*. The biofilm of the three *tonB* genes was reduced on polystyrene wells, but they were not to the level of significance to the wild type. However, the biofilm formation by the *tonB* Δ*BFO_0233* was significantly reduced in the presence of Neu5Ac (6 mM) compared to the level of the parental strain (P =0.0016) (Figure 3.28). Likewise, the ATCC 43037 Δ*nanO* was significantly reduced (P =0.0005) at levels comparable to the previously observed Δ*nanT*, which served as a negative control.

To demonstrate that *tonB* Δ*BFO_0233* is essential for sialic acid uptake, we checked the effect of additional sialic acid reported concentrations, 1.5 mM, and 3 mM. The results of the proportions of biofilm formation were reduced significantly on both concentrations. In addition, the role of Δ*nanO* for sialic acid transit significantly decreased the biofilm formation on 1.5 mM and 3 mM concentrations.

To identify whether TonB proteins can form biofilm in the presence of NAM. Our results showed that all mutants and wild-type strains could form biofilm on polystyrene wells, indicating these strains were not involved in the NAM uptake and that their phenotypes in the presence of sialic acid were not affected. Furthermore, data analysis showed no difference between mutants to the parental strain in biofilm formation in the absence of both NAM and Neu5Ac, indicating the role of sialic acid as an essential factor in the oral biofilm and that TonB BFO_0233 may be responsible for interaction with NanO rather than all TonB being interchangeable (Figure 3.28).
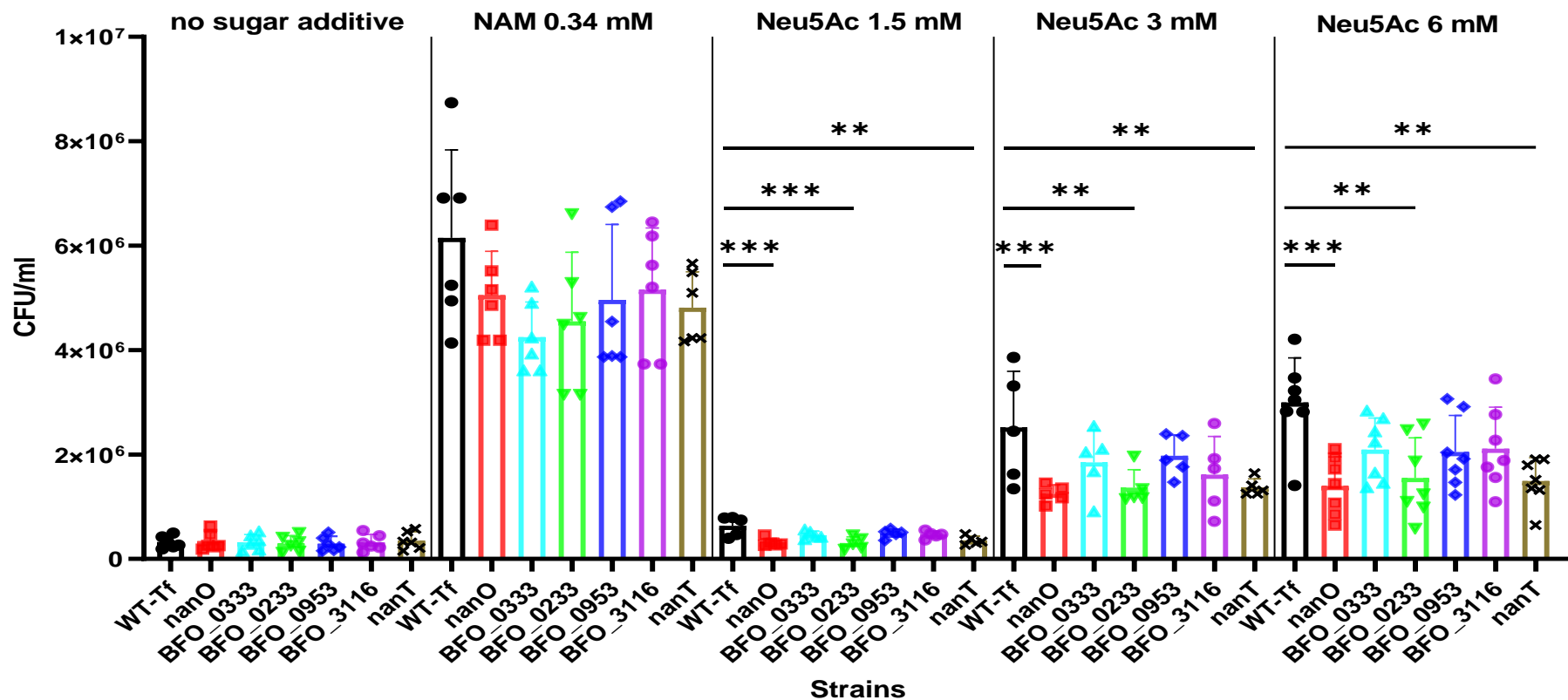
Figure 3.28 Growth biofilm of *T. forsythia* and isogenic mutant strains on Neu5Ac.
    Four-day-old plate of *T. forsythia* and other mutant strains was harvested, washed three times with PBS, and adjusted to a cell density at OD_{600} of 0.5. The growth of each strain was in TSB medium supplemented with either N-acetyl muramic acid (NAM) or sialic acid (Neu5Ac). No sugar additive was added to the control wells including bacteria in TSB. After seven days of anaerobic incubation, the 96-well plate was washed two times with PBS and the biofilm was harvested and assessed by cell counting. To eliminate bias, this study counted all 16 squares of Helber counting chamber (Hawksley). Means of three wells were calculated before applying the one-way ANOVA test, with $P < 0.05$ being taken as the level of significance (*); $P < 0.01$ being taken as the level of significance (**); $P < 0.001$ being taken as the level of significance (***); $P < 0.0001$ being taken as the level of significance (****). Enclosed shapes are data for every individual experiment.

3.2.4.4. Staining of Biofilm Growth of *tonB* Mutagenesis

Lastly, biofilm staining assay was performed to determine whether the Δ*nanO* and Δ*tonB BFO_0233* influence the Neu5Ac uptake ability of *T. forsythia*. The wild-type and isogenic mutant strains were tested in biofilm formation. At various concentration of Neu5Ac, the results showed that the biofilm formation of Δ*nanO* and Δ*BFO_0233* were significantly weaker than that of the wild-type strain. The biofilm formation of Δ*nanO* and Δ*BFO_0233* on 10 mM was reduced to 2.5-fold and 2.7-fold compared to the parent strain (P=< 0.001) (Figure 3.29). In support of this observation, biofilm formations on Neu5Ac (3 mM and 6 mM) were reduced to 2-fold and 2-fold and 1.9-fold and 2-fold due to Δ*nanO* and Δ*tonB BFO_0233*, respectively (P=< 0.001). This finding indicates that Δ*BFO_0233* plays a more dominant role than other *tonB* genes in the Neu5Ac uptake of *T. forsythia*, as the case for Δ*nanO*. All tested isogenic mutants were able to form biofilm in the presence of NAM to levels comparable to the parent strain (Figure 3.29).
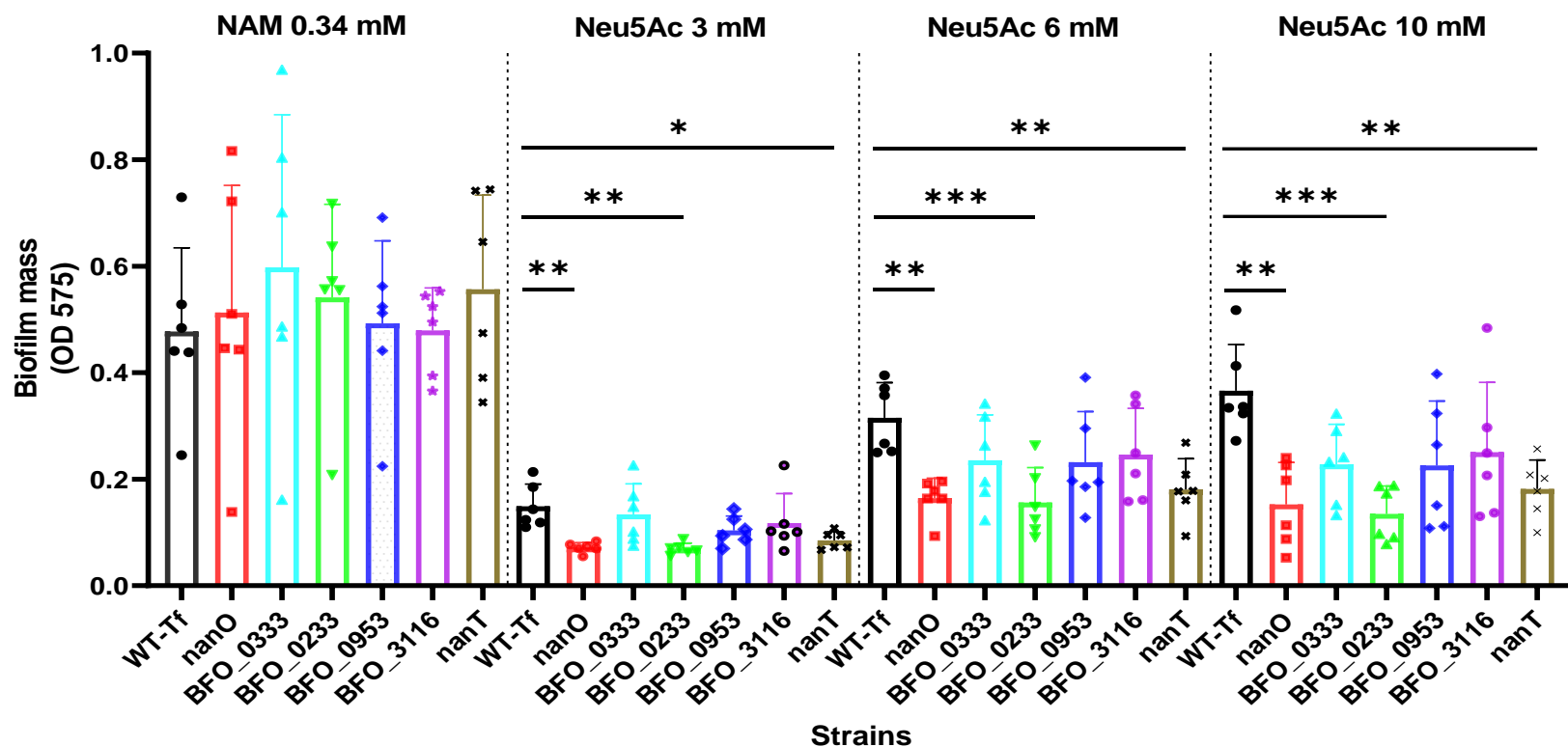
Figure 3.29 Biofilm formation of *T. forsythia* and isogenic mutant strains on Neu5Ac.
Growth in TSB containing three free Neu5Ac concentrations was prepared for WT *Tannerella*, Δ*nanO*, Δ*BFO_0333*, Δ*BFO_0233*, Δ*BFO_3116*, Δ*BFO_0953*, and Δ*nanT*. These strains were harvested, washed three times with PBS, and adjusted to a cell density at OD_{600} of 0.1. These strains were set up in triplicate wells supplemented with either 0.34 mM of NAM and 3 mM, 6 mM, and 10 mM of Neu5Ac (Total 200 µl). After seven days of growth, the 96-well plate was washed with PBS and stained with a crystal violet before being washed and measured at OD 575 nm. Means of three wells were calculated before applying the one-way ANOVA test, with $P < 0.05$ being taken as the level of significance (*); $P < 0.01$ being taken as the level of significance (**); $P < 0.001$ being taken as the level of significance (***); $P < 0.0001$ being taken as the level of significance (****). Enclosed shapes are data for every individual experiment.

3.3.4.5. Determination of Sialic acid Uptake by the Thiobarbituric Acid

To compare Neu5Ac uptake between the wild-type and the constructed mutants of this study, an old classic method for Neu5Ac uptake was performed.[209] Cells removed from plates of *T. forsythia* wild-type and the mutant's strains were washed three times in PBS and adjusted to OD_{600} of 2. Each strain was placed in a small Eppendorf containing either 500 or 1000 µM of Neu5Ac, a total 100 µl, before incubation in an anaerobic cabinet at 37 for three hours. Of the reaction mix, 50 µl of cleared media spent solution was oxidised by Sodium Periodate in 60 mM H_2SO_4 (25 µl) and placed in PCR tube before incubation for 30 minutes at 37 °C. The addition of 20 µl of Sodium Arsenite was used to halt the oxidation before adding 200 µl of Thiobarbituric Acid to label oxidised Neu5Ac with a thiol group (100 mM / pH 9.0). The reaction was then heated at 95 °C for 7.5 minutes. The whole reaction was added to a clear 96 well plate and the absorbance was measured at OD 549 nm. As a result, the deficient phenotype of Δ*nanO* and *tonB* Δ*BFO_0233* showed depletion of Neu5Ac uptake compared to parental and other mutants' strains. In the presence of 500 µM and 1000 µM of Neu5Ac, the uptake of *tonB* Δ*BFO_0233* was reduced to 48 µM and 77 µM compared to 105 µM and 184 µM of the wild-type strain. Furthermore, the role of sialic acid outer membrane Δ*nanO* was confirmed in the Neu5Ac uptake similar to the established sialic acid inner membrane gate (Δ*nanT*). In the presence of 500 µM and 1000 µM of Neu5Ac, the isogenic mutant Δ*nanO* uptake was reduced to 18 µM and 80 µM compared to 105 µM and 184 µM of the wild-type strain, respectively. We have noticed a strong depletion for Δ*nanO* and Δ*nanT* during Neu5Ac uptake of 500 compared to 1000 µM, suggesting the importance of this porin for Neu5Ac uptake in a sequestered environment (Figure 3.30).
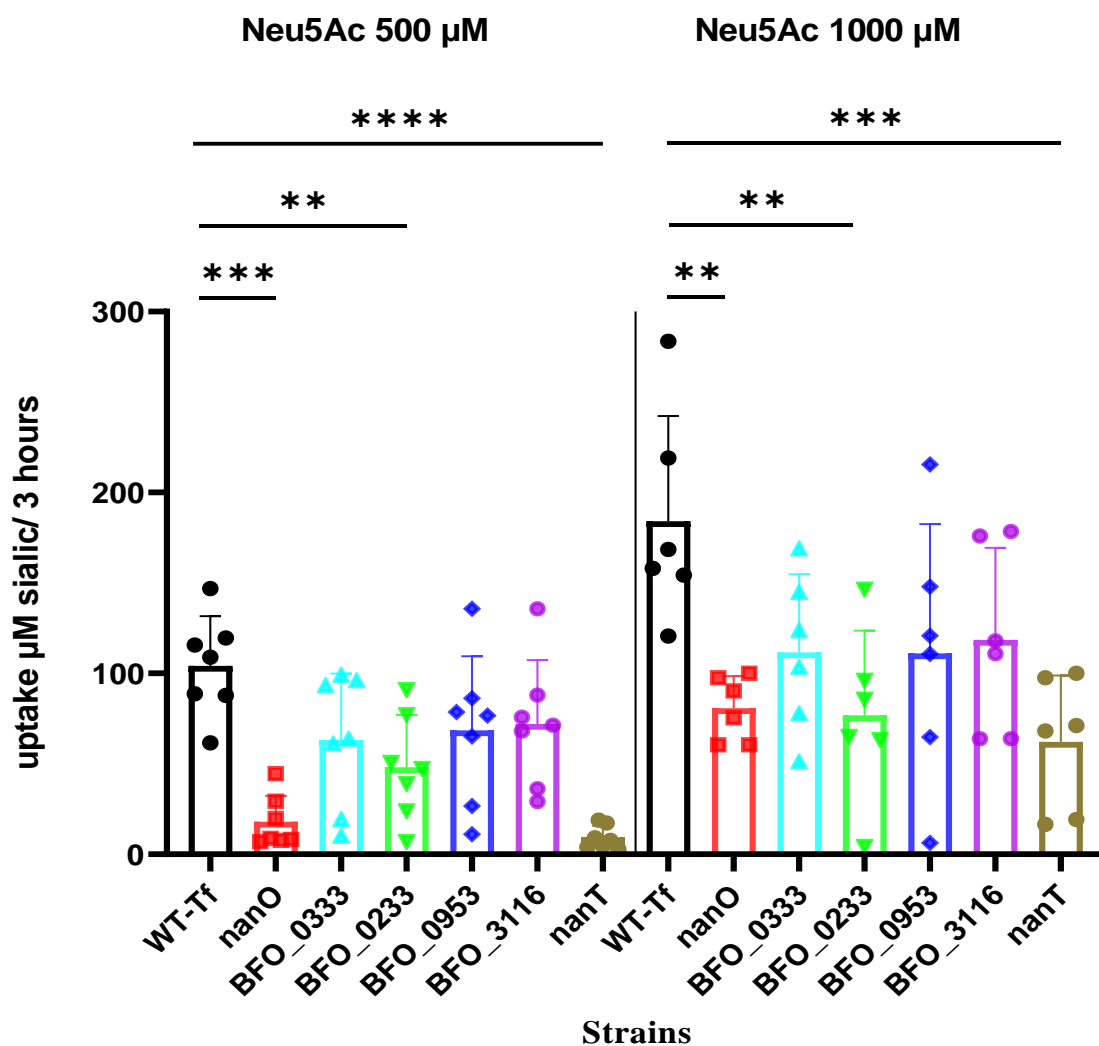
Figure 3.30 Neu5Ac uptake by *T. forsythia* strains

Four-day-old plate of *T. forsythia* and other mutant strains was harvested, washed three times with PBS, and adjusted to a cell density at OD_{600} of 2. The Neu5Ac uptake was set up in triplicate supplemented with either 500 µM or 1000 µM concentration (Total 100 µl). After three hours of incubation, the remaining Neu5Ac in the medium was going under the TBA assay before reading at OD 549 nm. Means of three wells were calculated before applying the one-way ANOVA test, with $P < 0.05$ being taken as the level of significance (*); $P < 0.01$ being taken as the level of significance (**); $P < 0.001$ being taken as the level of significance (***); $P < 0.0001$ being taken as the level of significance (****). Enclosed shapes are data for every individual experiment.

3.2.4.6. Biofilm Growth on Alternative Sialic Acid Source

Several studies have shown that the adherence and internalisation of *T. forsythia* to host cells are mediated by several extracellular matrix glycoprotein.[88,91] On the heavily sialylated salivary glycoprotein mucin, the contribution of the TonB proteins to deliver energy for the sialic acid uptake was tested using the parental strain ATCC 43037 and the negative control Δ*nanT* compared to the isogenic Δ*nanO* and four *tonBs* insertion mutants. The level of biofilm formation, while reduced, on polystyrene wells coated with mucin (60 μM) was not significantly different for Δ*BFO_0333*, Δ*BFO_3116*, and Δ*BFO_0953* from that of the parental strain ($P =>0.05$). In contrast, a significant decrease in the proportions of biofilm formation to mucin from the parental strain were observed with the Δ*nanO* ($P =0.013$) and Δ*BFO_0233* ($P =0.0139$) (Figure 3.31).

On other mucin concentrations, such as 15 μM and 30 μM, we observed a significant difference in forming a biofilm for Δ*nanO* and Δ*BFO_02333*. The control Δ*nanT*, an isogenic ATCC 43037 derivative with impaired sialic acid uptake, to form biofilm was significantly reduced as seen for Δ*nanO* and Δ*BFO_02333*. Likewise, on multiple mucin constructions (15 μM and 30 μM), the same three TonBs (Δ*BFO_0333*, Δ*BFO_3116*, and Δ*BFO_0953*) showed no significant difference to the level of the wild type. These results indicate that the *tonB* Δ*BFO_0233* has a greater role compared to other *tonB*s in the ability of ATCC 43037 to deliver energy and form biofilm on heavily sialylated salivary glycoprotein mucin (Figure 3.31).
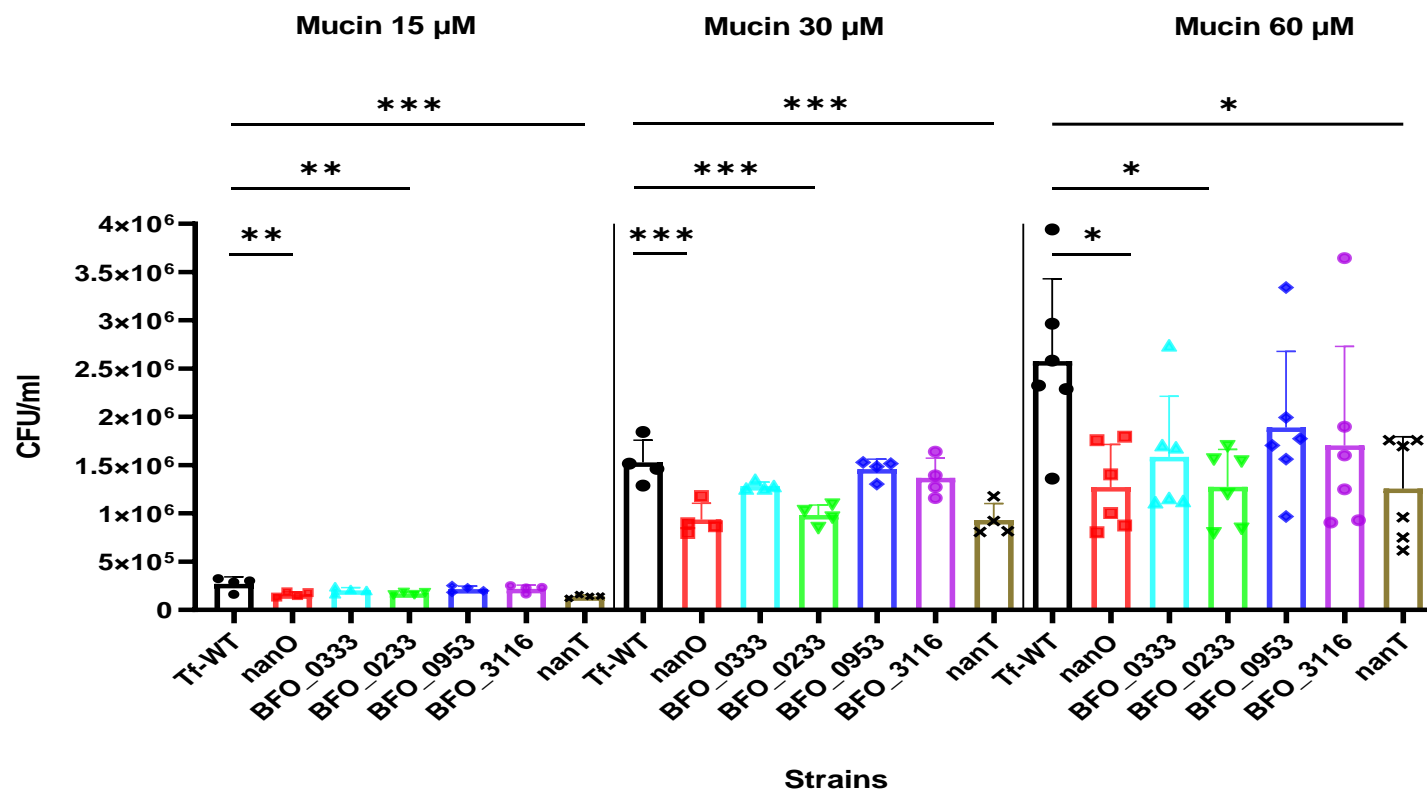
Figure 3.31 Biofilm formation of *T. forsythia* and isogenic mutant strains on glycoprotein-coated surfaces.
Different concentrations of mucin were coated on the 96-well plate overnight at 4 °C. Next day, these wells were washed off with PBS two times before inoculation of WT *Tannerella*, Δ*nanO*, Δ*BFO_0333*, Δ*BFO_0233*, Δ*BFO_3116*, Δ*BFO_0953*, and Δ*nanT* in TSB medium. These strains were harvested, washed three times with PBS, and adjusted to a cell density at OD_{600} of 0.05. Biofilm formation on mucin was set up in triplicate wells for each strain (Total 200 μl). After seven days of growth, the 96-well plate was washed with PBS, harvested, and counted. To eliminate bias, this study counted all 16 squares of Helber counting chamber (Hawksley). Means of three wells were calculated before applying the one-way ANOVA test, with $P < 0.05$ being taken as the level of significance (*); $P < 0.01$ being taken as the level of significance (**); $P < 0.001$ being taken as the level of significance (***); $P < 0.0001$ being taken as the level of significance (****). Enclosed shapes are data for every individual experiment.

3.3. Discussion

During infection, bacteria are often exposed to different environments. These diverse environments can arise from multiple stressors and nutritional restrictions, which in turn force these microorganisms to adapt to survive changing conditions. In severe conditions, the ubiquity of microbes and its survival are due to the availability of multiple transport systems to the microbe for an intake of nutrients.[235] Sialic acid is an important element to numerous organisms and the bacterial sialic acid transport systems are necessary in increasing fitness and overcoming sialic acid deficiency taken by a host. The TonB systems in Gram-negative bacteria play an essential role in the transport of many elements. Of these elements, iron, hemin, B[12], starch, serum, and other metals were established in many microorganisms.[236] In *E. coli*, several studies concluded that *tonB* gene transported iron and vitamin B[12] from the outer membrane into the periplasmic space.[131,142] In addition, work from the Stafford lab suggested that *tonB* energises sialic acid uptake in *T. forsythia* and is essential for survival within epithelial cells.[214] Similarly, to several other bacteria, *T. forsythia* contains multiple *tonB* encoding genes; therefore, we were interested in determining the role of TonB proteins in sialic acid transport during the adhesion and invasion into epithelial cells as well as asking if increased knowledge of this might lead to design of small peptides to target inhibition of TonB to prohibit this transit.

In *T. forsythia*, sequence analysis indicated that four *tonB* genes are located at separate loci, and there is weak DNA homology among these *tonB* genes. These *tonB* genes were difficult to predict *in silico* as analysis of the neighbourhood of genes encoding these four *tonB* genes does not clearly suggest the individual roles, specifically, in the presence of sialic acid. Presence of multiple *tonB* genes raises a few essential questions: Do all four have similar protein domains and structures? Are all four *tonB* genes involved in sialic acid transport? If not, which *tonB* gene indicates a huge role in the survival of *Tannerella* during biofilm formation?

The presence of multiple *tonB* genes contrasts with the many well characterised organism in which *tonB* is known to be important for small molecule transport, such as *E. coli*. However, the bioinformatic analysis is an important step in explaining the TonB protein. Division of TonB protein into the N-terminal and C-terminal domains showed conserved and similar motifs compared to the control *E. coli* TonB.

The average of TonB residues was analysed in 263 TonB sequences from 144 Gram-negative bacteria to conclude a range between 225 and 300 residues.[149] Of these organisms, almost half of them encoded more than one TonB protein in their genomes, with numbers ranging between two and nine different TonB proteins. Of these organisms with TonB

proteins, *Pseudomonas syringae pv. syringae B728a* was found to have nine TonB proteins.[149] In *T. forsythia*, the short TonB is 278 residues compared to 621 residues of the long TonB protein.

Following the established nine clusters of 28 TonB proteins representing 263 TonB sequences, the phylogenetic analysis of the C-terminal domains of *T. forsythia* TonB proteins showed three of TonB proteins fell into the 3A cluster and one fell into the 1C cluster. These clusters do not depend on the taxonomy of the organism, but they are based on the C-terminal domain.[149] Previous analysis of these clusters showed that TonB protein from the same species can be aligned within different clusters of CTD-domains. *P. syringae* pv. *Syringae* B728ABh has nine different TonB proteins where they were classified as several clusters.[149] Likewise, neighbour-joining bootstrap tree aligned the three *R. anatipestifer* TonBs into different clusters.[237]

The sequence and predicted secondary structures elements of the C-terminal domain of TonB are well conserved. Based on the conserved 90-residue C-terminal domain in the *E. coli* TonB protein, majority of TonB proteins in Gram-negative bacteria were found to share a predicated secondary structure consisting of 3 β-strands and 2 α-helices.[237] We highlighted residues of *Tannerella* TonB proteins that were either perfectly conserved or highly similar to the control TonB of *E. coli*. Among *tonB* families in Gram-negative bacteria,[149] YP motif is the most conserved feature and is conserved in all TonB proteins of *T. forsythia*. Residues with less similarity were more notable within α-helix 2 and β-strand 3 compared to *E. coli* TonB protein. The identical alignment of these loops is not inherently obvious when the loop lengthening is a general feature of many TonB proteins.[149] Shultis et al and Pawelek et al suggested in their crystal structure of the complexes of TonB protein with the BtuB and FhuA transport that the TonB is distal from the interaction site between the outer membrane receptors.[156,157] Alignment of all secondary structures of *E. coli* TonB to other organisms with TonB proteins showed a range of conserved to less similarity residues. These organisms that were bioinformatically inspected with less similarity residues were *S. ruber* DSM 13855, *H. influenzae* 86-028NP, *Xanthomonas campestris* pv. *campestris* str. ATCC 33913, *M. magneticum* AMB-1., *Silicibacter* sp. TM1040, and *E. carotovora subsp. atroseptica* SCRI1043.[149]

The second domain of the TonB protein is the periplasmic linker domain, which consists of three regions.[149] In *E. coli* TonB, the amino acid length of this linker is 117 residues, and the established region of this linker is the 33 amino acid Pro-rich region.[149] Although the Pro-rich is speculated to transduce mechanical energy from the cytoplasmic

membrane to the outer membrane, the Pro-rich domain was titrated with a ^{15}N-labeled sample of TonB-CTD with peptides representing the EP and KP repeats. As a result, this confirmed that the pro-rich domain does not have an influence on the flexible structure near the C-terminal domain, which was consistent with previous work on *E. coli* TonB and *V. cholerae* TonB1.[238,239] In contrast, several investigations reported the necessity of this region to span the periplasmic space and to form the TonB:FhuA and TonB:FhuD complexes.[240,241] Carter et al concluded also from their work that the TonB has more than one interaction site with TBDTs, including the Pro-rich domain.[241] With the limitations presented in the previous studies, more investigations about the Periplasmic linker domain, specifically the Pro-rich domain, are necessary. In this study, we noticed a conserved amino acid sequence among the four TonB proteins (GGQxxLMxxIA), but nothing is known about it.

The third domain of TonB protein is the N-terminal domain. The size of N-terminal domains ranges in length, however, each TonB has at least one N-terminal transmembrane α-helix. An alignment of N-terminal domains from 144 organisms concluded at least the presence of one predicted N-terminal transmembrane α-helix, which can be characterised with the conserved (SXXXH) motif containing many hydrophobic residues. In *E. coli* TonB, the spatial relationship between the Ser and His residues (SXXXH) can define the least required transduction element.[219] It also indicated the importance of this motif for coupling TonB to the cytoplasmic membrane electrochemical gradient.[219] In *Tannerella*, the presence of SXXXH motif was found in all four TonB proteins, but with less consensus among three TonBs and high consensus in TonB BTO_3116 protein. The *Tannerella* TonB (BTO_3116/TF0783) has ExbB and ExbD proteins within its genetic context, which may suggest its role with either or both iron and BtuB transport. However, Zhao and Poole concluded that the TonB1 of *P. aeruginosa* does not have Ser residue, but a site-directed mutagenesis showed the Ser residue is not essential for TonB1 functions.[242] Likewise, *Vibrio vulnificus* TonB lacks the conserved His residue, whereas *P. syringae phaseolicola* TonB3 lacks the entire motif.[149]

Several TonB proteins have large sized N-terminal domains (291–348 residues) compared to the N-terminal domain of *E. coli* TonB (124 residues). Of which, TonB proteins with homology to MecR1 and BlaR1 proteins were found in *S. aureus* and *S. sciuri* to deliver energy and are characterised with longer N-terminal domains.[149] Both MecR1 and BlaR1 were reported with a highly conserved loop region coding for an M56-Zn^{2+} peptidase involving in antibiotic resistance. This peptidase can regulate expression of genes encoding penicillin-binding protein 2A and β-lactamase. The profile alignment of longer N-terminal

domains from MecR1 and BlaR1 predicted a perfect conservation of a putative zinc-binding motif (HEXXH).[243] This zinc-binding motif was found in the N-terminal domains of *T. forsythia* BFO_0333 (621aa) and BFO_0233 (480aa), but this domain has unknown function in the regulation of other proteins in the periplasm and may influence regulation or post-trans effects. This domain also might be a cofactor for these two TonB proteins, specifically, the TonB BFO_0233 and sialic acid uptake. Further laboratory work and bioinformatics assessments are need for more clarification about this domain.

T. forsythia encodes two sets of TonB systems and two monocistronic transcription unit of TonB proteins. The location of TonB proteins can be in the same operon for hemin/iron uptake-regulated genes (ExbB and ExbD) (co-transcribed direction), such as *E. coli*, *Vibrio anguillarum* and *P. damselae*.[137,244] The presence of monocistronic transcription unit of TonB proteins was mentioned previously in *B. fragilis* 638R, *R. anatipestifer* and *A. baumannii*.[245–247] In Gram-negative bacteria, several studies reported two ExbD proteins that were encoded within the same TonB cluster. Of which, the presence of two ExbD was found in *A. hydrophila* NJ-35, *R. anatipestifer. A. baumannii*, *Flavobacterium psychrophilum*, and *X. campestris pv. Campestris*.[235,246–249] All previous TonB clusters with encoded two ExbD proteins were found to deliver energy for uptake. The presence of two ExbD proteins encoded within the same TonB cluster were needed for iron uptake in *R. anatipestifer* and *A. baumannii*, whereas presence of one ExbD without other ExbD was able to restore the growth of *F. psychrophilum*, and *X. campestris pv. Campestris* in the presence of iron. It is suggested that the additional ExbD component could enhance energy transducing ability due to specific communications between the components of TonB system.[247]

This study was able to amplify and clone the sialic acid outer membrane porins from the ATCC 43037 and 92A.2 strains. We inserted *Tannerella* ATCC43037-*nanO*, ATCC43037-*nanOU*, *Tannerella* 92A.2-*TF0033*, and *Tannerella* 92A.2-*TF0033,34* into the pBAD18 expression plasmid and transferred each into Δ*nan*Δ*tonB* MG1665 strain. This strain is devoid of outer membrane sialic acid porins and is unable to grow in the presence of sialic acid as a sole carbon and energy source.

Before testing the role of *Tannerella* TonB and starting the process of heterologous complementation, this study examined *E. coli* MG1665 Δ*nan*Δ*tonB* strain in several steps. Of which, this strain was gram staining and grown on MacConkey agar plate to test its ability for utilising the lactose available in the medium. Next, a combination of designed primers for PCR amplifications were used to check the location of all reported deletion to this strain (*nanC*/*ompR*/*tonB*). We also sequenced the sialic acid outer membrane (*nanC*) to check and

confirm the location of deletion. Following this, in M9 minimal medium assay, the role of Δ*nan*Δ*tonB* MG1665 strain in the presence of sialic acid was completely abrogated growth, as opposed to the presence of glycerol. This demonstrates the established protocol of proceeding to the next level of complementation. Transformations of empty vectors, that were used for either outer or inner membrane interest of genes ligation, negatively affect the growth of Δ*nan*Δ*tonB* MG1665. Although these vectors were compatible with each other, as well as pBAD18 vector was found previously to function normally with the same strain, we sought more clarification by complementing every TonB along with the sialic acid outer membrane. Complementation strains were unsuccessful in the presence of sialic acid as a sole carbon and energy source. Although a heterologous host like *E. coli* was not successful for this study, Zhao and Poole tried to define additional requirements for TonB protein by completing TonB-swapping experiments between *E. coli* and *P. aeruginosa*. They concluded that the *P. aeruginosa* TonB can function in *E. coli*, but *E. coli* TonB cannot function in *P. aeruginosa*.[242] In addition, partial functionality was noticed when expression of a chimeric *P. aeruginosa* TonB1 containing the N-terminal extension used with a chimeric TonB containing the C-terminal domain of *E. coli*.[242] Several studies from various Gram-negative bacteria were able to characterise the activity of TonB proteins using *E. coli*, such as *A. actinomycetemcomitans*, *Bartonella birtlesii*, and *A. baumannii*.[247–249] In contrast, complementation of three TonB proteins of *R. anatipestifer* were unable in the presence of hemin to retain growth, as opposed to δ-Aminolevulinic acid.[246] Of the three TonB proteins presence in *R. anatipestifer*, one of which required its complete ExbB1-ExbD1-TonB1 complex to show phenotype in complementation. This might explain the importance of polypeptide sequence identity between TonB proteins, which is needed to retain a strain's growth.[246]

Several assumptions can be predicted for non-functionality of the heterologous complementation *in E. coli* (Δ*nan*Δ*tonB* MG1665). One of which is that the presence of two chosen vectors for ligations may defeat one another once they are expressed inside the *E. coli* strain. We are not sure if these two vectors can produce a high-copy-number of expressions resulting in an altered stoichiometry between vectors and the strain. Similarly, an altered stoichiometry of the protein components from the ligation vectors was observed from heterologous complementation *in E. coli* for *A. baumannii* TonBs.[247] Although we had difficulty establishing complementation with empty vectors, we complemented Δ*nan*Δ*tonB* MG1665 with *Tannerella* TonB. As a result, complementation with *Tannerella tonB* genes showed no difference, and this further failure might be due to the functionality of *Tannerella*

tonB genes that cannot be transplanted and restored outside the *Tannerella* due to the genetic contest around each *tonB*. Low amino acid sequence homology, especially within the C-terminal domain between *Tannerella* TonB and *E. coli* TonB may cause unbalance growth within the *E. coli* strain. We also suspected that the site-directed mutagenesis might not be fully non-functional once we transform the outer membrane genes of *Tannerella*. Thus, we attempted to construct a full deletion for *nanC* (sialic acid main porin) in *E. coli* to investigate this matter, but with the time and laboratory restrictions, attempts were unsuccessful. Additional time and efforts are needed to inspect this strain for future experiments.

Microorganisms can respond to environmental changes and adapt their gene expression appropriately. This response leads to a process effecting cellular metabolism.[247] To further examine and confirm the function of individual *tonB* genes with sialic acid transport, this study constructed gene deletion for these *tonB* genes from the parent organism *T. forsythia*. Based on *Tannerella's* sequence analysis and on previous *Tannerella's* studies, we further checked whether *Tannerella* genome encodes for an erythromycin-resistance gene. We managed to delete all *tonB* genes from *Tannerella*, additional to the sialic acid outer membrane of (Δ*nanO*) using erythromycin cassettes. These mutants were established using an efficient method for construction of gene deletion, based on electroporation and natural competence. However, generation of combined *tonB* mutants was not possible in this study due to the bacterium fastidious difficulty along with time restrictions. With the range of planktonic and biofilm growth experiments, we found that three *tonB* (*BFO_0333*, *BFO_0953*, and *BFO_3116*) were not involved in the transport of sialic acid, indicating that they may accomplish different functions by serving as a polyvalent energy coupler for functioning of various TonB-dependent transporters. In particular, the mutant *BFO_0233* was significantly affected in the presence of sialic acid among other *tonB* compared to other wild type *Tannerella*. These data indicate that the *tonB BFO_0233* is a major contributor to the sequestration of sialic acid under the biofilm restricted condition.

The role of *T. forsythia* adhesion and nutrition to sialic-acid-like structures is well established.[203] Mucin can provide nutrient source for bacteria and promote their biofilm growth. Mucin coating can also provide the advantage of introducing both hydrophilic and hydrophobic properties to polystyrene surface enhancing adhesion and interaction.[250] We assessed the role of *tonB genes* on three different concentrations of the heavily sialylated salivary glycoprotein mucin. As expected, we found that the mutant *BFO_0233* was consistently at low levels of biofilm under restricted environmental condition as *Tannerella* tried to sequester sialic acid from the environment. This indicates that the *tonB BFO_0233*

can provide energy for sialic acid uptake from mucin where the sialic acid is α-2,3- and α-2,6-linked but largely 9-*O*-acetylated sialic acid. The effect of mucin on *T. forsythia* growth was introduced.

The growth of *T. forsythia* is strictly dependent on *N*-acetylmuramic acid (NAM) for axenic growth under laboratory conditions. NAM is to support construction of the peptidoglycan (PGN) of bacterial cell walls. Absence of NAM from laboratory growth of *Tannerella* results in an absence of growth. In addition, *Tannerella* is auxotrophic for NAM as it lacks general enzymes required for the *de novo* synthesis of precursors of PGN. In the oral habitat, *Tannerella* depends on cohabiting bacteria for the provision of nutrients including NAM, which it drives through cell wall turnover or decay of cells.[251] We additionally hypothesise an additional role of *tonB* genes in delivering energy during uptake of NAM. The constructed mutants *tonB* genes were assessed in biofilm assays including viable cell counting and bacterial cell density. Accordingly, data from this study demonstrated that deleted *tonB* genes do not impair *Tannerella* biofilm growth on NAM, growing at a rate comparable to the wild-type strain. Intriguingly, these phenotypes could be attributed to another energy system for the NAM uptake during biofilm formation.

Pathogens have several requirements for needed compounds that can contribute to the pathogen's environmental adaptability and spread in host tissues.[138] Remarkably, *Tannerella* was found to require TonB energy for the transfer of sialic acid across the outer membrane. *B. fragilis* contains six *tonB* gene homologs, and only *tonB3* gene was required for growth in the presence of soluble starch, serum, vitamin B_{12}, heme, and iron.[245] In addition, the well-established required substance for many pathogens is iron where the iron complexes can be transported though specific outer membrane receptors, which they are dependent on the TonB energy transduction system. Iron plays a huge role in a host-pathogen interplay through enabling its survival within host tissues and its contribution to virulence factors. In *Bradyrhizobium japonicum*, TonB1 was needed for the heme TBDT HmuR whereas TonB2 was required for the uptake of FhuE (Fe^{3+}-chelates), EntR (enterobactin), and FegA (ferrichrome).[252] In *A. hydrophila* NJ-35, the TonB2 system of was involved in the utilisation of iron compared to TonB1 and TonB3.[235] In hemin transport as a source of iron, both TonB systems were found to function in *V. anguillarum*.[137] Likewise in iron and hemin utilisation, *R. anatipestifer* was shown to use TonB2, but not TonB1.[253] In *Campylobacter jejuni* NCTC1116, both TonB1 and TonB3 were not only required for hemoglobin and hemin utilisation, but also for colonisation in chickens.[254]

Besides requirement of energy for utilisation, the role of TonB was explored in several pathogenic aspects. In *A. baumannii*, it is responsible for local and systemic human infections, and it possesses three TonB proteins.[247] An intention was carried out to identify three isogenic *tonB* genes in iron utilisation by growing these mutants in LB broth containing enough free iron. Surprisingly, the second TonB of *A. baumannii* was a monocistronic *tonB* locus, with absence of both *exbBD*, found to affect adhesion to polystyrene wells coated with fibronectin and human alveolar epithelial cells when it should have attached due to the presence of free iron. Likewise, the third TonB of *P. aeruginosa* was involved in motility and pilus assembly compared to the other two TonB proteins.[255] In *A. hydrophila,* TonB3 compared to other two TonB proteins has a major role in anti-phagocytosis properties and antibiotic resistance.[235] Only the TonB2 system in *V. anguillarum* was found to be essential for its virulence factors.[137]

Furthermore, *tonB* mutants can be suitable backbone strains for future vaccine development. In mice models, two out of multiple other studies indicated that targeting the *tonB* genes by a common mutagenesis approach produced effective vaccine candidates with protection against *Burkholderia mallei* and *B. pseudomallei*.[237,256] Cystic fibrosis patients suffer from chronic and eventually fatal infections caused by *Burkholderia cenocepacia*. The lack of useful treatments pushes the need to develop alternative therapies. Targeting *tonB* gene of this bacterium generated attenuation to this strain by encoding energiser protein of *tonB* gene. This protein is responsible for the activation of iron transport through the outer membrane. The vaccination elicited the immune response in clearing *B. cenocepacia* from the infected mice. While the mechanisms for acquisition of the needed compounds for survival and colonisation are closely between pathogenic process in several bacterial species, vaccine development by targeting the *tonB* gene seems a potential target.[237]

Taken together, the result of this study suggested the lack of functional redundancy or overlapping in *Tannerella* four *tonB* genes. There was a noticeable effect on biofilm formation due to each *tonB* mutant, and this emphasises the important role of interaction for each *tonB* during the biofilm formation. *B. fragilis* has a TonB protein that does not involve in the uptake of several carbon sources, but *in silico* is located to a surface antigenic lipoprotein family of cell wall/membrane (TamA/BamA/YaeT). This family is responsible for the biogenesis proteins forming β-barrel structures (TBDTs).[245] Previously, TonB was shown to be able to interact with nonreceptor proteins and it was found that there is absence of PFM in the periplasmic space.[87,250] This may suggest a new role of TonB protein in delivering energy for β-barrel protein assembly and/or OMV segregation.[245] This noticeable

effect on biofilm formation due to each *tonB* mutant may also suggest an interaction of TonBs as a heterodimer to link the periplasm between the ExbB and ExbD complex at the inner membrane and the TBDTs in the outer membrane.[158] Nevertheless, only the *tonB BFO_0233* was shown to be significance for the sialic acid utilisation through its interaction with the TonB box. Since the *tonB BFO_0233* mutant is not completely abolished, this may suggest additional energy for sialic acid via other *tonB* genes in the absence of *tonB BFO_0233*, but they cannot substitute functionally for *BFO_0233*. Moreover, the fact that each isogenic *tonB* mutant derivative was fully functional in the biofilm growth of NAM, it states the complete function for these strains when culture under laboratory conditions, which facilitates the growth defect of *BFO_0233* is due to inability to transit sequestered sialic acid effectively. Identifying how and why the TonB protein can be selective and yet sometimes promiscuous about which TBDTs or OMV they can energise is a significant step in our understanding of the TonB mechanism. Nonetheless, abilities of TonBs in transducing energy have specific requirements that can interact with a variety of TBDTs and/or ExbB/ExbD, although the molecular interactions of TonB with TBDTs have not been fully confirmed.[152,153] Furthermore, no specific elements have yet been identified to define the specific requirement of a TonB protein although we have the reflection of conserved motifs in both N-terminal and C-terminal domains. Yet, a consensus sequence of amino acid (5 to 8 amino acids) at the N-terminal plug region of TBDTs (i.e., TonB box) can suggest the need of energy from TonB protein. In organisms with multiple TonB proteins, each TonB may serve and function with a subset of the TBDTs. In contrast, organisms with only one TonB portion can use it to serve many different TBDTs.[149] *B. fragilis* has 104 predicted TBDTs and six TonB proteins, however, only one TonB was found to deliver energy for uptake.[245] Stork et. al. examined *V. anguillarum* that has two TonBs and less TBDTs compared to other organisms to find only one TonB importing the native siderophore anguibactin.[137] Nierman et. al. found only one TonB in *Caulobacter crecentus* that has 65 TBDTs.[257] Further research studies could investigate conserved motifs on the N-terminal and C-terminal domain and reflect on these clustering to substances.

The sialic acid *nan* operon of *Tannerella* was defined and tested in several studies.[91,105,110,128,258] The main gate of sialic acid entrance *nanO* was tested in heterologous complementation in *E. coli* system.[88] In this chapter, the mechanism of sialic acid transit from the outer membrane into the inner membrane was directly tested in the *T. forsythia*. A deletion of outer membrane sialic acid (*nanO*) directly affected the uptake of sialic acid compared to the wild type *Tannerella*, but not completely abolished in its ability to take up

salic acid. This suggests entrance of sialic acid as well via other non-specific outer membrane transporters. However, the mutant *nanO* does not affect the *Tannerella's* ability to use NAM as it retained a growth similar to the wild type.

Like multiple members of the *Bacteroidetes*, *T. forsythia* has a plethora of glycosidase and sugar-focussed metabolic systems.[259] When the source of sialic acid is only conjugated to glycoproteins, this study examined the behaviour of ΔnanO on multiple concentrations of the salivary glycoprotein mucin. The mutant *nanO* affected the biofilm growth of *Tannerella*, indicating a huge role of the sialic acid outer membrane porin. Likewise, the biofilm lifestyle of *Tannerella* was changed with deficient mutants including nonulosonic-acid (ΔpseC) and S-layer (ΔtfsAB).[37] It is tempting to suggest that the ability of mutant *nanO* to decrease biofilm setting is mirrored in the multispecies consortium where *Tannerella* may exhibit direct or indirect effect on other oral bacteria.

Sialic acid as a carbon and nitrogen source is important for bacteria where bacteria can transport this sugar into the peptidoglycan pathway or decorate their surfaces to avoid host immune surveillance.[87] Aside from heterologous complementation in *Tannerella*, the construction of mutations was introduced to confirm the role of several gene behaviours.[105,110,260,261] Construction of isogenic mutant of *wecC* gene responsible about the biofilm formation resulted in a reduction in the biofilm formation.[260] In *nan* operon of *Tannerella*, Honma et al's study tested two sialidase enzymes of *T. forsythia* to confirm their roles during biofilm process. This study mutate both sialidase enzymes to indicate their roles in the host cell interactions as well as to target host sialoglycoproteins to release and utilise sialic acid.[102] Afterwards, NanU from *T. forsythia* was characterised biochemically to indicate its optimisation for sialic acid uptake.[214] In comparison to the sialic acid outer membrane gate, is the role of *nanT*, the inner membrane gate of sialic acid, which was established on direct and indirect transit of sialic acid.[261] Furthermore, the *Tannerella nan* operon of sialic acid highlights its capabilities of enabling adhesion and biofilm growth upon different forms of sialic acid, such as recycled sialloglycoproteins and mucin. *T. forsythia* attenuation by *nanO* mutant was consistent with the result from other *nan* operon of sialic acid studies and confirmed the role of this operon in sialic acid transport. Of the sialic acid *nan* operon, both NanO and NanT represent the main entrance of sialic acid from the outer and inner membranes before metabolism indicating a role in *Tannerella* persistence on and inside the epithelial cells.[203]

Adapting to various environments is a huge challenge for the survival of any microbes. The variety of environmental conditions encountering bacteria requires quick

adjustments of metabolic processes to adapt to ecological changes. This study highlights a role that *tonB BFO_0233* plays in *T. forsythia* pathogenesis along with the sialic acid outer membrane porin (*nanO*). This study also suggests that the other three TonB may play a role in *T. forsythia* environmental adaption where TonB can control virulence-associated mechanisms and compensate for one another to affect bacteria phenotype. Besides the current understanding of the TonB role in the environmental adaption and pathogenic role, further studies are needed to clarify more about the exact mechanisms associated with TonB functional variations. Sialic acid as an essential nutrient requires energy for uptake through TonB, which indicates potential targets for the improvement of novel antimicrobial therapeutics due to increases in antibiotic resistance.

Future work:

This work used a defined medium to test the role of TonB in the presence of either NAM, sialic acid, or mucin; each carbon and energy source was supplemented with hemin and vitamin B_{12}. However, future work can test all these mutants in media where the vitamin B_{12} is changed to l-methionine (75 μg/mL) and the hemin is changed to ferrous ammonium sulfate (100 μM) and protoporphyrin IX (5 μg/mL). This will further illustrate whether these TonB were affected due to the presence of hemin and vitamin B_{12}. With the presence of four *Tannerella* mutants, a further conclusion can be drawn about the involvement role of each TonB in *Tannerella* virulence. Evaluation of biofilm formation and survival within oral epithelial cells for *Tannerella* and for *Tannerella* with other established oral biofilm model bacteria will clarify more subgingival biofilms and also during infection of human epithelial cells *in vivo*. Additional work can be concluded about the effect of these TonB from the intracellular survival in macrophage cells and cytotoxicity assay. Sensitivity of the wild-type and the *tonB* mutant strains can be tested against several antibiotics. The co-transcription of these TonBs should be confirmed by RT-PCR using total RNA. Designed primers for the intergenic regions located between the predicted ORFs should be used to amplify these regions. This process will help to clarify more about the regions around each TonB and whether these intergenic regions can be co-transcribed as an operon or influence each TonB. Differential expression of each TonB can further also be investigated by qRT-PCR analysis of RNA. In addition, the phenotype of each mutant was observed in this study through involvement with sialic acid transit. However, complementation of one TonB to another was seen in different studies. Double or multiple deletion mutants can exhibit further conclusion about all four TonB in *Tannerella* determining their physiological function and pathogenic significance.

Chapter IV

Bioinformatic analysis of TBDTs in *T. forsythia* and laboratory assessment of the TonB box peptide

4.1. Introduction

The role of TonB dependent transporters (TBDTs) were discovered in *E. coli* K12 during the resistance of bacterium to infection by bacteriophage T1. Later this TBDT was named FhuA and its role in the uptake of iron–siderophore complexes. Further explorations of homologous transporters defined a large role in transferring various macromolecules that are too large to diffuse via porins from the outer membrane. Besides that, TBDTs are usually important for sensing and adjusting to environmental signals and are linked to pathogenicity in bacteria like *E. coli, Serratia marcescens*, and *P. aeruginosa*.[262] An essential part of TBDTs is a small sequence that is located immediately at the N-terminal to the plug domain and identified as the TonB box. The TonB box is necessary for coupling between the TBDT and TonB system upon conveyed signal and allowing the movement of the uptake from the outer membrane.[262] The presence of TBDTs in *Bacteroides* was proposed by Abigail Salyers and others.[263] After 10 years, the presence of TBDTs in both *X. campestris* and *C. crescentus* confirmed their role in the transport of carbohydrates including sucrose and maltodextrins.[262] This led to multiple surveys of TBDTs across bacteria noting key differences of these TBDTs between *Bacteroidetes* and *Enterobacteria*.[262]

The *Bacteroidetes* have powered the capacity for complex carbohydrate degradation like mucin layer, plant fibers, and glycosaminoglycans. Dozens to several hundred distinct operons that are termed polysaccharide utilisation loci (PUL) are encoded within the *Bacteroidetes* for complex carbohydrate degradation and can influence a bacterium's metabolic niche and fitness within the host.[264] The PUL in *Bacteroidetes* encode genes targeting and importuning polysaccharides before completing hydrolysis to its component sugars. These PUL can be divided into generic or specific PUL targeting different glycan substructures.[265] In *Bacteroides uniformis*, there is a single generic PUL that can recognise and target discrete β1,3-glucan structures.[266] In *B. ovatus*, recognition and uptake of different types of xylan require the bacterium to deploy and activate multiple PUL.[267]

Within the past two decades, the sequencing of bacterial genomes has increased clarifying more about the *Bacteroidetes* SusC/D proteins for carbohydrate uptake. The TonB-dependent transporter (TBDT) and a surface lipoprotein are hallmark features of the *Bacteroidetes* PUL.[122] The presence of both TBDT and a surface lipoprotein is usually called out as a *susC*/*susD* homolog pair, following the Starch Utilisation System (Sus). Later

biochemical and functional studies provided a better understanding of the systems' conserved and novel features in the human oral cavity, gut, and the environment. In *B. thetaiotaomicron* genome, sequencing results showed more than 100 homologous *susC/D* pairings where these Sus are flanked by genes encoding predicted polysaccharide lyases, glycoside hydrolases, and other enzymes for carbohydrate degradation. However, the presence of only TBDT (SusC) without SusD is the key to the outer membrane porin of the host.[125]

The TBDTs and/or Sus concept for nutrient uptake and their mechanistic features in *T. forsythia* have not been fully explored. A recent discovery of a functional shufflon suggested TBDTs recombination can determine additional subclasses of TBDTs.[268] The main aim of this study was to predict and characterise these TBDTs in all available *T. forsythia* strains. The Pfam database has an advantage of linking similar domains with a Pfam family code, enabling prediction and analysis of TBDTs domains. Thus, the Pfam domain analysis was used to predict the broader role of TBDTs in all *T. forsythia* strains to review what is presently known and what work remains to be performed to understand these TBDTs. To clarify, four Pfam domains were used to locate homologies for the following: TonB-dependent receptor barrel domains (Pfam code: PF00593), TonB-dependent transporter plug domains (Pfam code: PF07715), N-Terminal extension TBDT domains (Pfam code: PF13715), and Signal transduction domains (Pfam code: PF07660). After we annotated each *Tannerella* genome using the Pfam database, a manual search approach using the Pfam family codes was used to predict the domain structure for each TBDT. In addition, the Pfam tool was used to locate and identify the surface lipoprotein (SusD protein). Today, we have four recent classifications to characterise the structure of each surface lipoprotein: SusD_RagB (Pfam code: PF07980), SusD_like (Pfam code: PF12741), SusD_like_2 (Pfam code: PF12771), and SusD_like_3 (Pfam code: PF14322). Additionally, a manual search approach was used to locate the domain for SusD protein and its cognate with SusC/TBDT. In this chapter, we described for the first time the structural features of TBDTs and Sus with an emphasis on the novel predicted features of TBDTs in *T. forsythia*.

Furthermore, the TonB box is a signature of a TBDT that is powered by the TonB system where the TonB C-terminal domain (a member of TonB system) interacts with the β-strand of TonB box on the TBDT (section 1.10). Three decades ago, there was a suggestion of using a TonB box peptide as an anti-bacterial inhibitor targeting the direct interaction of TonB system with the TBDTs in the context of iron transport.[269] Thus, the second aim of this

study was to predict the TonB box of sialic acid NanO and design a TonB box peptide targeting the interaction between the NanO and TonB system.

4.2. Results

4.2.1. TonB-dependent Transporters in *T. forsythia*

4.2.1.1 Structural Domains of TBDTs within *Tannerella* Species

        Initially, the core structure of the host is to identify and characterise the TBDTs where it transits a ligand into the periplasm. The previous conclusion of TBDTs defines its core structure with two key domains.[112] The first domain spans the outer membrane and is defined as a TonB-dependent receptor barrel domain. The second domain is housed within the barrel domain and is defined as a TonB-dependent receptor plug domain (Figure 4.1).[112] Today, there are 13 *T. forsythia* strains that are available within the NCBI database; thus, we searched these 13 genomes of *Tannerella* species and concluded the presence of at least 58 or more of TBDTs among 12 *Tannerella* human and non- human genomes (Table 4.1). In contrast, the BU0063 strain (*Tannerella serpentiformis*) is a periodontal health-associate isolate as it was isolated from a healthy patient with no signs of periodontal disease (periodontitis). This strain was found to have only 35 TBDTs compared to other strains. Today, the recent discovery of a functional shufflon uncovered additional subclasses of TBDTs.[268] These subclasses are resulted from the recombination of multiple TBDTs that generated three additional domains/subclasses for characterising the TBDTs.[268] The three domains/subclasses are Classical TBDT, N-Terminal Extension TBDT, and Signal Transduction TBDT. Some of TBDTs domains/subclasses are found with glycan-capturing lipoprotein (SusD) and are part of polysaccharide utilisation loci (PUL). Therefore, we are going to characterise each predicted TBDT following the reported domains/subclasses.
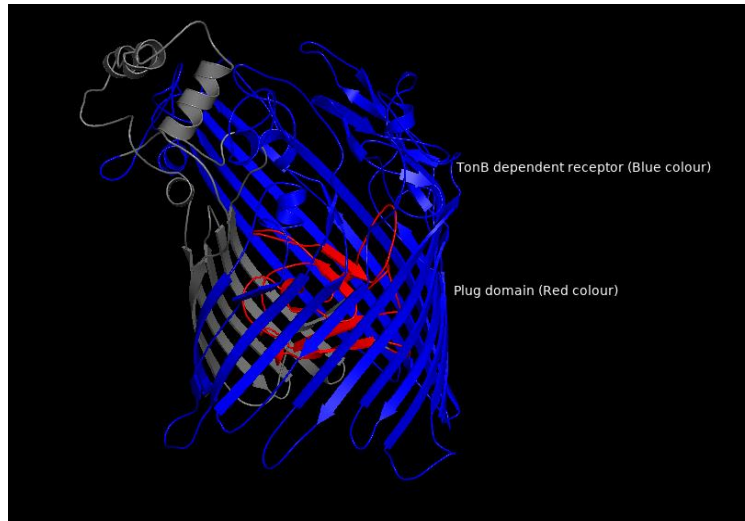
Figure 4.1 Structure of TonB-dependent transporter.
PyMol software was used to highlight the *T. forsythia* NanO structure. The barrel domain is displayed in blue, and the plug domain is displayed in red. This structure was predicted using Protein Homology/analogY Recognition Engine (Phyre) based on the NanO amino acid sequence.

Table 4.1 TonB-dependent Transporter (TBDT) Subclasses in *T. forsythia*

This table presents TBDTs prediction in 13 *T. forsythia* from human and non-human strains. It also presents also the three TBDT subclasses in each strain. The BU0063 is a periodontal health-associate isolate and it was isolated from a healthy patient. Both *T. forsythia* OH 1426 and OH 2617 strains are non-human genomes and were isolated from a dog (**NB**. more details about these strains are in the next chapter-section 5.3.1.2.).

Tannerella Strains	Total of TBDTs	Types of predicted TonB-dependent Transporters (TBDT)												
		Classical TBDT				N-terminal extension					Signal Transduction			
		Total	In TBDT	In PUL	Protein length range	Total	In TBDT	Protein length range of TBDT	In PUL	Protein length range of PUL	Total	In TBDT	In PUL	Protein length range
ATCC 43037-Human	62	11	11	0	268 – 791	46	9	744 – 1001	37	984 – 1261	5	0	5	1134 – 1176
92A.2-Human	60	9	8	1	663 – 783	45	17	551 – 1103	28	984 – 1176	5	0	5	1125 – 1176
3133-Human	55	10	10	0	230 – 791	40	11	744 – 1049	29	984 – 1261	5	1	4	1131 – 1176
9610-Human	63	12	12	0	219 – 975	46	19	744 – 1139	27	984 – 1190	5	2	3	1100 – 1179
KS16-Human	56	9	9	0	263 – 783	42	11	744 – 1080	31	984 – 1261	5	0	5	1125 – 1176
UB4-Human	58	11	10	1	219 – 972	46	13	672 – 1038	33	999 – 1261	6	0	6	1125 – 1176
UB20-Human	58	11	11	0	230 – 818	47	10	744 – 1001	37	984 – 1261	5	0	5	1131 – 1179
UB22-Human	63	10	10	0	219 – 780	46	12	744 – 1154	34	242 – 1261	7	0	7	1116 – 1176
WW 10960-Human	61	10	10	0	672 – 1139	45	16	709 – 1144	29	984 – 1193	6	1	5	1100 – 1179
WW 11663-Human	60	8	8	0	669 – 783	46	17	744 – 1139	29	1006 –1189	6	0	6	1131 – 1179
BU0063-Human	35	11	10	1	672 – 911	24	17	724 – 1111	7	999 – 1169	0	0	0	0
OH 1426-non-Human	58	11	10	1	641 – 926	45	17	737 – 1141	28	970 – 1163	2	0	2	1097 – 1135
OH 2617-non-Human	58	11	11	0	641 – 902	44	19	737 – 1173	25	1004 –1188	3	0	3	1116 – 1135

4.2.1.2. Classical TBDT

The presence of a classical TBDT is characterised by two key domains, barrel domain and plug domain. While the plug domain is housed within the barrel domain, it was noticed that the barrel domain may not be present or found within the TBDTs. In few cases however, Pfam database could predict the presence of Pfam plug domain of the classical TBDT without the Pfam barrel domain. Further investigation was conducted on *B. thetaiotaomicron* TBDTs suggesting that the presence of only Pfam plug domain is likely TBDTs.[262] The reason that the classical TBDT has a homologous to the plug domain without the barrel domain might be due to some significant changes to either sequencing or overall structure limiting the domain prediction.[262] As a result, the presence of the Pfam plug domain only without the barrel domain was found between one to five times within each *Tannerella* genome (Table 4.2). In contrast, this study noticed the presence of only the Pfam barrel domain but not the Pfam plug domain. The average presence of the Pfam barrel domain without the plug domain occurs between two to three times within a *Tannerella* genome (Table 4.2). Thus, this study included all classical TBDT that contains both key domains, or only the plug domain. We excluded all TBDTs that showed the presence of only the Pfam barrel domain from our analysis.

Table 4.2. Misannotation of Pfam domain families from TBDTs identification.

Strain	Plug domain only	Barrel domain only	NTE domain only
ATCC 43037	2	5	10
92A.2	1	0	15
3313	2	5	9
9610	3	3	12
KS16	3	1	14
UB4	4	0	14
UB20	3	0	14
UB22	3	0	10
WW 10960	2	2	13
WW 11663	2	3	13
BU0063	0	0	8
OH 1426	2	1	12
OH 2617	1	1	13

The homologues to these two key domains were previously characterised by Pfam domain families.[262] The 22 β-strand barrel domain is characterised by a Pfam code of PF00593, whereas the plug domain is characterised by a Pfam code of PF07715 (Figure 4.2). As a result, we predicted the presence of classical TBDT structure within each genome of *Tannerella*. The range of classical TBDTs was between nine to 12 with a *Tannerella* genome. For instance, both ATCC 43037 and 92A.2 strains contained 11 and nine classical TBDTs, respectively (Table 4.1; Figure 4.2; Figure 4.3-A). Likewise, the periodontal health-associate isolate was found to contain 11 classical TBDTs.

Next, we inspected the presence of surface lipoprotein within classical TBDTs. We found only three human isolates (92A.2, UB4, and BU0063) to have one of classical TBDT in PUL structure (SusD) (Table 4.1; Figure 4.3-A). The range of amino acids length of the classical TBDTs was between 219 and 791 base pairs. However, the range of amino acids length of the classical TBDTs in 92A.2, WW 10960, WW 11663, and BU0063 strains were between 663 and 1134 base pairs (Table 4.1).

		ATCC 43037	92A.2	3313	9610	KS16	UB4	UB20	UB22	WW 10960	WW 11663	BU0063	OII 1426	OII 2617
A	Classical TBDT	11 sequences	9 sequences	10 sequences	12 sequences	9 sequences	11 sequences	11 sequences	10 sequences	10 sequences	8 sequences	11 sequences	11 sequences	11 sequences
B	N-terminal extension	46 sequences	45 sequences	40 sequences	46 sequences	42 sequences	46 sequences	47 sequences	46 sequences	45 sequences	46 sequences	24 sequences	45 sequences	44 sequences
C	Signal Transduction	5 sequences	6 sequences	5 sequences	5 sequences	5 sequences	6 sequences	5 sequences	7 sequences	6 sequences	6 sequences	0 sequences	2 sequences	3 sequences

Figure 4.2 Distribution of Pfam architectures and subclasses of TonB-dependent transporters (TBDTs).

This figure summarizes the total of TBDT domains based on the TBDT subclasses in 13 *Tannerella* strains. A) This presents the first TBDT subclass, and it consists of the barrel and plug domains. Both barrel and plug domains are coloured in blue and red, and they have a Pfam code of PF00593 for the barrel domain and a Pfam code of PF07715 for the plug domain. B) This presents the second TBDT subclass that includes both the domains of subclass A and the N-terminal domain (PF13715- coloured in yellow). C) This presents the third TBDT subclass that includes the signal transduction domain (PF07660-coloured in green) along with the other TBDT A and B subclasses.
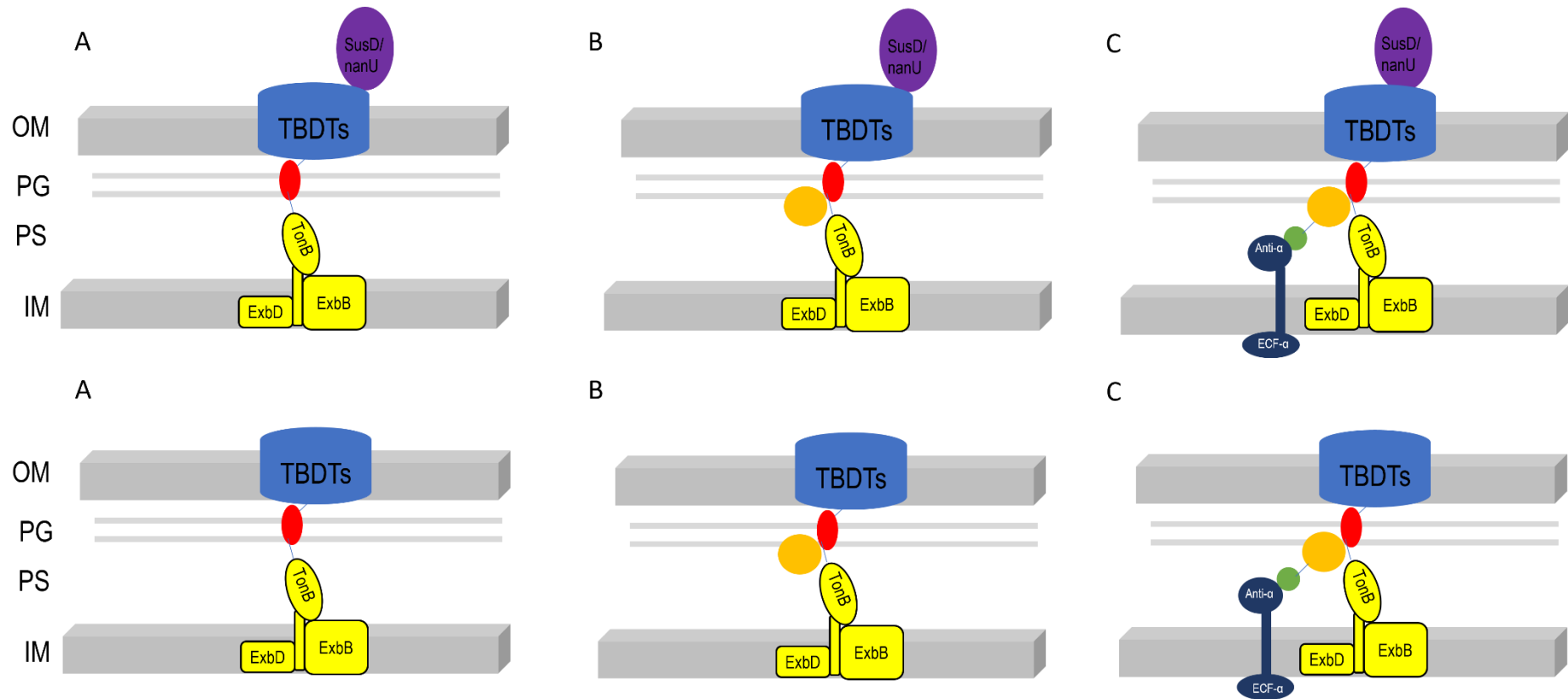
Figure 4.3 General TBDT architectures and subclasses.
Six different architectures represent TonB-dependent transporters found in Gram-negative bacteria. Both the TonB system and anti-sigma factor are shown. The TonB system includes TonB/ExbB/ExbD complex. The barrel domain is displayed in blue, and the plug domain is coloured red, representing the classical TBDT (A). The susD (nanU) is shown in purple. Both B and C represent the N-terminal extension domain (gold colour) and the signal transduction domain (green colour), respectively. Abbreviation: OM: outer membrane; PG: Peptidoglycan; PS: periplasmic space; IM: inner membrane.

171

The new classification of TBDT is characterised by the presence of barrel and plug domains in addition to carboxypeptidase D regulatory-like (or annotated additionally as DUF4480 domain) (Figure 4.2-B). The N-terminal extension domain is located on the N-terminal to the plug domain and TonB box and is coded with a Pfam code of PF13715 (Figure 4.3-B). We noticed the presence of the only Pfam carboxypeptidase D regulatory-like domain but without the Pfam plug domain. The average of presence of Pfam carboxypeptidase D regulatory-like domain without the plug domain was 12 times within the *Tannerella* strains (Table 4.2). Thus, this study excluded from analysis any TBDT containing only Pfam carboxypeptidase D regulatory-like domain without at least the presence of the Pfam plug domain.

Of the N-Terminal extension domain dissemination, the pathogenic *Tannerella* isolates from humans showed a range between 40 and 47 times, whereas around half of the N-Terminal extension domains from pathogenic *Tannerella* were absent from the non-pathogenic human *Tannerella* isolate (BU0063) (Table 4.1 and Figure 4.2-B). The architecture of the N-Terminal extension is found in transport to associate with and without the surface lipoprotein (SusD/PUL). Of the human pathogenic *Tannerella* isolated genomes, this domain was found to be associated with SusD/PUL in a range between 27 and 37 times.

Furthermore, we inspected the size of amino acids among the N-Terminal extension domains of the TBDTs. The size of 672 to 1154 base pairs was seen in the human pathogenic genomes that are not encoded within a SusD/PUL. This domain was encoded within a PUL showing a range of amino acid between 984 and 1261 base pairs. Interestingly, we found a lipoprotein (SusD/PUL) with 551 base pairs among this domain structure in the 92A.2 strain (BFO_0495), suggesting a possibility of different function. Likewise, a TBDT (SCQ25174.1) from the UB22 strain and its SusD (SCQ25176.1) has an unusual amino acid size of 242 base pairs, indicating a novel function.

4.2.1.4. Secretin and TonB N-terminus Domain (STN)/Signal Transduction TBDT

The third TBDT subclass is similar to the previous N-Terminal extension domain, but it is characterised by an additional N-terminal domain and is located in the N-terminal to the TonB box (Figure 4.3-C). This domain can be identified and recognised as a Signal Transduction domain and is characterised with a Pfam domain family of PF07660 (Figure 4.2-C). In addition, this domain was found to enable and facilitate the interaction in transit between the TBDT and the associated anti-sigma factor through signalling (Figure 4.3).[270]

The presence of this domain was not found within the periodontal health *Tannerella* isolate (BU0063). In contrast, the STN domain was encoded among other *Tannerella* genomes with a range between three and seven times (Table 4.1; Figure 4.2-C). For instance, the presence of STN domain within the most studied *Tannerella* ATCC 43037 and 92A.2 strains was five and six times, respectively. Furthermore, the STN domain was found among *Tannerella* genomes with and without the surface lipoprotein (PUL). The *Tannerella* 9610, WW 10960, and 3313 strains showed only that the STN domain was associated with and without PUL structures compared to other *Tannerella* strains. The range of amino acids length of STN domain was noticed to be between 1100 and 1179 base pairs.

4.2.1.5. Presence of the SusD/NanU Protein

The evidence of an interaction between the SusC and SusD proteins came directly from the early work on the *B. thetaiotaomicron*. Later validation in the recent crystal structures presented several interactions of SusC and SusD, such as the transport of Levan and peptide-targeting complexes.[123,262] This study concluded 357 of SusD sequences among human *Tannerella* isolates and 59 of SusD among non-human *Tannerella* isolates (Table 4.1). Of the 357 reported SusD proteins, 319 of which were encoded with the N-terminal extension TBDT domain.

Today, bioinformatic enhancements characterised and described four Pfam structures for SusD proteins (Figure 4.4).[271] These families are SusD_RagB (PF07980), SusD_like (PF12741), SusD_like_2 (PF12771), and SusD_like_3 (PF14322). The presence of several contexts in which two of these families are combined was previously noticed.[262] Investigation of total sequences for each Pfam SusD family using the Pfam database showed the two largest families (SusD_like_3 and SusD_RagB) in terms of falling sequences. Both SusD_like_3 and SusD_RagB were found to contain more than 21808 and 22462 sequences, respectively (Search was in April, 2022). The other two families contained more than 4100 sequences for SusD_like_2 and 891 sequences for SusD_like. In this analysis, the majority of the SusD sequences from *Tannerella* genomes fell within the combined SusD_RagB and SusD_like_3 families (Figure 4.4). In *B. thetaiotaomicron*, two of the previous reported combined families 'SusD_like_2 and SusD_like_2', and 'SusD_like and SusD_like_2' were absent within the *Tannerella* species. In terms of a single family, the SusD_like_2 was found to contain the majority of SusD sequences among *Tannerella* genomes. The presence of these families may distinguish between human pathogenic and non-pathogenic *Tannerella* isolates. We found the presence of SusD_like_2 is the only single family that is shared between the human pathogenic and non-pathogenic *Tannerella* isolates as other reported single families may associate with pathogenicity.

	ATCC 43037	92A.2	3313	9610	KS16	UB4	UB20	UB22	WW 10960	WW 11663	BU0063	OH 1426	OH 2617
A PF07980 SusD_RagB	1 sequence	0 sequences	0 sequences	0 sequences	0 sequences	1 sequence	0 sequences	0 sequences	0 sequences	0 sequences	0 sequences	0 sequence	0 sequences
B PF12741 SusD_like	1 sequence	1 sequence	3 sequences	2 sequences	1 sequence	1 sequence	2 sequences	2 sequences	2 sequences	1 sequence	0 sequences	0 sequences	3 sequences
C PF12771 SusD_like_2	7 sequences	4 sequences	4 sequences	3 sequences	4 sequences	6 sequences	6 sequences	6 sequences	5 sequences	6 sequences	2 sequences	4 sequences	2 sequences
D PF14322 SusD_like_3	1 sequence	1 sequence	1 sequence	1 sequence	1 sequence	1 sequence	2 sequences	1 sequence	1 sequence	0 sequences	0 sequences	1 sequence	0 sequences
PF14322 SusD_like_3 + PF07980 SusD_RagB	32 sequences	28 sequences	25 sequences	24 sequences	30 sequences	31 sequences	32 sequences	32 sequences	26 sequences	28 sequences	6 sequences	26 sequences	23 sequences
PF12771 SusD_like_2 + PF12771 SusD_like_2	0 sequences	0 sequences	0 sequences	0 sequences	0 sequences	0 sequences	0 sequences	0 sequences	0 sequences	0 sequences	0 sequences	0 sequences	0 sequences
PF12771 SusD_like_2 + PF12741 SusD_like	0 sequences	0 sequences	0 sequences	0 sequences	0 sequences	0 sequences	0 sequences	0 sequences	0 sequences	0 sequences	0 sequences	0 sequences	0 sequences
Total	42 sequences	34 sequences	33 sequences	30 sequences	36 sequences	40 sequences	42 sequences	41 sequences	34 sequences	35 sequences	8 sequences	31 sequences	28 sequences

Figure 4.4 Distribution of Pfam families for SusD

This figure shows the four families and the additional combined reported families of SusD. The presence and characterisation of SusD/NanU from total 13 strains are shown. These families are SusD_RagB (PF07980) (Blue colour), SusD_like (PF12741) (Yellow colour), SusD_like_2 (PF12771) (Green colour), and SusD_like_3 (PF14322) (Red colour).

4.2.1.6. Sialic Acid Outer Membrane (NanO) of *T. forsythia*

After sialic acid and its derivatives transit through the NanO-type outer membrane protein, the microbe can metabolise these units of sugars as needed. Analysis of *T. forsythia* S-layer also illustrated the presence of nonulosonic acid biosynthetic pathways. It was also needed to understand the behaviour of this microbe in multispecies biofilm as well as to understand the invasion of gingival epithelial cells.[272,273] Other studies have described the passage of sialic acid through the inner membrane, whereas others illustrated the sialic acid metabolism in *T. forsythia*.[88,91,105] *T. forsythia* has a unique *O*-glycan with abundant glycoproteins that are equipped with a complex *O*-linked decasaccharide and linked to the surface of S-layer of this bacterium. *T. forsythia* can terminate glycan with a modified nonulosonic acid. Nonulosonic acid has structural and biosynthetic similarities to sialic acid including either pseudaminic acid (5-N-acetimidoyl-7-N-glyceroyl-3,5,7,9-tetradeoxy-L-glycero-L-manno-NulO) or legionaminic acid (3,5,7,9-tetradeoxy-D-glycero-D-galacto-NulO). For both pseudaminic acid (Pse) and legionaminic acid (Leg), there was a report of several occurring derivatives. These derivatives are at the C-5 and C-7 positions from the N-acyl groups, such as *N*-formyl (Fo), *N*-hydroxybutyryl (Hb), *N*-acetimidoyl or acetamidino (Am), and *N*-acetyl or acetamido (Ac).[259] In the previous chapter (Chapter III), the deletion of the sialic acid outer membrane *nanO* affected the uptake of Neu5Ac, as well as the biofilm formation on both Neu5Ac and mucin. Thus, a comparative genome analysis was carried out to align the ATCC 43037 sialic acid outer membrane NanO to all predicted sialic acid outer membranes (NanO) from other *Tannerella* strains. This analysis divided all *Tannerella* strains based on the previous nonulosonic acid derivative before comparing between their sialic acid outer membranes (NanO) using the NanO of the pseudaminic acid ATCC 43037 strain compared to the NanO of legionaminic acid UB4 strain. Alignment of both nucleotide acid and amino acid sequences showed higher identities between these sialic acid outer membrane proteins NanO (Table 4.3). Surprisingly, the non-human isolate microbes of *Tannerella*, as a control, showed higher homologues to the sialic acid outer membrane of the human isolates *Tannerella*. This suggests that the sialic acid NanO from the ATCC 43037 strain and the predicted other sialic acid NanO proteins might be generic sialic acid with multiple functions to other derivatives.

Table 4.3 Alignment of outer membrane NanO of *T. forsythia* sialic acid

Pseudaminic acid strains	Isolation	UB4_nanO		Sheffield ATCC 43037-nanO	
		DNA	Amino acid	DNA	Amino acid
ATCC43037-NanO	Human	99.02%	99.55%	-	-
UB20	Human	98.54%	99.73%	98.20%	99.20%
9610	Human	93.95%	44.73%	95.28%	44.73%
OH2617	Non-Human	89.83%	95.62%	90.83%	96.62%

Legionaminic acid strains	Isolation	UB4_nanO		Sheffield ATCC 43037 - nanO	
		DNA	Amino acid	DNA	Amino acid
92A.2	Human	95.12%	98.26%	94.32%	95.26%
UB4	Human	-	-	99.02%	99.55%
UB22	Human	98.25%	99.09%	98.73%	99.37%
3313	Human	99.295	99.64%	94.32%	99.82%
KS16	Human	99.26%	99.82%	98.73%	99.29%
WW10960	Human	99.08%	99.64%	99.73%	99.82%
WW11663	Human	99.49%	99.82%	99%	99.28%
OH1426	Non-Human	89.83%	33.92%	89.83%	32.32%

4.2.2. TonB Box

In general, the TonB box consists of five to eight amino acids which are located on the N-terminal domain of the TonB-dependent transporter (TBDT). The TonB box is a conserved hydrophobic region responsible for the activation of TonB protein. The activation of TonB protein is processed physically to force and transit the required element from the outer membrane into the periplasmic membrane. Analysis of TBDTs using both bioinformatic and laboratory approaches uncovered a TonB box-binding site on the N-terminal domain of several TBDTs. Previous alignment of finding the TonB box of RagA of *P. gingivalis* compared the RagA sequence to other established TonB box sequences from TBDTs of FepA, HemR, SusC, FyuA, IrfA, CirA, BfrA, Tla, and BfeA (Figure 4.5).[274] In light of close alignment to *Tannerella* sialic acid outer membrane sequence (NanO), there were two families merging with higher identities. These two similar families were the Sus-type transporter and iron-type transporter. An alignment of *Tannerella* NanO to these families (SusC and RagA) considering their identified TonB box sequences resulted in a predicted protein sequence for the TonB box of NanO (GASVVE) (Figure 4.5).

Figure 4.5 Alignments of TonB boxes with the highest similarity scores compared to RagA of *P. gingivalis*
This figure showed alignments for SusC (starch utilisation), FyuA (yersiniabactin siderophore), HemR (hemin-regulated protein), FepA (ferric enterochelin), IrgA (iron-regulated outer membrane), CirA (colicin 1), BfrA (exogenous ferric siderophore), Tla (TonB linked adhesion), and BfeA (ferrichrome-iron). Abbreviation for bacteria: bt, *B. thetaiotamicron*; ye, *Y. enterocolitica*; pg, *P. gingivalis*; ec, *E. coli*; vc, *V. cholera*; bp, *Bordetella pertussis*; pa, *P. aeruginosa*, bb, *Bordetella bronchiseptica*.
Edited with permission from (Ref:270); Appendix VIII: 8.12. Copyright (1999), American Society for Microbiology."

4.2.2.1. Alignment of the TonB Box of NanO

As outlined above, we predicted the TonB box of sialic acid NanO 'GASVVE' by aligning the previous prediction of other bacteria TonB boxes. Then we asked about the presence and conservation of the TonB box of NanO among other TBDTs in *T. forsythia* ATCC 43037 and whether a designed peptide of the TonB box of NanO may affect other TBDTs functions (Figure 4.6). Alignment of all the predicated *Tannerella* TBDTs within this study showed several proteins (4 to 6) with residue identities to the TonB box of NanO, however, only two predicted proteins with complete match to GASVVE (Figure 4.7). Three amino acids were almost found in each predicted TonB box of *Tannerella* TBDTs (glycine, alanine, and valine). This may suggest that a designed peptide mimicking the TonB box of NanO may attack more than one TBDTs once these TBDTs interact with the TonB protein.
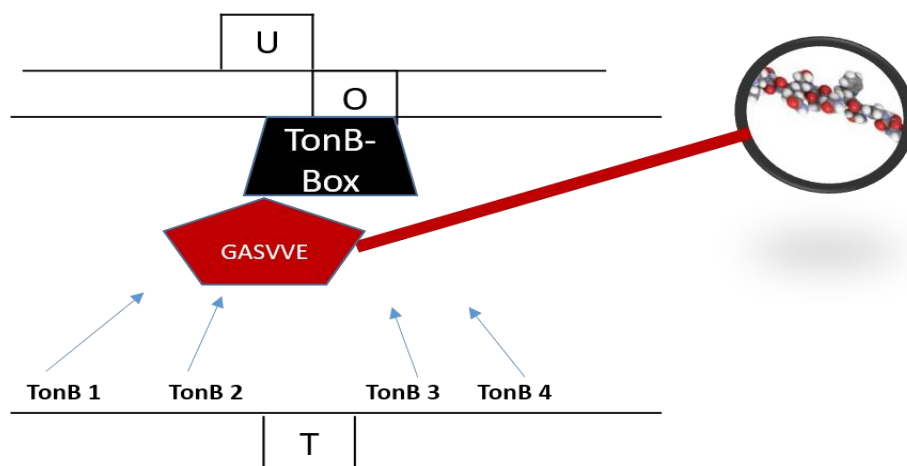
Figure 4.6 Schematic of the TonB box of NanO
This scheme illustrates the possible interaction of the TonB box of NanO with NanO-type transporter and other possible TBDTs that may require energy from the TonB system.

```
NanO           KNVSGTVTD----QSGEPVIGASVVEKGT-----------RNGTVTDIDGNYSL-S-TA--
BFO_0495       KTVSGRIVD----EQGEPVIGANVVEKGT-----------TNGTVTDADGRFSL-S-VA--
BFO_0499       KTVSGRIVD----EQGDPVIGANVVEKGT-----------TNGTVTDADGRFSL-S-VA--
BFO_2970       TQVQGTVVD----EQGEPVTGASIVLKTDR---------SKGTISDIDGKFTL-W-VP--
BFO_3110       VQIRGSVVD----ETGEPVIGANILVKGT----------TQGTTSDVDGQFSL-S-VDR-
BFO_1715       GKVIGTVVDA---MTGEPVIGANVTVKGT----------TVGTITDINGHFSL-E-AP--
BFO_1718       VVIKGKIID----ESGEPLPGVSVYTKDM----------TSGTISDPDGKFSL-P-VNH-
BFO_0588       KHITGKVSD----LQGEPIIGANVVEKANP---------ANGTITDVEGNFTI-D-IQ--
BFO_0298       VLIKGVVYD----DHNVPAIGVNVVKKGS----------TEGVITDLNGQFEI-R-AS--
BFO_0723       RPVKGRVAD----TSGEPVTGVTIVEKANP---------LNGTITDVDGHFTI-D-VQ--
BFO_0981       TSIGGRIVD----ASGEPVIGATIVEKGNT---------SNGTITDVDGQFTI-E-TP--
Tfor_00596     ITVRGIVTD----DGHEPVIGATVVVKGDP---------SKGTITDLDGNYVL-NNVP--
BFO_1399       QIISGQVMD----SEGFPVIGANIRVKNT----------TIGGITDIDGAYTL-E-VP--
BFO_1407       ITVRGTVVD----DTGLTVIGASVVEENNR---------TIGTITDIDGKYVL-ENVP--
BFO_2600       VTVRGTVVD----DTGLTVIGASVVESNR----------TIGTITDIDGKYVL-ENVP--
BFO_1456       KTISGIVRD----ERGEPVIGANIIEKGVP---------GNGTITDVDGMFSL-S-VP--
BFO_2286       ITVRGSVSD----TRKEPVIGATVIVKGNV---------NQGTITDMDGRYEL-RDVP--
BFO_2915       RSVSGRITD----AQGEPVIGANVVEAGT----------TNGTVTDIDGRFTL-E-IS--
BFO_2985       VAIRGTVTD----INDNPLIGVNVSIKGT----------TIGNITDSEGNYTM-E-APG-
Tfor_00600     ITVRGTVTD----DGHEPVIGATVVVKGDP---------SKGTITDLDGNYVL-NDVP--
BFO_2271       ITVKGTVVD----EMNEPVIGANVVEKGT----------INGIITDLDGHFTL-S-VD--
BFO_2635       ITVRGNVKD----VKGEPLIGVSVLISGT----------SLGTITDVDGNFTLNN-VA--
BFO_2801       RRISGKVVD----AEGQPVIGANVVEKGT----------GHGVITDVDGRFSL-E-MP--
BFO_2812       RRISGKVVD----AEGRPVIGANVVEKGT----------GHGVITDVDGRFSL-E-VP--
BFO_2743       QQVKGTVLD----ENGEPIIGANVVEKGT----------INGMVTDLDGHYTL-S-VSNV
BFO_0394       RTVSGTVVD----EQGVPLLGVNVSVSGT----------TTGTITDIDGKFTL-E-VP--
BFO_1191       ITVSGVVTEQ---ATGEPAIGVSILVKGT----------TNGTVTGIDGSYTL-SDVP--
BFO_0970       NVIIGAVVDA---KTGDPLAGASILVKNSDP--------PRGATADADGKFEL-T-AT--
BFO_2562       ITVRGIVTD----RNNEPLIGVTIRVEGYE---------NQGTVTDMDGKYMLVG-VR--
BFO_2013       KGITGVVKD----SKGEPVIGANVIEKGTP---------ANGTVTDIDGNFSL-S-VD--
BFO_1796       VNITGTVVD----HLNDPLPGVNIQVKGL----------SVGTITDAEGVFSI-E-VPY-
BFO_3314       RKITGTIKD----ELGEGMVGVNILLKGT----------NTGTTSDYDGNFSI-E-VP--
BFO_2107       TQVQGTVVD----EQGEPVIGASVVLKSDR---------TKGTVTDMDGKFTF-A-AP--
BFO_1298       RVVTGVVLAG---DDNQPLVGVSIQVPSDELKKAGISSNTLGTITDVNGNFSI-S-VPD-
BFO_1024       KEIKGTVKD----RSGEPIIGANVYVKGT----------TNGVVTGVDGDFSI-Q-VS--
Tfor_00675     KKVSGIVKEA---DTNEPLIGATVVVKGT----------TNGVITDLDGHYAI-E-VN--
BFO_0030       TRITGNVTD----EKGEPVIGASILVKGT----------DQGTVTDLDGNFSL-S-VL--
BFO_0033       TRITGNVTD----EKGEPVIGASILVKGT----------DQGTVTDLDGNFSL-S-VSSS
BFO_1412       AQVRGTVTD----ENGEPVIGASVVVKGT----------TVGTVTDMDGKFALPE-VPS-
BFO_2513       ITVRGTVTD----RNNEPLTGATVVVEGST---------SQGTVTDIDGNYTL-AGVN--
BFO_0188       VVVRGRITD----TQNEPVVGANIIEEGAS---------GNGTISNADGAFTL-E-VN--
BFO_1519       KAVTGVVMD----ASNTPIIGANIIEKGTA---------TNGTVTDHDGRFAL-S-VS--
BFO_1510       ------------NSMHKVVNIGVL-----------------------------------
BFO_1769       -----------------------------------------------LCL-------
Tfor_01497     ----------------------------------------------------------
Tfor_01271     ---------------------------------------------GLFTT-------
BFO_2076       YIIKGTVKN----RQDEPIPYATVMVTGT----------NIGMAANVEGKFQI-K-LNR-
BFO_2077       YTVKGTVVD----ENGDPLPGASVWVVGA----------TIGAGTNANGEFAI-R-LNEG
BFO_2621       ANITGHVQDA---TTKEHLPYINIQVKGT----------TIGVTTDATGHYML-KDLPV-
BFO_2639       -----------------------------------------------YWL-------
Tfor_02463     NTVELKVVDA---ENRSIVESATVQWKPV-----GVAKFKDGALTNRHGVVHI-S-MKEK
BFO_1598       AGIRGVVVDD---TNEEPLIGANIYIEQL----------RQGDAAGSNGEFAF-SGLAA-
BFO_2217       YVLSGAVTD----ATGHPVSMATVAIEHT----------SLGAFTDDKGHYTL-E-LSA-
BFO_1383       YEIRGIVVDA---DTQEPLPGAAVYDQAN---------YNGTATDLEGHFRL-TGLKS-
BFO_2605       ----------------------------------------------------------
BFO_3283       ----------------------------------------------------------
BFO_0521       --------------QSEAP---------------------------------------
BFO_1590       FSLRGTVYEMSTRHQKQPLPGASLIIEAS----------GIGTTSDVNGDYTL-TPIFE-
BFO_3014       FTITGKILNS---KSKEGVGFAAVSFADN----------RLWTVADEAGRFVL-RDVPG-
BFO_0736       ----------------------------------------------FTLYM-LS--
BFO_0418       ----------------------------I------------------GLFFLHSIVS--
BFO_3323       VKVSGNVRD----ADGNPLELVAVQVKHT----------LIGAMTDEKGFYSL-S-VSPG
```

Figure 4.7 Amino acid alignment of *Tannerella* TBDTs.
Multiple sequence alignment was performed for *Tannerella* TBDTs highlighting the similarities between the TonB box of NanO 'GASVVE' and other possible TonB boxes of TBDTs.

4.2.2.2. Optimisation of TonB Box Peptide

The introduction of TonB box peptide was presented in an *E. coli* study where the TonB box indicated its physical interaction with the TonB protein. The TonB box pentapeptide FhuE receptor showed inhibition *in vivo* for several TonB-dependent transporters of iron, phi 80 infection, and colicins B and Ia. Previous inhibition was reported with the addition of 100 μg/ml TonB box pentapeptide (Glu-Thr-Val-Ile-Val).[269] Here, we tested the *Tannerella* biofilm formation on Neu5Ac (6 mM) with 50 μg/ml, 100 μg/ml, and 150 μg/ml of the predicted and designed *Tannerella* TonB box peptide (GASVVE) as a preliminary test for more detailed experiments presented later in this chapter. As a result, a growth biofilm of seven days showed that there was a reduction, but not significant in the biofilm of *Tannerella* with the addition of 100 μg/ml and 150 μg/ml of the TonB box peptide (GASVVE) (Figure 4.8). Therefore, this study chose the 100 μg/ml of the peptide concentration although this experiment was performed one time.
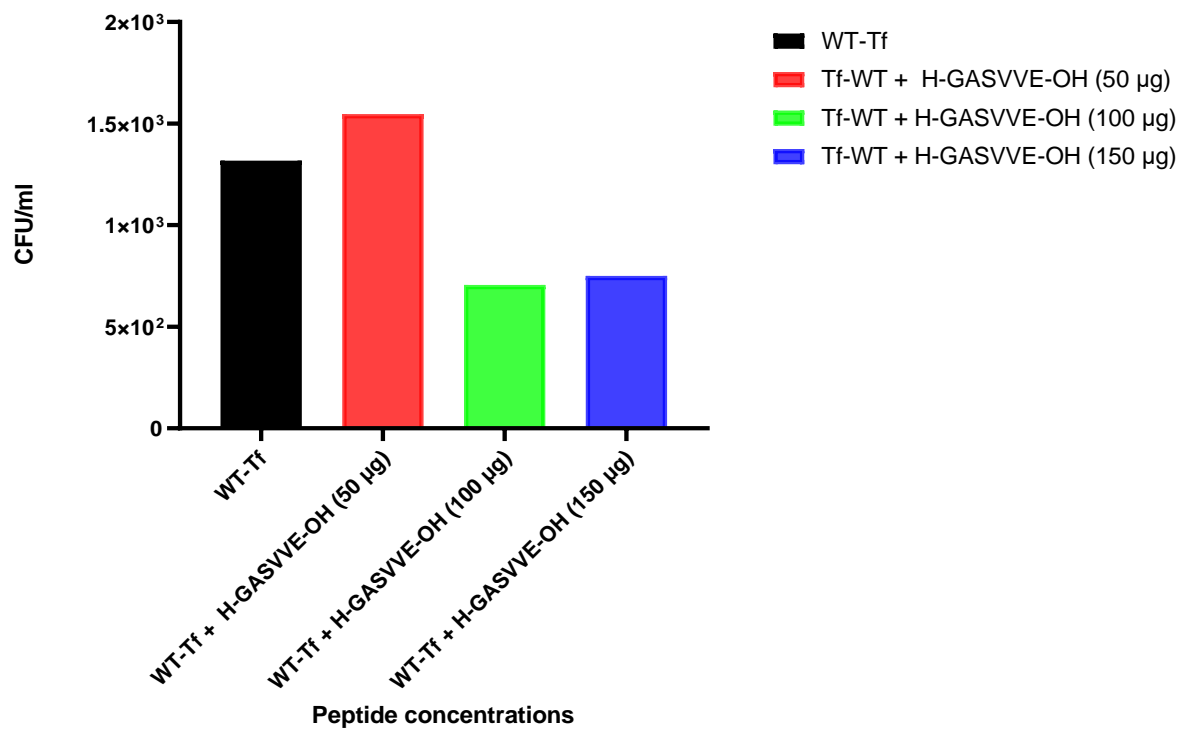
Figure 4.8 Effect of peptide concentrations on *T. forsythia* biofilm formation.
        The figure shows the ability of peptide concentration to inhibit biofilm growth of *T. forsythia* after seven days. *T. forsythia* was inoculated in a defined TSB medium with either NAM or Neu5Ac along with three different peptide concentrations, 50, 100, or 150 μg/ml. The samples were washed and resuspended in sterile PBS to be checked by cell counting. To eliminate bias, this study counted all 16 squares of Helber counting chamber (Hawksley).

4.2.2.3. Planktonic Growth of *Tannerella* on the TonB box peptide

In the presence of NAM or sialic acid as a sole source and carbon energy, the bacterial growth of *Tannerella* was observed after seven days with three different peptides. The designed TonB box peptide of NanO (GASVVE) was tested against a scrambled version (SEVAGV) and the *E. coli* FepA TonB box peptide (DDTIVE). The effect of GASVVE peptide (100 µg/ml) showed a reduction of growth but not significant in the presence of sialic acid compared to the parental strain and the other two peptides. No growth reduction was seen for GASVVE peptide compared or the other two peptides in the presence of NAM (EGGSVE and DDTIVE). This indicates not only the effect of this peptide on planktonic growth in the presence of sialic acid, but also the interaction of the sialic acid NanO with the TonB protein (Figure 4.9).
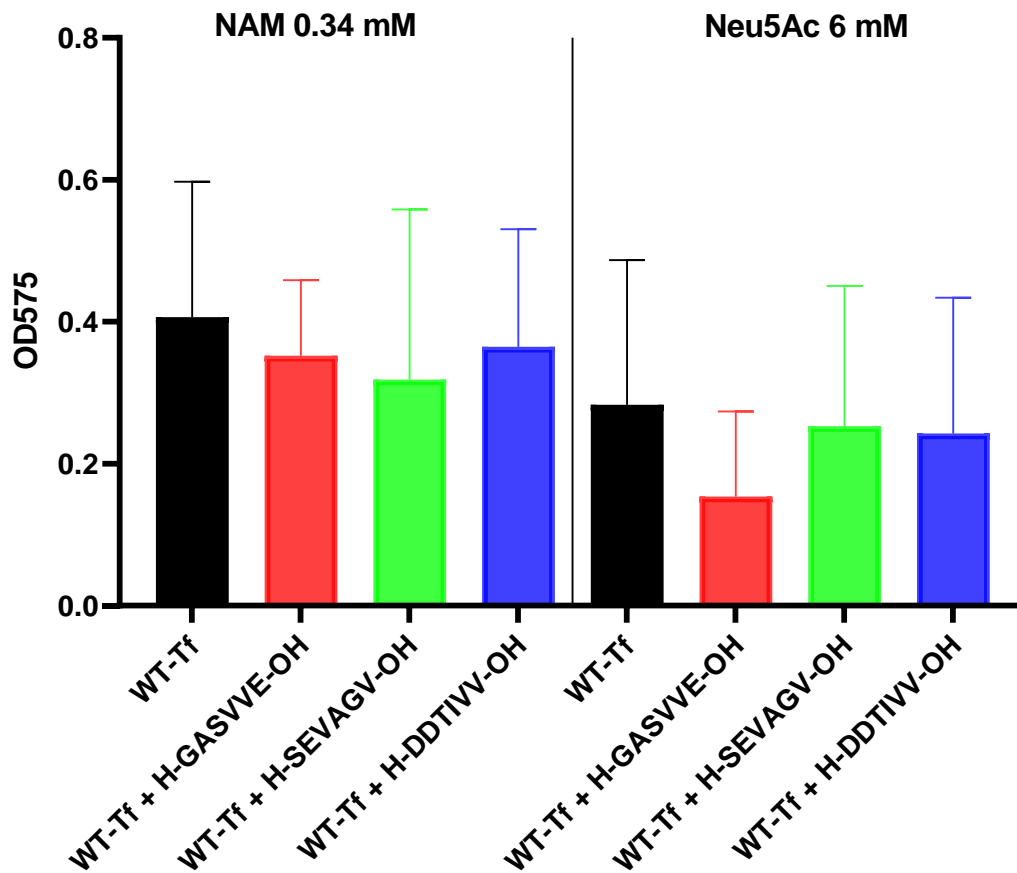
Figure 4.9 Growth of *T. forsythia* on the TonB box peptide.
       This figure shows the planktonic growth of *T. forsythia* on three different TonB box peptides. Defined TSB medium with either NAM or Neu5Ac containing 100 μg/ml of either peptide GASVVE, SEVAGV, or DDTIVE was used to test the planktonic growth. The inoculation of ATCC 43037 at OD_{600} of 0.05 was harvested from four days plate growth and washed three times in sterile PBS. The optical density of each culture was measured after seven days at OD 575 nm. Vertical bars represent the standard deviation of the means where experiments were conducted on three separate occasions in triplicate. In the presence of sialic acid, the effect of GASVVE peptide revealed a reduction of growth but not significant to parental strain and the other two peptides.

4.2.2.4. Effect of TonB box peptide on Biofilm Formation of *T. forsythia*

  Growth of *Tannerella* ATCC 43037 in the presence and absence of peptides in biofilm culture was assessed after seven days of incubation and across a range of Neu5Ac concentrations (3 mM, 6 mM, and 10 mM). The mean CFU/ml of GASVVE peptide (100 μg/ml) on 6 mM of Neu5Ac was approximately 2-fold lower than the parent strain and control peptides. On 3 mM and 10 mM of Neu5Ac, the effect of GASVVE peptide on biofilm growth was significantly different by 2.4-fold ($P<0.0001$) and 2.5-fold ($P<0.0002$) respectively compared to the parent strain along with SEVAGV and DDTIVV peptides. There was no detectable difference in the biofilm formation between both SEVAGV and DDTIVV control peptides compared to the parent strain. In the presence of NAM, the biofilm formation of *Tannerella* with GASVVE peptide showed slight, but not significant difference compared to the parent strain (Figure 4.10). This indicates a strong defect of growth biofilm on Neu5Ac to parent strain in the presence of the TonB box peptide.
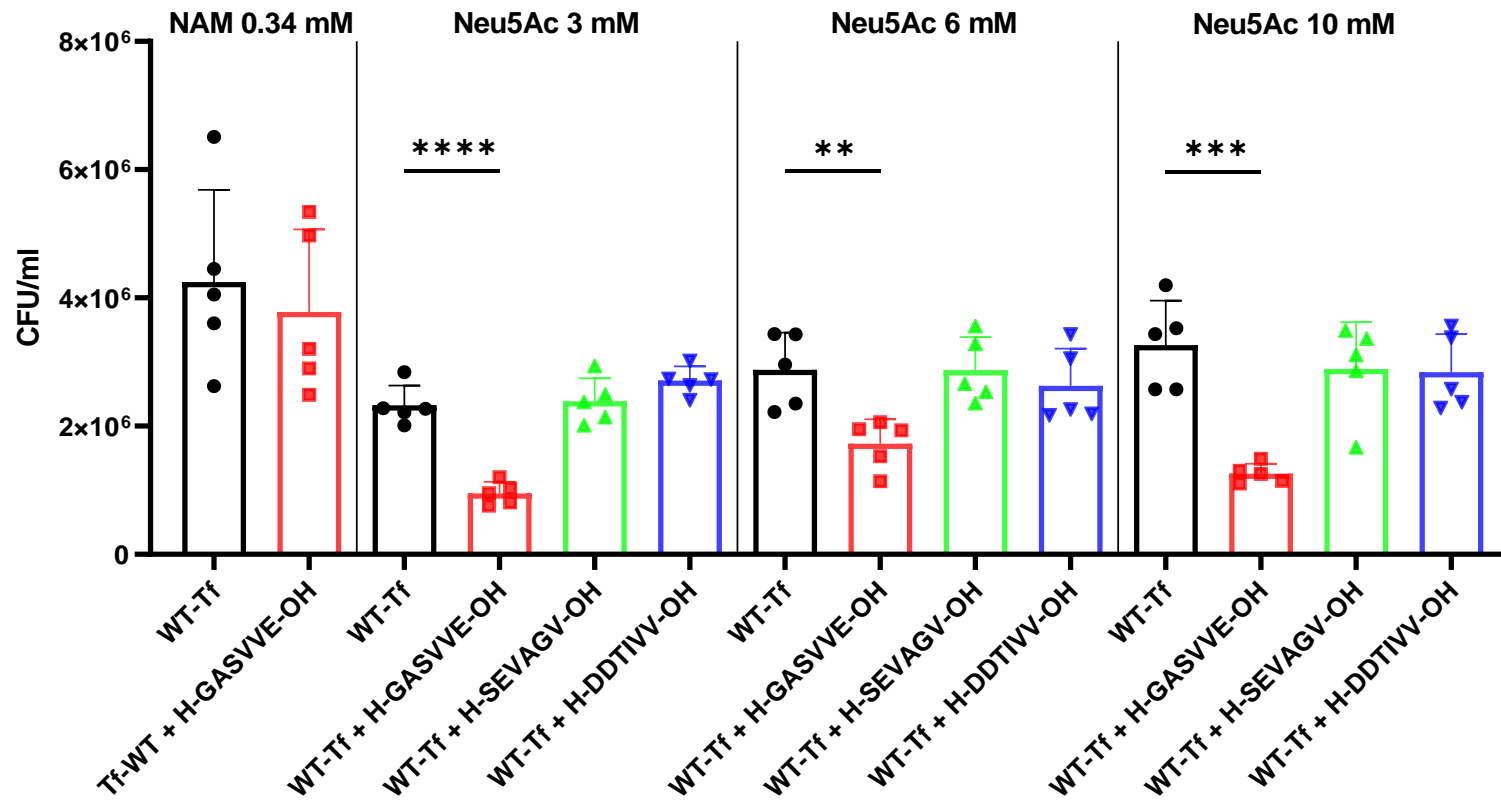
Figure 4.10 Effect of GASVVE peptide on biofilm formation in the presence of NAM or Neu5Ac by CFU counting.
Four-day-old plate of *T. forsythia* was harvested, washed three times with PBS, and adjusted to a cell density at OD$_{600}$ of 0.05. Biofilm formation was assessed by inoculating triplicate wells of a 96-well plate and harvesting the biofilm after seven days. These experiments used concentrations of 0.34 mM of NAM and 1.5 mM, 3 mM, and 6 mM of Neu5Ac. *T. forsythia* biofilms were supplemented with 100 μg/ml of either peptide, GASVVE, SEVAGV, or DDTIVV. After seven days of growth of growth, the 96-well plate was washed two times with PBS and the biofilm was harvested and assessed by cell counting. To eliminate bias, this study counted all 16 squares of Helber counting chamber (Hawksley). Means of three wells were calculated before applying the one-way ANOVA test, with $P < 0.05$ being taken as the level of significance (*); $P < 0.01$ being taken as the level of significance (**); $P < 0.001$ being taken as the level of significance (***). Enclosed shapes are data for every individual experiment.

4.2.2.5. Staining of Biofilm Growth on TonB Box Peptide

Having shown that there is a significant effect of GASVVE peptide on *Tannerella* biofilm growth by cell counting. We now tested the role of TonB box peptides in *Tannerella* biofilm formation using a crystal violet staining method. Due to lower effect of TonB box peptide on NAM, the biofilm formations were tested only in the presence of 3 mM, 6 mM, and 10 mM of Neu5Ac containing either GASVVE, SEVAGV, or DDTIVV peptide. The biofilm growth on 10 mM was reduced to 2-fold by GASVVE (P<0.0004) compared to the parent strain and the other two control peptides. In support of this observation, biofilm growths on Neu5Ac (3 mM and 6 mM) were reduced by 1.5-fold (P<0.0013) and 1.7-fold (P<0.0026) due to GASVVE, respectively (Figure 4.11).
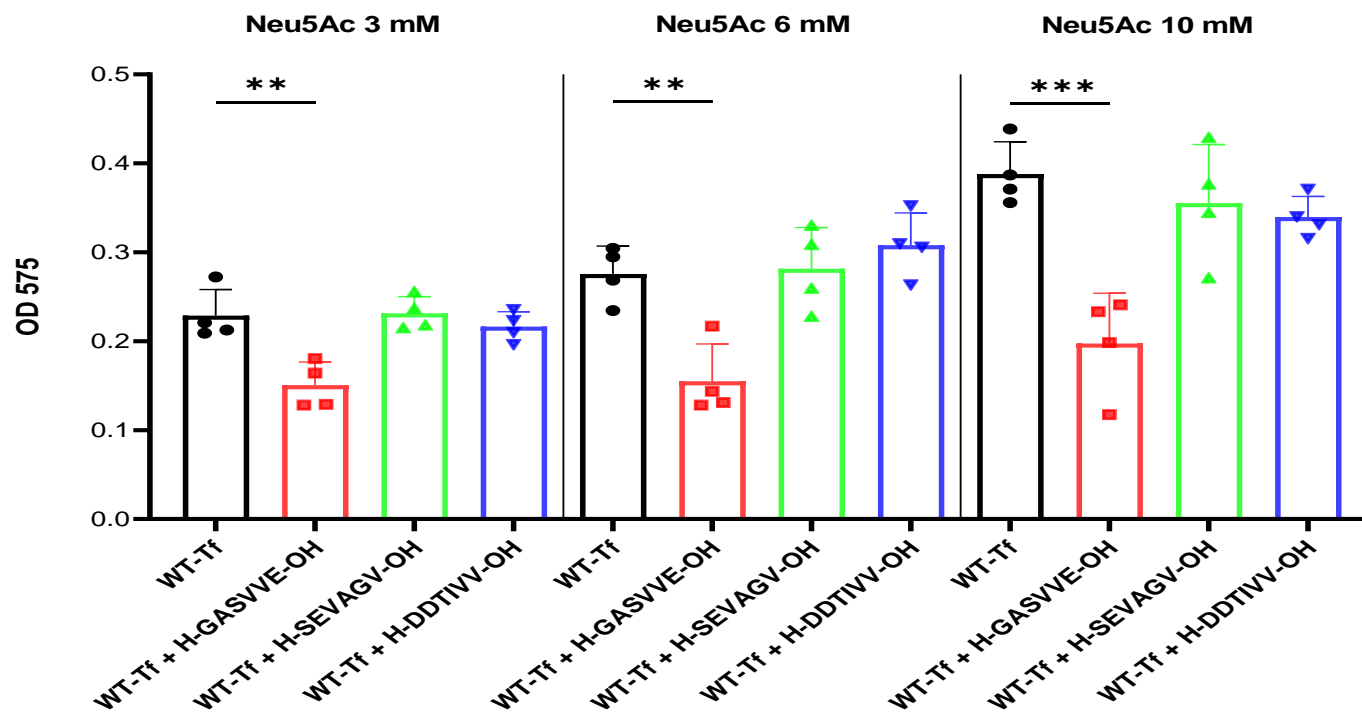
Figure 4.11 Effect of GASVVE peptide on biofilm formation in the presence of NAM or Neu5Ac by crystal violet staining.
Four-day-old plate of *T. forsythia* was harvested, washed three times with PBS, and adjusted to a cell density at OD_{600} of 0.1. *T. forsythia* biofilms were set up in triplicate wells supplemented with either 0.34 mM of NAM and 3 mM, 6 mM, and 10 mM of Neu5Ac of Neu5Ac (Total 200 µl). At the time of inoculation, 100 µg/ml from GASVVE, SEVAGV, or DDTIVV was added to the medium as indicated. After seven days of growth, the 96-well plate was washed with PBS and stained with a crystal violet before being washed and measured at OD 575 nm. Means of three wells were calculated before applying the one-way ANOVA test, with $P < 0.05$ being taken as the level of significance (*); $P < 0.01$ being taken as the level of significance (**); $P < 0.001$ being taken as the level of significance (***); $P < 0.0001$ being taken as the level of significance (****). Enclosed shapes are data for every individual experiment.

4.2.2.6. Effect of GASVVE Peptide on Sialic acid Uptake

As outlined above, the impact of TonB box peptides on sialic acid uptake was measured and evaluated. *Tannerella* wild-type and *Tannerella* wild-type with either peptide (GASVVE, SEVAGV, or DDTIVV) was incubated for three hours in the presence of 500 μM or 1000 μM of Neu5Ac. Next, the TBA assay was used to quantify sialic acid uptake in the presence or absence of each TonB box peptide. The GASVVE showed a significant effect on *Tannerella* uptake of Neu5Ac compared to the bacteria alone and other control peptides (SEVAGV and DDTIVV). In the presence of 500 μM Neu5Ac containing GASVVE (100 μg/ml), the uptake was reduced from 130 μM to 75.5 μM ($P < 0.05$). In the presence of 1000 μM containing GASVVE (100 μg/ml), the uptake of Neu5Ac was reduced from 186 μM to 62 μM ($P < 0.05$) (Figure 4.12).
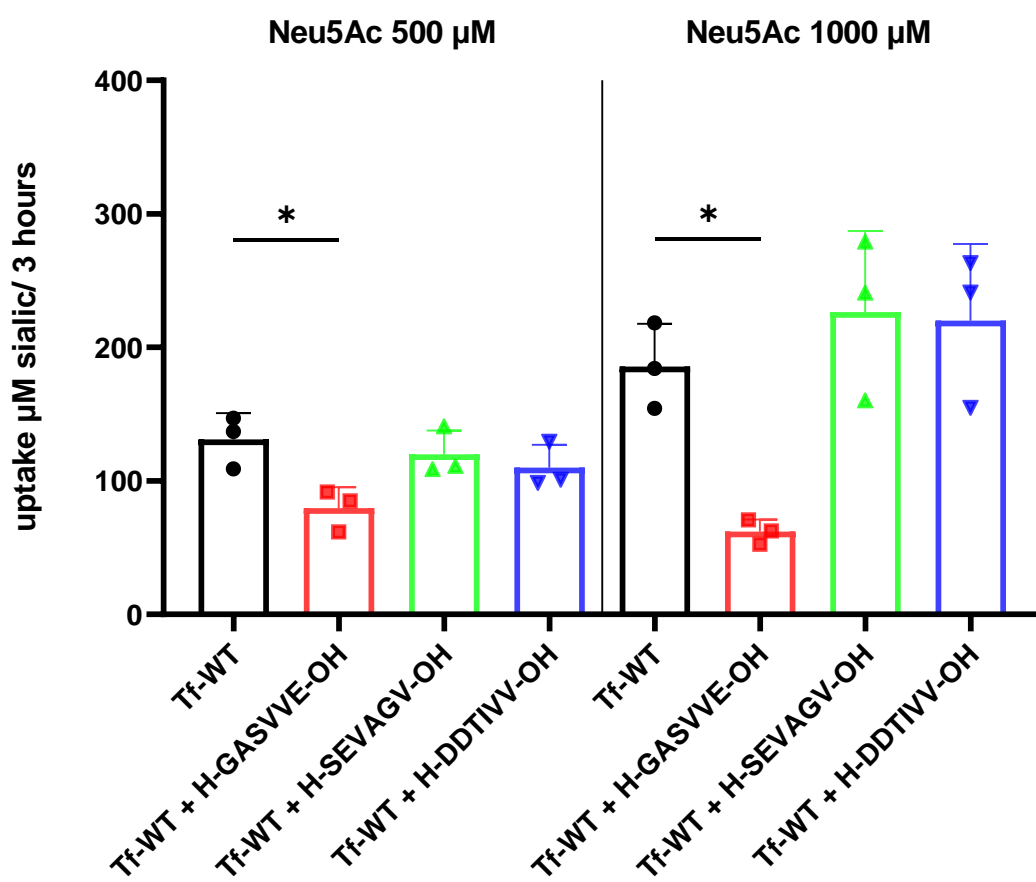
Figure 4.12 Effect of GASVVE peptide on sialic acid uptake by TBA assay.

Four-day-old plate of *T. forsythia* was harvested, washed three times with PBS, and adjusted to a cell density at OD_{600} of 2. The Neu5Ac uptake was set up in triplicate supplemented with either 500 μM or 1000 μM concentration (Total 100 μl). At the time of inoculation, 100 μg/ml from GASVVE, SEVAGV, or DDTIVV was added to the assay. After three hours, the remaining Neu5Ac in the medium was going under the TBA assay before reading at OD 549 nm. Means of three wells were calculated before applying the one-way ANOVA test, with $P < 0.05$ being taken as the level of significance (*). Enclosed shapes are data for every individual experiment.

4.2.2.7. Effect of GASVVE Peptide on Alternative Sialic Acid Source

Previous experiment of introducing the inhibitor peptide to *Tannerella* growth on sialic acid showed a significant growth defect compared to the parent strain. Several studies have shown that the adherence and internalisation of *T. forsythia* to host cells is mediated by several extracellular matrix glycoprotein. Next, the effect of this peptide was tested for *Tannerella* biofilm on the heavily sialylated salivary glycoprotein mucin. Equal growth of *Tannerella* in the presence of mucin only or mucin and an inhibitor peptide showed a difference in biofilm formation. The average bacteria count of GASVVE peptide on 60 μM of mucin was approximately decreased by two-fold compared to untreated. A biofilm growth reduction on 15 μM and 30 μM of mucin (one to two-fold and one to two-fold, respectively) was observed with an GASVVE peptide compared to untreated (Figure 4.13). Overall, these data indicate that the peptide targeting TonB-box-TonB interactions may have potential to inhibit growth, but also that NanO inhibition is an effective strategy.
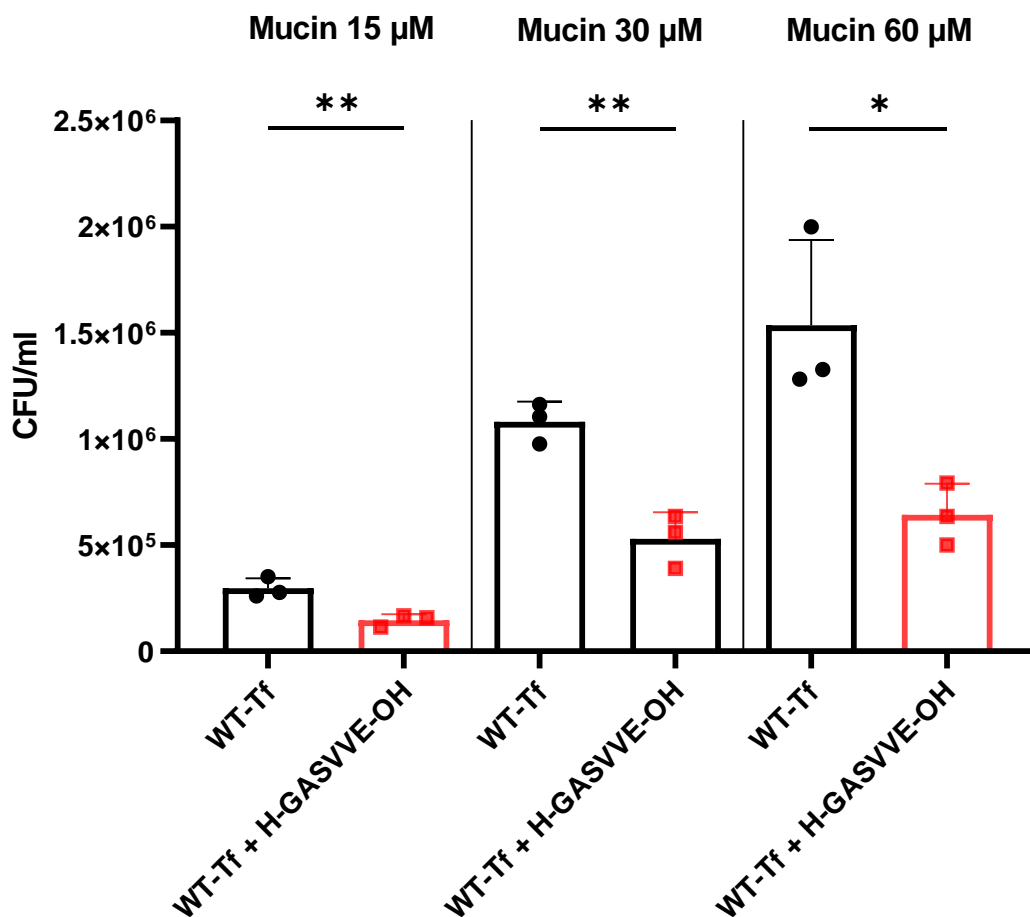
Figure 4.13 Effect of GASVVE peptide on biofilm formation of *T. forsythia* on alternative sialic acid source (mucin).

The ability of only GASVVE peptide effect on alternative sialic acid sources (mucin) was tested by inoculating triplicate wells of a 96-well plate and harvesting the biofilm after seven days. These experiments used concentrations were 15 μM, 30 μM, and 60 μM of mucin. After seven days, the 96-well plate was washed two times with PBS and the biofilm was harvested and assessed by cell counting. To eliminate bias, this study counted all 16 squares of Helber counting chamber (Hawksley). Means of three wells were calculated before valuated by *t* test, with $P < 0.05$ being taken as the level of significance (*); $P < 0.01$ being taken as the level of significance (**). Enclosed shapes are data for every individual experiment.

4.3 Discussion

The human microbiota provides the host with many metabolic functions. Degradation and fermentation of complex carbohydrates and glycan are examples of metabolic functions that are essential to a host's survival. In the bacterial phyla, *Bacteroidetes* are often characterised by a unique metabolism mediated through polysaccharide utilisation loci (PUL). PULs consist of two essential proteins, TonB-dependent transporter (TBDT/SusC) and glycan-capturing lipoproteins (SusD). More broadly, within the Gram-negative bacteria, it was established that the SusC is involved in the uptake of sugars, iron, and vitamins.[112] Subsequent literature exposed the functions of SusD for the uptake of glycans and other nutrients.[262] Although some TBDTs have not been experimentally validated, the use of Pfam characterisation to annotate and predict TBDTs was validated in two other studies.[262,271] One of which is the structural domains of TBDTs within *B. thetaiotaomicron*, which was annotated and characterised by using the Pfam database. As a result, the Pfam outcomes for *B. thetaiotaomicron* analysis were reliably compared to the annotations that were conducted by the CAZy database and compiled within the Polysaccharide-Utilisation Loci Database (PULDB).[262] Here, this study benefited from the Pfam database to conclude and characterise the TBDTs in *Tannerella* species. This analysis of TBDTs can also lead to a better understanding about the broader structure and potential role of these TBDTs in *Tannerella* species.

Classical TBDTs with the essential barrel and plug-domains were found in genomes as single TBDTs or as part of a PUL. However, in some cases in *Tannerella* genomes, the Pfam barrel domain without the Pfam plug domain was observed within this analysis. While biologically this is unlikely, it could be that a loss of genomic content or a miscalled start site might be the reason for the presence of only Pfam barrel domain.[262] An additional possible reason is that this protein might function as another type of outer membrane transporter.[275,276] This illustrates a limitation of *in silico* analysis where additional analysis of explaining the presence of only barrel domain was not possible, which resulted in the exclusion of the classical TBDT without the Pfam plug domain. Similar scenario of exclusion was noticed among the characterisation of TBDTs from *B. thetaiotaomicron*.[262] Furthermore, the presence of classical TBDTs, with only the two key domains, indicates that additional domains are not required for proper function of TBDTs and interaction with TonB system, mainly TonB protein, or other additional homologous.[112,262] In addition, the number of predicted amino acids can be a hint to predict similar transport of substances. Two studies were able to conclude that some single TBDTs (without SusD) were found to transport iron, thiamine, and vitamins.[277,278] In *B. thetaiotaomicron*, the classical TBDTs without SusD were found to be

less than 800 amino acids, as well as their previous predictions and further lab experiments suggested that this range of amino acids could show similar transport structures to iron, thiamine, and B_{12}.[262,278] Here, this study found classical TBDTs not associated with surface lipoprotein (SusD) were less than 1000 amino acids except for the WW 10960 strain (Table 4.1). This may suggest that the classical TBDTs not associated with surface lipoprotein (SusD) may fall in the same transport structures to iron, thiamine, B_{12}, or other small molecules. Additional studies found that the TBDTs with surface lipoprotein (SusD) can play a role in different substances. The classical TBDT with a surface SusD was found to transit starch in *B. thetaiotaomicron* and transferrin iron in *Neisseria*.[121,123,262] This study found that the classical TBDT of *Tannerella* genomes is more likely presence in a single structure without the surface lipoprotein (SusD).

A recent report about the classical TBDTs associated with the surface lipoprotein (SusD) suggested that they may form a dimer structure. This report of dimer formation was noticed in *P. gingivalis* (RagAB) and *B. thetaiotaomicron* (SusCD) including crystal structures and size-exclusion experiments.[262] In contrast, all classical TBDTs among *Tannerella* genomes did not show Pfam homologous to classical TBDT dimer associated with PUL/SusD. This agrees with the previous conclusion about the characterisation of TBDTs from Gram-negative bacteria where a dimeric complex is unusual among classical TBDTs.[279]

The DUF4480 domain (carboxypeptidase D regulatory-like domain) was recognised previously in a study by Glenwright that presented the X-ray crystal structures of the SusC/D complexes, but the function of this domain was not known.[275] A recent investigation in *B. thetaiotaomicron* removed the N-terminal extension from a TBDT (BT1763). As a result, this domain clarified that the N-terminal extension is the essential for growing of *B. thetaiotaomicron* on the cognate substrate Levan.[280] In *Tannerella* genomes, the presence of 40 or more of N-Terminal extension TBDTs was noticed where the most studied strain (ATCC 43037) was predicted to include 46 of this domain. In contrast, the strain BU0063, non-pathogenic *Tannerella* human isolate, showed only 24 of N-Terminal extension TBDTs. This might suggest that the N-Terminal extension TBDTs transport different metabolic functions, leading to a biological interest of further characterising the typical substrates among pathogenic *Tannerella* genomes within this domain.

The N-Terminal extension TBDT is found in transporters associated with and without PUL of *Tannerella* species. In *B. thetaiotaomicron*, it was noticed that the N-Terminal extension TBDT without SusD can transport ferric iron and B_{12}, whereas the N-Terminal

extension TBDT with SusD can transport polysaccharides and other host-glycan.[262] Furthermore, in *B. thetaiotaomicron*, the prediction of amino acid long of the N-Terminal extension TBDT without SusD was between 700 and 953 base pairs compared to the N-Terminal extension of TBDT with SusD (938 – 1120).[262] Here, we noticed that the range of 672 to 1154 amino acids and 984 to 1261 amino acids between this domain without SusD and this domain with SusD, respectively. The range of amino acids length of this domain from *Tannerella* genomes was quite similar to other amino acids long of *B. thetaiotaomicron*, predicating a similar transport structure. Interestingly, a novel of different functions based on unusual amino acid long (< 700) is predicted within this analysis for BFO_0495 and SCQ25174.1 from 92A.2 and UB22, respectively.

The STN domain was found in this study to be N-terminal to carboxypeptidase D regulatory-like domain. Likewise, the STN domain in *B. thetaiotaomicron* was found to be N-terminal to carboxypeptidase D regulatory-like domain.[262] The term of Signal Transduction TBDT came from the concerted interactions of signalling between TBDTs and extra cytoplasmic function sigma factors and anti-sigma factors in *B. thetaiotaomicron*. This signalling can differ and switch due to the mediated *O*-glycan degradation as well as capsular polysaccharide synthesis.[270] This domain was also found and characterised outside *Bacteroidetes*, but the STN domain has not always been noticed in the N-terminal to carboxypeptidase D regulatory-like domain. From other organisms, this domain was seen in *E. coli* (FecA), *P. aeruginosa* (FoxA), and *S. marcescens* (HasR). The structure of this domain was investigated to show a small globular domain of two α-helices and five β-sheets.[161,281,282] Further investigation of FoxA from *P. aeruginosa* showed no impact on FoxA-TonB binding upon removing this domain, indicating a need to confirm the conservation of this structure, specifically, in *Bacteroidetes*.[161]

The STN domain within this study was found with and without the surface lipoprotein. Interestingly, this domain was absent from the BU0063 strain, suggesting the presence of this domain as a metabolic enzyme activity can initially differentiate pathogenicity of *Tannerella* species. All human *Tannerella* isolates showed the presence of some or all of this domain as a PUL. The presence of this domain was also noticed in *B. thetaiotaomicron* as a PUL.[262] For the first time in *Bacteroidetes*, we noticed the presence of STN domain was not associated with the PUL structure in three different *Tannerella* strains, which may correspond to a novel functioning (Table 4.1).

The presence of the STN domain seems limited compared to the presence of other TBDT domains. The frequency range of this domain was between three to seven times

among *Tannerella* species. Another investigation for this domain found more than 17 STN with a single strain of *B. thetaiotaomicron*.[262] Characterisation of STN domain concluded a favour to target the host glycans, such as sulfated host glycans, complex N-glycans, *O*-glycans and glycosaminoglycans, and other polysaccharides.[283–286] These characterisations encourage the need to investigate more about the STN domain within the *Tannerella* genomes. This function may not only reveal additional supportive substances for *Tannerella* biofilm formation, but also may present some differential characteristics between pathogenicity of *Tannerella* species.

Unlike the TBDT structure, the presence of SusD protein was first characterised in *B. thetaiotaomicron* as an essential part of binding starch.[123] The recent structure of SusC/D transporter showed that the SusD protein sits like a lid over the SusC transporter.[280] Further investigation of SusD structure found that glycans bind both host and dietary polysaccharides.[280] However, the function of SusD to recognise glycans is still unknown as its recent defined structure of SusD showed no obvious precise feature in its functional domains.[262] Thus, the SusD proteins are reported to have eight tetratricopeptide repeat (TPR) forming a right-handed superhelix that frames the rest of the structure.[287] Therefore, it suggested that the SusD protein contains a ligand-binding face where it resides opposite of the TPR domains.[262] The typical size of SusD protein within this study range between 390 to 700 amino acids long, which is in a similar range to the previous indication about *Bacteroidetes* Sus-like systems.[287]

This study found 357 SusD proteins that are encoded as part of a TBDT-susD pair. We noticed many of SusD sequencing that were falling in a combined SusD_RagB and SusD_like_3 families. Likewise, an assessment of SusD sequences from *B. thetaiotaomicron* showed that many of the SusD sequencing falling on the same two families.[262] Regarding a single family, the SusD_like_2 family was the only common family among *Tannerella* genomes as well as the only family encoded more SusD sequences. Unlike *Tannerella*, the majority of SusD sequences from *B. thetaiotaomicron* were falling in the SusD_like family.[262]

With this in mind, there are some questions to be answered about the TBDTs. The TBDTs are unique with individual discrete domains, but it is unknown if the size range of SusD protein influences interactions with its cognate TBDT. How glycoprotein attachment is facilitated between the TBDT plug domain and SusD protein. In addition, it is undetermined whether SusD proteins can pair with other TBDTs within the same or different domains.

Prediction of TBDT/PUL favour to substrate was not highly reliable as several TBDT/PUL can favour the same substrate.

In this section and for the first time, we have bioinformatically predicted presence of TBDTs among *Tannerella* genomes. We have noticed some differences in the presence of TBDTs among human pathogenic and non-pathogenic *Tannerella* isolates. The presence of SusD and PUL structures were defined and characterised within this study. In addition, we note that the Signal Transduction domain and some SusD protein families can further differentiate between *Tannerella* pathogenic and non-pathogenic isolates. This analysis concludes novel domains of the reported TBDTs, which they encourage for further work to differentiate types of TBDT rearrangements.

Next, on common with several other bacteria, *T. forsythia* contains multiple *tonB* encoding genes; our data, in chapter III, showed that the *tonB BFO_0233* can deliver energy in the Neu5Ac uptake of *T. forsythia* during biofilm growth. We then hypothesised to design small peptides to target TonB energy through TonB box under conditions favourable to *T. forsythia* growth. There is no universally-accepted theory of the mode of action of the antimicrobial peptides.[7] Yet, the use of amino acid peptide to target the TonB box was suggested in a previous study with efficient inhibition of iron transport, where they treated TonB box of *E. coli* with the pentapeptide Glu-Thr-Val-Ile-Val to conclude the inhibition of TonB-Dependent Transporter.[269] Thus, the use of pentapeptide to target the TonB box seems valuable, which may inhibit the sialic acid transit. As the TonB box linked the TBDTs (Literature chapter-section 1.10), the greatest homology to the N- terminus and C-terminus could determine the TonB box. Initial identification of TonB box homologues was in *P. gingivalis* where alignment of RagA regions to SusC (*B. thetaiotamicron*) showed greatest homology on N-terminus (in the first 300 residues) and on the C-terminus of the protein (30 residues).[274] This suggested a relationship between RagA and other TonB-linked receptors in Gram-negative bacteria. A multiple alignment of TonB box regions of the RagA to SusC, FyuA (*Y. enterocolitica*), HemR (*P. gingivalis*), FepA (*E. coli*), IrgA (*V. cholera*), CirA (*E. coli*), BfrA (*B. pertussis*), FepA (*P. aeruginosa*), Tla (*P. gingivalis*), and BfeA (*B. bronchiseptica*) showed similarity values between their homologues. Based on the distinctive clade among the TonB-linked outer membrane proteins RagA (*P. gingivalis*) and SusC (*B. thetaiotaomicron*), biochemical evidence showed the following TonB box GANVVV for *P. gingivalis* and *B. thetaiotamicron*.[274] In *T. forsythia*, the way in which the amino acid alignment of *nanO* and *TF0033* compared to the TonB box in *P. gingivalis* and *B. thetaiotaomicron* predicted the *Tannerella* NanO TonB box, GASVVE. Thus, we aimed to

inhibit the biofilm growth of *T. forsythia* using the 'GASVVE' peptide, a scrambled version, and the *E. coli* FepA TonB box peptide 'DDTIVE' as controls.

Experimental evidence from planktonic growth, biofilm counting cells, biofilm staining, and TBA assays indicated for the first time not only the presence of *Tannerella* TonB box, but also the effect of the designed peptide inhibitor. Examination of peptide inhibitor *in vitro* affected *Tannerella* fitness, hindering its survival, and showing a clear attenuated phenotype. The attenuation of *Tannerella* in the presence of sialic acid and peptide inhibitor was evident *in vitro*. This was observed in the survival of *Tannerella* in the presence of both NAM and peptide inhibitor, in which the peptide did not significantly affect the growth. Furthermore, the presence of scrambled version and *E. coli* TonB box peptides showed no effect on the growth of *Tannerella* in the presence of Neu5Ac. Sequence of NanO TonB box was aligned with other transporter families leading to target this porin with a small peptide, which is consistent with pervious work of TonB box in the *E. coli* FhuE. In *E. coli*, the small amino acid peptide for *E. coli* FhuE receptor did influence the transit of iron as well as protect the bacterium from the invasion of phage and colicins B and Ia. Being impaired in sialic acid uptake, NanO was shown to interact with TonB BFO_0233, and this physical interaction can be disrupted with a synthetic peptide containing amino acids derived from the TonB box.

Furthermore, the effect of a short peptide from TonB boxes of three TBDTs binding to the TonB protein was assessed in *E. coli*. The short peptides for TonB boxes of BtuB (PDTLVVTANR), FhuA (EDTITVTAAP), and FepA (DDTIVVTAAE) were examined with a sensitive technique by the NMR spectroscopy to observe the chemical environment change with a backbone amide for each peptide. These peptides exhibited discernible chemical shift and affected the binding of TBDT and TonB protein.[155] Aside from the TonB box peptides, the function of TonB box was tested, inhibited, and confirmed through the mutation of one or more residues of TonB box. Previous studies showed impaired transport, less specificity of TonB box, and alteration through TonB box by promoting its unfolding.[176,288,289] A substitution of several positions of the TonB box amino acid using L8P or V10P mutations was performed to test the binding of TonB protein. Using fluorescence anisotropy, this study showed that the introduction of L8P or V10P mutation results in no high-affinity and a failure of BtuB to bind the TonB protein.[158] The importance of the TonB box was further explored in *Sphingobium* sp. SYK-6. The DdvT TBDT of this bacterium contains a conserved sequence of TonB box and is essential for 5,5′-dehydrodivanillate (DDVA) uptakes. Alanine mutations

were introduced to the highly conserved V42 and T43 residues of this TonB box of DdvT TBDT resulting in abolishing the uptake of DDVA.[290]

In the last decade, antibiotic therapy is the current treatment for many human infections. Growing resistance to the antibiotic therapy generates a pressing need against bacterial biofilm growth. The possibility of developing antibacterial strategies were examined, initially in siderophores-iron utilisation, and it was decided there was a need to target siderophores uptake where some studies targeted the siderophores biosynthesis.[139,291] In many pathogens, there are several iron acquisition systems and targeting these biosynthesises are exploited to the possibility of developing antibacterial inhibitors due to the problematic and multiplicity of these biosynthesis systems.[292] Rather than targeting multiple biosynthesis systems as the case in iron, it was suggested to attack the common players in the iron acquisition systems (i.e., TonB). In uropathogenic *E. coli*, a wide variety of iron acquisition systems is expressed in the urinary tract due to extremely low iron availability. A screen of 149,243 compounds was carried out to identify iron chelators and compounds to inhibit bacterium growth rather than inhibiting iron acquisition. Under iron-limiting conditions, there were only 16 compounds with abilities to arrest growth of this bacterium in a strain with the presence of TonB. These 16 compounds were bacteriostatic and did not affect proton motive force.[293] In *A. baumannii*, a bacterium can cause many human infections, further assessment for existing chemicals in larger compound libraries showed two chemicals: ebselen and ST0082990. The effect of these two chemicals on *A. baumannii* were due to blocking TonB-dependent transport via TonB-ExbBD.[293]

Given the increasing antibiotic resistances in Gram-negative bacteria, developing antibacterial strategies targeting TonB showed a difference in arresting growth of the uropathogenic *E. coli* and the bacteremia *A. baumannii*.[293,294] Of these strategies, the designed amino acid peptide inhibitor of TonB box in this study demonstrated a promising inhibition against a member of the oral biofilm pathogens as an attractive strategy since sialic acid is essential for colonisation and scarce in the oral biofilm.

Taken together, it was reported that the interaction between TonB protein and TonB box are mostly hydrophobic due to the highly conserved Valine (Val) on the TonB box.[159] The salt bridge also between the TonB box and TonB protein is driven by the presence of negative and positive charge of residues, respectively. The negative charge residues on the TonB box were investigated to conclude Glu-E in the TonB box of FhuA and Asp-D in the TonB box of BtuB.[176] Further residues mutation for (D-L-V) in the TonB box of FecA, (D-V-V) in the TonB box of the BtuB, and (V-T) in the TonB box of the DdvT in *Sphingobium* sp.

SYK-6 abolished the uptake of these TBDTs.[155,158,290] A previous designed peptide (ETVIV) abolished the uptake of iron in *E. coli*, whereas the designed TonB box peptide in this study (GASVVE) inhibited the uptake of Neu5Ac in *T. forsythia*. Previous alignments of TonB boxes with the highest similarity scores compared to the TonB box of RagA of *P. gingivalis* highlighted high conservation of three residues, namely, gly-G, asp-D, thr-T, and val-V (Figure 4.5-this chapter). Furthermore, we noticed a conserved presence of glu-E, asp-D, and val-V residues among several TonB boxes that were laboratory tested,[155,158,290] indicating a huge potential for designing a selective TonB box peptide by combining these residues targeting multiple TBDTs uptakes.

In conclusion, the designed TonB box peptide of this study required future evaluation and assessments. This peptide should be tested again with all *tonB* mutants that were generated in this study and whether we should observe any further change to the Neu5Ac uptake and overall genotype of these mutants. Besides that, this will clarify more about the role of this TonB box peptide whether it can target more than one TBDT in *T. forsythia* in case of testing GASVVE with either *nanO* or *BFO_0233* mutant (Chapter III in this study). *B. fragilis*, for instance, has a sialic acid NanO and NanU that was previously established, suggesting the possibility of testing the GASVVE peptide on this bacterium, as well as whether this peptide may inhibit the uptake of Neu5Ac in *B. fragilis* and increase its selectivity. In addition, we realised an identical TonB box sequence between *Tannerella* NanO and *B. thetaiotaomicron* and *P. gingivalis*, suggesting a huge possibility of similar results to this study upon testing GASSVE peptide on these bacteria (Figure 4.5-this chapter). Future bioinformatic alignment of TBDTs in *P. gingivalis* and *T. denticola* may shed light on a new function of this peptide in oral biofilm and increase its selectivity.

Chapter V

.

Bioinformatic analysis of sialic acid utilisation genes in bacteria

5.1. Introduction

The number of human pathogens with the ability to use sialic acid as a growth factor or sole carbon source are increasing and being uncovered, inhabiting a range of biological niches.[84] Sialic acid is found on the surface of gangliosides, glycoprotein, and sphingolipids and its role in biology is well known. *T. forsythia* can transport and form a biofilm on Neu5Ac to support its growth and produce cell wall peptidoglycans substituting its requirement for *N*-acetyl muramic acid (NAM). *T. forsythia* is dependent on a large *nan* gene cluster (16kb) consisting of nine genes. This cluster consists of two established genes for sialic acid catabolism *(nanA, nanE)*, a transporter gene to move small molecules across this bacterium inner cytoplasmic membrane (*mfs/nanT*), two outer membrane genes for sialic acid uptake (*nanO-nanU*), three auxiliary genes that play a role in scavenging sialic acid from the environment (*nanH, hexA,* and *nanS*), and a mutarotase (*nanM*) to support the efficient use of α-*N*-acetylneuraminate (Figure 5.1). Although there is a diversity in mechanisms of sialic acid transport from the extracellular to intracellular environments, transport of sialic acid across the outer membrane of Gram-negative bacteria was reported in *E. coli* via the NanC porin and in *T. forsythia* via NanO (Chapter III-this study) and NanOU.[202,214] In contrast, the transport of sialic acid across the inner membrane detailed four families of transporters in bacteria (Figure 5.2). First, the Major Facilitator permeases and is known as MFS-NanT (for *N*-acetylneuraminic transporter) when it was recognised in *E. coli* transporting the Neu5Ac as the sole source of carbon.[295] Second, the tripartite ATP-independent periplasmic (TRAP) and was recognised in the human pathogen *H. influenzae*.[296] This transporter consists of one binding protein component (*siaP*) and two other membrane components (*siaQM*). Third, the ABC transporters and was identified in *Haemophilus ducreyi* encoding a 4 gene operon for sialic acid uptake, named *satABCD*.[297] Fourth, the Sodium Solute Symporter family (SSS) and was recognised in *Salmonella enterica* serovar Typhimurium.[298] Other sialidases and catabolism genes are likely to share the same functions across microbes.[203]

Figure 5.1. Sialic acid *nan operon* .
This figure illustrates some confirmed or predicated sialic acid gene clusters in the *nan* gene operons of the annotated organisms. Key: inner membrane transporter: *nanT* (major facilitator superfamily permease); TonB-dependent sialic acid transport (TBDTs): *nanO, nanU, nanC*; Catabolic genes: nanE (N-acetylmannosamine-6P epimerase), *nanA* (neuraminate lyase); Scavenger gene: *nanH*; other accessory genes: *nanS* (sialic acid 9-O-acetylesterase), *nanM* (sialic acid mutarotase), *nahA*/*hexA* (beta hexosaminidase), *estA* (sialyl transferase), *yhcH* putative (glycolyl sialic acid processing enzyme). This figure is gifted to this study by Prof. Graham Stafford.
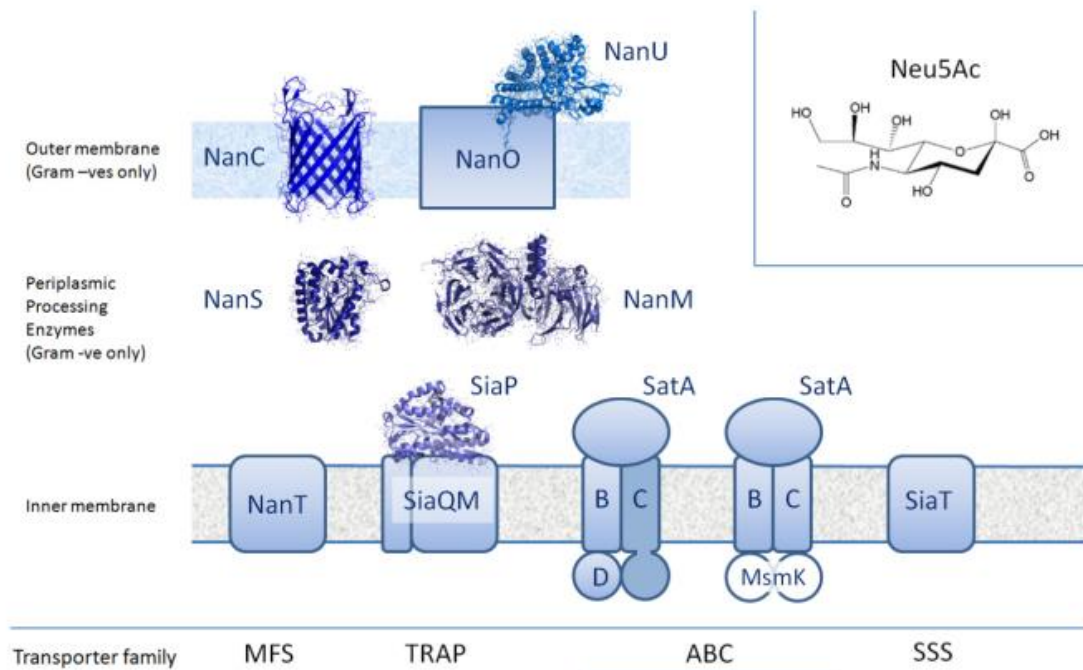
Figure 5.2 Families of sialic acid transporters.
This figure represents the established two outer membrane (NanC and NanO) and four inner membrane families (NanT, SiaPQM, ABS, and SSS) of sialic acid uptake. "Represent with permission from (Ref:296). Copyright (2016) eWhiterose."

The role of sialic acid was highlighted in several periodontal pathogens enabling their adhesion to human cells. Of which, *Fusobacterium nucleatum spp. polymorphum* contains only the sialylation part of the *nan* operon, but it was established to form synergistic biofilms via co-aggregation dependent mechanisms with *T. forsythia*.[203] Both members of the red complex *P. gingivalis* and *T. denticola* have at least a putative sialidase encoding gene, although the full *nan* operon is not present in their genomes.[203,299] Furthermore, the presence of only the catabolic and inner membrane transport genes not clustering with the sialic acid outer membrane genes was reported in other periodontal bacteria, namely *A. actinomycetemcomitans* and *F. nucleatum*.[203]

While the sialic acid *nan* operon was found to be essential for *T. forsythia* survival and biofilm formation in the absence of *N*-acetylmuramic acid (NAM), the understanding of the *nan* operon and its behaviour mechanism behind bacterium survival and interaction is therefore critical in the prevention of targeting the transport of sialic acid. This chapter aimed to further explore and predict the distribution and presence of sialic acid *nan* operon in other human and non-human bacterial species. This was achieved by extracting, filtering, and classifying all completed genomes on the National Center for Biotechnology Information (NCBI) and PULDB databases using the *nan* operons from *T. forsythia* and other reported bacteria as an input. With this in mind, previous identification and characterisation of genes and enzymes in *T. forsythia* were generally based on the lab ATCC 43037 strain from the American Type Culture Collection. However, there were many inconsistent results and sequence mismatches leading to a recent reinvestigation of this strain. The ATCC 43037 strain was found to be incorrectly labelled and deposited in the NCBI databases.[300,301] Thus, a further aim was projected to sequence the ATCC 43037 strain and account for the strain attribution error before continuing any additional laboratory work.

5.2. Materials and Methods

5.2.1. DNA sequencing

The extraction of genomic DNA was performed using a Promega Wizard Kit (section 2.3.1). The genomic DNA was then purified with AMPure XP following the kit's protocol (Beckman). Pipetting up and down the DNA product with beads was performed for 10 minutes before incubating the DNA product for five minutes on a magnetic stand. The DNA product was precipitated with 80% ethanol two times, each time for 30 second. A special grade water was used to resuspend the DNA genomic provided by the kit.

Next, the genomic DNA of *T. forsythia* was quantified using NanoDrop, with a minimum required concentration 10 ng/uL. This study was aware of overestimating DNA concentration by NanoDrop (sometimes up to 10-fold). Therefore, multiple extractions were performed accounting for such a problem. Next, we checked the DNA integrity by running the DNA genomic on a 0.75% agarose gel with a molecular weight ladder, accounting for DNA degradation.

5.2.2. Approach, tools, and databases

5.2.2.1. Searching for virulence factors

Predicted and tested virulence factor genes were retrieved from all reported *T. forsythia* literatures. While some of these genes were might be incorrectly annotated and confused between both strains of *Tannerella* (ATCC 43037 and 92A.2), a search of orthologs was used to apply comparative genome analysis between *Tannerella* strains. Next, Pfam identification and annotation were applied for *Tannerella* genomes before linking Pfam family code between genomes.[302] A manual filtration approach was applied to exclude repetitive genes or incorrect virulence factor gene.

5.2.2.2. Prediction of *nan* operons

This study was intended to identify bacterial species with similar *nan* operon to *T. forsythia*. Unlike previous conclusions about sialidases or sialic acid catabolism gene clusters, the *nan* operon in *T. forsythia* is unique in clustering inner membrane transporter, sialidases, and sialic acid catabolism genes within the outer membrane sialic acid gene (*nanO*). After exploring previous sialic acid literature, this search was determined by listing all bacteria meeting the inclusion criteria for this analysis: 1) The bacterium must contain one of the inner membrane transporter (*mfs-nanT*, *sss*, *ABC*, and *TRAP*) and one of the outer membrane transporter (*nanC*, *nanO*, *susC*, or *ragA*) of the *nan* operon along with either or both of the catabolism genes (*nanA*, *nanE*_NCBI search) (*unk*, *EPI*_PULDB search), and either or all the outer membrane supporting genes (*nanH*, *nanS*, *hexA*, and *nanM*- NCBI key words) (*GH33*, *GH20*, *EST*, *unk*- PULDB key words). 2) the inner and outer membrane of

sialic acid and their catabolism and auxiliary genes must be encoded within 30 ORFs from the start to the end of each cluster as it was concluded from the best bidirectional hits from the NCBI to *T. forsythia nan* operon.

5.2.2.3. Sequence retrieval and *nan* operon identification

The National Centre for Biotechnology Information (NCBI) and CAZy-Polysaccharide-Utilisation Loci (PULDB) databases were used to identify homologues for sialic acid *nan* operons in bacterial species.[303,304] Besides *T. forsythia*, we searched the literature for more similar *nan* operon to *T. forsythia* to conclude reported bacteria with SSS inner membrane transporter, namely *Pseudoalteromonas translucida* and *Shewanella pealeana* ATCC 700345.[203,298] Furthermore, we only found a similar *nan* operon to *T. forsythia* with ABC transporter in the PULDB database, namely *Arachidicoccus sp.* 5GH13-10, *Pedobacter sp.* BS3 BS3-1, and *Sphingobacterium s*p. C459-1T.[304]

This analysis started with NanT and NanO sialic acid sequences from *T. forsythia* ATCC 43037, the SSS inner membrane and NanO transporters sequences from *P. translucida* and *S. pealeana* ATCC 700345, and ABC inner membrane and NanO transporters sequences from *Arachidicoccus sp.* 5GH13-10, *Pedobacter sp.* BS3 BS3-1, and *Sphingobacterium s*p. C459-1T as templates for the Nucleotide Blast (Blastn) and Protein BLAST (BlastP) to extract data from the NCBI database. We also expanded our structured search in the NCBI for more bacteria with *nan* operon using *T. forsythia* sialic acid catabolism genes (*nanE* and *nanA*). In addition to the NCBI, we searched the PULDB database using key words of *nan* operon, namely *susC* (predicted sialic acid outer membrane transporter), and MFS and ABC (predicted sialic acid inner membrane transporters).

The sialic acid inner membrane transporters were used to predict and locate all possible sialic acid outer membrane transporters. Likewise, the sialic acid outer membrane transporters were used to predict and locate all possible sialic acid inner membrane transporters. Next, an analysis of the chromosomal clusters was used to locate the catabolism pathway genes, as well as to locate all possible outer membrane auxiliary genes.

The rationale for searching both Blastn and Blastp was as the following: A) to avoid false positive results, the sialic acid TonB-dependent transporter (NanO) was annotated with SusC, RagA, and NanC. All these annotations for the sialic acid TonB-dependent transporter were retrieved, and they were *in silico* examined while searching for NanO homologues. Likewise, four families representing the inner membrane sialic acid transporters and all other reported annotation for the inner membrane transporters (SSS, ABC, and TRAP) were included. B) BLAST search tool for *T. forsythia* ATCC 43037 allowed us to avoid false

negatives due to low spans of DNA sequences between its *nan* operon. Similarly, *P. translucida* strain and *S. pealeana* ATCC 700345, with sodium solute symporter (SSS) transporter, were found to contain low spans of DNA sequences between their *nan* operons, as the case for all three bacteria with ABC transporters.

5.2.2.4. Annotation of homologous gene/protein

Pfam and UniProt were searched to annotate all possible protein domain structures and functional annotations for the literature reported bacteria used within this study.[302,305] The InterPro and Pfam databases were used to annotate non-annotated genes responsible for sialic acid feeding pathways.[306] Misannotation of outcome sialic acid genes from the InterPro and Pfam databases was blasted using the BLAST algorithm with the default parameters to deliver the best bidirectional hits for orthologs.[307] In addition, we used NCBI alignment search tool to calculate the similarity between sequencing from this analysis. The NCBI Taxonomy database was used to clarify the taxonomic affiliations of the analysed genomes.[307]

5.2.2.5. Sequence alignment and phylogenetic tree

This study collected *nan* operon from each bacterium to carry out multiple alignments of amino acid sequences. ClustalW was used to align all orthologs and compile phylogenetic trees.[308] The phylogenetic trees were constructed with the default parameters using the maximum-likelihood method implemented in ClustalW. The obtained tree was visualised and edited using the Interactive Tree of Life (iTOL) version 6.[309]

5.3. Result

5.3.1. Genomes of *Tannerella* species

5.3.1.2. Genome Sequence of *T. forsythia* ATCC 43037 strain

In 1986, *T. forsythia* ATCC 43037 was deposited to the NCBI database after it was isolated by Tanner et al. Years later, an additional strain of *Tannerella* was isolated to be characterised and reported as *T. forsythia* FDC 92A.2 strain. This strain was sequenced and annotated by Los Alamos National Laboratory before it was made publicly available in 2005.[300] In 2013, the FDC 92A.2 strain was realised in the NCBI as the GenBank accession number CP003191. Unfortunately, the genome of FDC 92A2 strain was incorrectly attributed through NCBI to the type of strain ATCC 43037. Numerous cases of finding mismatches between both genomes were reported, exploring the confusion between both genomes. In 2015, Friedrich and others performed shotgun sequencing of ATCC 43037, but it was a draft genome assembly and not fully available.[215] Table 5.1 presents the list of 2022 NCBI *Tannerella* strains and their genomes assemblies. Today, there are 13 strains of *Tannerella* organisms, two of which are only non-human isolates. The other 11 human isolate strains were reported as a pathogenic microbe, except for one putative periodontal health associated isolate *T. serpentiformis* (BU0063). In 2019, we first decided to re-sequence *T. forsythia* ATCC 43037 strain and account for mismatching between both genomes, specifically, this confusion has persisted in other databases.

We sent our DNA genomic to MicrobesNG, UK where the *T. forsythia* ATCC 43037 was sequenced using Nanopore and Illumina. The advantage of this service was at a minimum coverage of 30x and a very competitive rate to generate whole genomes using 2x250bp paired end reads. The reference genome for this service was chosen using Kraken (Taxonomic Sequence Classification system) along with constructed ATCC 43037 reference that was assembled by Prof. Graham Stafford. The reads were mapped using BWA (Burrows-Wheeler Aligner) to assess the quality of the data. The performed sequence was subsequently error corrected including assembly, annotation, and variant calling. The N50 length was increased after scaffolding and gap filling, resulting in a contig that covered 99% of the whole genome with over 3.28 Mbp. The total sequence length of the assembly was 3,287,589 bp containing three assembled contigs (3.28 Mbp, 1425 bp, 261 bp). The new assembly ATCC 43037 strain results in a total of 2,755 genes, 46 tRNA and 1 tmRNA.

Table 5.1. *T. forsythia* genome assembly and annotation report

Strain name	Isolate	Assembly	Genome size (Mbp)	GC	Number of Scaffolds	Release date
Sheffield ATCC 43037	Human	-	3.28	47.10	1	Not yet but it was sequenced in March, 2019.
NCBI- ATCC 43037	Human	GCA_001006485.1	3.28	47.10	141	13-May-2015
92A.2	Human	GCA_000238215.1	3.41	47	1	15-Dec-2011
3313	Human	GCA_001547875.1	3.35	47.10	1	11-Apr-2015
9610	Human	GCA_001938785.1	3.20	47.30	79	06-Jan-2017
KS16	Human	GCA_001547855.1	3.39	47.20	1	11-Apr-2015
UB4	Human	GCA_900096725.1	3.23	47.20	71	01-Oct-2016
UB20	Human	GCA_900096735.1	3.25	47.10	93	01-Oct-2016
UB22	Human	GCA_900096715.1	3.27	47.10	98	01-Oct-2016
WW10960	Human	GCA_002529295.1	3.31	47.20	98	11-Oct-2017
WW11663	Human	GCA_002529085.1	3.30	47.10	140	11-Oct-2017
OH1426	Non-Human	GCA_003860135.1	3.46	47.70	290	03-Dec-2018
OH2617	Non-Human	GCA_003860025.1	3.32	47.80	226	03-Dec-2018
Tannerella serpentiformis W11667 (BU0063)	Human	GCA_001717525.2	2.9	56.50	1	31-Aug-2016

5.3.1.3. Comparative analysis of *T. forsythia* genome assemblies

The new sequence ATCC 43037 strain was used to perform whole-genome comparisons with other *Tannerella* assemblies. This comparison can show genomic structural differences and gene order conservation. This study chose the 92A.2 strain as a genome reference to compare both ATCC 43037 genome and other human *Tannerella* genomes. Only three strains (92A.2, 3313, and KS16) were found as one scaffold (i.e. complete vs non-contiguous). DNA fasta was prepared for the other genomes with multiple scaffolds, benefiting from the enhancement of bioinformatic tools. Blast Ring Image Generator (BRIG) was chosen to construct this comparison (Figure 5.3).[310] As a result, this alignment showed that the ATCC 43037 genome and other *T. forsythia* genomes have higher similarities compared to the reference genome (92A.2). In addition, there were significant differences between the putative periodontal health-associate isolate (*T. serpentiformis*; BU0063) and other pathogenic strains.
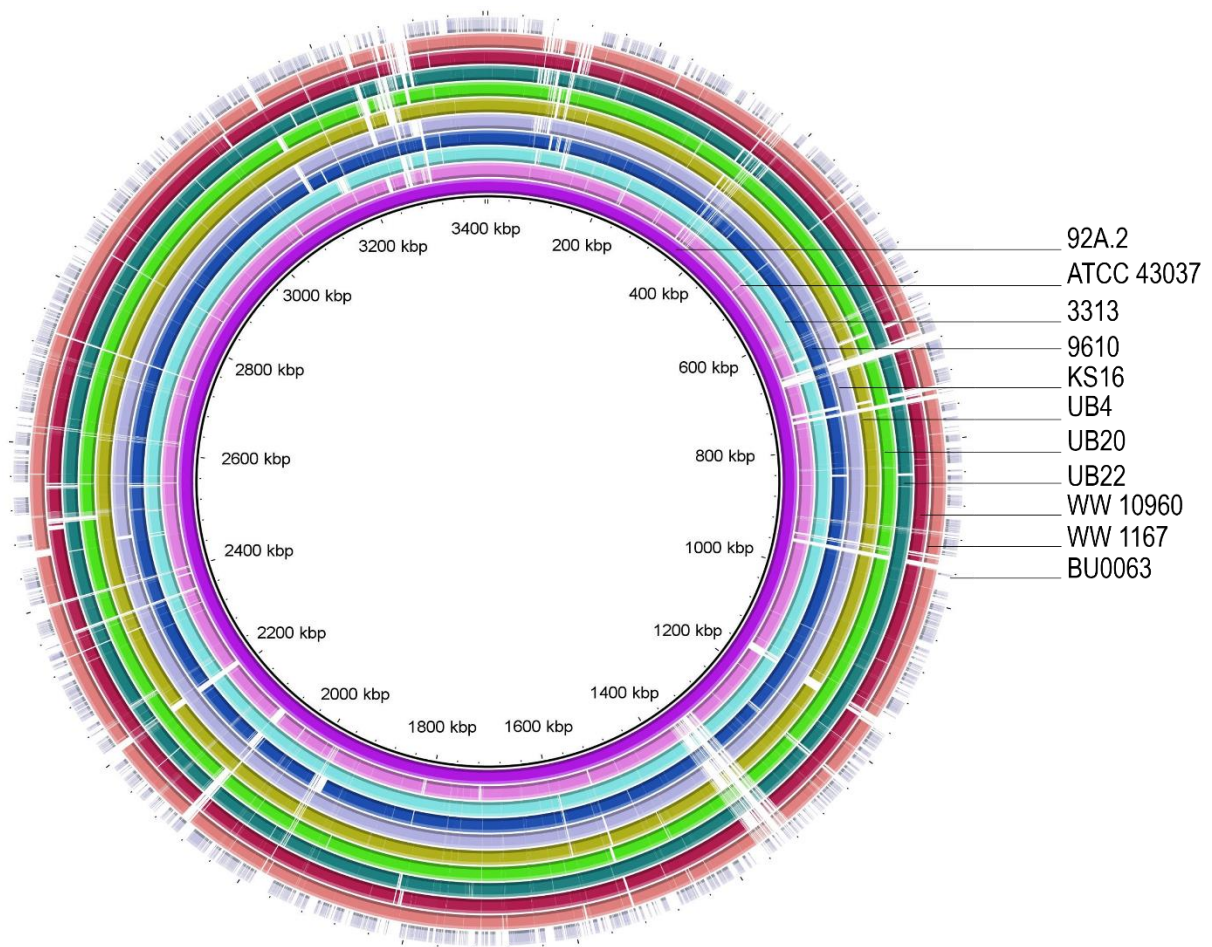
Figure 5.3 Multiple whole genome alignment of 11 human *Tannerella* genomes.
Blast Ring Image Generator (BRIG) was used to align these strains, with each colour
representing a genomic strain. All 10 *Tannerella* strains were aligned to the reference
established genome of 92A.2. BRIG is incorporated with a blatn service, shaping the overall
DNA sequence of each genome. Higher identities were noticed between all *T. forsythia*
strains compared to non-pathogenic isolate strain BU0063. The genome sequence ATCC
43037 showed over 98 % identity to the 92A.2.

Furthermore, we aligned *Tannerella* strains to the reference sequence 92A.2 genome using NCBI blast. Here, we have collected every strain's contigs and combine them using 'combine FASTA' bioinformatic tool, generating one DNA fasta. This enables alignment of all human and non-human *T. forsythia* isolate strains to the constructed database 92A.2. Blastn with default 'Algorithm parameters' and optimisation for 'Discontiguous megablast' that ignores some bases was chosen for search analysis and comparison. By using nucleotide sequences, we found higher identity between pathogenic *Tannerella* strains compared to the database reference genome (92A.2), with an average of 99.2% (Table 5.2). In contrast, non-pathogenic BU0063 and the other two non-human isolate showed lower identity to others, indicating some differences to the 92A.2 genome structure. Likewise, an average of 99% was found when we compared all human isolate strains to this study sequenced ATCC 43037. Then, we combined all contigs of the previous reported ATCC 43037 and aligned *Tannerella* strains to conclude on average 99.1% similarity.

Table 5.2. Alignment of *T. forsythia* isolate strains

Alignment of nine human *T. forsythia* isolate strains, periodontal health associate isolate *Tannerella* sp. BU063, and non-human *Tannerella* isolate genomes compared to the FDC 92A2 strain as a reference sequence. Results are based on NCBI blastn output in whole-genome alignments.

Tannerella forsythia strains	Isolate	Identity to 92A.2 strain	Identity to Sheffield ATCC 43037 strain	Identity to 2015 NCBI ATCC 43037 stain
ATCC 43037	Human	98.91%	-	100%
92A.2	Human	-	98.91%	99.17%
3313	Human	99.22%	98.95%	99.29%
KS16	Human	99.16%	99.25%	99.25%
9610	Human	99.35%	98.63%	99.04%
UB4	Human	99.55%	99%	99%
UB20	Human	99.31%	98.66%	98.66%
UB22	Human	98.94%	99.66%	98.66%
WW10960	Human	99.28%	99.04%	98.99%
WW11663	Human	99.25%	98.99%	99%
BU0063	Human	78.47%	78.48%	78.48%
OH1426	Non-Human	89.59%	89.49%	89.49%
OH2617	Non-Human	89.55%	89.48%	89.48%

5.3.1.4. Virulence factors

T. forsythia is a pathogenic organism, and to understand the mechanisms causing the pathogenesis correlated with this microbe, it is essential to determine and characterise associated virulence functions. In the literature, numerous genes were identified *in vitro* and *in vivo* assays to be associated with the pathogenicity of *T. forsythia*.[44,102,311–319] The majority of these virulence genes were identified in ATCC 43037, suggesting the need to check the relevance and significance of these genes among other strains. We assessed the presence or absence of virulence factor among *Tannerella* species including periodontal health-associated isolate. This study retrieved each virulence gene from the reported literature after checking the protein domain and annotation of every virulence factor using bioinformatic tools concluding 45 genes (Figure 5.4; Genes IDs are shown in Appendix VIII: Figure 8.13). Then we identified Pfam domain for each virulence gene before assessing the presence or absence of such a gene, as well as using the NCBI protein blast among *Tannerella* species.

The presence or absence of virulence factor gene between pathogenic human isolate strains was investigated. The S-layer is an indication of a general protein *O*-glycosylation system in a species. The presence of S-layer can assist a species-host interactions and impact its lifestyle. The S-layer of *T. forsythia* are found in all strains containing two glycoproteins.[319] Furthermore, the KLIKK proteases (Karilysin, mirolase, miropsin-1, mirolysin, miropsin-2, and forsilysin) were known for their ability to degrade an array of host proteins and their role in preventing host defences and fuelling local inflammation.[318] Strain KS16 showed no homologues to Karilysin, whereas strain WW10960 showed no homologues to karilysin, mirolysin, and miropsin-2. Absence of miropsin-1 and mirolysin is found in 3313 strain. In complex oligosaccharides and proteoglycans, glycosidases can hydrolyse terminal glycosidic linkages that are found in saliva, gingival crevicular fluid and periodontal tissue. The α-arabinofuranosidase (abfA/TF0780) is absent in the 3313 strain. Furthermore, the hemagglutinin (TF1723) that can expose protein epitopes for adherence and colonisation, is absent from 3313 strain.

Next, we have compared the presence and absence of virulence genes between *Tannerella* pathogenic strains and non-pathogenic strain. Of the reported 45 putative virulence genes, 21 of these factors are absent from the putative periodontal health-associated specie (*T. serpentiformis*; BU0063). Leucine-rich repeat BspA protein is a surface and secreted protein. BspA can traffic bacteria proteins to the outer membrane and can mediate *T. forsythia* interactions with the host factors. This essential protein is missing from BU0063. Two of KLIKK proteases (forsilysin and prtH) are absent from this non-pathogenic species.

216

Some enzymes that promote biofilm in patients with periodontitis found to be active only *in vivo*. Of these enzymes, TF1705 was found to be active *in vivo* assay only compared to *in vitro*, indicating pathogenicity of this microbe.[315] β-hexosaminidase (nahA) is part of sialic acid *nan* operon and is essential for *Tannerella* species to cleave glucosamine and galactosamine.[319] Both TF1705 and nahA are missing in BU0063. Furthermore, the other 15 virulence factors are missing including part of sialidases (NanH and SiaHI), glycoside hydrolase (TF1320, TF1502, Bgix, and Hexa), α-galactosidase (TF2392), β-galactosidase (TF0229, TF1468, TF2386, TF0636, TF0960, TF0371), and α-1,2-mannosidase (TF0617, TF1335). Strain BU0063 does not have a homologous for hemagglutinin (TF1723), which exposes protein epitopes for adherence and colonisation.[319] In contrast to the human isolate species, the strains of non-human isolate did not contain virulence factors of hemagglutinin, and some of the KLIKK proteases and glycoside hydrolases families.
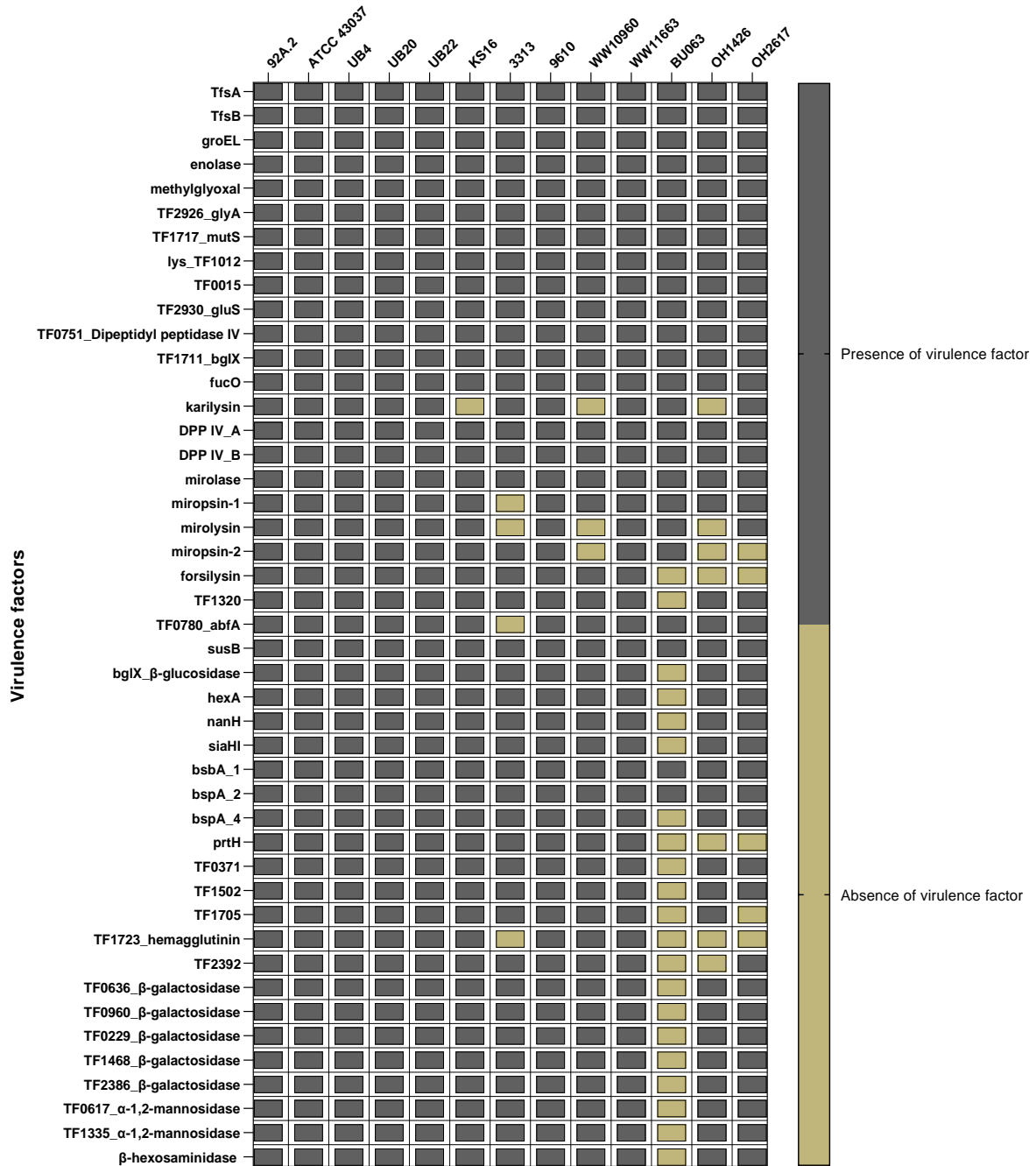
Figure 5.4 *Tannerella* species virulence factor genes.
Pfam domain code was used to assess the presence or absence of virulence gene between *Tannerella* species along with NCBI blast. These virulence factor genes were retrieved from the literature for ATCC 43037 and 92A.2 strains. *Tannerella* sp. BU063, non-pathogenic microbe, was used as a reference for comparison of a presence or absence of a virulence gene between human *Tannerella* species. In addition to the human isolate microbes, this figure included two non-human isolates *Tannerella* microbes: OH1426 and OH2617.

5.3.2. Distribution of *nan* operons

5.3.2.1. Initial results of *nan* operons

To make this study tractable, we proposed to bioinformatically predict bacteria with *nan* operons similar to *T. forsythia*; therefore, the analytical approach of this study was only to include: 1) bacterial species that encoded the main components of the sialic acid *nan* operon (*nanO*) and (*nanT*, *sss*, or *ABC*) along with at least one member of either the auxiliary or catabolism genes. 2) The predicted sialic acid *nan* operons of genes were within 30 ORFs from the beginning of each cluster. All *T. forsythia* strains have full-identically organised *nan* operons, however, we only included in this study the *T. forsythia* ATCC 43037 strain. Furthermore, we updated our main template of search (i.e. *T. forsythia* ATCC 43037) by searching the literature to extract more bacteria transporting sialic acid before using their nucleotide and protein sequences to search databases and identify more potential *nan* operons feeding pathways for sialic acid among microbes (Appendix VIII: Table 8.1). The analysed genomes were selected from the NCBI and PULDB databases between 1ˢᵗ.05.2020 to 1ˢᵗ.12.2021. About 2600 bacterial genomes from the NCBI and PULDB databases were examined and filtered for the presence of *nan* operons. We also noticed that most bacterial species included multiple strains that were in full or partial *nan* operons. However, this study only included one strain from each bacterium to simplify the outcomes. Therefore, the filtered and analysed genomes from these bacterial species represent 12 orders, eight classes, and four phyla. These phylum bacteria are *Bacteroidetes*, *Proteobacteria*, *Verrucomicrobia*, and *Gemmatimonadetes*. Of the four phyla, this study concluded the presence of *nan* operons in 222 bacterial species, belonging to 31 bacterial families (Table 5.3).

Interestingly, the dissemination of *nan* operons that is isolated from human microbes was 67 (Figure 5.5) compared to 155 non-human microbes (Appendix VIII: figures 8.15- A, B, C). The bacterial families of the human microbes were *Bacteroidaceae*, *Barnesiellaceae*, *Weeksellaceae*, *Tannerellaceae*, *Prevotellaceae*, *Rikenellaceae*, *Dysgonomonadaceae*, *Flavobacteriaceae*, *Sphingobacteriaceae*, and *Porphyromonadaceae* (Table 5.3).

We also examined the intraspecies distribution of the putative *nan* operons to identify whether all strains from species encoded the *nan* operon. Examination of the intraspecies noted a few exceptions to the distribution of the *nan* operon among all strains within the same bacterium.[303,304] Of twenty-eight strains in *Capnocytophaga sp*., only six of which encoded sialic acid *nan* operon (H6253, 878, W10638, FDAARGOS 1468, DSM 7271, and 864). Among the twenty-four sequenced *B. thetaiotaomicron*, only one did not encode sialic acid *nan* operon (*B. thetaiotaomicron* CL06T03C18). Similarly, for *B. stercoris*, two out of four

sequenced strains do not encode the *nan* operons (ATCC 43183 and CL09T03C01). From the soil and rhizosphere of Soybean, *Hymenobacter sp*. and *Sphingobacterium sp*., each bacterium has 10 sequenced strains. Of the *Hymenobacter sp.* multiple trains, three strains were found only to encode sialic acid *nan* operons (BRD128, NBH84, and PAMC 26628). Likewise, two strains only of the *Sphingobacterium sp*. (ML3W and PM2-P1-29) encoded sialic acid *nan* operons.

Table 5.3: List of phylum and families of species with *nan* operons.

Phylum	Family	Isolation		Total of
		Human microbes	Non-human microbes	species
Bacteroidetes	*Bacteroidaceae*	26	8	34
	Barnesiellaceae	1	0	1
	Chitinophagaceae	0	26	26
	Cytophagaceae	0	6	6
	Cyclobacteriaceae	0	9	9
	Dysgonomonadaceae	4	4	8
	Flavobacteriaceae	0	28	28
	Fulvivirgaceae	0	2	2
	Flammeovirgaceae	5	1	6
	Haliscomenobacteraceae	0	1	1
	Hymenobacteraceae	0	4	4
	Marinilabiliaceae	0	4	4
	Muribaculaceae	0	1	1
	Mangrovivirgaceae	0	1	1
	Rhodothermaceae	0	1	1
	Porphyromonadaceae	3	1	4
	Prevotellaceae	12	2	14
	Proteiniphilum	0	2	2
	Prolixibacteraceae	0	6	6
	Rikenellaceae	3	3	6
	Saprospiraceae	0	1	1
	Spirosomaceae	0	3	3
	Sphingobacteriaceae	4	28	32
	Tannerellaceae	8	1	9
	Weeksellaceae	1	1	2
Gemmatimonadetes	*Gemmatimonadaceae*	0	1	1
Proteobacteria	*Chromatiaceae*	0	1	1
	Pseudoalteromonadaceae	0	2	2
	Shewanellaceae	0	4	4
	Woeseiaceae	0	1	1
Verrucomicrobia	*Opitutaceae*	0	2	2
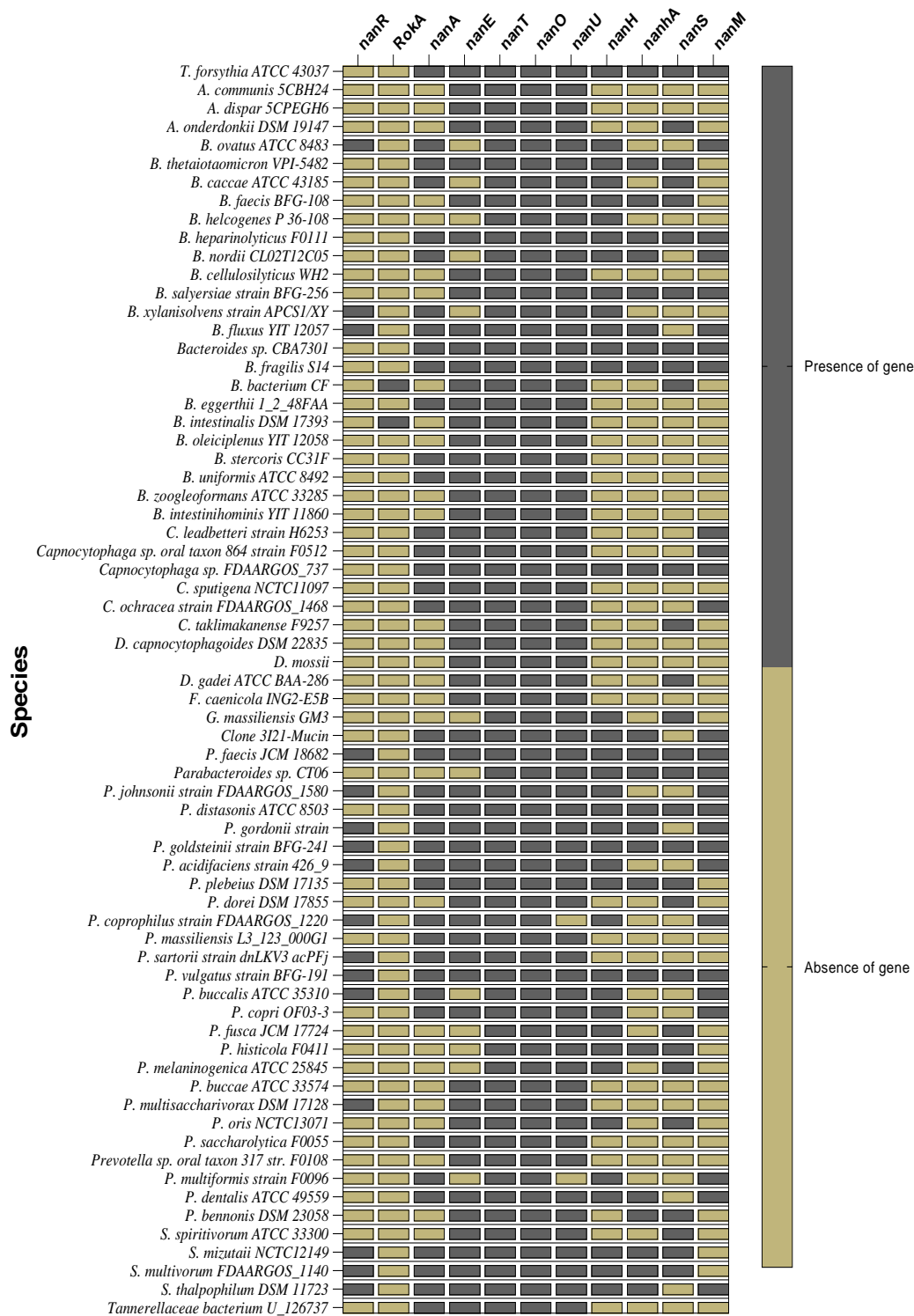
Figure 5.5 Sialic acid utilisation, catabolism, and transport clusters from human isolated microbes. Predicted sialic acid *nan* operons from the isolated human genome. All predicted microbes were structured to the order of *nan* operon of *T. forsythia*, although many bacterial genomes showed reverse start and end of *nan* operons. Key: dark gray colour means presence of a gene within the *nan* operon structures.

5.3.2.2. Sialic acid inner membrane transporters within the *nan* operon

Before sialic acid catabolism, sialic acid must be transported into the cell unless there is either of the endogenous biosynthesis sialyation or biosynthesis.[298] The inner membrane transporters for sialic acid are four diverse families. These families are the Major Facilitator Superfamily (MFS-NanT), Sodium Solute Symporter (SSS), ATP-binding Cassette (ABC), and Tripartite ATP-Independent Periplasmic (TRAP-NeuT).[84,297,320,321] These families were found and identified in 222 bacterial genomes except for TRAP family. This study used six bacteria species that were reported within the literature containing TRAP family. These bacterial genomes were used to identify if they can meet this study inclusion criteria for sialic acid *nan* operon (Appendix VIII: Table 8.1). However, none of the literature reported TRAP bacterial genomes and their extension blast results were found to contain the sialic acid NanO/NanC within 30 ORFs of the putative *nan* operon cluster.

Three inner membrane sialic acid transporters were found within the structures of *nan* operon. The first inner membrane transporter, (MFS/NanT), was highly distributed in the analysed genomes of this study. The presence of NanT was 67 times in homo sapiens isolated bacterial genomes (Figure 5.5) compared to 110 of the non homo sapiens isolated bacteria (Appendix VIII: Figure 8.15- A, B, and C). All the 67 human microbe genomes with NanT belonged to *Bacteroidaceae, Barnesiellaceae*, *Weeksellaceae*, *Tannerellaceae*, *Prevotellaceae*, *Rikenellaceae*, *Dysgonomonadaceae*, *Flavobacteriaceae, Sphingobacteriaceae*, and *Porphyromonadaceae* families. The second distributed inner membrane transporter was the Sodium Solute Symporter (SSS). This family in the structure of *nan* operons was found only with non-human microbes, representing 42 bacterial genomes. Likewise, the third inner membrane transporter was the ABC, and the outcomes of this family showed only three bacterial genomes, which were isolated from non-human microbes. Additional analysis for these three microbes showed none meeting the inclusion criteria of this study (Appendix VIII: Table 8.1).

To examine the evolutionary distribution of the *nan* operons within the reported bacterial genomes, we performed a phylogenetic analysis for the sialic acid inner membrane transporters (NanT) and (SSS) (Appendix VIII: Figures 8.16 and 8.17). Each assembled tree was quite compact for every family of transporters. None of the trees of the two families showed long branches away from other members, and each tree's structure was in agreement with microbial taxonomy.

5.3.2.3. Sialic acid inner membrane transport and catabolism within the *nan* operon

The *nanA* gene encodes the enzyme *N*-acetylneuraminic lyase and can degrade intracellularly sialic acid to the amino sugar *N*-acetylmannosamine (ManNAc) once transit across the inner membrane sialic acid transporter.[298] The presence of *nanA* was found in 127 bacterial genomes within the analysed genomes (Table 5.4). Of the presence of *nanA*, 29 microbes of the 67 human isolated bacterial genomes did not encode for *nanA* within their *nan* operons. The presence of *nanA* gene was found to be associated with both *nanT* and *sss* types of inner membrane transporters. In the analysed human and non-human isolated genomes, we found *nanA* to be clustered with *nanT* in 88 bacterial genomes (38 in human sapiens), whereas to be clustered with *sss* in 39 bacterial genomes, indicating sialic acid utilisation using SSS inner membrane uptake via this protein domain.

Table 5.4 Total presence of inner membrane catabolism genes

Genes	Total presence of a gene	Total presence within human sapiens isolation	Total presence within non-human sapiens isolation	Clustered with NanT	Clustered with SSS
RokA	10	2	8	7	3
NanR	72	16	56	46	26
NanA	127	38	89	88	39
NanE	152	55	97	137	15

The *nanE* gene is characterised by the presence of *N*-acetylmannosamine-6-phosphate 2-epimerase converting *N*-acetylmannosamine to *N*-Acetyl-D-Glucosamine 6-phsphate (GlcNAc-6-P).[298] The presence of *nanE* was found in 152 bacterial genomes within the analysed genomes (Table 5.4). Of the presence of *nanE*, 11 of 67 human isolated bacterial genomes did not encode for *nanE* within the *nan* operons. The *nanE* gene with *nanT* transporter was found to be clustered 137 times in the analysed genomes compared to 15 times with *sss* sialic acid transporter.

The *RokA* gene is a member of the sugar ROK family, which can act as a repressor, open reading frame, and kinase. The *RokA* gene (kinase) is to activate *N*-acetyl-d-glucosamine (NAG) in the pathway of amino sugar utilisation.[322] In *T. forsythia*, kinase is not clustered with *nan* operon; however, it encodes kinase separately for phosphorylation before amino sugar utilisation. Within the putative *nan* operon structures, the distribution of this gene was found only in *B. bacterium* CF and *B. intestinalis* DSM 17393 of the analysed isolated human bacterial genomes compared to eight kinases of the isolated non-human bacterial genomes. Of the 10 bacterial genomes with *RokA*, we found *RokA* to be clustered more with *nanT* sialic acid transporter seven times compared to three times with *sss* transporter (Table 5.4).

The *nanR* repressor gene regulates the expression of *nan* operon for sialic acid uptake and metabolism.[298] Among the analysed bacterial genomes of this study, we identified 16 genomes from isolated homo sapiens with *nanR* repressor compared to 56 isolated non homo sapiens bacterial genomes. We also found that the presence of *nanR* was 46 times and 26 times clustering with *nanT* transporter compared to *sss* transporter, respectively (Table 5.4).

5.3.2.4. Sialic acid outer membrane transporter within the *nan* operon

The TonB-dependent transporters play a crucial role in passing substrates, forming links between the outer membrane and the inner membrane of a bacterium. This linkage between the two membranes is mediated in sialic acid transportation by a porin that was found to be essential for the uptake of sialic acid. Several annotations for this porin were found within the databases (RagA, NanC, SusC, and NanO), but this study included what is involved with sialic acid transit.[203] To do so, all inner membrane sialic acid transporters were used to locate the outer membrane of sialic acid porin, confirming its presence within the *nan* operon structures. Bacterial genomes without the predicted NanO type outer membrane protein were excluded from the study analysis. A constructed phylogenetic tree was used to illustrate bacteria families and their members from human microbial species (Figure 5.6). Another a phylogenetic maximum-likelihood tree based on NanO was built for the analysed human (Figure 5.7) and non-human microbial genomes (Appendix VIII: Figure 8.18). Both phylogenetic trees showed homologous between the NanO *T. forsythia* compared to the other predicted NanO type outer membrane proteins.

The X-ray crystal structure shows a similar complex architecture of SusCD and RagAB where SusD and RagB tightly capped SusC and RagA, respectively. This suggests that the structure of both SusD and RagB acts like a lid that can open and close separately.[323] The SusD/NanU is found to be an important mechanism for nutrient uptake in a range of carbohydrates, as well as to enhance the uptake. Homologous for SusD, RagB, or NanU within the sialic acid NanO/SusC were searched and included in this analysis. Only 14 bacterial genomes were found without the sialic acid outer membrane protein (NanU/SusD). Of these 14 bacterial genomes, the NanU/SusD was absence from two isolated homo sapiens bacterial genomes in this study, namely *P. multiformis* strain F0096 (*Prevotellaceae* family) and *P. coprophilus* strain FDAARGOS_1220 (*Bacteroidaceae* family). Overall, the presence of NanU/SusD was reported for its necessity for the uptake of sialic acid, but our finding from these bacteria suggested their lost/absence, which indicates a similar shape to the *E. coli* NanC, suggesting a unique structure.
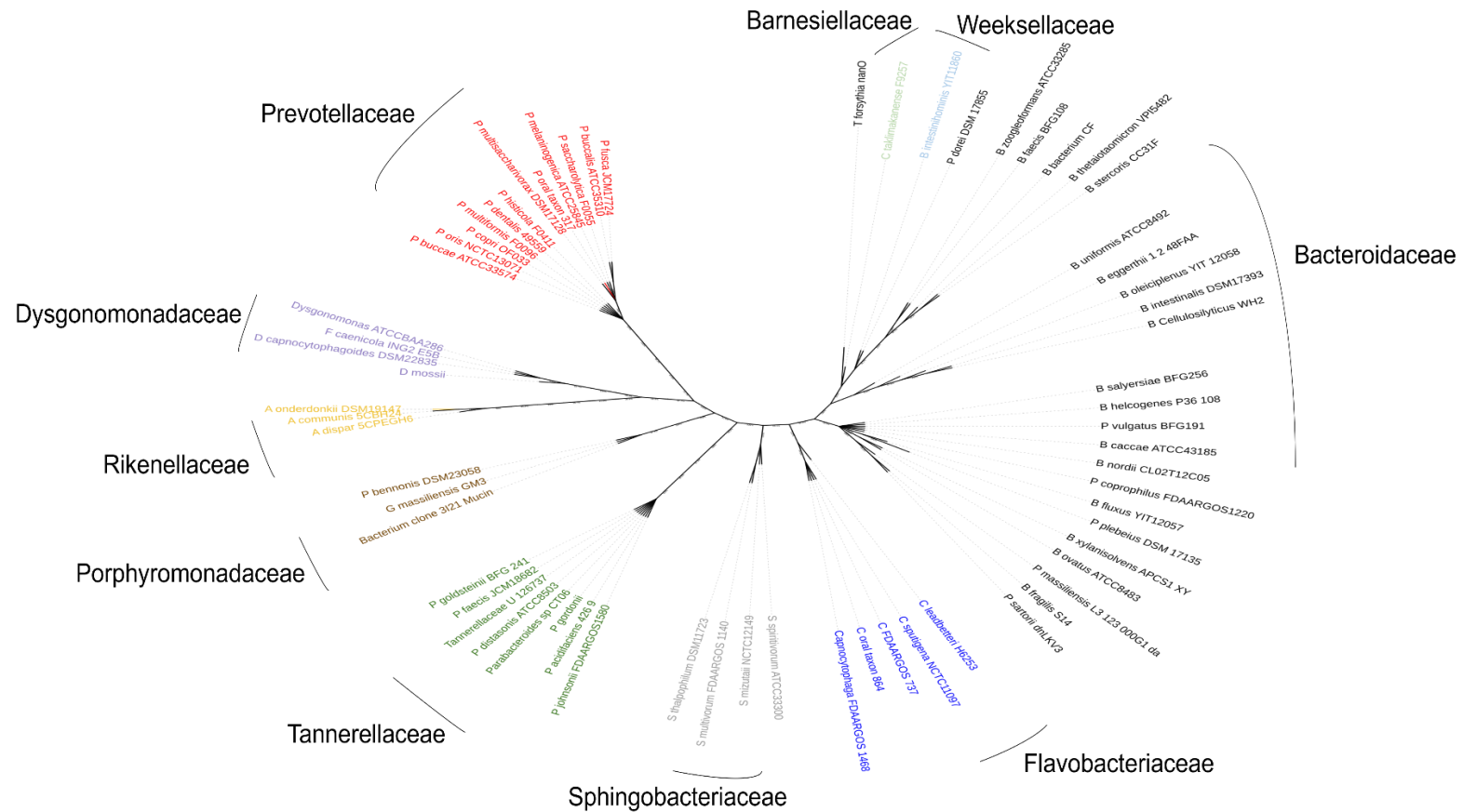
Figure 5.6 Phylogenetic tree of predicted NanO type outer membrane proteins.
The tree was obtained using NanO sequences of all human microbes with *nan* operons. This tree is edited to highlight families of the reported bacterial genomes and it was not based on bootstrap. The tree is viewed and built by Interactive Tree of Life, version 6.5.2.
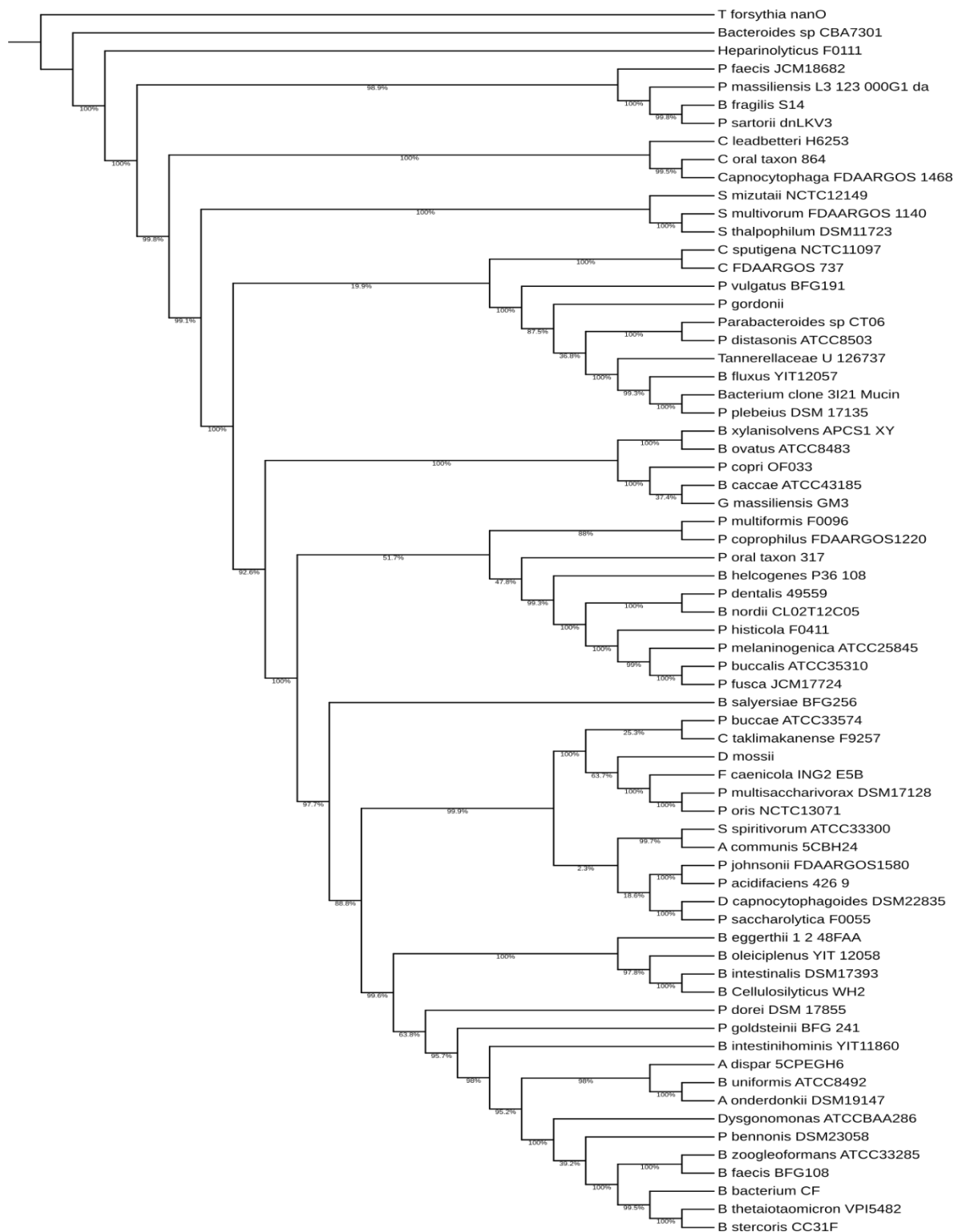
Figure 5.7 Phylogenetic tree using NanO amnio acid sequences of all human isolated bacterial species with *nan* operons. The tree is viewed and built by Interactive Tree of Life, version 6.5.2. Bootstrap analysis was performed and calculated based on the alignment of the amino acid sequences by ClustalW.

5.3.2.5. Sialic acid outer membrane auxiliary genes within the *nan* operon

       Investigation about the auxiliary genes showed varied order of these genes. Sialidases are a set of GH33 Glycosyl blade β-propeller structure enzymes and contribute to the virulence factors of many pathogenic organisms.[203] The number of bacterial genomes that encode the *nanH* within our results, the key enzyme for the sialic acid cleavage, were 129 compared to 93 bacterial genomes without *nanH* (Table 5.5). Of the bacterial genomes with *nanH* gene, 35 were found in homo sapiens isolated bacteria of the following families *Bacteroidaceae*, *Tannerellaceae*, *Prevotellaceae*, *Sphingobacteriaceae*, and *Porphyromonadaceae*. In the analysed bacterial genomes, the *nanH* gene was found to be clustered with *nanT* in 95 bacterial genomes, whereas to be clustered with *sss* in 34 bacterial genomes. Phylogenetic analysis of sialidases encoded by the *nan* operons of the predicated bacteria showed closely several clusters compared to the *nanH* of *T. forsythia* ATCC 43037 (Appendix VIII: Figures 8.19 and 8.20).

Table 5.5 Total presence of auxiliary genes

Genes	Total presence of a gene	Total presence within homo sapiens isolation	Total presence within non homo sapiens isolation	Clustered with nanT	Clustered with SSS	Clustered with ABC
nanH	129	35	94	95	34	0
nanS	78	28	50	65	10	3
nahA	87	23	43	56	10	0
nanM	38	26	12	36	2	0

Bacteria with sialidase enzymes, that are known to access human sialic acid, cannot efficiently cleave diacetylated sialic acid without the sialate-*O*-acetylesterases (*nanS*).[203] The distribution of *nanS* gene in this analysis resembles that the *nanS* was part of the *nan* operons in 78 bacterial genomes (Table 5.5). The dissemination of *nanS* was found in 23 of the human microbe organisms compared to 43 of the non-human microbe organisms. Furthermore, the presence of *nanS* was found in nine of the human bacterial genomes without the sialidase enzyme to capture this sialic acid (*nanH*). Within the distribution of *nanS* gene, this analysis found that the presence of *nanS* with *nanT*, *sss*, and *ABC* transporters was seen in 65, 10 and three bacterial genomes, respectively.

Genes for both the auxiliary *nahA* and *nanM* genes within the *nan* operons were found in the analysed genomes. The β-hexosaminidases (*nahA*) are a group of enzymes that can cleave β-linked glucosamine or galactosamine and assist adhesive molecules upon the removal of terminal sialic acid residues.[324] The sialic acid mutarotase (*nanM*) is able to quickly balance solutions of sialic acid to the resting stability position.[325] Both *nahA* and *nanM* genes were found in 23 and in 26 of the genomes belonging to human microbes respectively (Table 5.5). In contrast to the isolated human microbes, both *nahA* and *nanM* genes were found in 43 and 12 genomes, respectively. In all species examined in this study, both *nahA* and *nanM* were clustered with *nanT* in 56 and 36 bacterial genomes, respectively. Likewise, both *nahA* and *nanM* were encoded on the same strand of sialic acid inner membrane (*sss*) in 10 and two bacterial genomes, respectively.

5.3.2.6. Possible feeding pathways for the analysed human microbes

Human microbe strains involved in sialic acid utilisation were divided into two groups. First, bacteria that one would predict can cleave sialloglycans and breakdown sialic acid (glycans). Second, bacteria with only genes that are required for cleavage of the sialic acid or for catabolism of the derived sialic acid.[326] Previous work on fucose predicted the feeding pathways for human gut microbes based on histidine sensor kinase.[327] Further computational study predicted feeding pathways of human gut microbes based on the secreted glycosyl hydrolases.[328] Thus, the presence of only sialidases (*nanH* or *nanS*) and, or β-hexosaminidases (*nahA*) with outer and inner membranes transporters compared to the presence of only metabolism genes (*nanA* and, or *nanE*) with outer and inner membranes transporters can predict possible feeding structures using the *nan* operons. Therefore, we analysed and predicted all human microbes in this study with only one or two sialidases and, or β-hexosaminidases pathways to be classified as an 'accumulator bacterium'. Also, we predicted all human microbes in this study only with catabolism pathways to be classified as a 'utiliser bacterium', whereas human microbes with both sialidase and catabolism pathways to be classified as a 'bifunctional bacterium'.

Of the 67 human microbes outside of *T. forsythia*, 13 bacterial genomes were isolated from oral cavity compared to 49 bacterial genomes from stomach. The rest were isolated from wound or blood cultures. Of the oral human isolated bacterial genomes, we classified six species as utilisers, one as an accumulator, and six as bifunctional species (Figure 5.8). In addition, the isolated gut bacterial genomes were classified as following: 15 species were utilisers, five species were accumulators, and 29 species were bifunctional bacteria (Figure 5.9).

In summary, data on possible sailic acid feeding pathways demonstrated two characterisations in the human genomic microbes. First, larger numbers of the analysed human microbes were utilisers, suggesting a common cohabit of different bacteria within a biofilm. Second, this may suggest an additional feature for accumulator and bifunctional bacteria which can cleave, utilise, and donate sialic acid, suggesting their importance to utiliser bacteria in establishing a biofilm community.
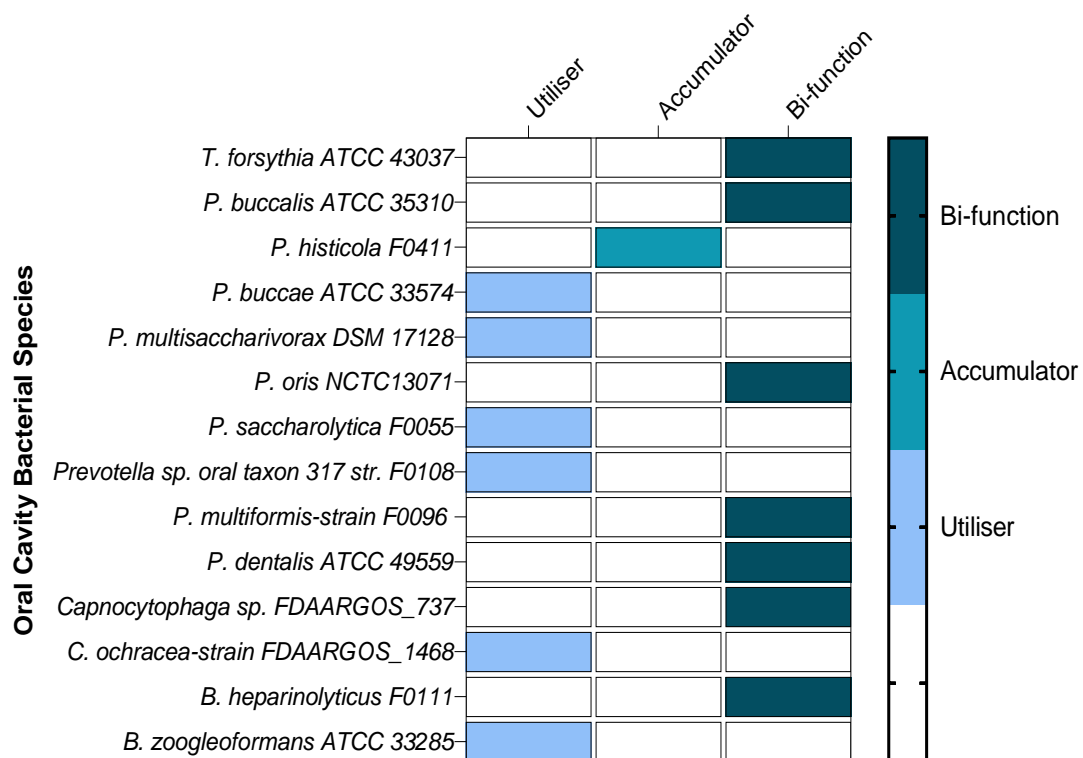
Figure 5.8 Classification of oral cavity bacterial species feeding pathways.
This figure shows the classification of the isolated oral genomes based on their *nan* operon:
"bi-function", "accumulator", or "utiliser". The presented sialic acid microbes are clustered in
agreement with their isolations. Four colours were chosen to illustrate each species
classification.
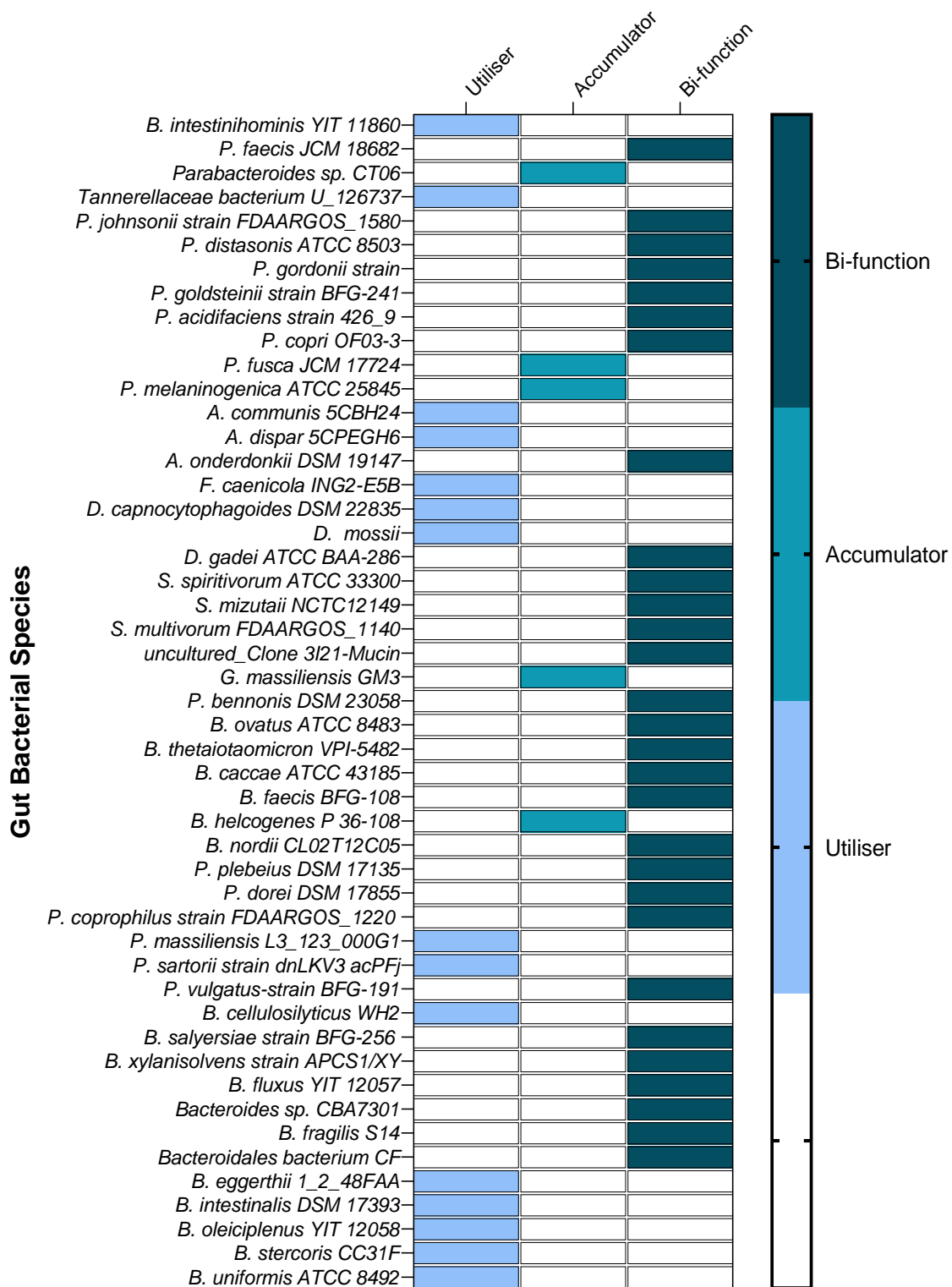
Figure 5.9 Classification of gut bacterial species feeding pathways.
The list of gut microbes is classified in either "bi-function", "accumulator", or "utiliser"
based on their *nan* operons. Four colours were chosen to illustrate each species classification.

5.3.2.7. Alignment of TonB boxes of NanO from human microbes with *nan* operons to the TonB box of NanO of *T. forsythia*

In chapter IV of this study, we were able to predict the TonB box of NanO (GASVVE) and design a TonB box peptide inhibiting the uptake of Neu5Ac. We then asked about the selectivity of this peptide and whether GASVVE peptide may inhibit the uptake of sialic acid in other human microbes. Here, we aligned all NanO amino acids from 63 human microbes within the predicted feeding pathways (section 5.3.2.6.) to predict and compare their TonB boxes to the TonB box of *Tannerella* NanO. As a result, we noticed a high conservation of residues between these predicted TonB boxes to the TonB box of *Tannerella* NanO, namely gly-G and val-V (Figure 5.10). This suggests a selectivity of GASVVE peptide targeting other sialic acid NanO-type outer membrane proteins. Further research studies can investigate the possible feature of GASVVE peptide on other microbes and reflect on these conserved motifs for a universal antimicrobial peptide.

```
   T. forsythia_NanO    ....NVSGTV TDQ.SGEPVI GASVVEKGTR .NGTVTDIDG NYSLSTASNA
Bacteroides_sp_CBA73    ....TITGVV SDA.TGEPII GASVVEVGTT .NGTITDIDG KFVLNMNPDG
B_heparinolyticusF01    ....TCKGVV VDA.TGETVI GASVVKGTT .NGTITGLDG DFSLTGVKKG
P_faecisJCM_18682(GG    ....SIQGVV YDD.LNTTIP GASVIEKGTT .NGTVTDIDG KFTIQ.VTRG
P_massiliensis_L3_12    ....SCTGIV KDT.TGETVI GASVVKGTT .NGTITGIDG DFSLNNVKQG
P_sartorii_strain_dn    ....SITGIV VDV.TGEPVI GASVVELGTT .NGTITDFDG NFSLS.VESN
B_fragilis_S14(ANQ60    ....SITGVV TDA.SGESVI GASVVEVGTT .NGVITDISG KFTLM.VDPN
C_orchraceaFDAARGOS1    ....VKGTVS SDE.M..PLP GAAVVKGTT .HGTQTDFEG NYTLT.AKEG
S_mizutaii_NCTC12149    ....ALKGTV TDP.LGNPLE GVTVSNQTQG .HVTKTDEQG GFSVQ.AKVS
S_multivorum_FDAARGO    ....AVRGKI TDAKTGELLG NVTVKVKGSS .QSLQTNAVG TFSIS.TNPN
CapnocytophagaSp_FDA    ....IVKGTV SSD..ELPLP GAAVVKGTT .HGTQTDFDG NYTLT.VKEG
Parabacteroides_sp_C    ...TVSGTV LDT.YNMGVP GASIVEKGTT .NGTITDFDG NYTLTVSNKN
P_distasonis_ATCC850    ...TVSGTV LDT.YNMGVP GASIVEKGTT .NGTITDFDG NYTLTVSNKN
Tannerellaceae_bacte    ...TVKGVV YDD.LNMTVP GASVVEKGTQ .NGIVTDMDG NFVLSVS.PG
P_gordonii_strain(RG    ...TITGTV MDE.SNFSVP GASIVEKGTT .NGTVTDLDG NFSIIVSNKN
P_copri_OF033(DXA63_    ...TCTGVV VDD.LGETVI GASVIIKGTS .NGTVTDLDG KFSLSGVKNG
uncultured_Clone_312    ...ICQGIV KDA.TGESVI GASVIVKGTT .NGTITDIDG NFSLSNVKKG
P_plebeius_DSM17135(    ...ICQGIV KDA.TGESVI GASVIVKGTT .NGTITDIDG NFSLSNVKKG
B_fluxus_YIT12057(EG    ...TITGTV VDP.MGP.VI GASVLQVGTT .NGTITDIDG KFTLT.VPEN
P_vulgatus_strain_BF    ...TITGVV VDA.TGEPVI GASVLEVGTT .NGTITDVDG NFTIQ.VPVG
 P_buccalisATCC35310    ...TVTGTI LDE.NQEPLV GVTVSTKDGK .GNTVTNIDG KFTIQ.TASN
G_massiliensis_GM3(G    ...TVNGTV LDD.AKEPLL GVSVQVKGTT .QGSMTDING KFSVS.APSD
B_caccae_ATCC43185(B    ...TVNGVV LDD.TREPLI GVSVLVKGSQ .QGTVTDLDG KYSIT.APAD
B_ovatus_ATCC8483(BA    ...KVSGVV QDQ.TGVELP GVSISVKGMD .KGTITNASG EFSIM.VPPN
B_xylanisolvens_stra    ...KVSGIV QDQ.TGVELP GVSISVKGVD .KGTITNASG EFSIM.VSPD
Prevotella_oraltaxon    ...KISGTV VSAGDNEPVI GATITVVGTK .TAAVTDFDG KFSLTTDVPN
P_dentalisATCC49559(    ...NVAHGT VIDENGEPLI GVTVTIEGQK GTGTVTDLDG RFELRGVSRG
B_nordii_CL02T12C05(    ...RIIKGV VTDENGEPLI GVSVLVKGTT .TGTITDMDG IYSLE.VPED
B_helcogenes_P36_108    ......CTGI VKDTTGETVI GASVVKGTT .NGVITGLDG DFTLTGVKKG
P_histicolaF0411(J5A    PVIYTFQGQV TDTD.GMPLI GTTITQKGGD KVYAVADLEG RYVIKSDNPH
P_fusca_JCM17724(ADJ    PVLYTFNGQI QDVA.GNPLI GATVMVKGAS KAHAVADLEG KFVLKTENPN
P_melaninogenica_ATC    PTLYTFNGQV QDVE.GNPLI GATIMVKGSS KVRAMSDLEG RFVLKSETPN
P_buccaeATCC33574(HM    ....KVTGTV TDGSSGEPII GVTVK.IKGG ATGAVTDLDG NYSVNAK.PG
P_multisaccharivorax    ....KVTGTV SDSQG..PII GATVR.VKGT QIATVTDMDG KYSISIK.PG
P_orisNCTC13071(NCTC    ....KITGKV TDGKE..TII GATVK.QKGT NNATITDLDG NFTLSVP.QG
F_caenicola_ING2E5B(    ....RITGTV TDSRDNSAMI GVNIM.EKGT LNGTVTDIDG NFSLNVTTQD
D_mossii(HMPREF9456_    ....QISGKV TDAKDGSPMP GVTVQ.TQ.T KVGVMTDFDG NYTIKAN.KG
P_saccharolyticaF005    ....QISGKI SDTK.AEPIV GAVVR.IKNK T.AVVTDVNG YYTINAAPGD
D_capnocytophagoides    ....QIKGKV TDAKDGSPLV GVTVK.TQKN S.GVMTDLDG KYTINAQKGD
P_johnsonii_strain_F    ....KVIGVV KSAH.NEPLI GVNVV.VKGS TQGTVTDLEG HYSIDVPDSK
P_acidifaciens_strai    ....KVTGVV KSAY.DEPLI GVNVV.VKGS TQGTVTDFNG RYSIEVPNNK
S_spiritivorum_ATCC3    ....QVQGTV RDQD.GKPLA GVSVS.VVES QKQTFSDGKG NFSIAVKPSD
B_eggerthii_1_2_48FA    ....TITGTV IDTR.HEPVI GASVL.EKGT SNGTITNLDG EYSLKVS.PG
A_communis_5CBH24(A5    ....EIQGRV T..ADGQPLV GASVI.IKGT TTGTSTDIAG NFTVRAK.NG
B_cellulosilyticus_W    ....TVQGKV VSKTDGESII GATVIETNQT SNGTITDIDG NFTLSVP.QG
B_intestinalis_DSM_1    ....TVQGKV VSKTDGEPII GATVIETNQT GNGTITNIDG DFTLSVP.QG
B_oleiciplenus_YIT_1    ....TVTGTV VD.TDGLTIP GVNIL.VKGT TIGVITDIDG NFKISVPDKD
B_zoogleoformansATCC    .....IKGTV KDSQ.GEPLP GASIQ.VKGS TVGTITNIEG GFSINAKSRD
B_faecis_BFG108(K6V3    .....IKGSV SDGG.GEPLP GVSVA.VKGT TSGTITDMDG KFSINARNND
B_thetaiotaomicron_V    .....VQGTV TDET.GESLP GVSVV.VKGT TNGTVTDVNG KYSIQATSKD
B_stercoris_CC31F(HM    ....TVTGRV SDAM.G.TLP GVSVS.VKGT TSGTVTDADG KYSLNADRNA
Bacteroidales_bacter    .....IKGSV SDTK.GAPIA GVAIV.VKGT NNYTTTGQNG EYQITANKGD
D_gadei_ATCCBAA286(H    .....ITGKV TDSH.QEALA GVSVV.LKGA TVGTFTDIDG AYKITVPSGD
P_bennonis_DSM23058(    .....VTGQV TDNT.GLPLT GVMVK.VKNT GNGTATDQDG RYSIRVAKSD
B_intestinihominisYI    .....VRGIV TDDN.GEPLP GVAVV.KDGT SLAVSTDIDG RFAIKASEGD
P_goldsteinii_strain    ....NVKGVI VDET.GEPVI GANVV.VSGT TNGTVTDFNG NFSLEVPENG
P_dorei_DSM17855(GKD    ....TVTGQV VDVT.SEPII GASVV.VKGT ANGTITNFDG KFTLSVQKGE
A_onderdonkii_DSM191    ....KMTGRV LDET.KAGVP GATVI.VKGS TRGVITGNDG TFSIDVKPTD
B_uniformis_ATCC8492    ....KITGVV KDNK.GEPII GANIVELNKK GNGTITDIDG KFTLSVSKNA
A_dispar_5CPEGH6(A5C    .......... .......... .......... .......... ......MNRH
B_salyersiae_strain_    .......... .......... .......... .......... ......MKQH
P_multiformisF0108(Q    ....GIGTST NGDGAFGGSV SLATHAPNRE PGVEVSGSYG SYSTYNTGVK
P_coprophilus_strain    ....FKGILN YKTNHFG... .LQQFDLNVS GGIGNNWLY. SGSVYQTFDP
           Consensus    .......g.v .d...g.p.. g..v.....t ..gt.td.dG .%........
```

Figure 5.10 Alignment of predicted TonB boxes to the TonB box of *Tannerella* NanO
Multiple sequence alignment was performed between *Tannerella* TonB box and other
predicted TonB boxes from the reported bacteria within this study, highlighting the
similarities between the TonB box of NanO 'GASVVE' and other possible TonB boxes.

5.3.2.8. Signatures of horizontal gene transfer

There are several methods to detect putative horizontal gene transfer event. One such methods is the presence or absence of integrons and/or transposase within a cluster of genes. Presence or absence of integrons and/or transposases within the *nan* operons may suggest a mode of gene transfer.[298] This study inspected all reported *nan* operons to locate transposases and, or integrons. Of the analysed human microbes, the presence of transposases and, or integrons was found in six genomes, from *Rikenellaceae*, *Prevotellaceae*, *Dysgonomonadaceae*, and *Bacteroidaceae* families. These microbes are *B. stercoris* CC31F, *B. salyersiae* BFG-256, *B. thetaiotaomicron* VPI-5482, *F. caenicola* ING2-E5B, *P. melaninogenica* ATCC 25845, and *A. communis* 5CBH24. Of the analysed non-human microbes, the presence of transposases and, or integrons was found in 12 genomes, from the following families: *Rhodothermaceae*, *Saprospiraceae*, *Prolixibacteraceae*, *Shewanellaceae*, *Flavobacteriaceae*, and *Chitinophagaceae*.

5.4. Discussion

Many sequence genomes are interrupted with short reads, even at very high coverage. These short readings cannot close all gaps. Previous 2015 sequencing of *T. forsythia* included 141 contigs with N50 size 114 kilobase pairs (kbp). Of these contigs, 487 kbp represented the largest sequence.[215] Likewise, a recent study reported improved assembly of ATCC 43037 with a larger contig of 1.85 Mbp, in 2020.[329] In contrast, our study was able to sequence the genome of *T. forsythia* into three contigs using Nanopore and Illumina service from MicrobesNG, UK. After scaffolding and gap filling, the largest contig covered over 99.5 % of the home genome with over 3.28 Mbp. The sequence service used in this study was sufficient to close many gaps using sequencing technology with long reads. This finding of ATCC 43037 sequencing from this study is consistent with only three strains from *Tannerella* with completed assemblies, each with one contiguous sequence (92A.2, 3313, and KS16).

This study chose 92A.2 as a reference genome to compare other *Tannerella* strains as it was reported with the confidence of completeness.[300] BRIG, over several reported tools, provides an overall picture with colours of the completed *Tannerella* genomes, indicating their similarities.[310] As expected, we found differences to the genome structure of the putative periodontal health-associated isolate (BU0063). Likewise, we found non-human isolate to reveal differences to all three reference genomes (92A.2 and ATCC 43037). This suggests the need of sequencing technologies that can provide medium-sized or long reads, accounting for interruption with short reads, even at very high coverage.

Many genomes of *Tannerella* lack functional annotation where on average a strain contains between 2400 to 2700 protein-coding genes.[259] Comparative assessment of *Tannerella* virulence factors can increase knowledge about the presence or absence of such factor gene. It can also confirm and characterise the presence and absence of candidate genes involved in pathogenesis as the cultivation of *T. forsythia* and the systematic knock-out approach are very challenging.[329]

Differences in the presence of virulence factor genes between strains might indicate truncated gene sequences due to mutations of start or stop codons. However, a comparison of the entire genome to identify the presence of a virulence factor to periodontal health-associate isolate can indicate a virulence factor pathogenicity. This study concluded 21 virulence factor genes that are absent within the putative periodontal health-associated species *Tannerella* BU0063. Surprisingly, four of KLIKK proteases (mirolase, miropsin-1, mirolysin, and miropsin-2) are present within the periodontal health-associate isolate. This was confirmed by the comparative genome analysis along with Pfam domain in this survey. We detected homologous for mirolase, miropsin-1, mirolysin, and miropsin-2 from BU0063 strain with

46.10 %, 57 %, 68.99 %, and 57.48 % to the wild-type ATCC 43037 strain. Unlike the previous conclusion about the absence of KLIKK protease from BU0063, the enhancement of bioinformatic tools concluded homologues of these protease within BU0063.[318] The previous conclusion about the absence of KLIKK protease from BU0063 has to be revised, encouraging further investigation for this strain.

Several of glycoside hydrolase along with neuraminidase (*nanH*) are needed for sailic acid cleavage and hydrolyses of glycosidic linkages in complex oligosaccharides. These enzymes can contribute to pathogenicity in different ways, and they play vital functions in the pathobiology of periodontal diseases.[319] BspA is absent in BU0063, but it is one of the important virulence factor with several functions as it was reported with a function in oral biofilm formation and mediation of the innate immune system.[319] The absence of these genes indicated these virulence factors genes pathogenicity. Further laboratory assays are needed for characterising absence of the virulence factors from the periodontal health-associate isolate (BU0063) and their presence in other pathogenic strains as potential targets for therapeutic strategies

Next, the bioinformatic finding enhances the picture of overall sialic acid utilisation among bacteria. In this study, we examined and investigated the distribution and spread of the *nan* operon from *T. forsythia* to thousands of microbes that are available within the NCBI and the PULDP databases.[303,304] The analysed *nan* operon from *T. forsythia* encodes extracellular genes for the cleavage of sialic acid (*nanH*, *nahA*, *nanS*), extracellular and intracellular uptake genes for the transport (*nanO*, *nanT*), and intracellular catabolism genes of sialic acid (*nanA*, *nanE*). This study search against each member of *T. forsythia nan* operon to predict other bacteria with similar structures. In addition, data were also collected on using the inner membrane transporter (SSS) from *P. translucida* and *S. pealeana*, as well as the inner membrane transporter (ABC) from *Arachidicoccus sp.* 5GH13-10, *Pedobacter sp.* BS3 BS3-1, and *Sphingobacterium s*p. C459-1T. This comparative genomics approach indicated the huge capability of bacteria for delivering and utilising many sialic acids structures. Within the human microbe organisms, this analysis highlighted the mutualistic relationships between these microbes based on terms of cleavage and, or catabolism of sialic acid. Similar techniques were applied for the identification of monosaccharides utilisation in different microbial taxa.[328]

Initial results from this analysis showed thousands of bacteria, which were filtered and removed for one or more reasons. (1) We have included bacteria that were meeting the inclusion criteria for this analysis. (2) For each analysed bacterium, multi-strains were found

with full or partial structure of the *nan* operon. Here, we presented one strain from each bacterium limiting repetitive results. (3) Duplicate submission for the same bacterium was widely found within both databases. Thus, we performed a comparison for duplicate submission, filtering out extra submission. (4) Preference of inclusion was given to a bacterium with a completed genome over a draft genome.

Sialic acids are simply a convenient source of food for many bacteria. Sialic acid, for some pathogens, can also be used in immune evasion strategies. Transit of sailic acid across the inner membrane was identified and characterised, presenting four families. The four families of the inner membrane sialic acid transporters are: (1) Major Facilitator Superfamily (MFS/NanT), (2) Sodium Solute Symporter (SSS), (3) ATP-binding cassette (ABC), (4) tripartite ATP-independent periplasmic (TRAP). These inner membrane transporters were best studied and reported in the *E. coli*, *S. enterica*, *H. ducreyi*, and *H. influenzae* respectively.[84,297,320,321] The inner membrane of sialic acid can be used to predict and conclude the structure of the *nan* operon cluster and its function for a bacterium. In this study analysis, the Major Facilitator Superfamily (MFS/NanT) showed to be widely clustered with the predicted NanO type outer membrane protein, particularly in humans, compared to other inner membrane transporters for sialic acid. Furthermore, the NanT was found to be highly clustered with at least one member of sialic acid catabolism genes. However, the TRAP family from the reported bacterial genomes that were used in this analysis, did not contain the predicted NanO type outer membrane protein within its cluster of genes. This family was reported to transport a variety of substrates under different contexts. Thus, this may speculate that this family can serve for more than one substrate indication and its clustering with sialic acid porin may not be possible.[330] Likewise on the sialic acid family (ABC), this study found no further than three members of the ABC transporter meeting the inclusion criteria of this study, which they were environmental organisms. In addition, all human microbes with the ABC sialic acid transporter were found, to the nature of our analysis, with no inclusion for the NanO/NanC within the same cluster.

A comparative genomic approach was applied additionally to *nanA* and *nanE* from *T. forsythia* to locate bacterial genomes with *nan* operons. These two genes are the inner membrane catabolism cluster,[203] which resulted in identifying more bacteria that were isolated from human sapiens with the *nan* operon structures. In addition, inclusion of both *nanA* and *nanE* genes in our analysis compared to other members of the auxiliary genes was found to reliably locate the inner membrane of sialic acid, which was highly clustered with the predicated NanO type outer membrane protein.

Data of the bacterial genomes without the NanO/NanC were removed from this study. Search analysis using the NanO/NanC predicated many bacterial genomes, but the majority of them were not in the cluster of *nan* operons. Generally, the utilisation of sailic acid was reported in several bacteria where the NanO/NanC did not cluster with the inner membrane sialic acid transporter.[204] Nevertheless, it is evident that mucin-degrading species are categorised into mucin specialists or mucin generalists.[331] Specialist mucin-degrading species such as *A. muciniphila* and *B. intestinihominis* were found only to grow on mucin *O*-glycans (as a sole polysaccharide source). In contrast, generalist mucin-degrading species like *B. thetaiotaomicron* and *B. caccae* were found to grow on several other polysaccharides.[331] Here, we can suggest that the mucin categories or terms (specialists or generalists) might apply to the sialic acid utilisation and transportation, explaining the presence of NanO/NanC without the inner membrane sialic acid transport in some bacteria. For instance, several gut microbes like *R. gnavus*, *B. breve* and *L. sakei* were reported to have catabolism inner membrane genes, but not clustering with the NanO/NanC.[326]

The NanU 'lid pin' and its clustering with NanO was monitored within the analyzed genomes of this study. In *T. forsythia*, NanU was found and reported to enhance transit of sialic acid through the NanO.[128] Similarly, *B. fragilis* NCTC 9343 has BF1720 (NanU) that is required for maximal function of sialic acid transit.[128] Both NanO and NanU were homologous to several members of the *Bacteroidetes*, such as *P. diastonis* (ATCC 8503: BD_2944 and 2945) and *B. vulgatus* (BVU_2430 and 2431).[128,203] Within our analysis, NanU was found in almost all the predicted bacterial genomes.

Absence or disorganisation of the catabolism inner membrane and the auxiliary genes showed insignificance *nan* operon structures. These insignificance structures were mentioned previously in a review reporting number of species that only encoded the key enzyme in the first step of sialic acid degradation *nanA* but did not encode *nanE* or *nanK*.[298] Several members of *Verrucomicrobia* and *Planctomycetes*, and two genera from *α-Proteobacteria*, some of which are commensal or pathogens of humans, did not encode all members of sialic acid catabolism genes *nanE* or *nanK*.[298] Commensals human gut bacteria like *Bacteroides* that can in some cases become opportunistic pathogens were reported with absence of and different distribution of the catabolism inner membrane genes, although sialic acid in *Bacteroides* has been known to be an important carbon source for these organisms.[332,333] The *nanK* was not required for *Bacteroides* in the presence of sialic acid for its catabolism.[332] *B. fragilis*, for instance, does not encode *nanK* that adds a phosphate group yielding *N*-acetylmannosamine-6-P (ManNAc-6P) as the *nanE* (Nacetylmannosamine-6-P) of *B. fragilis*

found to not require a phosphate group to perform sialic acid metabolic reaction.[332] Here, two additional genes (*nagA* and *nagB*) were reported with functions in the catalysing sugars and yielding glucosamine 6-phosphate and acetate, and forming fructose 6-phosphate (Fru6P) and ammonium ion, respectively. However, the location for *nagA* and *nagB* genes were highly variable between the reported bacteria, in some cases being part of the cluster, whereas in almost all cases were scattered around the genome. Thus, this study excluded both *nagA* and *nagB* genes to be a member of *nan* operon structures.

Sialic acid and its derivatives can vary from a bacterium to another in terms of catabolic and scavenging genes, generating a cohabit of different bacteria within a biofilm. The presence or absence of certain genes can predict possible feeding pathways of bacterial genomes. *F. nucleatum*, the orange complex bridging organism, has similar catabolic genes for sialic acid to *H. influenzae*. Although *F. nucleatum* is present in healthy individuals, but it was found to fully utilise sialic acid and is considered an opportunistic pathogen.[203] *F. nucleatum* lacks sialidase enzymes, but a synergistic biofilm in a contact-dependent manner was formed between both *T. forsythia* and *F. nucleatum*. *In vivo* using a mouse model, both bacteria showed an enhancement of abscess formation with a greater alveolar bone loss compared to each species alone.[334] Sialidases are not only to provide a nutritional and adhesive function to organisms, but also can modulate immune responses and immune evasion. Sialidase enzymes are not encoded within gut bacteria, namely *C. difficile* and *S. typhimurium*. Yet, these gut two bacteria were reported to utilise mucosal carbohydrates as a common strategy of catabolising microbiota. A sialidase deficient *B. thetaiotaomicron* was used *in vivo* colonising gnotobiotic mice along with *C. difficile* and *S. typhimurium* showing a downregulation of sialic acid catabolic pathway and impaired expansion.[335] Furthermore, *S. pneumoniae* sialidase can inhibit and manipulate killing by neutrophils with extensive deglycosylation.[336] Besides the possible feeding pathways predictions, cross-feeding between mucin glycan degraders and partner organisms were established in human gut microbes. The sensed or absorbed metabolic interactions can increase the total metabolic capacity. On germ-free mice, *Eubacterium rectale* ATCC 33656 was found to consume the acetate produced by the glycan degrader *B. thetaiotaomicron* VPI-5482, converting the acetate to butyrate.[337] The glycan degrader *B. thetaiotaomicron* VPI-5482 can also benefit from *Desulfovibrio piger* GOR1, which can reduce the liberates sulfate of cleaving glycan. In contrast, the process of glycan degradation can be impaired without hydrogen sulfide, leading *B. thetaiotaomicron* VPI-5482 to release sulfate to the genera of sulfate (like *D. piger* GOR1) to disrupt sulfur

bridges and increase access to mucus layer.[338] These indications of feeding pathways lead to an additional understanding of the complete and diverse metabolite pool.

In similar order and structure to *T. forsythia nan* operon, the presence of full *nan* operon (total nine genes), particularly in human microbes, was noticed in 12 of 67 homo sapiens bacterial genomes with eight to nine genes. These bacteria species belonged to the following families *Bacteroidaceae*, *Tannerellaceae*, and *Sphingobacteriaceae*. The list of these species is *P. faecis* JCM 18682, *P. distasonis* ATCC 8503, *P. goldsteinii* BFG-241, *S. mizutaii* NCTC12149, *S. thalpophilum* NCTC11429, *B. heparinolyticus* F0111, *B. nordii* CL02T12C05, *P. vulgatus* BFG-191, *Bacteroides sp*. CBA7301, *P. plebeius* DSM 17135, *B. thetaiotaomicron* VPI-5482, and *B. fragilis* S14. The sialic acid *nan* operon was reported in human-dwelling gastrointestinal anaerobes. Of the 12 bacterial genomes with full *nan* operons, *B. thetaiotaomicron* VPI-5482 was laboratory reported to use sialic acid as a growth factor or as the sole carbon source.[338] Likewise, *B. fragilis* S14, *B. fragilis* NCTC9343, *B. nordii* CL02T12C05 and *P. vulgatus* BFG-191 were found to utilise sailic acid.[339–341] On the other hand, results from this study proposed 'partial sialic acid *nan* operons. Partial sialic acid involves a cluster of NanO/NanC and inner membrane transporter with some catabolism or accessory genes. Partial *nan* operons highlighted the differences in scavenging and catabolising sialic acid (glycoprotein) among bacterial genomes. The wide distribution of the partial *nan* operon structures indicates the existence of multiple potential feeding pathways for sialic acid and its derivatives not only among isolated human genomic microbes, but also among isolated animals and environmental genomic microbes. One possible explanation to the different structures of the *nan* operons pathways is due to dissemination of sialic acid in nature, mainly in humans. In humans, *N*-acetyl-D-galactosamine (GalNAc-with side-chain oxygen atoms of Ser/Thr residues) is widely disseminated and lined between the glycan and protein of *N*- and *O*-linked proteoglycans.[342] Likewise, the GlcNAc can be found with human-glycoproteins as well as various human milk oligosaccharides.[343] In addition, GalNAc was found in plant-synthesized polysaccharides. On a terminal units of carbohydrate chains, both fucose (Fuc) and *N*-acetyl-D-neuraminic acid (Neu5Ac) are linked to proteins or lipids in animals, indicating they derive from the dietary components of animal origin.[343] The dissemination of sialic acid may illustrate several pathways for bacteria utilisation for such substrate, which might be a reason behind different *nan* operon structures and a possible indication of intensive feeding to the dissemination of sialic acid. Another possible explanation is, in prokaryotes where operons are pervasive, and their mechanisms of distribution are not fully understood. Some authors suggested that operon of genes within a

single co-transcribed region is more regulated and efficient as these rearrangements of genes are within a bacterium host genome.[344,345] However, the structures of operons are due to functional horizontal gene transfer event, where a cluster of genes is co-dependent functions and is derived by a major force within the bacterial evolution.[344,345] A possibility of the horizontal gene transfer might be reflecting on an adaptation for such organisms to the environment of the human microbiota. In addition, full or partial distribution of operons might be due to the fragmentation of well-adapted ancient operons. This fragmentation might lead to a new acquisition of the genes by the bacterial kingdom requiring only the evolution of regulatory elements and ignoring a selective advantage to the organism.[346] This led to a consideration and a description of "destructed" operon in prokaryotes, implying a limitation to the structure of *nan* operons.[347,348]

The ability of bacterial genomes to cleave and/or catabolism sialic acid have been previously determined by several studies.[298,349–351] Among several studies that were able to review, summarise, or predict utilisation and/or catabolism of sialic acid, this study was dedicated to survey and analyse the inclusion of the sialic acid NanO/NanC within the sialic acid feeding pathways. Many microbes are able to mimic host cell surfaces and avoid host immune attack by decorating their surface molecules (LPS and capsular polysaccharide).[203] Unlike other sialic acid studies, there was a big difference between results from this work compared to previous literature studies. Many of those previous results were structured on a large scale including either the catabolism of sialic acid pathways or identifying and classifying the extracellular sialidases.[298,349–351] Nevertheless, this work was conducted to address the term of specialist sialic acid species that were clustering with the sialic acid NanO/NanC protein. In addition, previous work from these studies predicated or summarised their results based on either NCBI database or the available literature. This study used a bi-directional NCBI search using both the nucleic acid and amino acid sequences along with analysis of the protein domain structures.[303] A further prediction of *nan* operons included searching the CAZy (PLUDB) database.[304]

In conclusion, this study presented a comprehensive computational analysis of the distribution of sialic acid *nan* operons. Many human microbes and their causes diseases can benefit from assemblages of microorganisms and their ability to drive sialic acid. The computational analysis can enhance the diagnostic tool and increase the ability of comparing closely and distantly multiple bacteria to each other. This will not only help find the presence or absence of certain feeding pathways, but it will also help future antimicrobial therapy targeting microorganism causing various human diseases. In addition to the human microbes,

this analysis predicted the structure of *nan* operons among isolated non-human microbes. However, the idea of exploring more about microbes with the *nan* operons is still limited to be discovered in this field. We used limited sequences as a template for the NCBI blast while blasting additional nucleotide and amino acid sequences from the resulting bacteria may generate extra bacterial genomes with sialic acid *nan* operons. Furthermore, future studies exploring the sialic acid *nan* operons will benefit from expanding their search about additional co-clustered metabolic and possible feeding pathways.

Chapter VI

Transcriptome Analysis of *T. forsythia* under Neu5Ac and mucin conditions.

6.1. Introduction

Periodontal pathogens must have the ability to obtain important growth factors, and the ability to reside and survive in their environmental niche. *In vitro* and *in vivo*, evidence showed the impact of sialic acid on establishing and providing an environmental niche for human commensal or pathogenic bacteria.[203] The periodontal pathogens, *T. forsythia*, is nutritionally fastidious, and requires *N*-acetylmuramic acid (NAM) to support its growth and to produce cell wall peptidoglycans. This bacterium can substitute its requirement for NAM with sialic acid in biofilm culture. This bacterium has a unique system (*nan* operon) compared to many other periodontal pathogens, consisting of scavenging enzymes and sialic acid transport and catabolism genes. The *nan* operon enables *T. forsythia* to not only scavenge for or transit sialic acid but can also provide all needed growth factors for processing sialic acid.[203]

In the past two decades, *T. forsythia* was characterised in terms of mono- and multi-species biofilm. In the previous two chapters, the role of TonB and NanO and the potential for their inhibition have been investigated. However, to date there is no knowledge of how *T. forsythia* responds to growth on NAM versus Neu5Ac or a physiologically relevant glycoprotein substrate. Here, to define the overall impact of Neu5Ac and mucin on *T. forsythia* and to investigate global gene expression in *T. forsythia*, transcriptome sequencing was carried out using RNA extracted from *T. forsythia* that was grown on Neu5Ac or mucin. Over > 95 % of the data were successfully aligned to the provided *T. forsythia* reference genome (ATCC 43037) provided by us before comparing the gene expression levels in both conditions based on normalised reads per kilobases of transcript. High sequencing reliability was validated by correlation analysis from three independent biological replicates for each condition. The transcriptomic data showed an overall picture of this bacterium's biological and pathway activities, enlightening our understanding of this bacterium's behaviour and potentially improving future antimicrobial therapy.

6.2. Results

6.2.1. Quality of Transcriptomics (RNA-seq) data

Transcriptomics was conducted on *T. forsythia* ATCC 43037 which was grown on either Neu5Ac or mucin compared to NAM. The RNA extraction was performed seven days after the incubation of this bacterium in triplicate (section 2.10). The extracted RNA samples had an OD260/280 of 1.8-2.2 and a concentration above 200 ng/μL. These extractions were further tested by Genewiz using Qubit and capillary gel electrophoresis. Next, the library preparation and Illumina sequencing were performed by Genewiz. Here, we used several webservers and databases to analyse and interpret differentially expressed genes (section 2.10.1).

6.2.2. Gene Expression Data Before and After Normalisation (Diagnostic plots)

In a typical RNA-sequence dataset, around 20 % of the genes in the genome are almost undetectable with zero read counts. Normalisation methods are needed for large scale expression data like RNA sequencing to preserve biological variation coming from the low end of the detection limit. The default of gene filtration is to keep genes in every library with expression of 0.5 counts per million (CPM). Several methods of the global normalisation factors were established to normalise gene expression. One of which is the rlog transformation where it minimises differences of small counts between samples for each row by transforming the count data to the $\log_2$ scale. $\log_2$ is a more robust method and can ensure low expressed genes not introducing much noise in the uploaded data. Also, the average-bulk normalisation is built with the advantage of taking the median ratios of observed counts as it assumed the expression of genes across all the samples are not significantly different. With the default normalisation method within the iDEP tool,[210] 2343 (84%) of the total 2803 genes from the RNA sequence counting data passed the iDEP filter of NAM vs Neu5Ac for the comparison. In NAM versus mucin transcriptomics, 2335 (83.2%) of the total 2803 genes passed the iDEP filter. These filtered genes were automatically converted to Ensembl IDs for further annotation and pathway databases. Figure 6.1 shows the data normalisation by boxplots and its similar distribution in the density plot across samples due to effect of the rlog transformation. These indicate high quality data comparable across the samples and make us confident in our analysis going forwards.
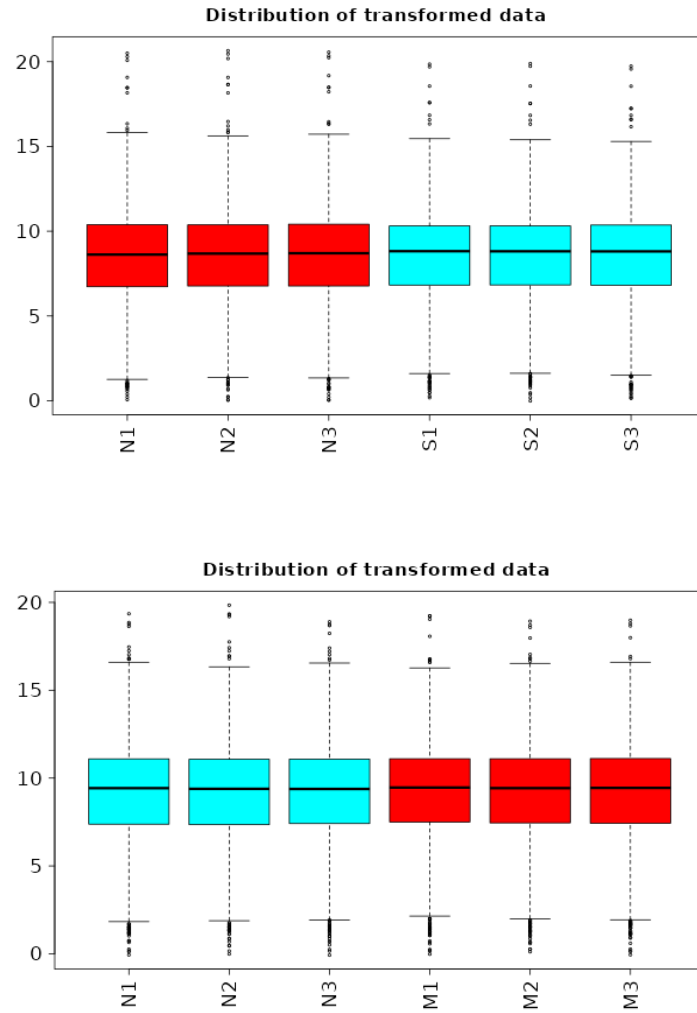
Figure 6.1. Box plots of expression data before and after normalisation.
This figure shows post-normalisation boxplots from NAM (N), Neu5Ac (S) and mucin
conditions (M). The post-normalised data were used for further analysis.

6.2.3. Sample Variability and Outliers-Exploratory Data Analysis (EDA)

6.2.3.1. Determining intra- and intergroup RNA samples

After normalisation, examining data analysis can reduce potential errors from the biological experiment steps, such as RNA extraction, digestion of DNA, removal of ribosomal RNA, RNA sequencing process, and removal of adapters. It also gives a general insight into the trend in gene expression change. This study conducted the intra and intergroup variability assessment for examining any potential error within the *T. forsythia* biofilm and the subsequent RNA extraction and sequencing. It is preferable that the intragroup assessment, representing technical or biological variability, is smaller than intergroup assessment, representing differences between experimental and control conditions. Of the normal and effective data examination is to design a heatmap and hierarchical clustering. The mean centring by genes method was applied to designing this heatmap to observe relative fold-changes. A big picture of the overall patterns can be shown from gene expressions by both heatmaps of NAM vs Neu5Ac and NAM vs mucin. As shown in figures 6.2 and 6.3, we observe that the genotype of *T. forsythia* under each substance condition differs significantly in the expression profiles of many genes. Replicates on the other hand, are grouped together as projected. To clarify these figures, the data were divided in the heatmap into three sections: top, middle, and bottom. The top quarter shows all up regulated genes in the presence of Ne5Ac (Figure 6.2) or in the presence of mucin (Figure 6.3), whereas the bottom quarter in both figures shows down-regulated expression genes in the presence of either NAM vs Neu5Ac or NAM vs mucin. The middle quarter in both figures shows the most variable genes due to substance condition.
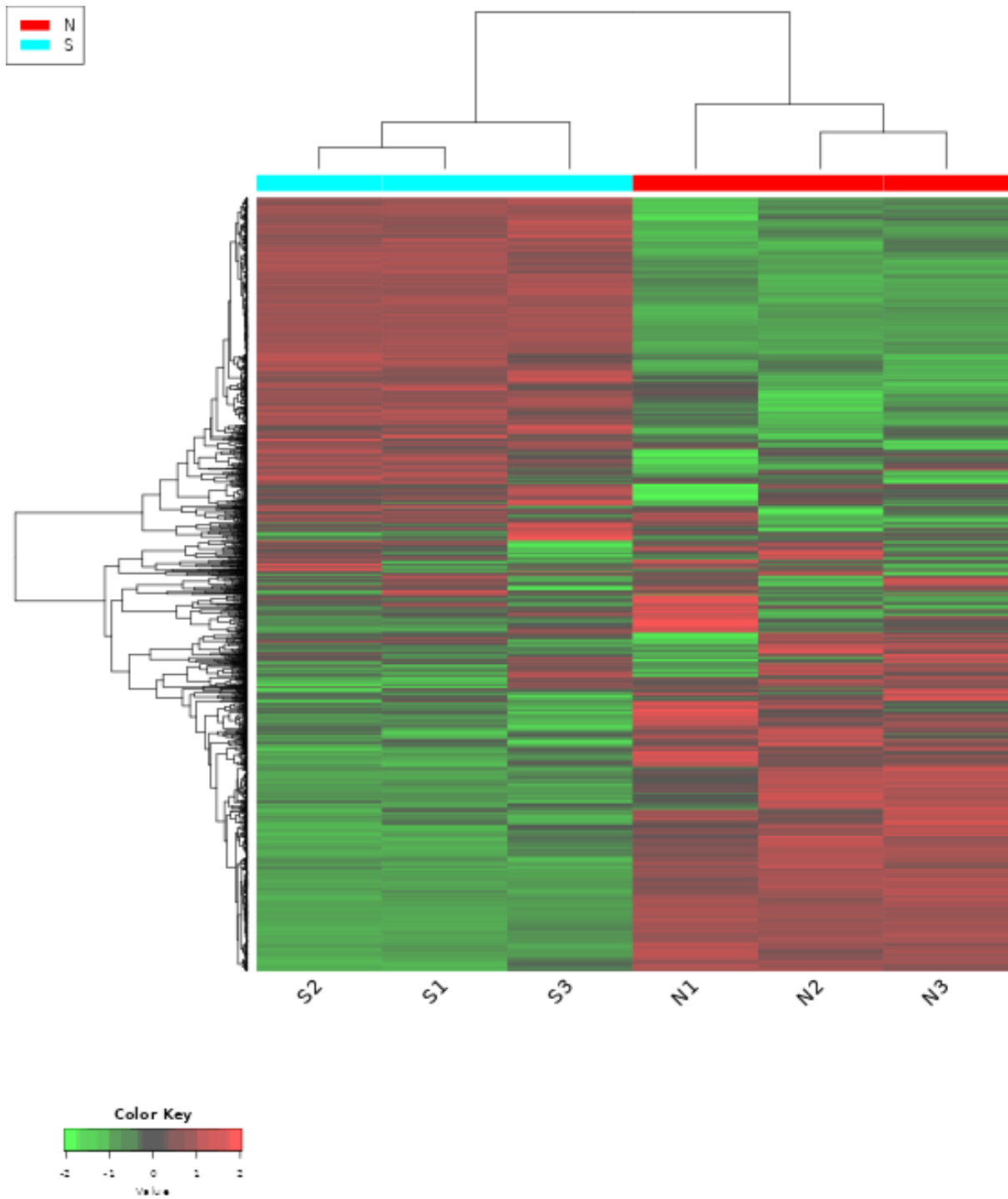
Figure 6.2. Hierarchical clustering heatmap for Neu5Ac conditions.
This figure shows transcriptomics data of *T. forsythia* on Neu5Ac. This figure was structured based on the average mean using corelation. NAM is abbreviated with 'N' (N1-3) and Neu5Ac is abbreviated with 'S' (S1-3).
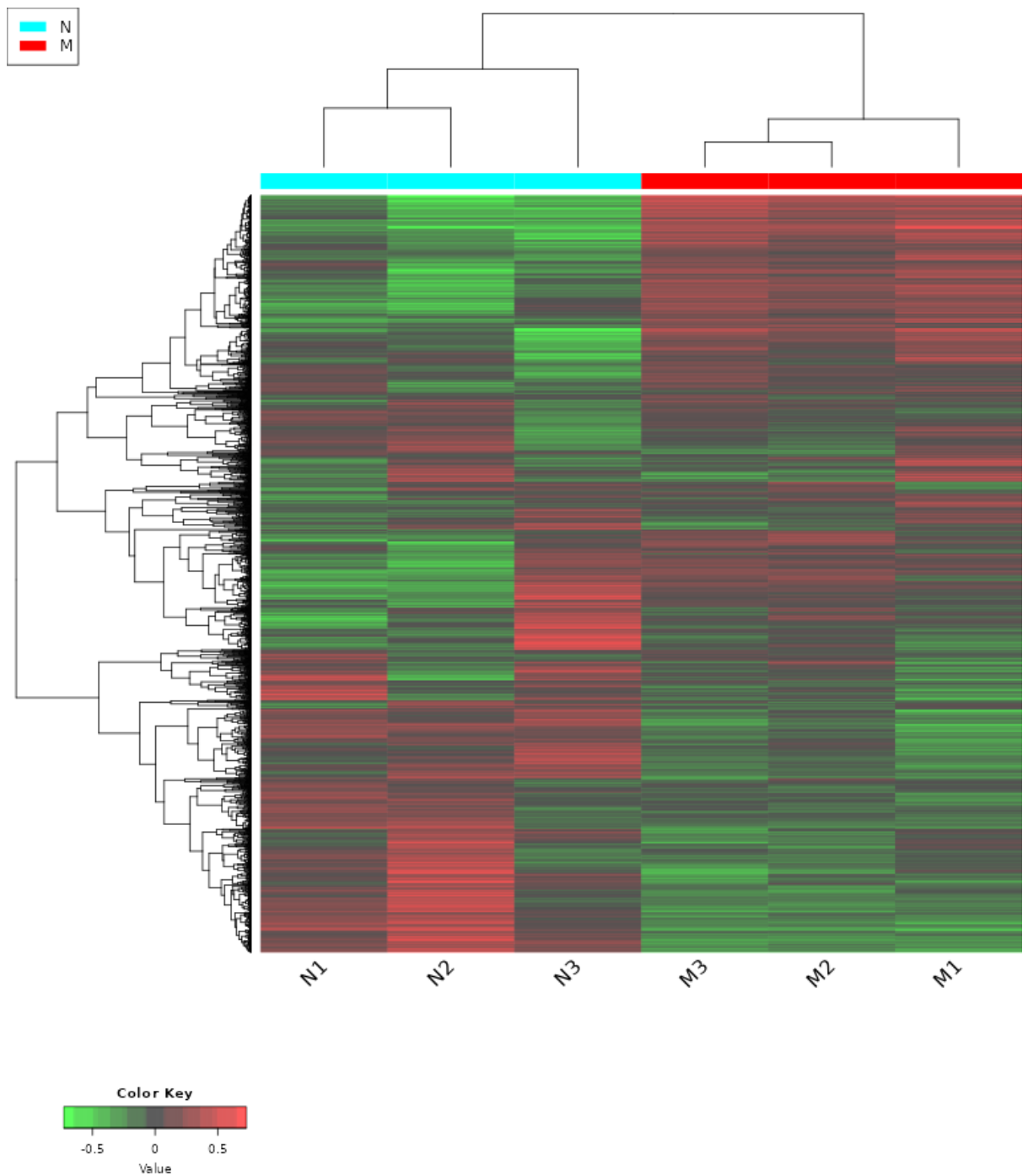
Figure 6.3. Hierarchical clustering heatmap for mucin conditions.
This figure shows transcriptomics data of *T. forsythia* on mucin. This figure was structured based on the average mean using corelation. NAM is abbreviated with 'N' (N1-3) and mucin is abbreviated with 'M' (M1-3).

Another way to assess the intra and intergroup variability is through Principal Component Analysis (PCA). The PCA was generated using normalised reads per kilobases of transcript per million mapped reads (RPKM) counts. Here, the PCA shows the total variation of *T. forsythia* biofilm on N-acetylmuramic acid (MurNAc/NAM) or N-acetylneuraminic acid (Neu5Ac). The PCA shows technical replicates within NAM in comparison to Neu5Ac across the two principal components: PC1 and PC2 (Figure 6.4). PC1 accounts for 87% of the variance (between two experimental conditions) compared to 7% for an additional variance on PC2 (representing technical or biological variability) (Figure 6.4). Likewise, in a comparison of *T. forsythia* biofilm between NAM and mucin, the technical replicates are plotted close to each other and coordinated on each of PCs as we expected. Samples of the biofilm of *T. forsythia* on NAM are on the left side (56%), whereas samples of the biofilm of *T. forsythia* on mucin are placed on the right side (25%) (Figure 6.5).
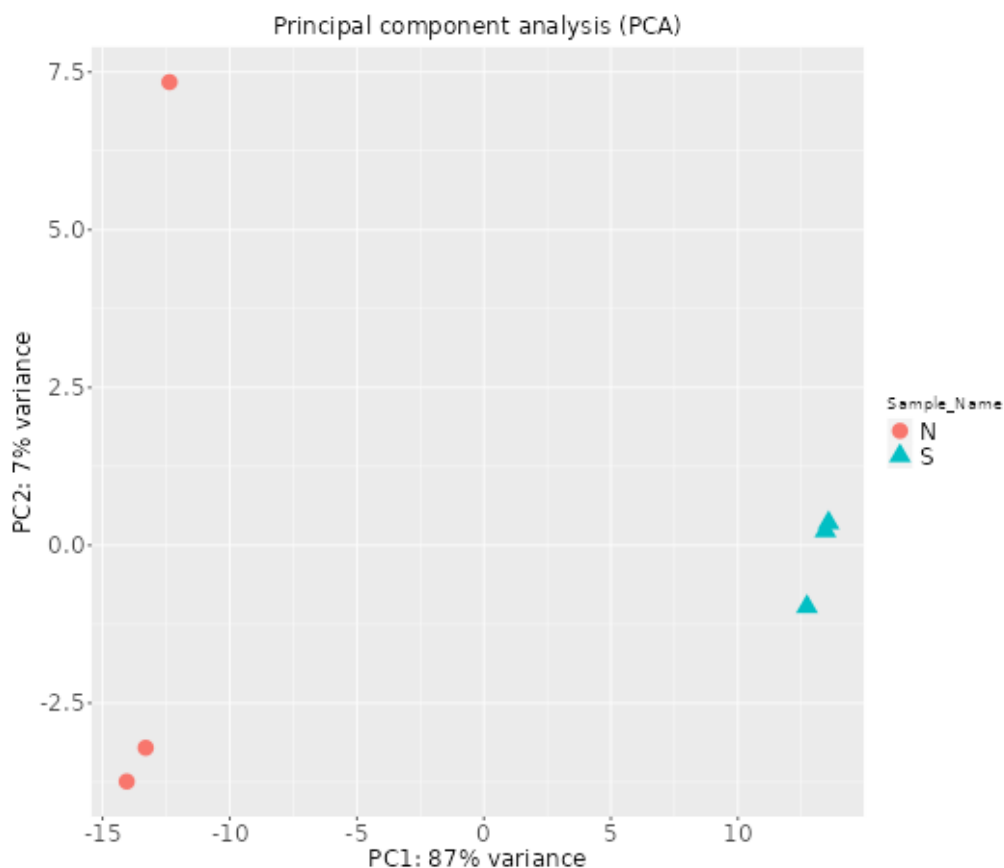
Figure 6.4. Principal component analysis 1.
This figure shows RNA sequence samples from biofilm of *T. forsythia* on NAM vs Neu5Ac on a scatter plot that describes general variability in expression profiles. NAM is abbreviated with 'N' and Neu5Ac is abbreviated with 'S'
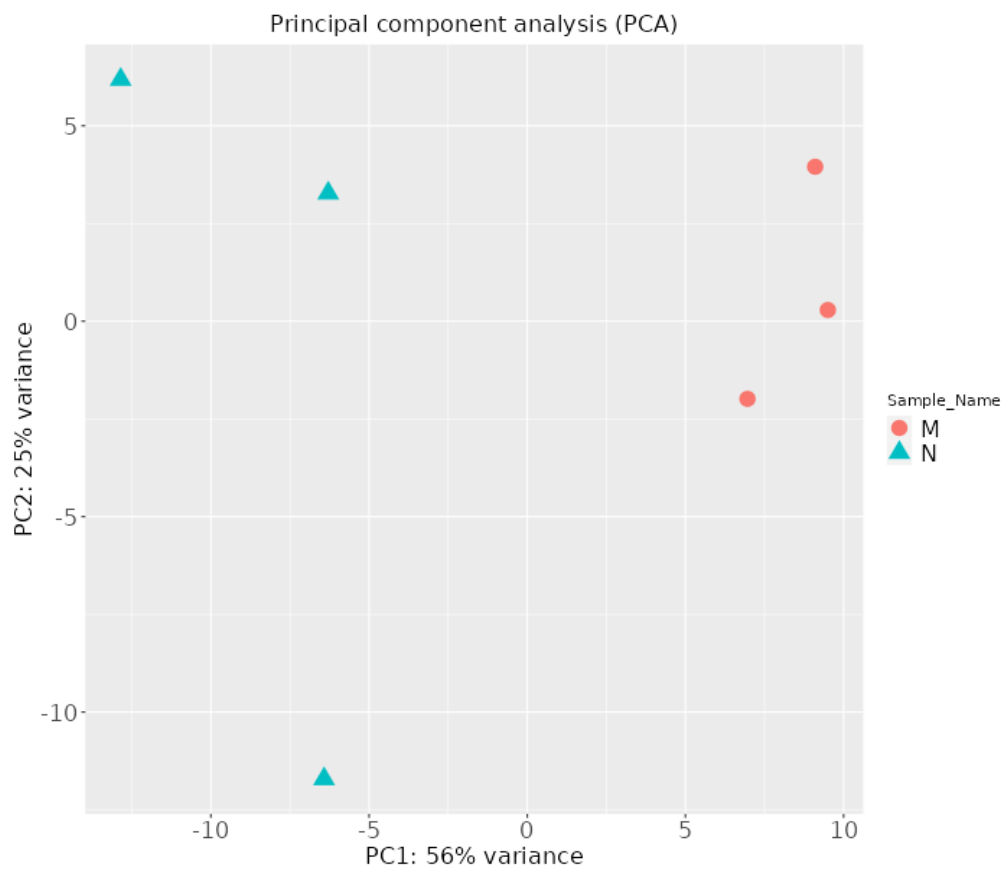
Figure 6.5. Principal component analysis 2.
This figure shows RNA sequence samples from biofilm of *T. forsythia* on NAM vs mucin on a scatter plot that describes general variability in expression profiles. NAM is abbreviated with 'N' and mucin is abbreviated with 'M'

6.2.3.2. Filtering out noise gene expression

After we excluded outliers and assessed inter- and intragroup variability, we checked the distribution of expressed genes between different experimental conditions. This can identify noise within the dataset that was caused by technical factors. One way to filter out the noise is to define a threshold for lower expressed genes based on the sample-to-sample variation. We first plotted and compared the distribution of gene expression for all samples with the same condition (Figure 6.7). Within a specific interval of normalised counts and a threshold of $|> 2|$ for lower expressed genes, the figures showed all samples starting to align and display similar distributions. These indicate high quality data, and it was possible to continue the analysis to consider gene regulation of individual genes.
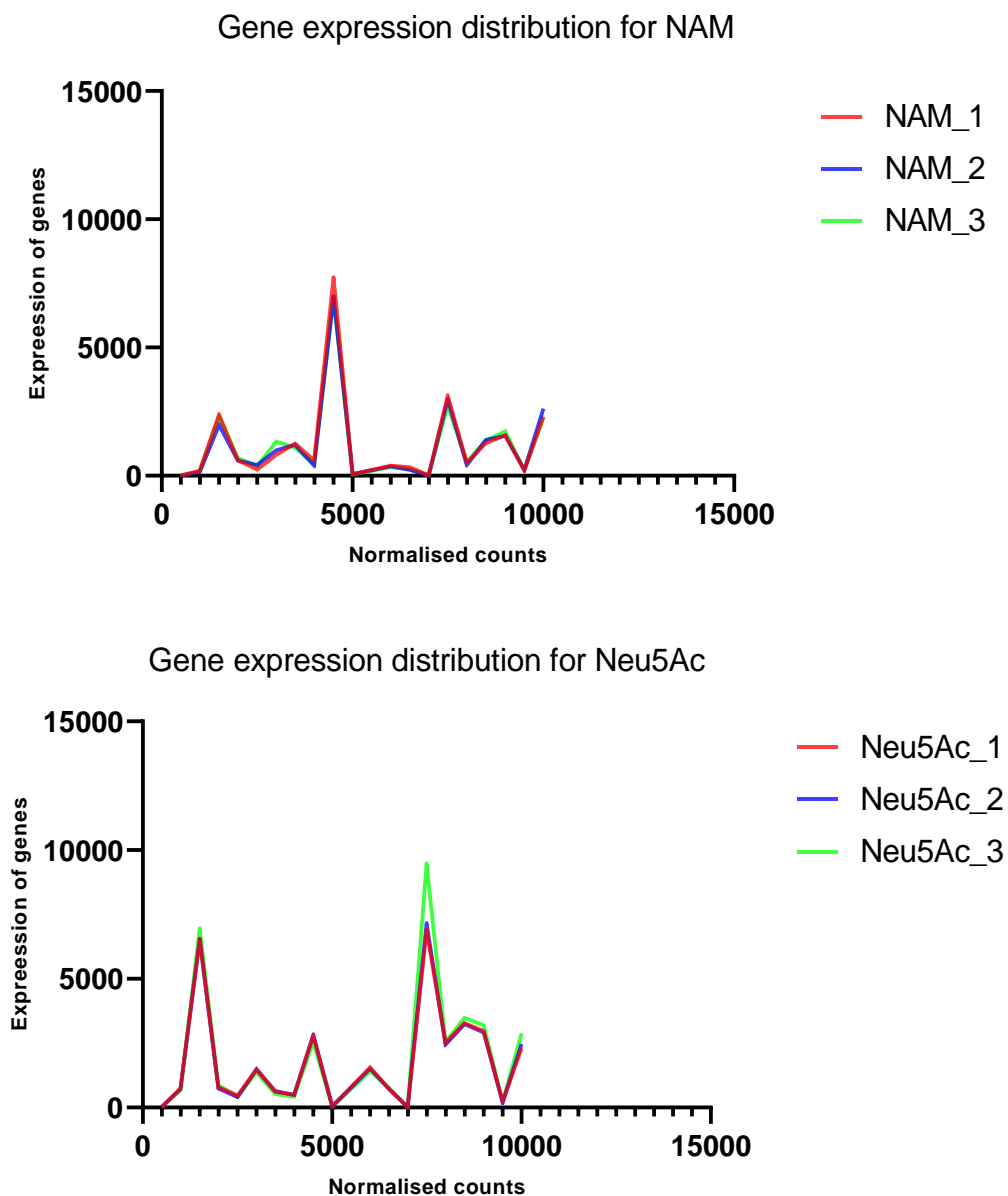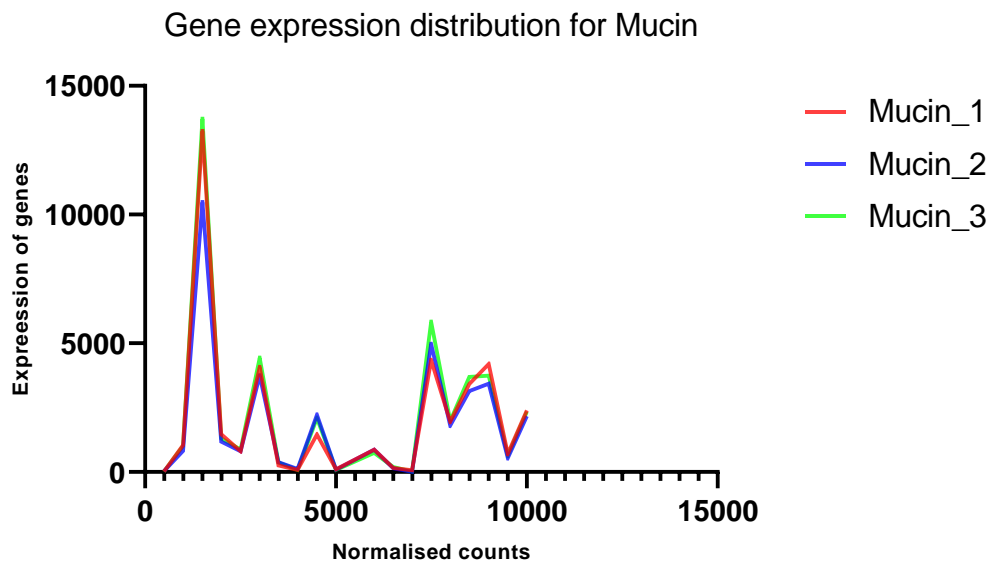
Figure 6.7. Determining a low count threshold.
Distribution of gene expression for all NAM, Neu5Ac, mucin biological sample replicates, detecting low count threshold (noise).

6.2.4. Identification of differential expression genes and visualization (DEGs)

6.2.4.1. Identification of the intensity ratio by the average intensity (MA-plot).

After filtering out noise in the gene expression data, a pairwise comparison between NAM and Neu5Ac, and NAM and mucin was performed. To scientifically identify differential expression genes, we hypothesised that each gene is differentially expressed between NAM and Neu5Ac. In contrast, the null hypothesis is that each gene between the two groups is not differentially expressed, and it has equal expression distribution. Thus, we ran pairwise comparisons to conclude $\log_2$ fold change for each gene, P values, and adjusted P values (padj). Setting the cut-off criteria as $|\log FC| > 2$ and FDR < 0.1, we plotted then $\log_2$ fold change for each gene against the base mean of normalised counts. A generation of an MA plot is used to visually assess the distribution of genes for each pairwise comparison between the two groups (Figures 6.8 and 6.9). As a result, a positive-and a negative- fold change indicated an increase, or a decrease of expression based on a different mean expression for each gene between each two experimental conditions (NAM vs Neu5Ac or NAM vs mucin).
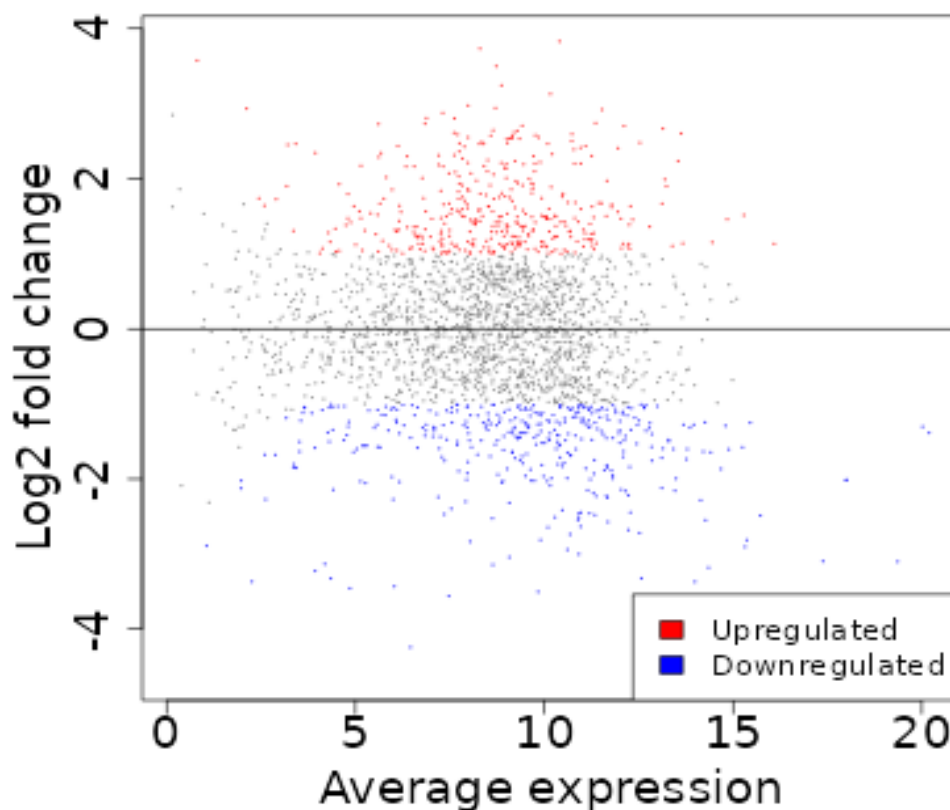
Figure 6.8. Testing for differential expression between NAM and Neu5Ac conditions.

Scatter plot is used to plot the mean of normalised counts against $\log_2$ fold change. Up- and down-regulated differentially expressed genes with a false discovery rate less than 0.05 are shown in red.
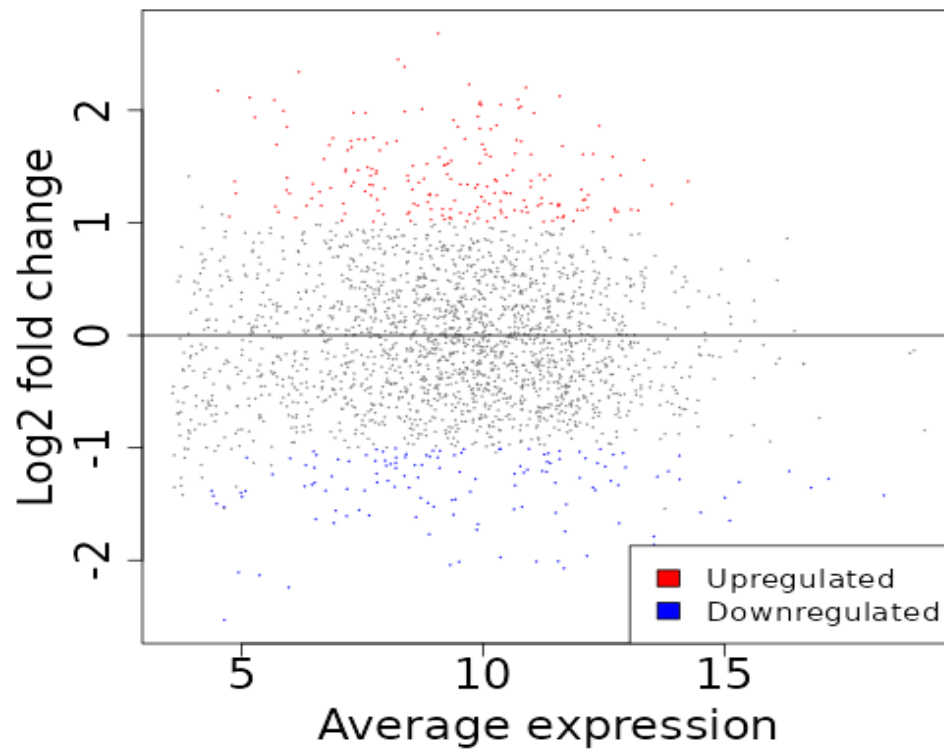
Figure 6.9. Testing for differential expression between NAM and Mucin conditions. Scatter plot is used to plot the mean of normalised counts against $\log_2$ fold change. Up- and down-regulated differentially expressed genes with a false discovery rate less than 0.05 are shown in red.

6.2.4.2. Expression of gene regulation (volcano plot).

The difference in expression is determined by identifying the gene regulation between experimental conditions. The parameter of *P* values at < 0.05 and a minimum fold change |log FC| > 2 and FDR < 0.1 was used to identify the differentially expressed genes. To investigate the replacement of NAM by Neu5Ac and NAM by mucin on the gene expression of *T. forsythtia*, we calculated the significantly up-regulated and down-regulated genes (*P*= < 0.05). The results of the study showed that there were 758 differential expressed genes in the Neu5Ac conditions compared to 363 in the mucin conditions. The up-regulated genes due to the replacement of NAM by Neu5Ac are 379 genes, whereas the down-regulated genes are 379 genes (Figure 6.10). Furthermore, the mucin conditions induces 200 genes in the presence of mucin only (Figure 6.11). Both Neu5Ac and mucin showed their importance in supporting *T. forsythia* biofilm in the absence of NAM.
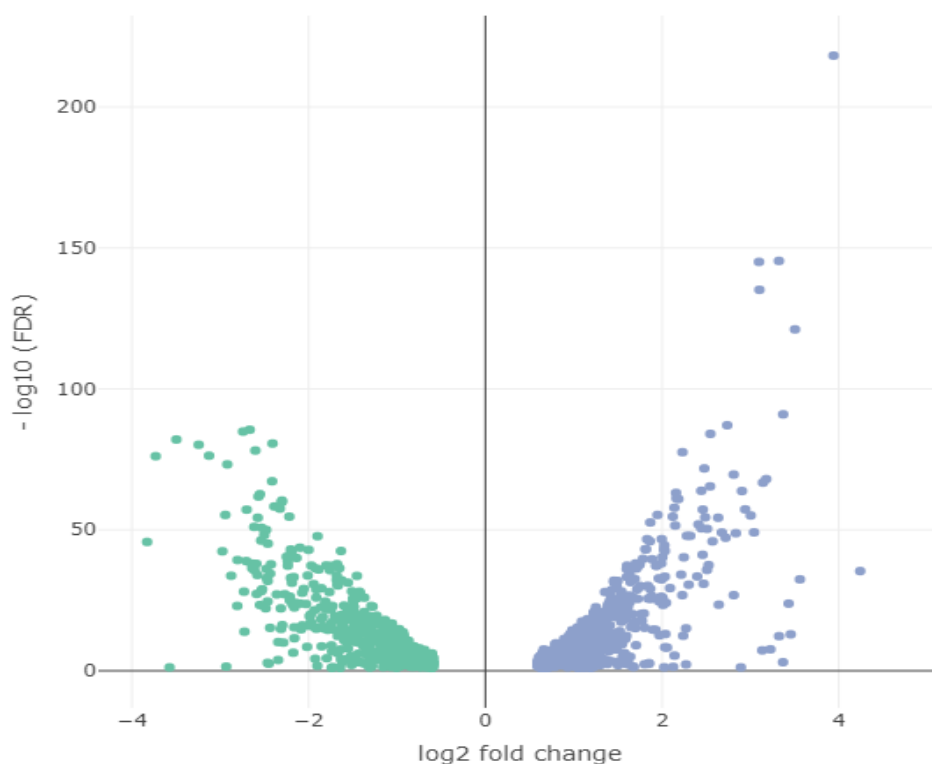
Figure 6.10. Expression of gene regulation in the presence of Neu5Ac versus NAM. In comparison with NAM replacement, the volcano diagram of differential genes presents the total of the up-regulated and down-regulated genes. The vertical axis illustrates the statistically significant degree of changes in gene expression levels. The points represent gene expression. Dark gray dots represent up-regulated differential expression genes and green dots represent down-regulated differential expression genes.
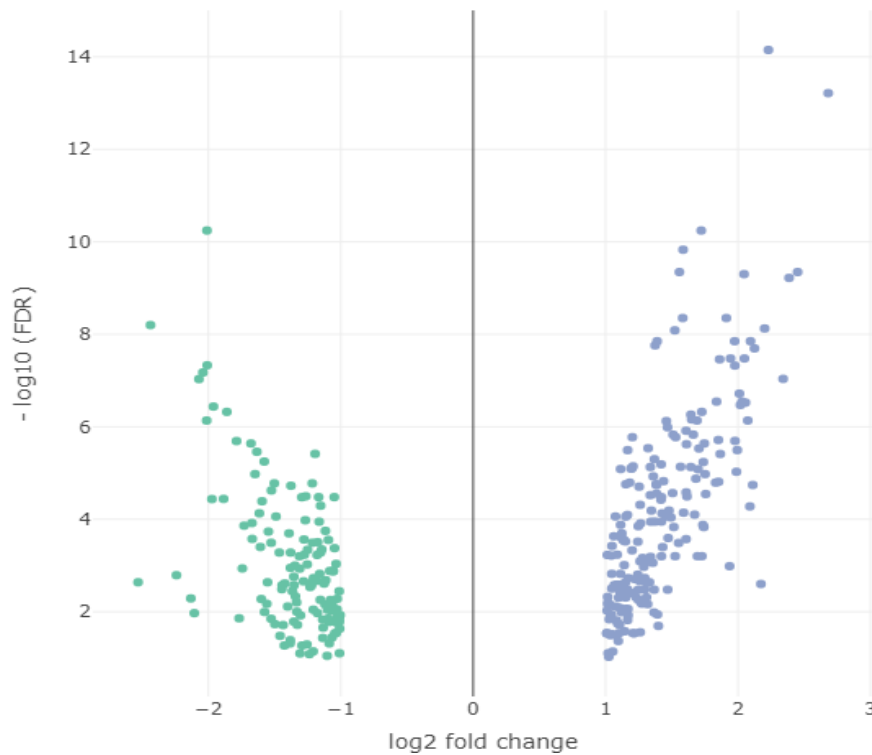
Figure 6.11. Expression of gene regulation in the presence of mucin versus NAM.
In comparison with NAM replacement, the volcano diagram of differential genes presents the total of the up-regulated and down-regulated genes. The vertical axis illustrates the statistically significant degree of changes in gene expression levels. The points represent gene expression. Dark gray dots represent up-regulated differential expression genes and green dots represent down-regulated differential expression genes.

6.2.5. Pathways analysis

6.2.5.1. *k*-Means Clustering

      k-Means clustering is a method of clustering each observation with the nearest mean.[210] Here, *k*-Means can divide all filtered genes from Neu5Ac or mucin conditions into multiple clusters before carrying out enrichment analyses. In the Neu5Ac transcriptomic data and after consulting sum of squares (WSS) on a proper *k*, we settled on $k = 4$ because they accounted for small variation and included all 2343 filtered genes. Genes in cluster A and B are down-regulated in the presence of Neu5Ac compared to clusters C and D (Figure 6.12). In the mucin transcriptomic data, we decided on $k = 6$, including 2335 genes. Genes in clusters A, B, and C are up-regulated compared to other clusters (Figure 6.13). The representation of the enrichment results for each cluster within Neu5Ac or mucin heatmaps is shown as hierarchical clustering trees. One disadvantage is that we noticed multiple annotations for the same gene due to linking between databases, impacting the overall conclusion.
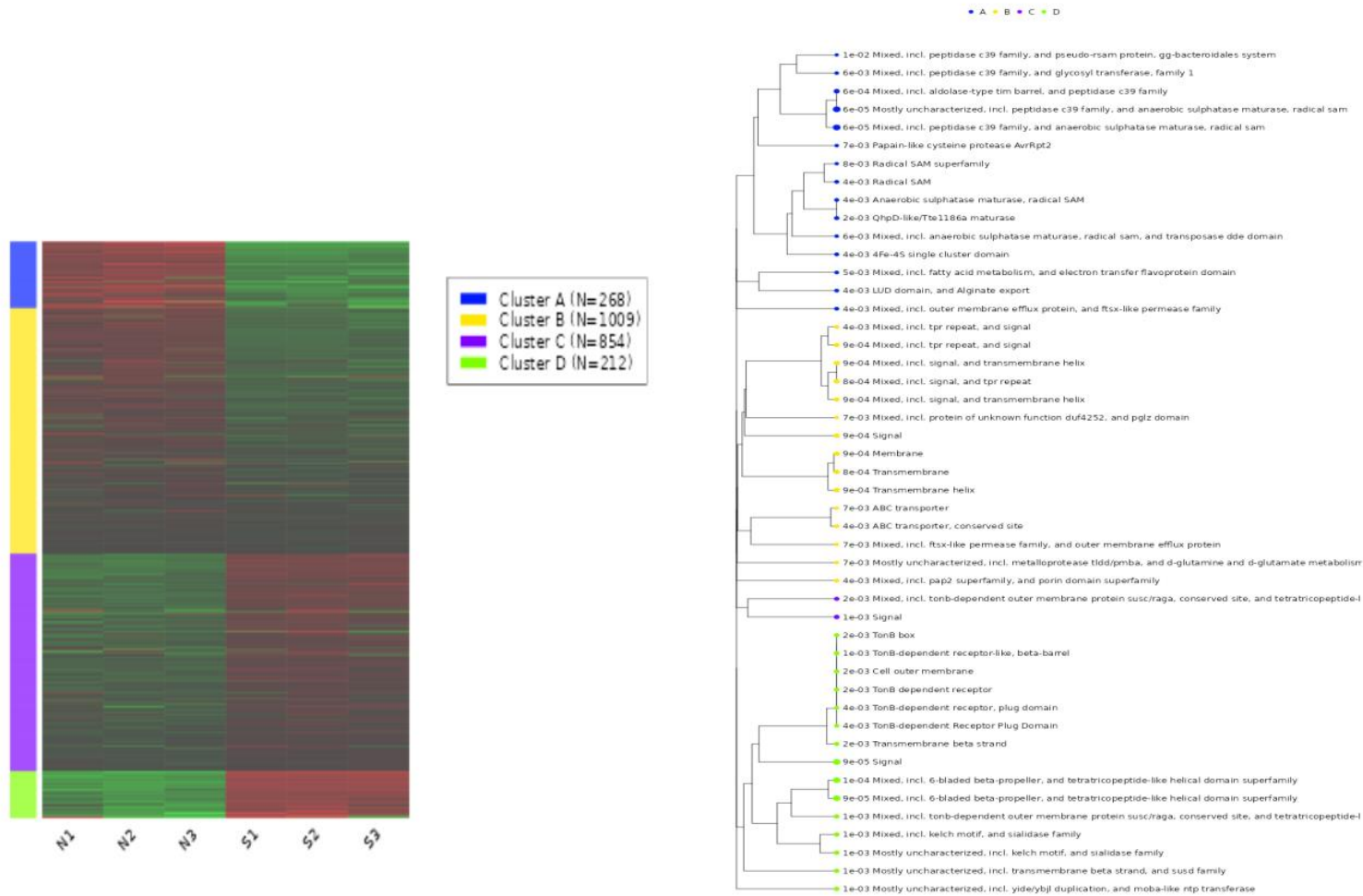
Figure 6.12. *k*-means clustering for Neu5Ac transcriptomic data.
This figure divided 2343 genes from Neu5Ac transcriptomic data into four clusters based on their nearest means before carrying out enrichment analyses. A hierarchical cluster tree shows the enrichment results of each cluster of this heatmap.
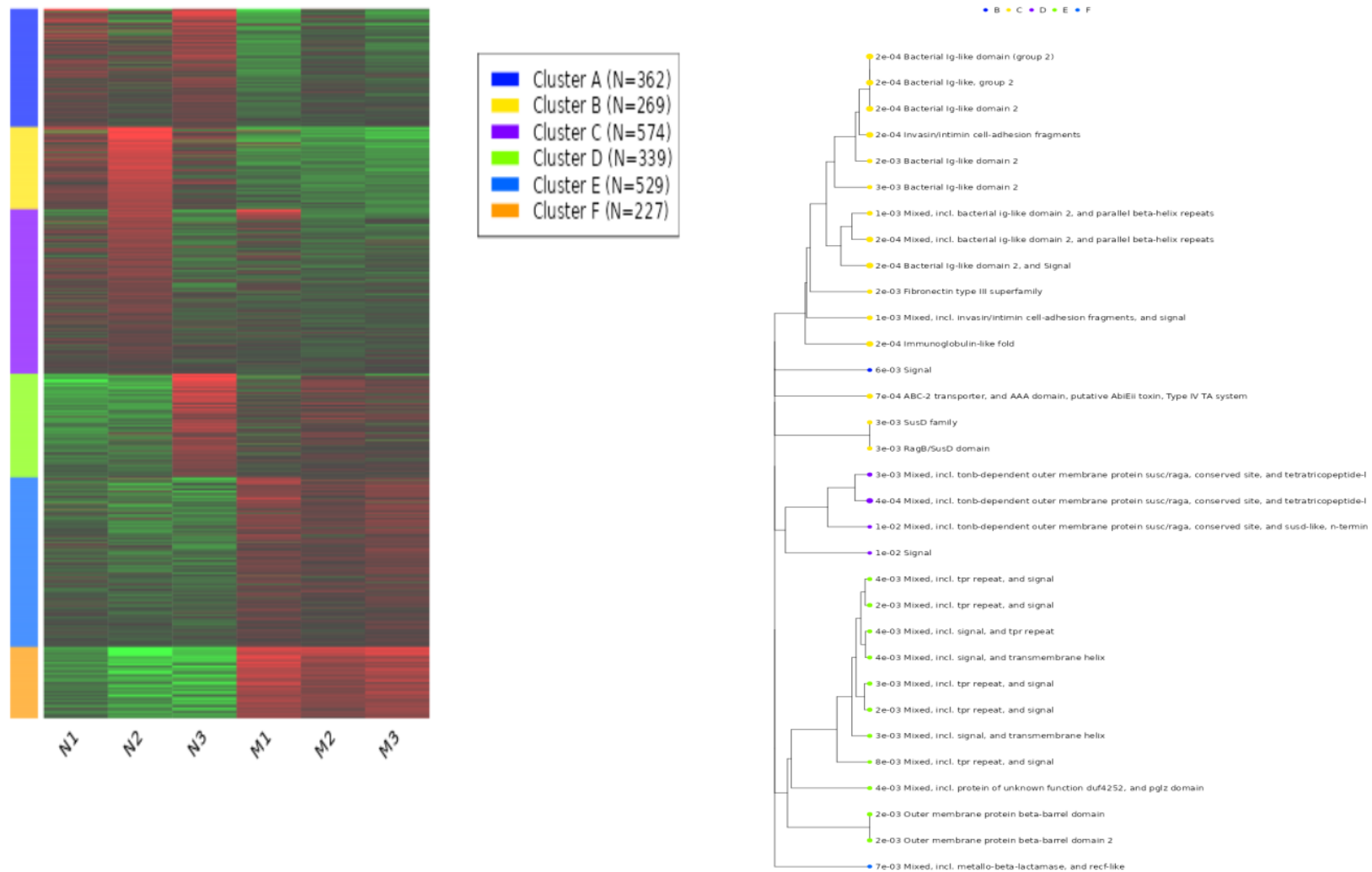
Figure 6.13. *k*-means clustering for mucin transcriptomic data.
This figure 6 shows the divided 2335 genes from mucin transcriptomic data into six clusters based on their nearest means before carrying out enrichment analyses. A hierarchical cluster tree shows the enrichment results of each cluster of this heatmap.

Furthermore, the enrichment analysis of the top 30 up-regulated genes from Neu5Ac transcriptomic data was investigated. The top three pathways with higher fold enrichments included papain-like cysteine protease AvrRpt2, peptidase C39 and bacteriocin processing, and glycosyltransferases (Figure 6.14). In the mucin transcriptomic data, the top three pathways with higher fold enrichments included β protein architecture, pepSY-associated TM protein, and kelch motif (Figure 6.15). Overall, enrichment analysis of top up-regulated genes from Neu5Ac and mucin transcriptomics data were different, suggesting the unique biology of this bacterium on different substrates. Unfortunately, we noticed multiple annotations for the same gene due to the linking between databases and the confusion between *Tannerella* strains.
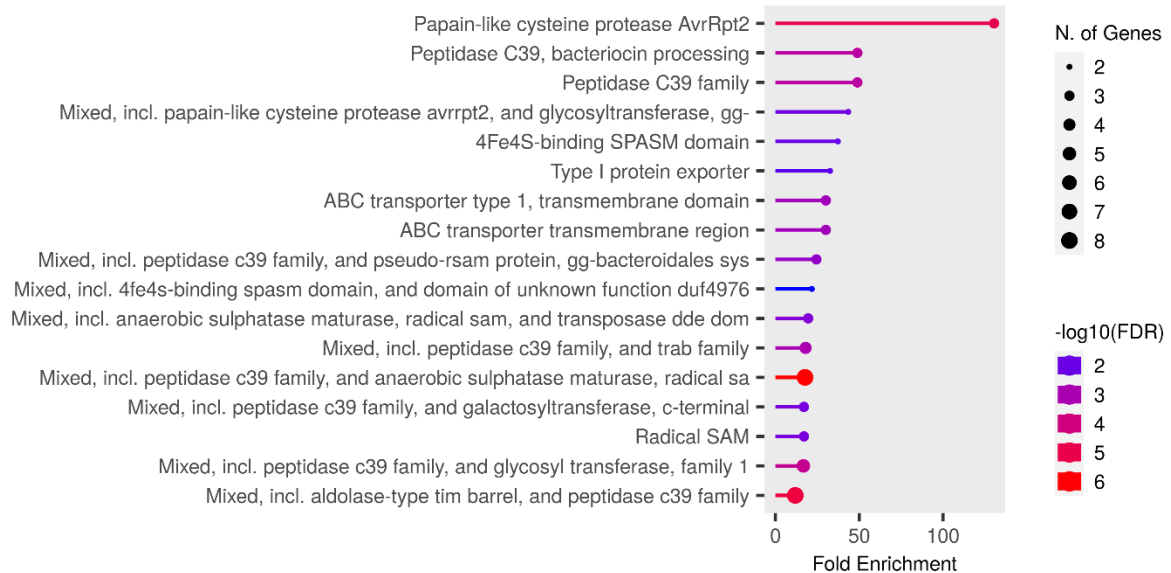
Figure 6.14. Top up-regulated genes from Neu5Ac transcriptomic data.
This figure illustrates the enrichment analysis of the top 30 up-regulated genes from Neu5Ac transcriptomic data.
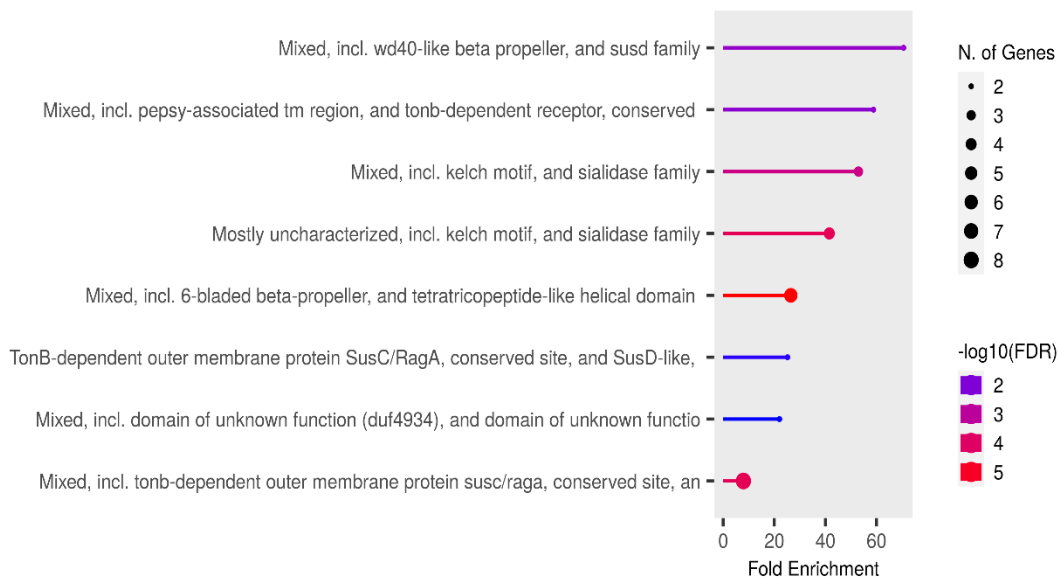
Figure 6.15. Top up-regulated genes from mucin transcriptomic data.
This figure illustrates the enrichment analysis of the top 30 up-regulated genes from mucin transcriptomic data.

6.2.5.2. Pathway activity (PGSEA)

Besides the enrichment analysis from *k*-Means Clustering, PGSEA method is a parametric gene set enrichment analysis (PGSEA) that can estimate the difference in response between treatment and control conditions.[352] This method of enrichment calculates the difference between treatment and control conditions before applying enrichment analysis. The effect of Neu5Ac or mucin on *T. forsythia* was estimated for DEGs. In response to Neu5Ac, we have detected additional significant pathways, such as helicase, zinc, DNA-binding, vitamin b6, immunoglobulin-like fold, PPIC-type, metallo-beta-lactamase, Ribnuclease Z, and porphyrin and chlorophyll metabolism (Figure 6.16). In response to mucin, we have detected additional significant pathways, such as galactose, histidine, and endonuclease (Figure 6.17). With this in mind, the annotation methods from these databases are still poor about *T. forsythia*.
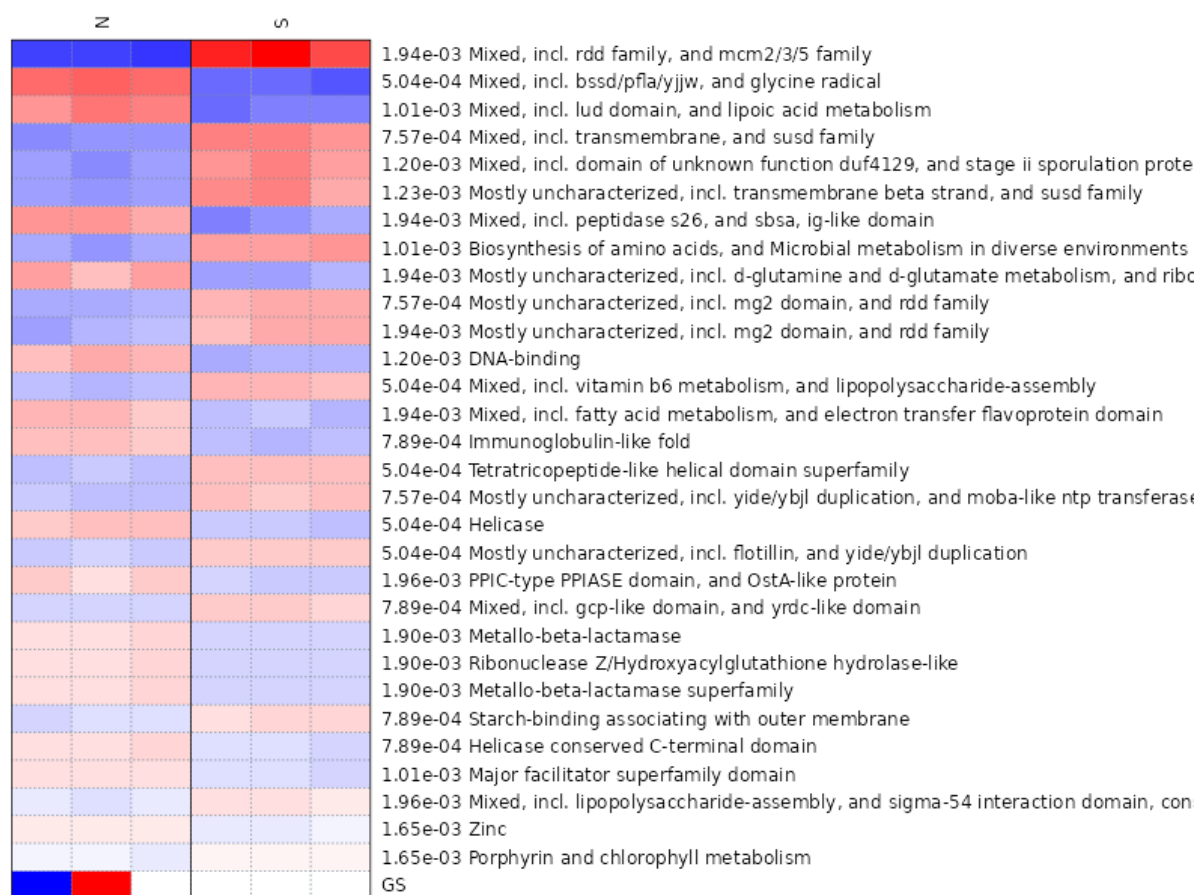
Figure 6.16. Pathways analysis associated with the interaction term in response to Neu5Ac. This figure shows the interaction term based on the fold-changes. All sample groups were analsysed with PGSEA method. Red colour indicates up-regulated pathway genes, whereas the blue colour indicates down-regulated pathway genes.
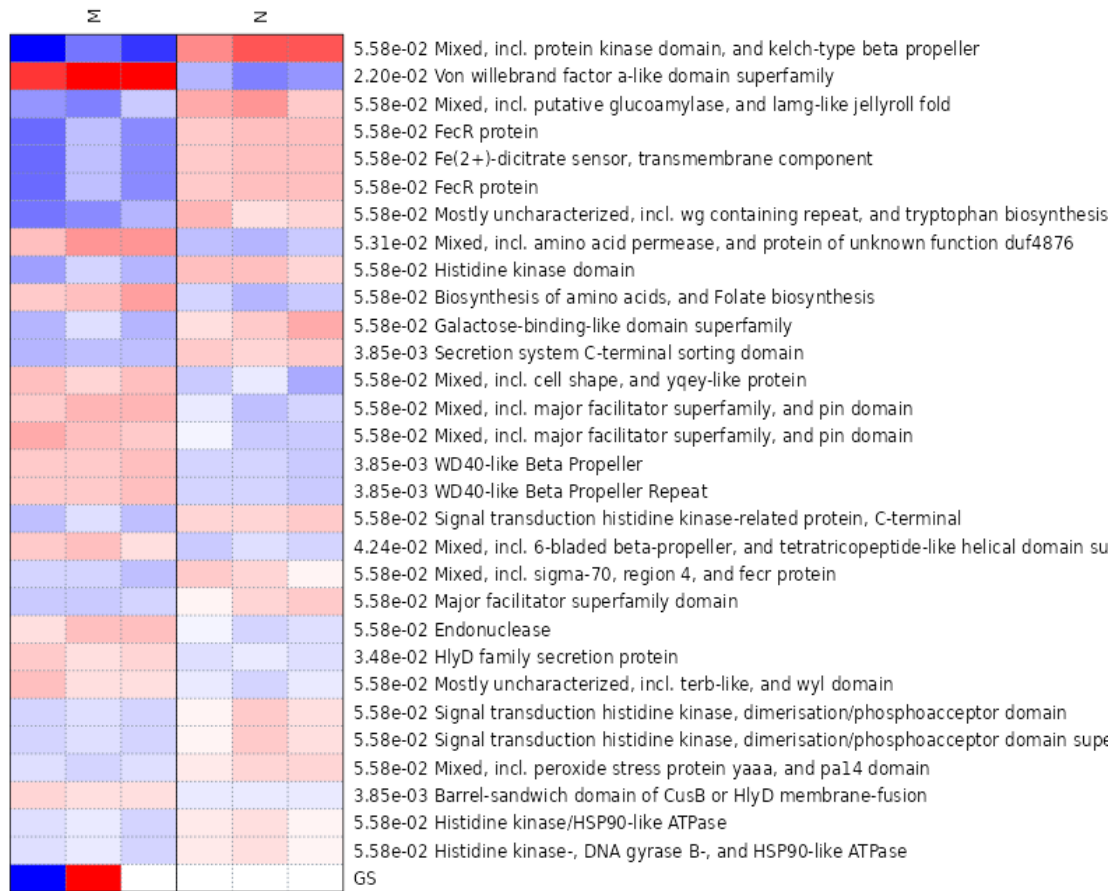
Figure 6.17. Pathways analysis associated with the interaction term in response to mucin. This figure shows the interaction term based on the fold-changes. All sample groups were analsysed with PGSEA method. Red colour indicates up-regulated pathway genes, whereas the blue colour indicates down-regulated pathway genes.

6.2.5.3. Functional Enrichment Analysis of DEGs- Gene Ontology

We sought the expression of Gene Ontology (GO) including biological process, molecular function, and cellular component of DEGs from Neu5Ac and mucin conditions. Based on the 'Comparative Go', webserver for comparative GO,[211] we performed GO analyses for all up-regulated and down-regulated DEGs in NAM vs Neu5Ac and NAM vs mucin. (Appendix VIII: Tables 2, 3, and 4 for up-regulated and tables 5,6, and 7 for down-regulated genes in Neu5Ac). (Appendix VIII: Tables 8, 9, 10, for up-regulated and tables 11, 12, and 13 for down-regulated in mucin). To probe more about the biology of *T. forsythia* during biofilm on Neu5Ac or mucin, the top 50 GO biological processes indicated that the up-regulated DEGs based on NAM vs Neu5Ac were enriched in metabolic process (50), cellular regulation (39%), and biological regulation (11%) (Figure 6.18-A), while the down-regulated DEGs were enriched in cellular process (47%), metabolic process (47%), biological regulation (3%), and localisation (3%) (Figure 6.19-A). Furthermore, the top 50 GO cellular components were enriched in integral components of membrane (72%), plasma membrane (11%), beta-galactosidase complex (7%), cytoplasm (5%), and cell periphery (5%) (Figure 6.18-B). The 50 GO cellular component of down-regulated genes were enriched in integral component of membrane (31%), ribosome (20%), cell outer membrane (16%), plasma membrane (16%), cytoplasm (14%), and ATP-binding cassette (4%) (Figure 6.19-B). The top 50 up-regulated enriched genes in molecular function include catalytic activity (43%), binding (33%), transcription regulator activity (11%), ATP-dependent activity (9%), transporter activity (2%), and small molecule sensor activity (2%) (Figure 6.18-C). The top 50 down-regulated enriched genes in molecular function include catalytic activity (39%), binding (36%), structural molecule activity (16%), translation regulator activity (6%), and transporter activity (3%) (Figure 6.19-C).

In mucin conditions, when we examined the top 50 enriched genes to biological process, we concluded the metabolic process (52%), cellular process (46%), and localisation (2%) (Figure 6.20-A). The down-regulated enriched genes in biological process include metabolic process (44%), cellular process (32%), biological regulation (12%), response to stimulus (5%), localisation (3%), and biological process in interspecies interaction (3%) (Figure 6.21-A). The cellular component of up-regulated includes cytoplasm (29%), ribosome (25%), plasma membrane (9%), cell outer membrane (7%), large ribosomal (5%), small ribosomal (5%), integral component of membrane (15%), ATP binding (2%), proton-transporting ATP synthase (2%), and proton transporting ATP synthase (2%) (Figure 6.20-B). The down-regulated genes include integral component of membranes (60%), cell outer

membrane (18%), plasma membrane (14%), and cytoplasm (8%) (Figure 6.21-B). The up-regulated genes in molecular function were assigned to binding (39%), structural molecular activity (29%), catalytic activity (27%), transport activity (4%), and translation regulator activity (2%) (Figure 6.20-C). The down-regulated genes in molecular function were assigned to catalytic activity (53%), binding (34%), ATP-dependent activity (6%), transporter activity (3%), small molecule sensor activity (2%), and transcription regulator activity (2%) (Figure 6.21-C). Overall, it was difficult to extract accurate information about *T. forsythia* under Neu5Ac and mucin conditions due to the confusion about this bacterium in databases.
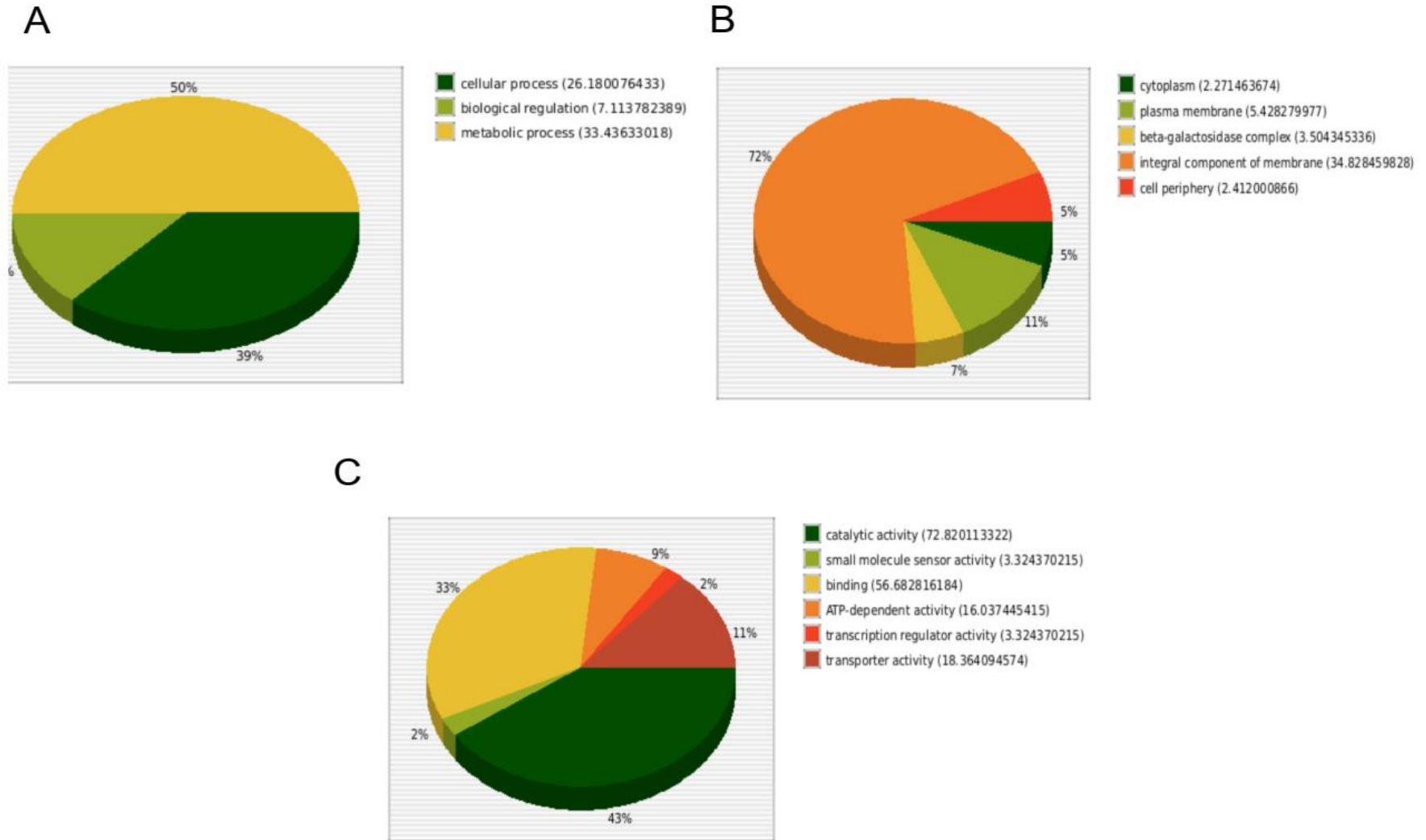
Figure 6.18. Gene ontology of top 50 up-regulated genes in Neu5Ac conditions.
This figure shows three main GO categories with details and proportion analysis: biological process (A), cellular component (B), and molecular function (C).

271
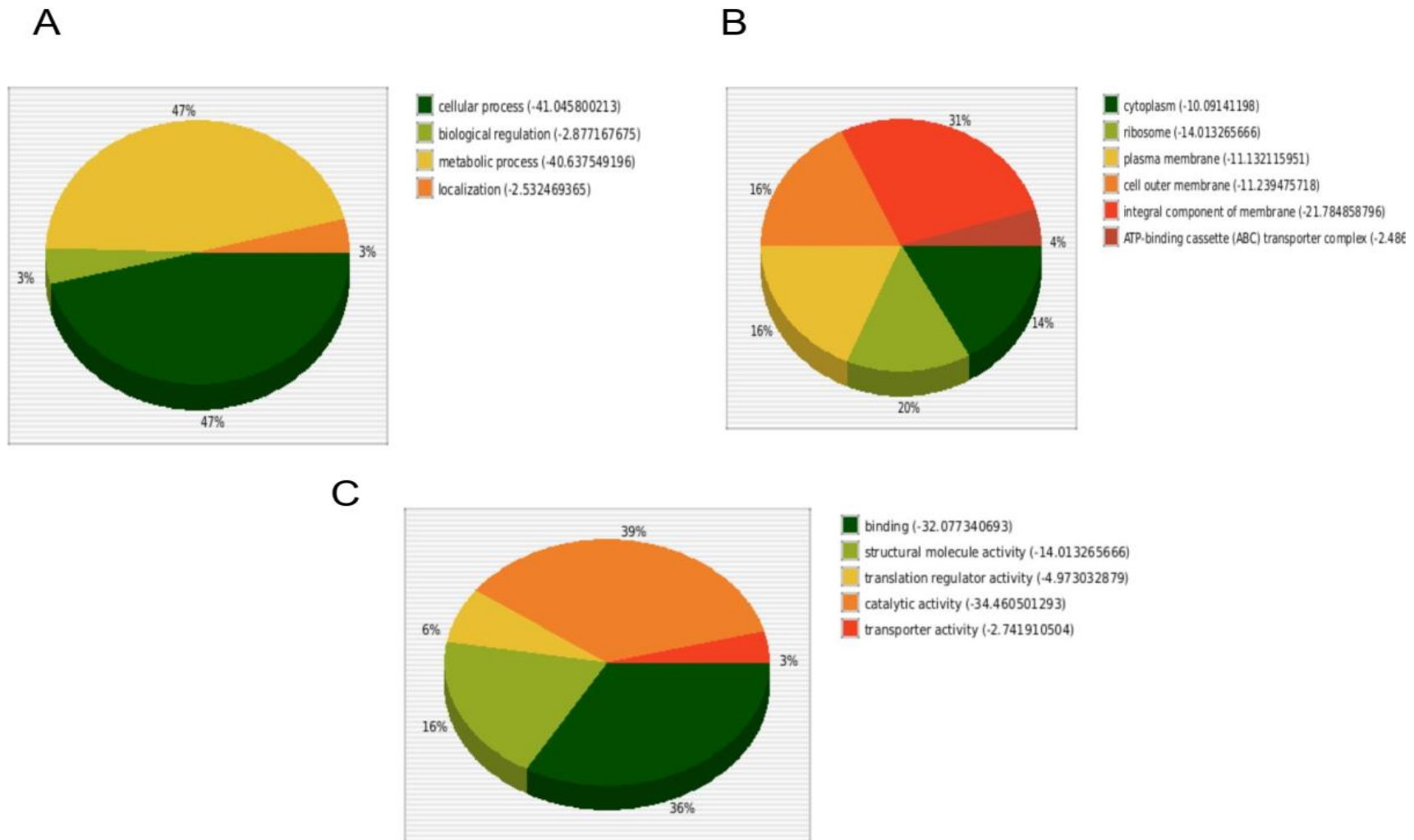
Figure 6.19. Gene ontology of top 50 down-regulated genes in Neu5Ac conditions.
This figure shows three main GO categories with details and proportion analysis: biological process (A), cellular component (B), and molecular function (C).
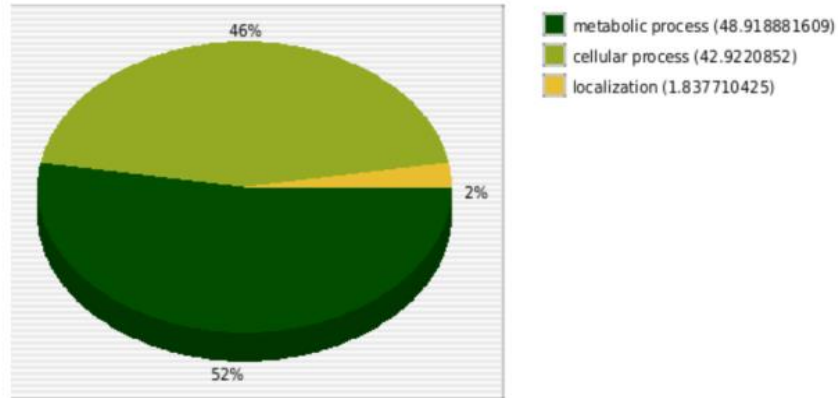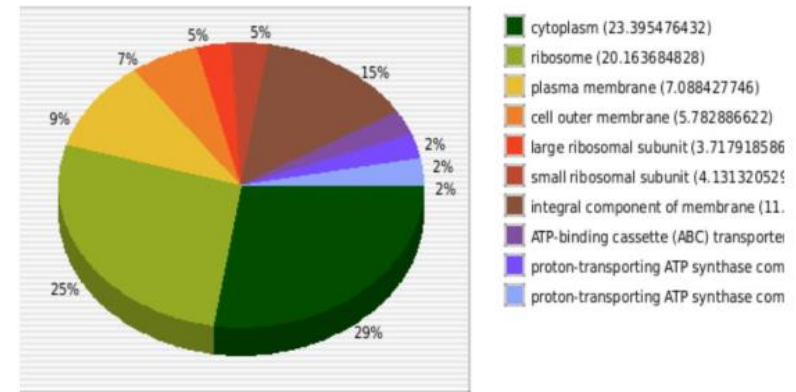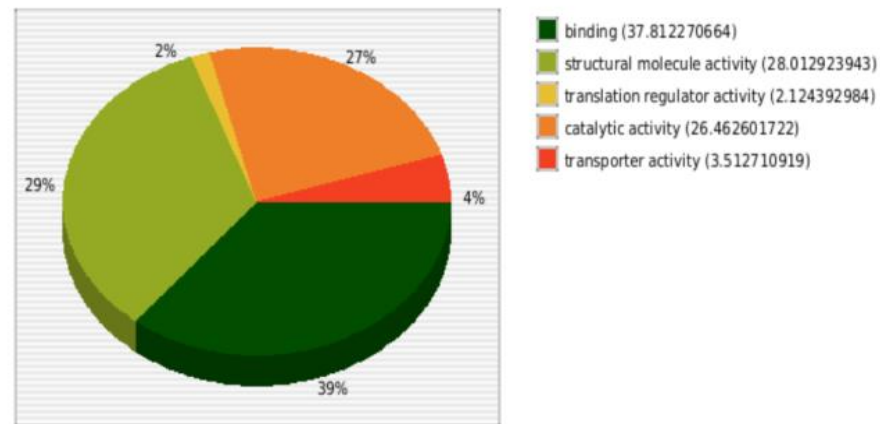
Figure 6.20. Gene ontology of top 50 up-regulated genes in mucin conditions.
This figure shows three main GO categories with details and proportion analysis: biological process (A), cellular component (B), and molecular function (C).

273
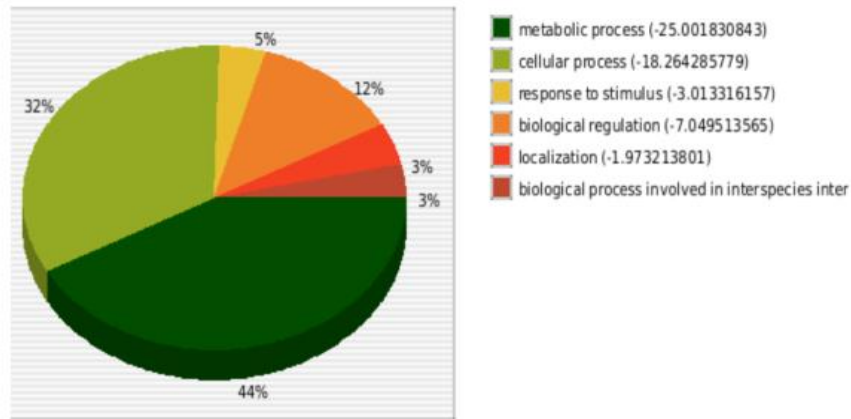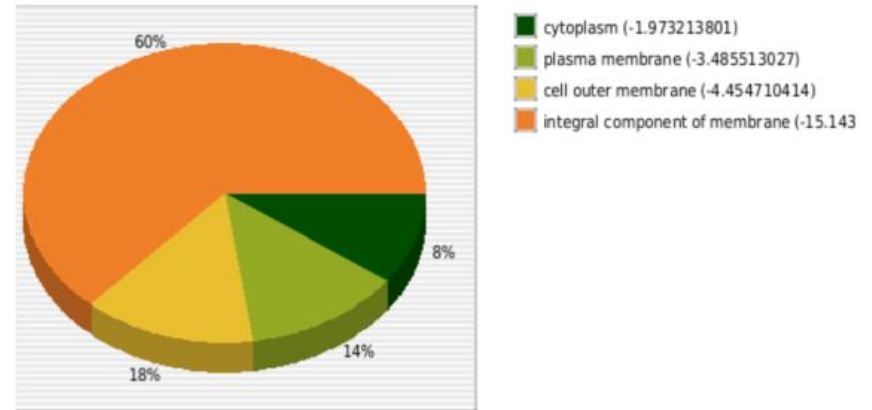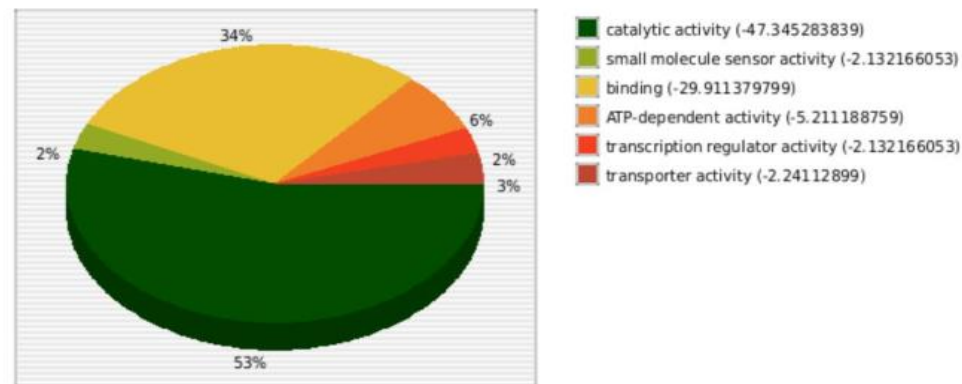
Figure 6.21. Gene ontology of top 50 down-regulated genes in mucin conditions.
This figure shows three main GO categories with details and proportion analysis: biological process (A), cellular component (B), and molecular function (C).

6.2.5.4. KEGG (Kyoto Encyclopedia of Genes and Genomes) Pathway Enrichment

Expression of significant data was mapped to KEGG database and Kobas' enrichment tools to observe which sections of gene functions were differentially expressed.[212,213] Here, the top 50 up-regulated or down-regulated genes to summarise KEGG pathway linked with higher order functional information. In response to Neu5Ac, five clusters were enriched in the metabolism of cofactors and amino acids (cluster 1), carbohydrate metabolism (cluster 2), cellular pathway (cluster 3), biological process (cluster 4), and genetic and environmental information process (cluster 5) (Figure 6.22). KEGG analysis revealed that the down-regulated top 50 DEGs were mainly enriched in metabolism (clusters 1 and 2), cellular pathway (cluster 3), carbohydrate metabolism (cluster 4), and genetic information processing (cluster 5) (Figure 6.23).

In response to mucin, the results from clusters 1 and 2 pinpoint the metabolism. Cluster 3 included cellular pathway (Figure 6.24). The down-regulated top 50 genes were generally enriched in the metabolism and environmental information process (Figure 6.25).
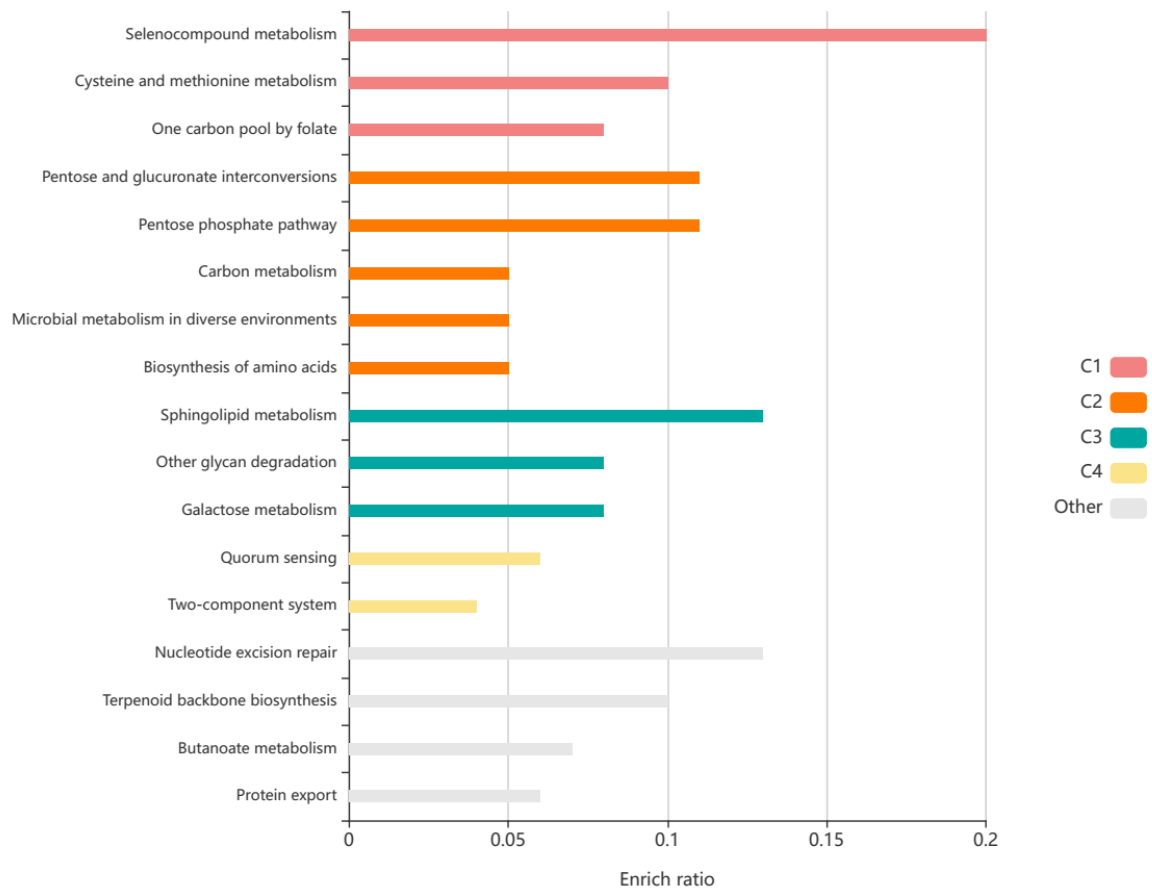
Figure 6.22. KEGG analysis of top 50 up-regulated DEGs in Neu5Ac.
This figure shows KEGG pathways classification in different clusters. The x-axis presents the enrich ratio, and the y-axis presents the name of the KEGG pathway classification. Five clusters were enriched in metabolism of cofactors and amino acids (cluster 1), carbohydrate metabolism (cluster 2), cellular pathway (cluster 3), biological process (cluster 4), and genetic and environmental information process (cluster 5).
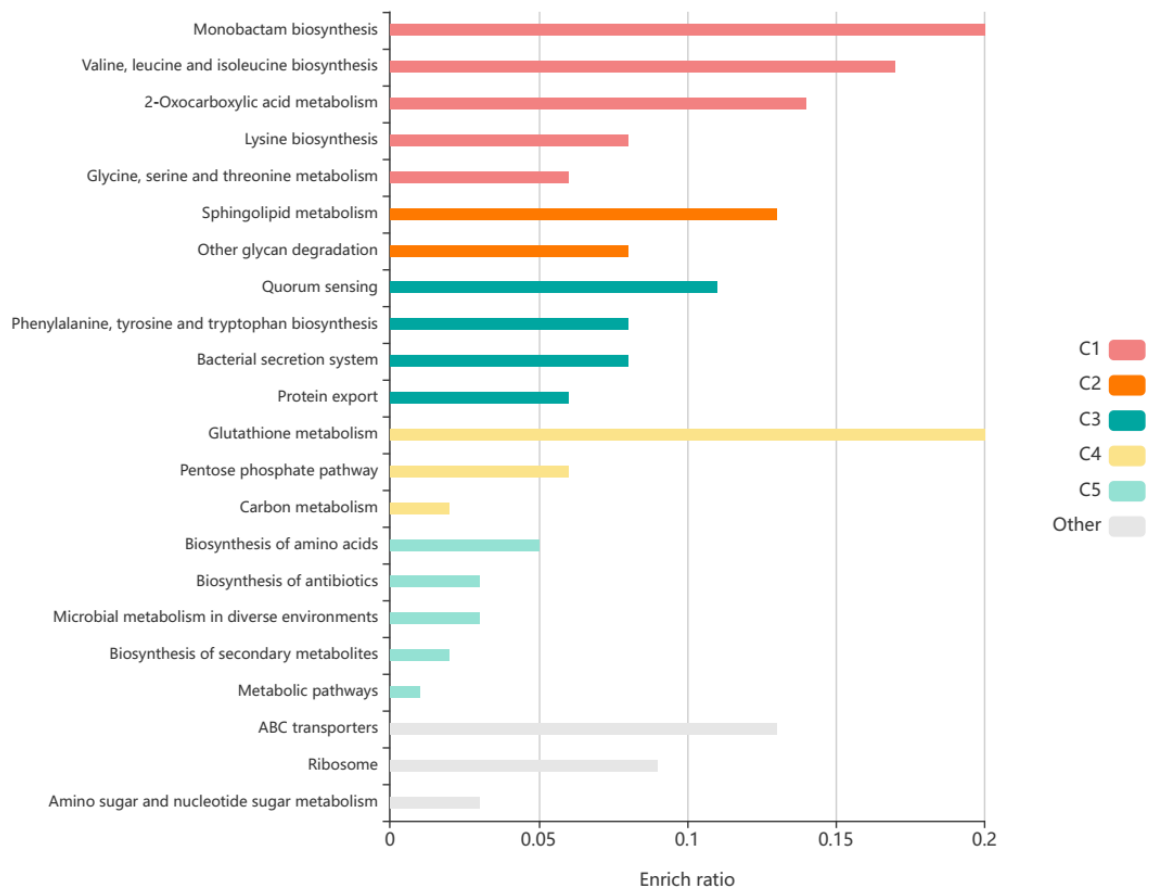
Figure 6.23. KEGG analysis of top 50 down-regulated DEGs in Neu5Ac.
This figure shows KEGG pathways classification in different clusters. The x-axis presents the enrich ratio, and the y-axis presents the name of the KEGG pathway classification. KEGG analysis revealed that the down-regulated top 50 DEGs were mainly enriched in metabolism (clusters 1 and 2), cellular pathway (cluster 3), carbohydrate metabolism (cluster 4), and genetic information processing (cluster 5).
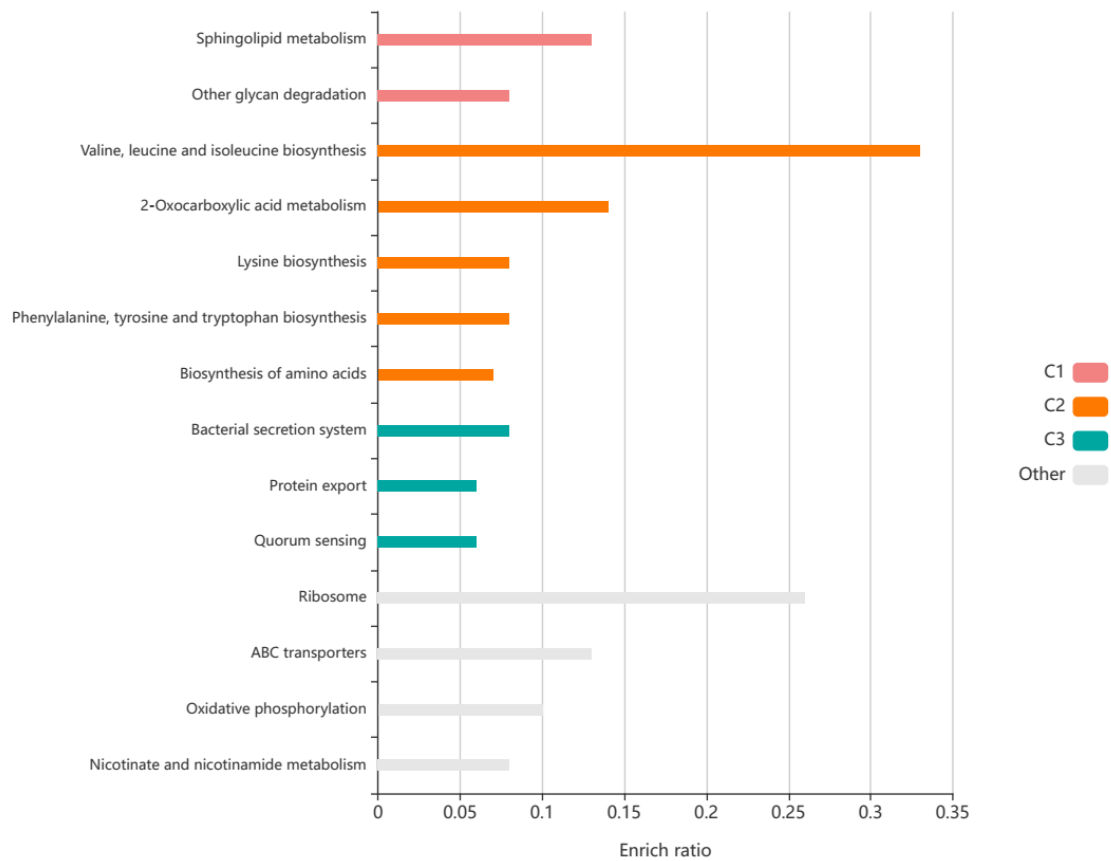
Figure 6.24. KEGG analysis of top 50 up-regulated DEGs in mucin.
This figure shows KEGG pathways classification in different clusters. The x-axis presents the enrich ratio, and the y-axis presents the name of the KEGG pathway classification. In response to mucin, the results from cluster 1 and 2 pinpoint the metabolism. Cluster 3 included cellular pathway.
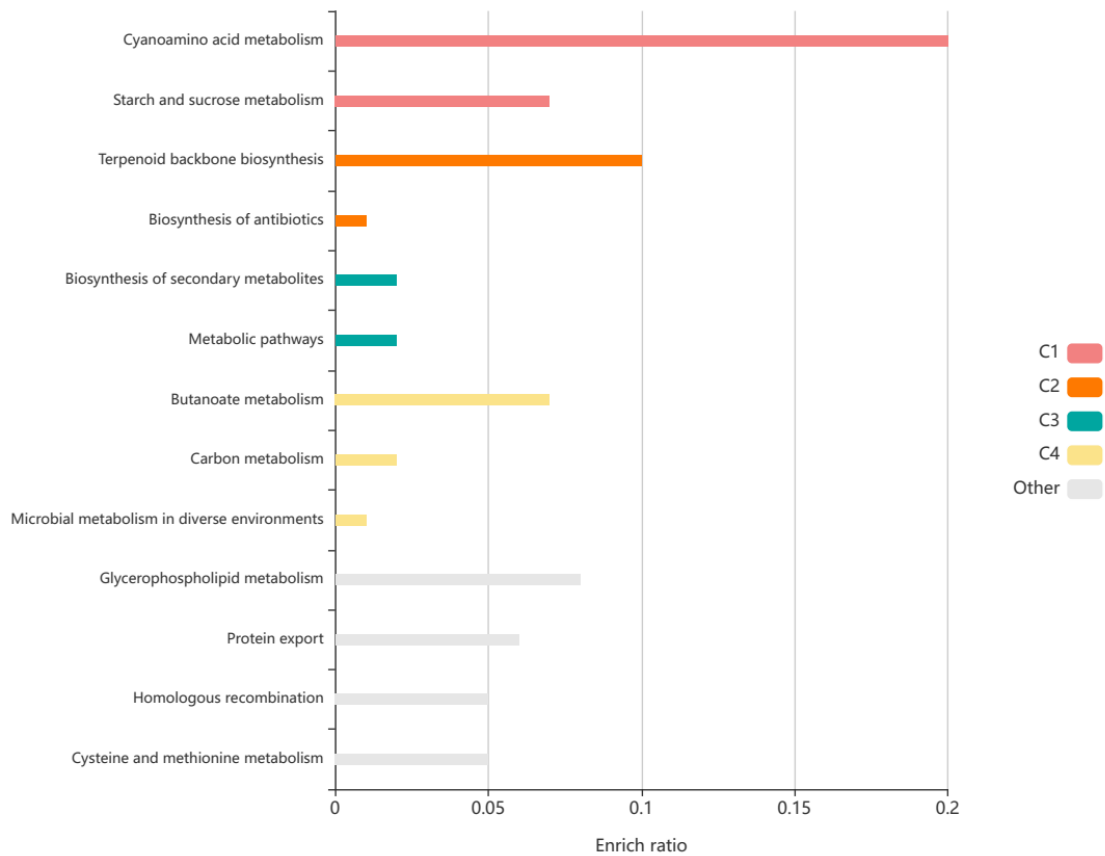
Figure 6.25. KEGG analysis of top 50 down-regulated DEGs in mucin.
This figure shows KEGG pathways classification in different clusters. The x-axis presents the enrich ratio, and the y-axis presents the name of the KEGG pathway classification. The down-regulated top 50 genes of were generally enriched in metabolism and environmental information process.

6.2.5.5. Construction of Weighted Correlation Network Analysis and Identification of Key Modules

Correlation networks can illustrate correlation patterns among genes across DNA/RNA samples. These correlation networks can enable gene screening methods to classify and recognise various biological contexts like bacteria, identifying biomarkers or therapeutic targets. Weighted correlation network analysis (WGCNA) is based on the eigengene network methodology to cluster and measure highly correlated genes before their clustering.[353] Here, we used all DEGs within this study from the results of NAM vs Ne5Ac or NAM vs mucin to produce co-expression modules. Two gene clustering trees were generated for topological overlap matrix (TOM) where each module was calculated based on its corresponding Neu5Ac or mucin characteristics. In the presence of Neu5Ac, correlation analysis revealed three different modules where the turquoise module has the highest relationship between genes, including 551 genes (Figure 6.26). In the presence of mucin, correlation analysis revealed nine different modules, where the blue module has the highest relationship between genes including 200 genes (Figure 6.27). The list of top 10 genes from both trees is shown in two network layouts, and they were further used for gene enrichment analysis. The top 10 Neu5Ac correlated genes were related to UDP-glucose/GDP-mannose dehydrogenase, regulator propeller, and other domains. The top 10 mucin correlated genes were related to RNA polymerase signa factor, DNA biding HTH domain, starch and sucrose metabolism, and other domains.
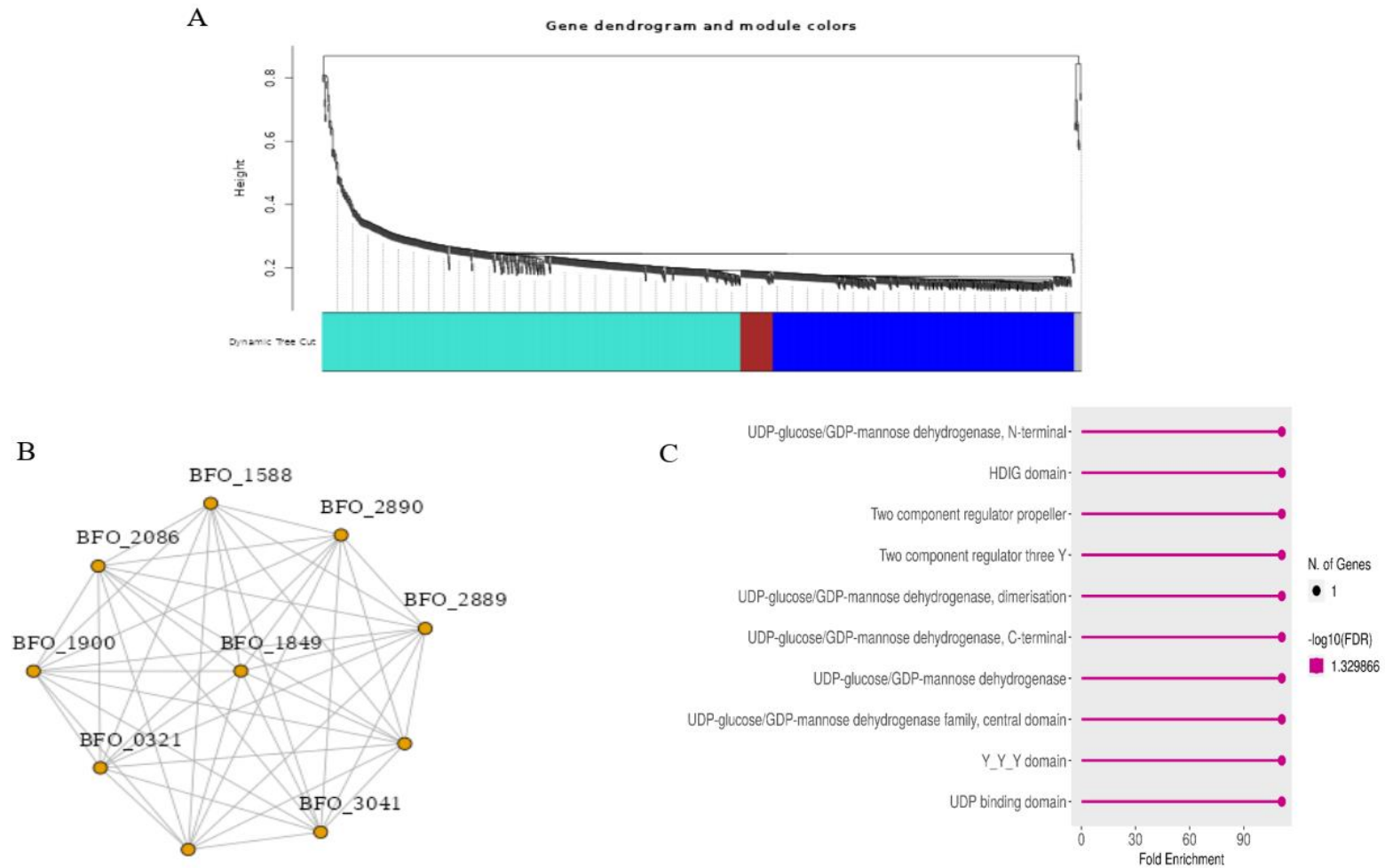
Figure 6.26. Correlation Network Analysis of Neu5Ac transcriptomic dataset.
This figure 6 illustrates the correlation between RNA sequence data based on WGCNA method. A. This panel shows the cluster diagram of co-expression network modules based on correlation patterns of genes. B. This panel illustrate the top 10 correlated Neu5Ac genes.
C. This panel shows the enrichment analysis of top 10 correlated Neu5Ac genes.
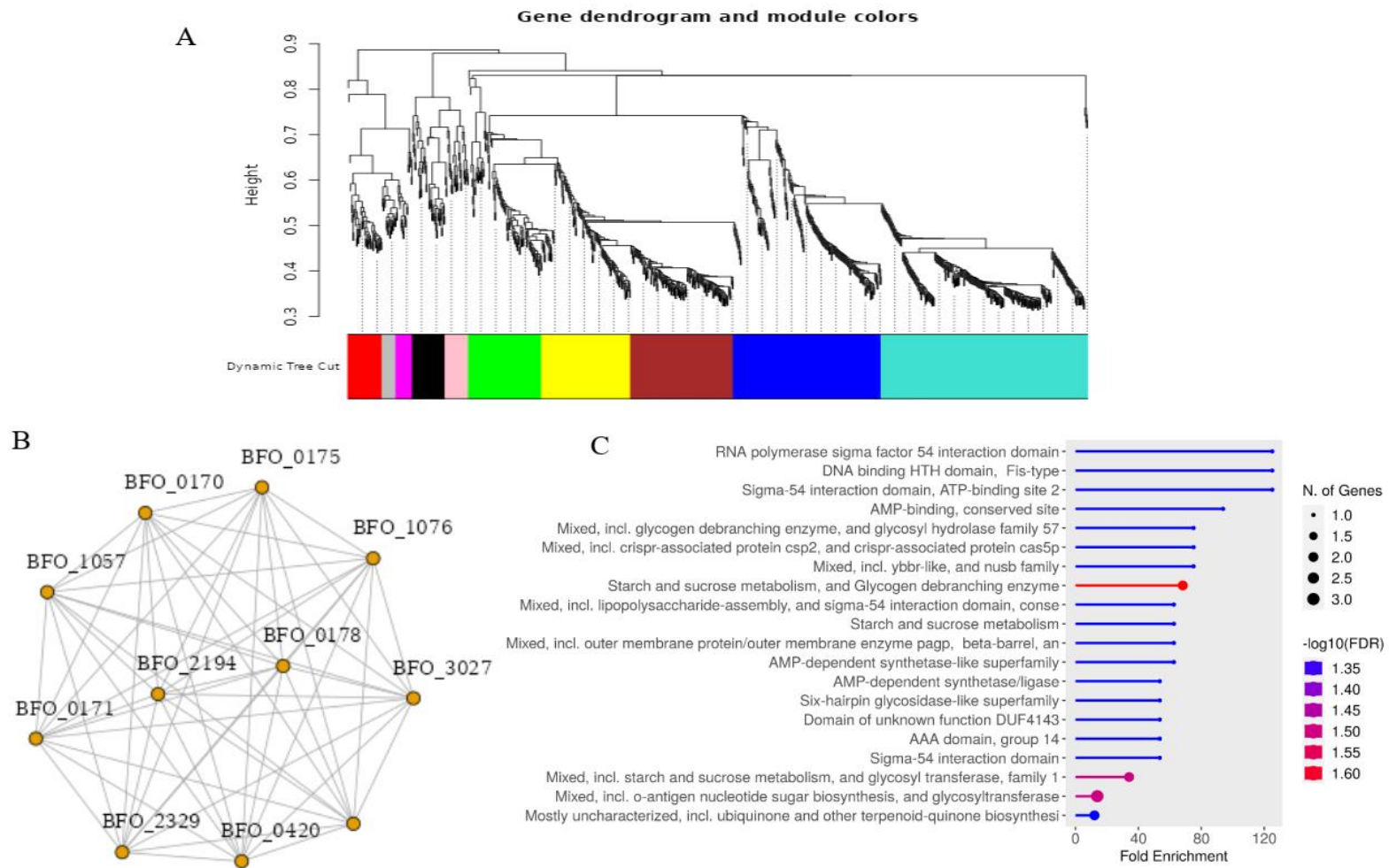
Figure 6.27. Correlation Network Analysis of mucin transcriptomic dataset.
This figure 6 illustrates the correlation between RNA sequence data based on WGCNA method. A. This panel shows the cluster diagram of co-expression network modules based on correlation patterns of genes. B. This panel illustrate the top 10 correlated mucin genes.
C. This panel shows the enrichment analysis of top 10 correlated mucin genes.

6.2.5.6. Protein–Protein Interaction (PPI) Network Analysis

To further explore the interplay among *T. forsythia* transcriptomics grown on Neu5Ac or mucin, we performed a PPI network based on STRING online database.[354] In the network, we assessed the top 50 up-regulated genes in Neu5Ac to show high significant interaction between these proteins. Interestingly, proteins highlighted in blue colour were enriched to be part of peptidase C39 family to the superfamily of ABC transporters. The C39 family can cleave glycine from various bacteriocins. For instance, the Gram-positive bacterium *Lactococcus lactis* can produce precursor peptide like Lanthipeptides (LanA) that has an additional domain (C39 peptidase) for the modification enzymes and the export protein.[355] Furthermore, proteins highlighted in red colour were enriched to be part of putative glucoamylase (lamg). These proteins are involved in the hydrolysis of starch (Figure 6.28).[356]

Likewise, we examined the top 50 up-regulated genes in mucin biofilm to show high significant interaction between these proteins. Proteins highlighted in red colour were enriched to be part of structural molecule activity. Proteins highlighted in green, yellow, and blue colours were described with initiation factor and Ribosomal protein S5, Ribosome, and kelch motif, and sialidase family, respectively. We noticed the presence of *nanH* and *nanhA* (part of *nan* operon) among the top 50 up-regulated genes (Figure 6.29).
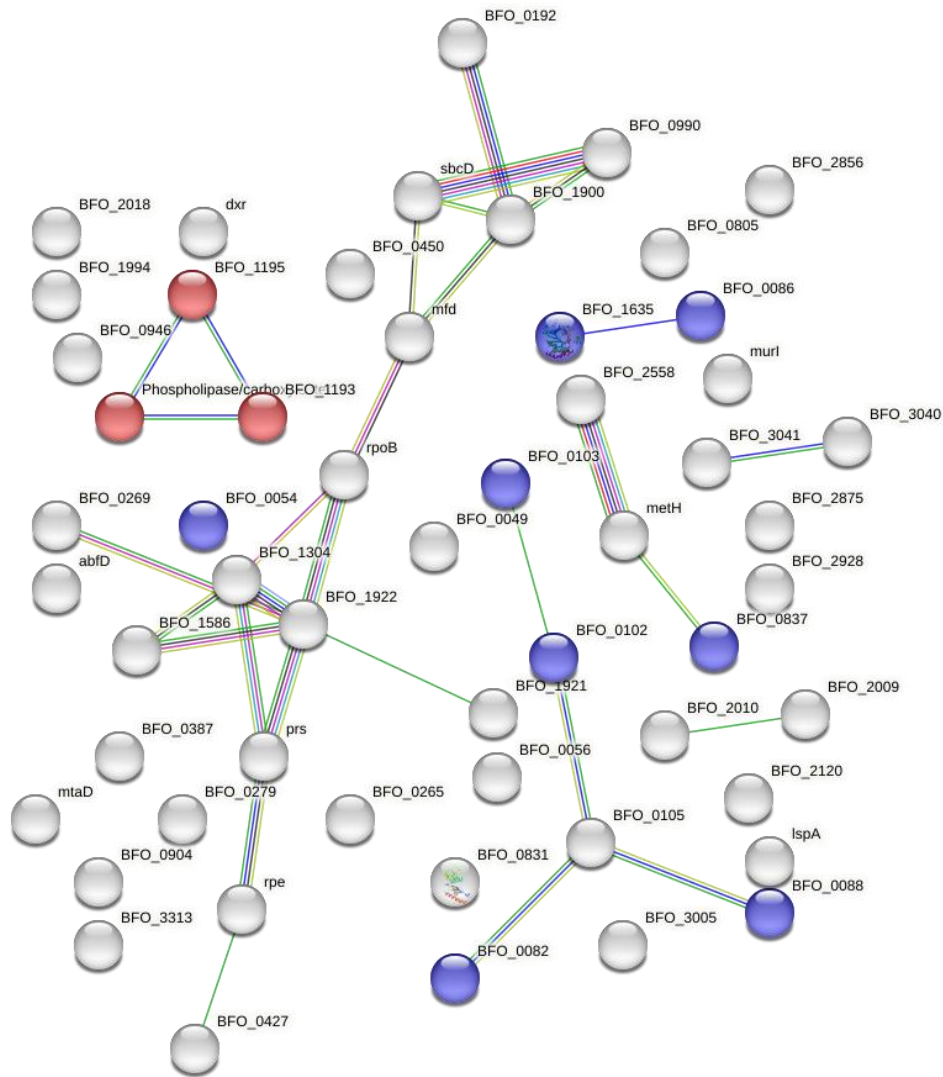
Figure 6.28. Protein–protein interaction in Neu5Ac.
This figure shows protein–protein interaction of the top 50 up-regulated genes in response to
Neu5Ac. The blue and red colour represent high interaction proteins belonging to C39 family
and hydrolysis of starch, respectively. The presence of edges between proteins indicates
protein interactions. More edges illustrate evidence base for identification of a functional
link. Turquoise edge indicates the interaction from curated databases. Pink edge indicates the
interaction from experimental data. Green, red, and blue edges are from predicted
interactions. Yellow, black, and light blue indicate text mining, co-expression, and protein
homology.
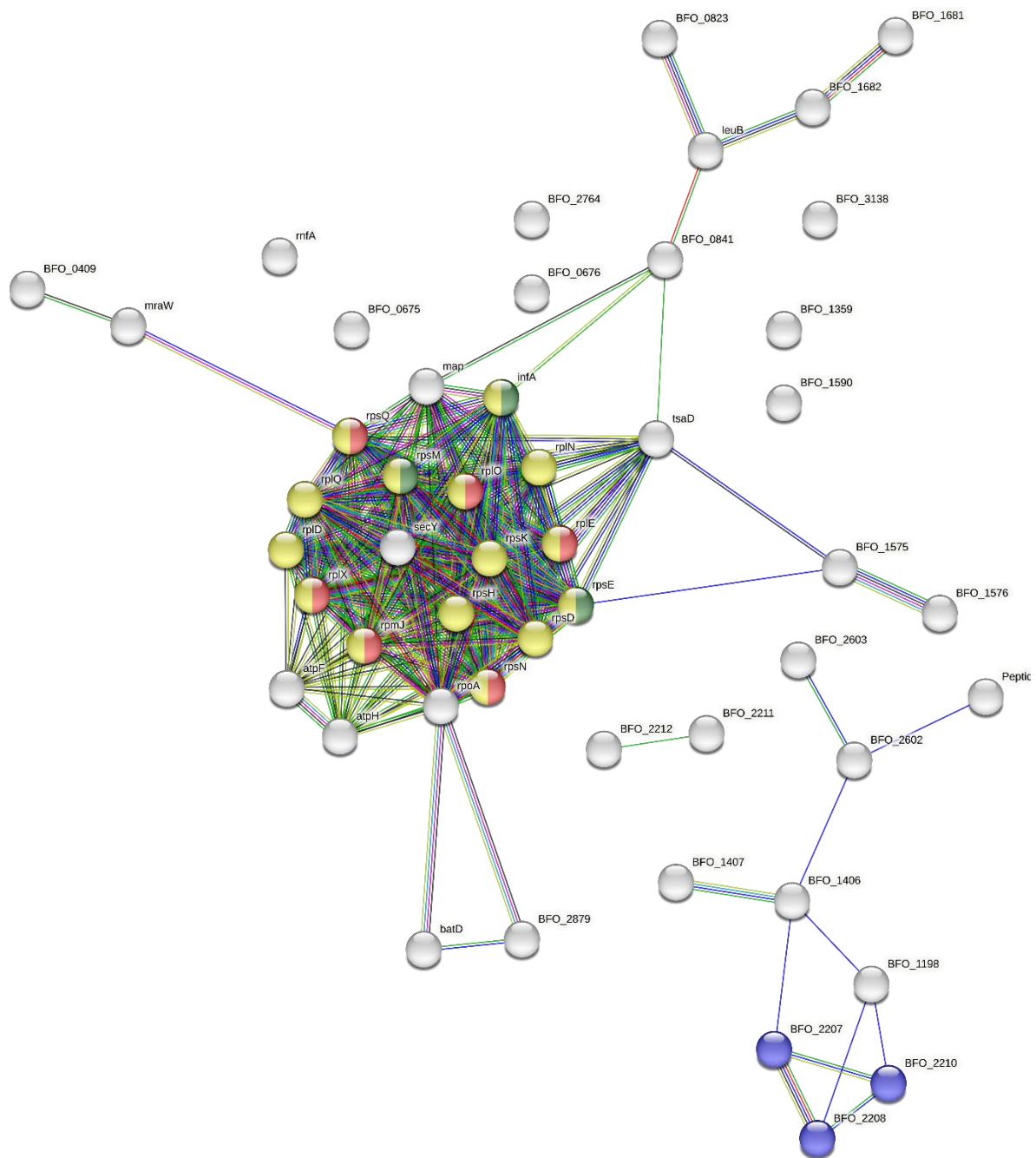
Figure 6.29. Protein–protein interaction in mucin.
This Figure shows protein–protein interaction of the top 50 up-regulated genes in response to mucin biofilm. The red, green, yellow, and blue colours represent high interaction proteins belonging to structural molecule activity, initiation factor and Ribosomal protein S5, Ribosome, and kelch motif and sialidase family respectively. The presence of edges between proteins indicates protein interactions. More edges illustrate evidence base for identification of a functional link. Turquoise edge indicates the interaction from curated databases. Pink edge indicates the interaction from experimental data. Green, red, and blue edges are from predicted interactions. Yellow, black, and light blue indicate text mining, co-expression, and protein homology.

6.2.6. Analysis of Differential Expression Genes

Expression of different transporters is an additional interest to this study as they are used to transit essential nutrient uptake and secrete multipe toxins and antimicrobial agents. Here, the uptake of and biofilm on Neu5Ac and mucin led to an increase and a decrease in the expression of transcripts that encode multicomponent transporters. The impact of Neu5Ac and mucin on transporters is briefly summarised based on the functional analysis of differentially expressed genes.

6.2.6.1. KEGG BRITE Database

We performed KEGG pathway enrichment and interrelation analysis for all up-regulated and down-regulated DEGs identifying more significant transporters.[212] In response to Neu5Ac, KEGG analysis concluded 16 up-regulated transporters compared to 29 down-regulated transporters (Tables 6.1 and 6.2).

Of the significant up-regulated transporters to Neu5Ac, electrochemical protein driven (TolB) (BFO_2640), insertion porin family (bamA) (BFO_1445), and lipopolysaccharide export (lptB) (BFO_1493). TolB is a periplasmic protein with a role in signalling and maintaining outer membrane integrity. TolB belongs to the cell envelope Tol complex and interacts with the peptidoglycan-associated lipoprotein. In *E. coli*, TolB interacts with the trimeric porins OmpF, OmpC, PhoE, and LamB in the absence of lipopolysaccharide.[357] Next, the bamA is an integral β-barrel protein and is a member of the total five Bam proteins in *E. coli*. BamA is an essential protein compared to other protein members. The Bam family is responsible for folding and inserting synthesised outer membrane proteins into the outer membrane once they are transported across the aqueous periplasmic membrane.[358] LptB, along with LptA, LptC, LptD, and LptE, transport the lipopolysaccharide from the periplasmic space to the outer membrane. LptB is associated with the cytoplasmic ATP protein in the inner membrane, and its deletions results in abnormal membrane structures in the periplasm.[359]

In response to mucin, KEGG analysis showed nine significant up-regulated transporters compared to eight down-regulated transporters (Tables 6.3 and 6.4). Of the significantly up-regulated transporters to mucin: BFO_1575 (mlaF) and BFO_1576 (mlaE), which were also significantly down-regulated in response to Neu5Ac. The mlaF forms nucleotide-binding domains (NBDs) of the ABC transporter, whereas the MlaE protein forms the transmembrane domains of the ABC transporter complex. These mlaF and mlaE along with MlaD, and MlaB can maintain outer membrane lipid asymmetry by providing energy by retrograde trafficking of phospholipids from the outer membrane to the inner membrane. The

MlaFEDB is predicted to derive this energy of transport from the breakdown of a molecule called ATP.[360] This may suggest that different sugars found in glycoproteins may drive additional transporters. The rest of the mucin significant up-regulated transporters are similar to Neu5Ac up-regulated transporters.

Table 6.1. KEGG BRITE up-regulated significant transporters in response to Neu5Ac.

Up-regulated Transporters	Annotation/Function/role
BFO_0102	peptidase
BFO_0088	peptidase
BFO_2043	ABC transporter
BFO_1493	Lipopolysaccharide transporter
BFO_2263	Drug efflux transporter
BFO_3105	putative membrane protein
BFO_3104	ABC-2 type transporter
BFO_0153	efflux ABC transporter
BFO_2801	TonB-linked outer membrane protein
BFO_1445	insertion porin family
BFO_2640	Electrochemical potential-driven transporter
BFO_0027	Electrochemical potential-driven transporter
BFO_1286	auxiliary transport protein
BFO_2106	hypothetical protein
BFO_1455	SusD family protein
BFO_0299	SusD family protein

Table 6.2. KEGG BRITE down-regulated significant transporters in response to Neu5Ac.

Down-regulated Transporters	Annotation/Function/role
BFO_1576	gamma-Hexachlorocyclohexane
BFO_1575	gamma-Hexachlorocyclohexane
BFO_2262	Drug efflux
BFO_1762	Cell division
BFO_2098	ABC transporter, ATP-binding protein
BFO_0151	Efflux ABC transporter, permease protein
BFO_0443	MATE efflux family protein
BFO_2040	transporter, major facilitator family protein
BFO_0985	heavy metal-associated domain protein
BFO_1741	hypothetical protein
BFO_2076	putative TonB-dependent hemin utilisation receptor HmuR
BFO_2077	outer membrane porin BtuB family protein
BFO_0521	TonB-dependent receptor
BFO_2639	TonB-dependent receptor
BFO_0459	Outer membrane efflux protein
BFO_0518	Electrochemical -TrkA N-terminal domain protein
BFO_3142	efflux transporter, RND family, MFP subunit
BFO_2813	SusD family protein
BFO_1406	SusD family protein
BFO_1299	SusD family protein
BFO_2206	SusD family protein
BFO_0395	SusD family protein
BFO_1202	SusD family protein
BFO_2914	SusD family protein
BFO_0023	Mg2+ transporter-C, MgtC family

Table 6.3. KEGG BRITE up-regulated significant transporters in response to mucin.

Up-regulated Transporters	Annotation/Function/role
BFO_1576	gamma-Hexachlorocyclohexane
BFO_1575	gamma-Hexachlorocyclohexane
BFO_0014	Cobalamin transporter
BFO_2640	Electrochemical potential-driven transporter
BFO_1406	SusD family protein
BFO_1299	SusD family protein
BFO_2206	SusD family protein
BFO_3111	SusD family protein
BFO_0023	Mg2+ transporter-C, MgtC family

Table 6.4. KEGG BRITE down-regulated significant transporters in response to mucin.

Down-regulated Transporters	Annotation/Function/role
BFO_0102	peptidase
BFO_1451	Zinc transporter
BFO_2299	efflux ABC transporter
BFO_1510	TonB-dependent receptor
BFO_2801	TonB-dependent receptor
BFO_0588	TonB-dependent receptor
BFO_2713	Transporter
BFO_2972	SusD family protein
BFO_0982	SusD family protein

6.2.6.2. TonB-dependent Transporters (TBDTs)

TBDTs are generally essential for sensing and adjusting to environmental signals and are linked to pathogenicity in bacteria.[112] In chapter IV (section 4.2.1.1), we concluded the presence of 62 in *T. forsythia* ATCC 43037 strain. The Neu5Ac conditions led to a significant increase in the expression of 20 TBDTs compared to down-regulated 30 TBDTs (Figure 6.30). The mucin conditions also significantly induced 12 TBDTs, suggesting that these differences in the expression of TBDTs genes could be explained by changes in favouring one condition to another, which indicates a different metabolism structure (Figure 6.31). Interestingly, there was a similarity between genes that are differentially expressed to either Neu5Ac or mucin. There were significantly seven up-regulated TBDTs that were expressed in both substance conditions: NanO, BFO_1590, BFO_0723, BFO_1198, BFO_1407, BFO_1412, and BFO_0188. The structure of these TBDTs is NTE, with high similarity to the NanO TonB box, and only BFO_1590 is not associated with SusD. In contrast, there were similar significant downregulations of these TBDTs in response to both substance conditions: BFO_1510, BFO_0418, BFO_2013, BFO_2107, BFO_2801 BFO_3314, BFO_0030, and BFO_0298. This may suggest their essential role under NAM conditions.
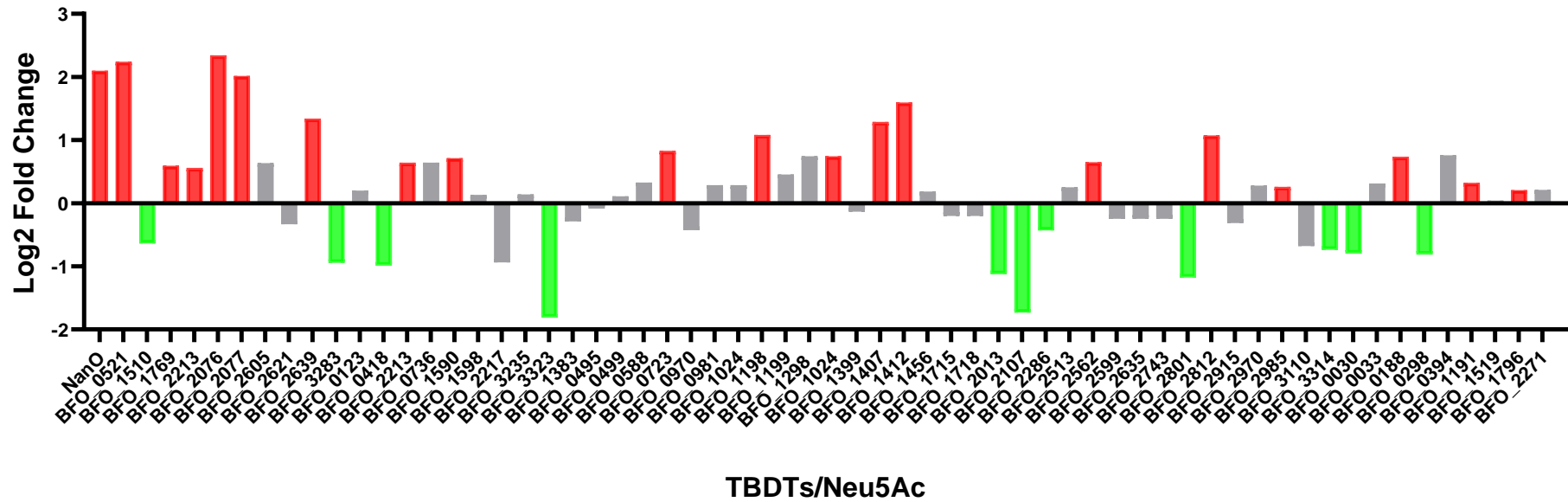
Figure 6.30. Differential expression of TBDTs in Neu5Ac transcriptomics.
Results are expressed as $\log_2$ fold change in Neu5Ac biofilm compared to NAM. Significant up-regulated levels are represented as red bars ($P < 0.05$), down-regulated levels are represented as green bars ($P < 0.05$), and those with no significant change are presented as gray bars.
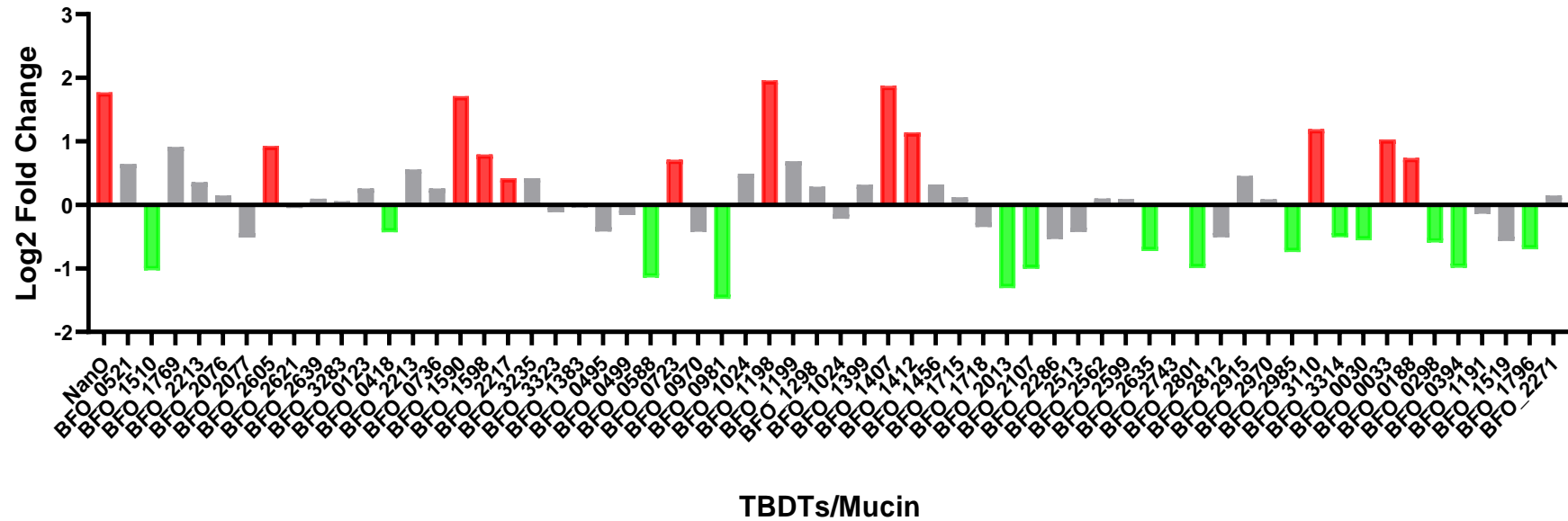
Figure 6.31. Differential expression of TBDTs in mucin transcriptomics.
Results are expressed as $\log_2$ fold change in mucin biofilm compared to NAM. Significant up-regulated levels are represented as red bars ($P < 0.05$), down-regulated levels are represented as green bars ($P < 0.05$), and those with no significant change are presented as gray bars.

6.2.6.3. Sialic acid uptake and biofilm formation

In chapter III, we constructed several TonB mutants to assess their role with sialic acid uptake, identifying the role of *T. forsythia* TonB BFO_0233 in Neu5Ac uptake and biofilm formation. To conclude more about the transit of sialic acid, the transcriptomics data of *T. forsythia* growth on Neu5Ac and mucin confirmed the role of the sialic acid *nan* operon.[203] In the presence of either Neu5Ac or mucin, the transcriptomic data of this study showed that *nan* operon were significantly up-regulated, except *nanE* or *nanAE* under mucin conditions, suggesting that *nanE* is constitutive (Figure 6.32).
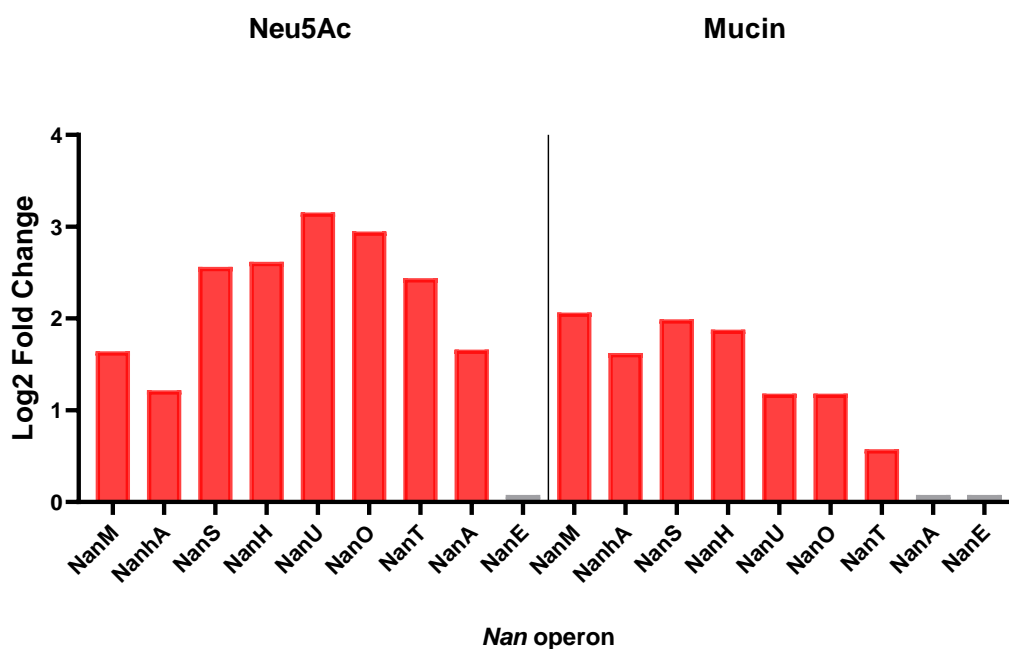
Figure 6.32. Differential expression of *nan* operon.
This figure illustrates transcriptomic change of *Tannerella* sialic acid *nan* operon. Results are expressed as $\log_2$ fold change in Neu5Ac, and mucin biofilm conditions compared to NAM. Significant up-regulated levels are represented as red bars ($P < 0.05$), and those with no significant change are presented as gray bars.

6.2.6.4. Virulence factors

Further to the main interest of this study, the role of Neu5Ac and mucin conditions confirmed the expression of several genes in *T. forsythia* virulence factors (section 5.3.1.4.). Of the summarised 45 putative virulence factors (section 5.3.1.4.), only 12 were significantly up regulated in Neu5Ac conditions while the rest were not differentially expressed (Figure 6.33). In the mucin transcriptomic data, there were only four significant up-regulated virulence factors (Figure 6.34). This indicates the role of the expression of these factors can differ from one condition to another. This is consistent with the previous statement concluding the difference of virulence factors expression between *in vivo* and *in vitro*.[319] Of the 45 factors, FucO is glycosidase, and it is dedicated to oligosaccharides, whereas β-hexosaminidase is dedicated to cleaving glucosamine and galactosamine. Moreover, forsilysin is known to have the ability to degrade an array of host proteins and to prevent host defences. These three factors were found to be significantly up-regulated in both Neu5Ac and mucin conditions, indicating their potential role in survival and pathogenicity. In contrast, the S-layer showed no change or response in both conditions.
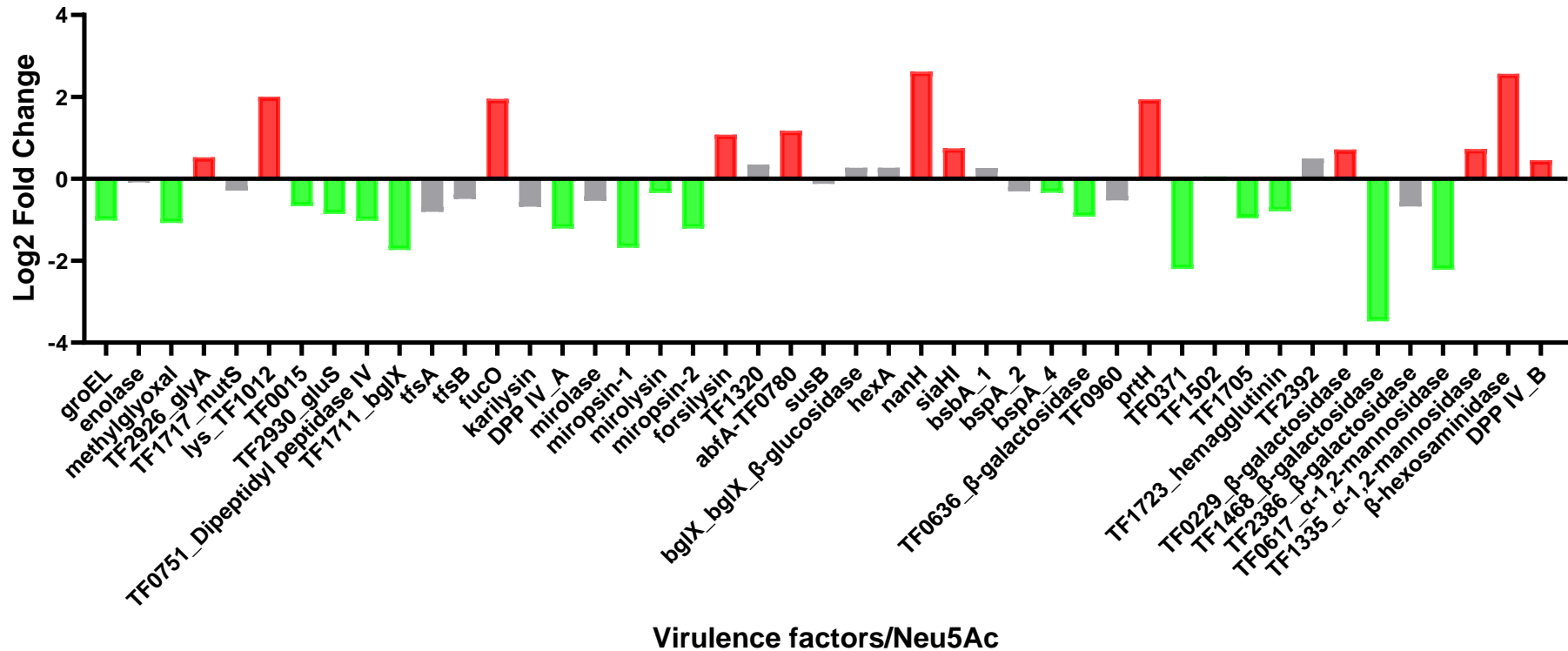
Figure 6.33. Differential expression of virulence factors in Neu5Ac transcriptomics.
This figure illustrates more about *Tannerella* virulence factors changes. Results are expressed as $\log_2$ fold change in Neu5Ac biofilm compared to NAM. Significant up-regulated levels are represented as red bars ($P < 0.05$), down-regulated levels are represented as green bars ($P < 0.05$), and those with no significant change are presented as gray bars.
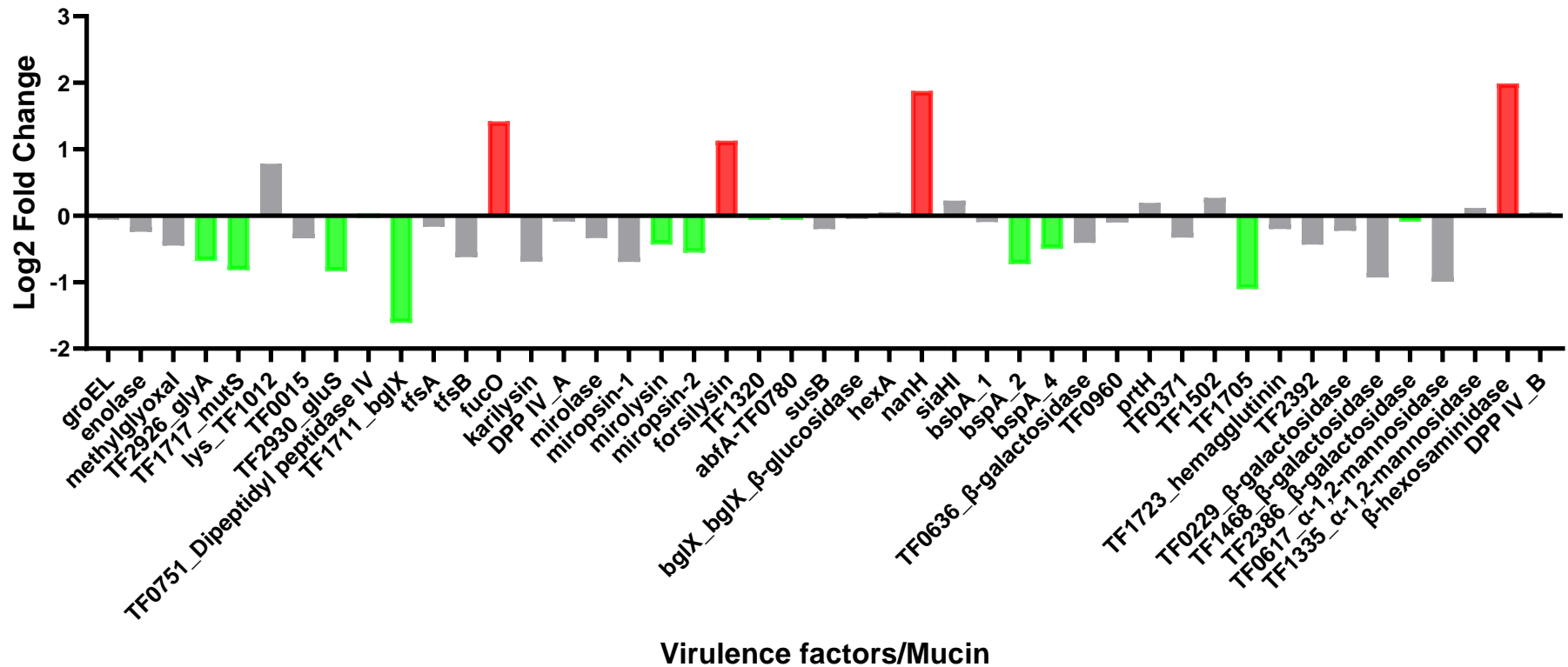
Figure 6.34. Differential expression of virulence factors in mucin transcriptomics.
This figure illustrates more about *Tannerella* virulence factors changes. Results are expressed as $\log_2$ fold change in mucin biofilm compared to NAM. Significant up-regulated levels are represented as red bars ($P < 0.05$), down-regulated levels are represented as green bars ($P < 0.05$), and those with no significant change are presented as gray bar.

6.3. Conclusion

T. forsythia grows in subgingival anaerobic conditions and forms part of the Red-complex of periodontal bacteria that contributes to the polymicrobial disease of periodontitis.[203] It is well established as having fastidious nutritional requirements, i.e. it only grows in mono-culture in the presence of NAM,[31] while it is thought that it also uses proteinaceous compounds alongside this as nutritional substrates and as such is considered to have asaccharolytic anaerobic physiology.[203] Previous work showed that *T. forsythia* could utilise sialic acid as an alternative nutritional source during biofilm growth, a process enabled by a dedicated sialic acid scavenging, catabolism, and uptake system.[88] The oral environment is mainly rich in sialylated glycoproteins like salivary mucins. Human salivary glycoproteins consist of various complex sugar substrates like mucin. The presence of sialic acid in mucin is linked through its terminal sugar by a 2-6′ glycosidic bond to N-acetylglucosamine.[361] Furthermore, the sialic acid uptake is important not only for growth in biofilm, but also for intracellular survival, suggesting a role for sialic acid uptake in virulence.[105] Indeed this ability may influence its ability to evade the immune system by a transient intracellular lifestyle that might allow persistence in the oral cavity.[203]

Sequencing of RNA fragments is used in a range of quantitative assays. Typically, these reads from the RNA-sequence are assigned to a class where this class is mapped to a common region of the target genome. Each mapped class represents a target transcript based on the number of reads in a class. However, the number of replicates in data sets of interest is often too small.[362] Thus, one test was proposed to test for differential expression is the Poisson distribution, which is uniquely determined by its mean.[363] This method can predict smaller variations than what is seen in the data, leading type-I error (the probability of false discoveries). This model was later extended and updated with the median of the ratios of observed counts to estimate the size factors. This model count is used in DESeq package and was used to adjust the values of the triplicate measurements per condition.[362]

Before analysing different expressed genes and pathway enrichment, the RNA dataset requires variability assessment to enhance the reporting outcomes. Of the reported variability assessments are the intergroup variability, which represents differences within control and non-control experimental conditions. Another variability assessments are the intragroup variability, which represents technical or biological variability. The assessment of variability can help identifying outliers that were not excluded during upstream steps. Heatmap and hierarchical clustering is an essential part of visualising RNA-sequence data. Heatmap with the mean centring by genes and dividing each row by its standard deviation can help reveal potential issues in the dataset. The designed heatmaps of this study showed a big picture of

gene expression clustering due to relative fold-changes of the most up-regulated and down-regulated genes. Both PCA and Pearson's coefficient can identify contaminated sample or mislabelled library by clustering such matter within another group.[364] The advantage of PCA is that it reduces the number of genes to a minimal set, reflecting the total variation from a large dataset. PC1 can describe the most variation within the data, whereas the PC2 can describe the second most variation. Here, we found PC1 and PC2 describing almost all variances within the dataset of NAM vs Neu5Ac and NAM vs mucin.

Filtering out lower expressed genes can account for additional variability between samples. Once a normalisation of reads per kilobases of transcript per 0.5 million mapped reads counts is performed, it is possible to compare two different samples in terms of their normalised expression values for all genes.[364] Thus, we first plotted and compared the distribution of gene expression for all samples. For NAM samples, a higher consistency in gene expression was noticed among distributed samples. Likewise, we noticed a higher consistency in gene expression among Neu5Ac and mucin samples. This is more likely due to less intragroup variability and a greater difference within intergroup variability. It is also more likely to be the result of the high read coverage across samples and the sequencing of all the samples on the same sequencing run with advanced RNA sequence instruments.

The Pairwise comparison tests were used to statistically differentiate between the two experimental groups, NAM and Neu5Ac and NAM and mucin. This type of analysis is used to test the hypothesis that every gene between the two experimental groups is in fact differentially expressed (a demonstration of significantly different expression distributions). It is also a rejection of the null hypothesis that every gene between the two experimental groups is not differentially expressed (a demonstration of equal expression distributions). The Pairwise comparison tests were used to differentiate expression distribution for the sequenced genes.[364] These tests generated for each gene the $\log_2$ fold change, *P* values, and adjusted *P* values. We found both a positive- and a negative- fold change in overall gene expression. This indicated a different mean expression for every gene between all experimental conditions. Next, we assessed the expression of gene regulation using the adjusted *P* values (padj) at < 0.05 to identify the differentially expressed genes. This study found 758 different expressed genes due to the survival of *T. forsythia* on Neu5Ac. The total of the up-regulated and down-regulated genes were 379 genes compared to 379 down-regulated genes, respectively. In mucin conditions, the total of the up-regulated and down-regulated genes were 200 genes compared to 136 down-regulated genes, respectively.

The Neu5Ac and mucin conditions resulted in general similarity outcomes from gene ontology, and pathway enrichment analysis. Furthermore, functional enrichment analysis indicated that the top 30 up-regulated genes under Neu5Ac were mainly involved in papain-like cysteine protease AvrRpt2, peptidase C39 and bacteriocin processing, and glycosyltransferase. The cysteine protease AvrRpt2 is a protein that enables a pathogen to target immune signalling components enabling infection.[365] In addition, bacteriocins are another defence weapon some bacteria use to inhibit the growth of other bacterial species and promote their growth.[355] The β-protein architecture, pepSY-associated TM protein, and kelch motif were assigned to the top 30 up-regulated genes under mucin conditions. The β-protein structure consists of hydrogen bonds and is responsible for anchoring proteins with multiple function for nutrient uptake and adhesion to the outer membrane.[366] The pepSY-associated TM protein is a family found in channels and transporters, introducing the flexibility for transport across the membrane.[367] The kelch motifs have diverse proteins with multiple functions, like hydrolysing sialic acid residues from glycoproteins.[368] It is clear that free Neu5Ac enables acceleration of *Tannerella* growth compared to mucin conditions, leading this bacterium to quickly produce defence proteins enabling its biofilm and infection. This was clearly established about this bacterium's ability to evade the immune system and increase persistence in the oral cavity.[319]

In recent years, there is a rapid development in advancing the co-expression network research. Understanding the complex interactions between different genes will lead to an in-depth interpretation of their regulatory networks and their implication for antimicrobial therapy.[369] Next, this study used the protein-protein interaction network to predict the interaction among the top 50 up-regulated genes from the RNA sequences analysis. Under Neu5Ac conditions, we noticed a higher correlation between the top 10 up-regulated genes and between the top 50 up-regulated genes in protein-protein interaction. This correlation concludes a higher acceleration of these top regulated genes in oxidising GDP-D-mannose to GDP-D-mannuronic acid as it is incorporated into some bacterial capsular polysaccharides. In contrast, we noticed among the top 10 correlated genes under mucin conditions that genetic information processing including RNA polymerase factor and DNA binding HTH domain were highly up-regulated. These genes are part of cellular pathway involving cell motility, cell growth, and death. This suggests that *T. forsythia*, under mucin conditions, generates many translations due to the additional process of adjusting on glycoproteins before benefiting from sialic acid. This was supported by protein-protein interaction, where the top

50 genes showed a lot of translation to form a biofilm on mucin. Of these top 50 translations in protein-protein interaction are the up-regulation of *nanH* and *nahA* of *nan* operon.

The outer membrane of Gram-negative bacteria consists of outer and inner leaflets including lipopolysaccharides and phospholipids, respectively. Asymmetric lipid distribution of the outer membrane in Gram-negative bacteria is influenced by numerous cellular functions and different biophysical properties. The input of energy in the form of ATP is one way to maintain the distribution of asymmetrical lipids across the membrane.[359] The transcriptomic data illustrated that *T. forsythia* under Neu5Ac conditions induced lptB/BFO_1493 to transfer the lipopolysaccharide from the periplasmic space to the outer membrane. This seems predictable as lipopolysaccharide is located on the outer leaflet of the outer membrane, maintaining the barrier function from the passive diffusion of hydrophobic solutes. In contrast, lptB/BFO_1493 was found with no change under mucin conditions. Instead, BFO_1575 (mlaF) and BFO_1576 (mlaE) are known to maintain the outer membrane lipid asymmetry by providing energy.[360] These two genes were down-regulated under the Neu5Ac conditions but up-regulated under mucin conditions. This implies additional effort from *Tannerella* under different sugar conditions to maintain its outer membrane through lipid arrangement.

Exploration of differential expression genes showed up-regulation of 20 TBDTs in Neu5Ac conditions compared to 12 in mucin conditions. In response to Neu5Ac and mucin conditions, we found seven TBDTs that were significantly up-regulated: NanO, BFO_1590, BFO_0723, BFO_1198, BFO_1407, BFO_1412, and BFO_0188. *In silico* analysis using NCBI and Uniport predicted the function of these TBDTs as the following: sialic acid uptake, belongs to phospholipases in lipid hydrolysis and to glycoside hydrolase family 13, serves enzymes that catalyse several chemical reactions, operates with bacteria lipases and peroxidases, arounds pathway transit folded proteins across the inner membrane, belongs to enzymes for the oxidative stress response, and operates with glycoside hydrolase family-16, respectively.[303,305] This is consistent with the role of TBDTs in bacteria, where they are responsible for adjusting to environmental signals and nutrient uptake, enhancing pathogenicity in bacteria.[112] Furthermore, we detected five TBDTs that were only up-regulated under mucin conditions (BFO_2605, BFO_1598, BFO_2217, BFO_0033, and BFO_3110). The *In-silico* analysis showed that they are generally involved in cell wall formation, export of lipid components, de-hydrogenation and hydroxylation reactions, the family of lyases, and the family of protease inhibitors, respectively.[303,305]

Surprisingly, we detected a few virulence factors that were up-regulated in response to Neu5Ac and mucin conditions, suggesting these factors are highly linked to pathogenicity, whereas the rest might be constitutive. Likewise, in a transcriptomic data of *P. gingivalis*, proteases like RgpA and Kgp, hemin uptake systems encoded by the *iht* and *hus* operons, and HaeS of iron acquisition were not differentially up-regulated.[370] Here, we found that the four *Tannerella tonBs* showed no change under Neu5Ac conditions, whereas both *BFO_0233* and *BFO_3116 tonBs* were significantly up-regulated under mucin conditions. However, it is confusing that the gene expression of *BFO_0233 tonB* was down-regulated under Neu5Ac conditions, as this study confirmed its role in the uptake of Neu5Ac and biofilm formation on mucin. This may imply the limitation of transcriptomics in detecting the "genes that matter" as normal expressed regulatory genes may play unknown and different roles under conditions.[361]

Overall, the transcriptomic data suggests that the growth of *T. forsythia* on mucin requires the bacterium to induce additional genes and translation in order to adjust its growth and biofilm formation. This study defined and predicted several up-regulated genes from the transcriptomic data play a major role in the survival of this bacterium. We also predicted and compared the functional enrichment analysis of several genes using gene ontology and pathways enrichment analysis, including biological process, molecular function, and cellular component. However, Western blot analysis, qPCR, and ELISA are essential for further clarifying the genes expression levels from this transcriptomic data.

Chapter VII

Overall Summary and Discussion

7.1 Summary of major findings

The work described in this thesis has illustrated a deeper knowledge of the molecular mechanism behind *T. forsythia tonB* genes and the TonB dependent transporter *nanO* and their roles in utilising the host-derived sugar sialic acid, which is necessary for *T. forsythia* biofilm growth and cell interactions. This was followed by bioinformatic identification and characterisation of TonB dependent transporters (TBDTs) in all genomes of *T. forsythia* before inhibiting the uptake of sialic acid through TonB box of these TBDTs.

Previous work on *T. forsythia* NanO and NanU (NanOU) was tested in a heterologous expression method where the outer membrane (NanOU) transport genes of *T. forsythia* were expressed in an *E.coli nanC* mutant (*E. coli* Δ*nanCnanR*[amber] Δ*ompR*::Tn10[tet]) to confirm its role with sialic acid transport. Both genes were able to grow using Neu5Ac or NeuGc as a sole carbon and energy source, indicating free sialic acids from glycoprotein are bound to the extracellular neuraminate uptake protein (*nanU*) before moving to the neuraminate outer membrane permease (*nanO*).[128] Then the it was necessary to examine whether these NanOUs require energy for sialic acid transport and whether they have a functional dependence on the TonB system (TonB–exbB–exbD) proteins. A further modification was performed to the previously designed system in *E. coli* missing *nanC* and *nanR* as they are sialic acid and its regulator genes, respectively (*E. coli* Δ*nanCnanR*[amber] Δ*ompR*::Tn10[tet]). Thus, the *tonB* of this *E. coli* system was deleted and replaced by inserting a kanamycin-resistance cassette (Δ*tonB*::FRT-Km-FRT). As a result, *nanO* and *nanU* were inserted into the *E. coli* system showing no growth was observed in the absence of *tonB*, indicating the need for energy through the TonB system (TonB–exbB–exbD proteins).[128]

While the health of the oral cavity is a particularly complex balance of host and bacterial factors, this study aimed to examine the disease-causing bacterium. One way of examining this matter is the molecular level behind *T. forsythia* to determine specific transporters involved in enabling this bacterium to avoid the immune system and persist in the oral cavity. Thus, a deeper understanding of the molecular level behind *T. forsythia* will allow the design of therapeutic agents to not only target defence against the pathogen and switching dysbiosis but also to enable the commensal bacteria to return to a healthy periodontium. This chapter is to summarise the major findings from each part of this study before relating them to the main aim of preventing periodontal disease.

Chapter III: Role of TonB transporters in *T. forsythia*.

This chapter aimed to characterise the role of multiple *tonB* genes with Neu5Ac using mutagenesis in *T. forsythia*.

- Electroporation and natural mutagenesis methods were used to introduce DNA to create the *nanO, BFO_0333*, *BFO_0233*, *BFO_3116*, *BFO_3116*, and *BFO_0953* mutants. Both methods showed a significantly higher number of colonies produced, implementing their efficiency as a universal transformation method for *T. forsythia* mutagenesis.

- The ability of *T. forsythia* to uptake sialic acid and form a biofilm is an essential mechanism of this bacterium through its virulence factors. This study investigated and described the role of TonB proteins in the uptake of sialic acid before leading to biofilm formation. Likewise, the role of NanO in the transit of sialic acid through the outer membrane was reviewed using mutagenesis. Through all four *tonB* generated mutants, the sialic acid uptake and biofilm formation were attenuated and reduced by *tonB BFO_0233*. Furthermore, *T. forsythia* attenuation by *nanO* mutant was observed in this bacterium's behaviour during sialic acid uptake and biofilm formation. Similarly, this generally describes the overall formation of biofilms on teeth and around gingiva before leading to the accumulation of dental calculus or plaque, and thus by periodontitis.

Chapter IV: Bioinformatic analysis of TBDTs in *T. forsythia* and laboratory assessment of the TonB box peptide.

This chapter focused on identifying TonB dependent transporter (TBDTs) structures among all *Tannerella* available genomes and comparing these TBDTs between human pathogen and non-pathogen *T. forsythia* genomes. Furthermore, this study aimed to examine TBDTs using both bioinformatic and other laboratory approaches, uncovering a TonB box-binding site on the N-terminal domain of NanO, as well as other TBDTs.

- The domain structures of all *Tannerella* TBDTs were assessed and determined using Pfam database, resulting in four established domains. Pfam database was able to locate homologies for the following well domains: TonB-dependent receptor barrel domains (Pfam code: PF00593), TonB-dependent transporter plug domains (Pfam code: PF07715), N-Terminal extension TBDT domains (Pfam code: PF13715), and Signal transduction domains (Pfam code: PF07660).

- Bioinformatics analysis concluded a huge difference in the total of TBDTs between human pathogen and non-pathogen *T. forsythia* genomes, indicating the presence of some unique domains associated with human pathogen strains. The well-studied human pathogen strain (ATCC 43037) has 62 TBDTs and compared to 35 TBDTs in the periodontal health-associate isolate (BU0063).

- *In silico* analysis along with two similar families, the Sus-type transporter and iron-type transporter, this study identified the protein sequence for the TonB box of NanO (GASVVE). The designed TonB box peptide was effective at not only the uptake of sialic acid, but also the biofilm formation when sialic acid or mucin was utilised as a growth substrate. We also concluded a higher similarity between the TonB box of NanO and other TBDTs from the ATCC 43037 strain, suggesting a further role of this peptide on other TBDTs.

- These data indicated that the designed peptide was sufficient to inhibit the uptake of sialic acid, as well as biofilm formation. It also confirmed there is a direct contact between the *nanO* and *tonB* for energy during the transit of sialic acid, which supports the reduction of sialic acid uptake and biofilm formation by the mutant *tonB BFO_0233*.

Chapter V: Bioinformatic analysis of sialic acid utilisation genes in bacteria

This chapter aimed to explore the distribution of sialic acid *nan* operon in microbes to gain an insight into conserved function, domains, and residues before predicting their feeding pathways through *nan* operons. We also re-sequence the *T. forsythia* ATCC 43037 strain and compare all known and established virulence factors between *Tannerella* strains.

- *In silico* analysis concluded 21 virulence factor genes that are absent within the putative periodontal health-associated species *Tannerella* BU0063 compared to other human pathogen strains.

- *In silico* analysis concluded the presence of *nan* operons in 222 bacterial species of 31 bacterial families. The dissemination of 222 *nan* operons isolating from human microbes was 67 compared to 155 non-human microbes. These emphasise the importance of *tonB* genes, *nan* operons, and TBDTs in Neu5A acquisition, which lays the foundation for future studies considering the survival mechanisms of *T. forsythia* in Neu5Ac-restricted environments.

- We predicted the feeding pathways of some human microbes based on their *nan* operons. Of the 13 oral human isolated bacterial genomes, we classified six species as

utilisers, one as an accumulator, and six as bifunctional species. We also demonstrated this further by examining 49 bacterial genomes from the stomach, concluding 15 species were utilisers, five species were accumulators, and 29 species were bifunctional bacteria.

- *In silico* NanO amino acid alignments from all concluded human microbes with *nan* operons showed high similarities between their TonB boxes. These data demonstrated a potential route for future therapeutic treatment through small amino acid peptides (TonB box), where sialic acid is the main derivative for many microbes, specifically oral pathogens.

Chapter VI: Transcriptome Analysis of *T. forsythia* under Neu5Ac and mucin conditions.

This chapter aimed to investigate global gene expression in *T. forsythia* in response to Neu5Ac and mucin by RNA-sequencing.

- In response to Neu5Ac, there were 758 differential expressed genes compared to 363 in mucin conditions. The upregulated genes due to the replacement of NAM by Neu5Ac are 379 genes, whereas the downregulated genes are 379 genes. Furthermore, the mucin conditions induce 200 genes in the presence of mucin only. Both Neu5Ac and mucin showed their importance of supporting *T. forsythia* biofilm in the absence of NAM.

- There were significantly seven upregulated TBDTs that were expressed in both substance conditions: NanO, BFO_1590, BFO_0723, BFO_1198, BFO_1407, BFO_1412, and BFO_0188. The structure of these TBDTs is NTE, with high similarity to TonB box of NanO. In contrast, there were similar significant downregulations of these TBDTs in response to both Neu5Ac and mucin conditions: BFO_1510, BFO_0418, BFO_2013, BFO_2107, BFO_2801 BFO_3314, BFO_0030, and BFO_0298. These seven shared TBDTs under both substances indicated their huge role in enabling this bacterium adjustment and interaction with the host cell, and therefore to its ability to contribute with other pathogens to periodontal disease.

In conclusion, the ability of *T. forsythia* to utilise sialic acid for its interaction, adhesion, and invasion is dependent on *nan* operon, which was found to be up-regulated. Host sialic acid (glycoprotein) is not freely available requiring this bacterium to present its accessory genes (*nanH*, *nanS*, *nanhA*, *nanM*). The action of this process starts with 9-*O*-acetylesterase (NanS) to deacetylate sialic acid at carbon position 9, as the acetyl group was shown to inhibit the role *T. forsythia* sialidase (NanH), whereas the NanhA works on cleaving NAG from host glycoprotein. The role of NanH is then activated to cleave sialic acid at α2,3 or α2,6 glycosidic bonds of the terminal sialyl residues before free sialic acid binding to NanU and then to NanO, the outer membrane gate of sialic acid. Here, the role of TonB BFO_0233 is then activated upon the signal from the TonB box of NanO delivering energy for transiting of sialic acid from the outer membrane into the periplasm membrane. The sialic acid is then transported via the inner membrane (NanT) before the catabolism pathways convert and reuse this transit.[203]

7.2. Periodontal disease and sialic acid: A possible Approach?

This study has highlighted the role of multiple transporters in the uptake of sialic acid and biofilm formation, suggesting the possibility of this bacterium in the adherence and invasion into the host. The identification of TonB BFO_0233 protein and confirmation of the previous NanO behaviour illustrates the influence of the disease progression in the host by *T. forsythia* along with coupling with the other two red complex bacteria.[203] Unfortunately, the research in this area is limited due to *T. forsythia's* fastidious growth requirement for culturing, and misannotation and confusion between its strains. Being the only red complex bacteria and among few of Gram-negative bacteria with an S-layer and the unique structure of its *nan* operon, there are many questions to be answered about this bacterium and its influential role in the progression of periodontal disease.[203] However, the presence of *nan* operons among several oral microbes indicated their need for sialic acid, where sialic acid can mediate cell-cell regulation, control membrane transport, and decorate their cell surfaces.[84] The presence of TBDTs among Gram-negative bacteria enable their ability to adjust to environmental signals and increase their pathogenicity.[112] In the oral cavity, the dysbiosis imbalance is due to the implication of Gram-negative bacteria which are strongly present in periodontal disease.[203] Unlike other specifically designed peptides, the designed TonB box peptide can target many microbes in the oral cavity with dependency on sialic acid. Thus, it is a possibility for therapeutic treatment development to target the uptake of sialic acid through NanO and its interaction with TonB protein.

Furthermore, in siderophores-iron utilisation, several studies developed antibacterial strategies that were examined and tested in some bacteria, such as *E. coli* and *A. baumannii*.[292,293] These studies concluded the effect of the chemicals and compounds with abilities to arrest the growth of bacteria with TonB systems. Although this is a great advancement in the therapeutic treatment development, these chemicals and compounds may attack commensal bacteria and restrict their benefits in our lives. With this in mind, the idea of TonB box peptide can develop with further research studies since it was reported that the interaction between TonB protein and TonB box is mostly hydrophobic due to the highly conserved Valine (Val) on the TonB box.[159] In addition, the interaction between the TonB box and TonB protein is driven by the presence of negative and positive charges of residues. Previous laboratory work with antimicrobial peptides targeting the TonB boxes concluded several peptides with a positive effect on targeting multiple pathogens.[155,158,290] Designing a selective TonB box peptide by combining these residues from previous antimicrobial

peptides may target multiple TBDTs uptakes in pathogens, leading to a realistic opportunity of tackling the burdening issue of periodontal disease.

Doctoral Development Plan

Attended Courses:

- HAR6045 Further Statistics for Health Science Researchers

- HAR6041 Health Needs Assessment, Planning and Evaluation

- BMS6007 Mass Spectrometry-based proteomics and metabolomics

- FCM6100 Research Ethics and Integrity

Attended Seminars:

- Tools for literature searching.

- Introduction to basic statistics for biologists.

- Mandatory CPD- an update on working with ionizing radiation in primary care dental

practice.

- Mental floss: restorative trivia

- Medical issues and conditions relating to oral surgery

- Academic publishing - editing and peer review.

- Introduction to microscopy and imaging.

- Introduction to advanced microscopy techniques

CoSHH Assessments

6046-Polymerase chain reaction PCR (phusion and dreamtaq)

6045-Colony PCR (dreamtaq)

6044-Bacteria genomic DNA extraction (gram negative - wizard kit)

6043-DNA ligation

6042-DNA band gel extraction and purification of PCR products

5983-Tryptone Soy Broth for TF

5982-Bacteria agar plate with supplements

5980-SCD - Growth media of Bacteria

5979-SCD Gel electrophoresis of protein samples

5978-SCD Agarose gel electrophoresis

5977-Restriction Digestion of DNA

5425-Minimal media

8291-Bugbuster- cell lysis and extraction of soluble and insoluble protein.

8290-Bacterial RNeasy RNA extraction using Qiagen kit.

7504-RNA isolation from Bacteria using Ambion RiboPure kit.

References

1. Sudhakara P, Gupta A, Bhardwaj A, Wilson A. Oral dysbiotic communities and their implications in systemic diseases. *Dent J*. Published online 2018. doi:10.3390/dj6020010

2. Bourgeois D, Inquimbert C, Ottolenghi L, Carrouel F. Periodontal Pathogens as Risk Factors of Cardiovascular Diseases, Diabetes, Rheumatoid Arthritis, Cancer, and Chronic Obstructive Pulmonary Disease—Is There Cause for Consideration? *Microorganisms*. Published online 2019. doi:10.3390/microorganisms7100424

3. Marsh PD. Are dental diseases examples of ecological catastrophes? *Microbiology*. Published online 2003. doi:10.1099/mic.0.26082-0

4. Eke PI, Dye BA, Wei L, Thornton-Evans GO, Genco RJ. Prevalence of periodontitis in adults in the united states: 2009 and 2010. *J Dent Res*. Published online 2012. doi:10.1177/0022034512457373

5. Meyer PA, Yoon PW, Kaufmann RB, Centers for Disease Control. Introduction: CDC Health Disparities and Inequalities Report - United States, 2013. *MMWR Surveill Summ*. Published online 2013.

6. Socransky SS, Haffajee AD, Cugini MA, Smith C, Kent RL. Microbial complexes in subgingival plaque. *J Clin Periodontol*. Published online 1998. doi:10.1111/j.1600-051X.1998.tb02419.x

7. Sztukowska MN, Roky M, Demuth DR. Peptide and non-peptide mimetics as potential therapeutics targeting oral bacteria and oral biofilms. *Mol Oral Microbiol*. Published online 2019. doi:10.1111/omi.12267

8. Hughes FJ. Periodontium and Periodontal Disease. In: *Stem Cell Biology and Tissue Engineering in Dental Sciences*. ; 2015. doi:10.1016/B978-0-12-397157-9.00038-2

9. Belstrøm D, Fiehn NE, Nielsen CH, et al. Differences in bacterial saliva profile between periodontitis patients and a control cohort. *J Clin Periodontol*. Published online 2014. doi:10.1111/jcpe.12190

10. Zijnge V, Van Leeuwen MBM, Degener JE, et al. Oral biofilm architecture on natural teeth. *PLoS One*. Published online 2010. doi:10.1371/journal.pone.0009321

11. Griffen AL, Beall CJ, Campbell JH, et al. Distinct and complex bacterial profiles in human periodontitis and health revealed by 16S pyrosequencing. *ISME J*. Published online 2012. doi:10.1038/ismej.2011.191

12. Rôças IN, Siqueira JF, Santos KRN, Coelho AMA. "Red complex" (Bacteroides forsythus, Porphyromonas gingivalis, and Treponema denticola) in endodontic infections: A molecular approach. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*. Published online 2001. doi:10.1067/moe.2001.114379

13. Lamont RJ, Jenkinson HF. Life below the gum line: pathogenic mechanisms of Porphyromonas gingivalis. *Microbiol Mol Biol Rev*. Published online 1998.

14. Yoshimura F, Murakami Y, Nishikawa K, Hasegawa Y, Kawaminami S. Surface components of Porphyromonas gingivalis. *J Periodontal Res*. Published online 2009. doi:10.1111/j.1600-0765.2008.01135.x

15. A. S, T. W, C. A-B, et al. The capsule of porphyromonas gingivalis leads to a reduction in the host inflammatory response, evasion of phagocytosis, and increase in Virulence. *Infect Immun*. Published online 2011. doi:10.1128/IAI.05016-11 LK - http://hz9pj6fe4t.search.serialssolutions.com.proxy.cc.uic.edu/?sid=EMBASE&sid=EMBASE&issn=00199567&id=doi:10.1128%2FIAI.05016-11&atitle=The+capsule+of+porphyromonas+gingivalis+leads+to+a+reduction+in+the+host+inflammatory+response%2C+evasion+of+phagocytosis%2C+and+increase+in+Virulence&stitle=Infect.+Immun.&title=Infection+and+Immunity&volume=79&issue=11&spage=4533&epage=4542&aulast=Singh&aufirst=Amrita&auinit=A.&aufull=Singh+A.&coden=INFIB&isbn=&pages=4533-4542&date=

16. Ôgawa T, Yagi T. Bioactive mechanism of porphyromonas gingivalis lipid A. *Periodontol 2000*. Published online 2010. doi:10.1111/j.1600-0757.2009.00343.x

17. Díaz L, Hoare A, Soto C, et al. Changes in lipopolysaccharide profile of Porphyromonas gingivalis clinical isolates correlate with changes in colony morphology and polymyxin B resistance. *Anaerobe*. Published online 2015. doi:10.1016/j.anaerobe.2015.01.009

18. Kadono H, Kido JI, Kataoka M, Yamauchi N, Nagata T. Inhibition of osteoblastic cell differentiation by lipopolysaccharide extract from Porphyromonas gingivalis. *Infect Immun*. Published online 1999.

19. Kato H, Taguchi Y, Tominaga K, Umeda M, Tanaka A. Porphyromonas gingivalis LPS inhibits osteoblastic differentiation and promotes pro-inflammatory cytokine production in human periodontal ligament stem cells. *Arch Oral Biol*. Published online 2014. doi:10.1016/j.archoralbio.2013.11.008

20. Bliska JB, Galán JE, Falkow S. Signal transduction in the mammalian cell during bacterial attachment and entry. *Cell*. Published online 1993. doi:10.1016/0092-8674(93)90270-Z

21. Hajishengallis G, Darveau RP, Curtis MA. The keystone-pathogen hypothesis. *Nat Rev Microbiol*. Published online 2012. doi:10.1038/nrmicro2873

22. Naylor Kathryn; Douglas Ian; Murdoch Craig; Stafford Graham. The Role of Outer Membrane Proteins of Porphyromonas gingivalis in Host-Pathogen Interactions. Published online 2016. http://etheses.whiterose.ac.uk/15759/1/Kathryn Naylor Thesis.pdf

23. Visser MB, Ellen RP. New insights into the emerging role of oral spirochaetes in periodontal disease. *Clin Microbiol Infect*. Published online 2011. doi:10.1111/j.1469-0691.2011.03460.x

24. Leung WK, Haapasalo M, Uitto VJ, Hannam PM, McBride BC. The surface proteinase of Treponema denticola may mediate attachment of the bacteria to epithelial cells. *Anaerobe*. Published online 1996. doi:10.1006/anae.1996.0005

25. Mathers DA, Leung WK, Fenno JC, Hong Y, Mcbride BC. The major surface protein complex of Treponema denticola depolarizes and induces ion channels in HeLa cell membranes. *Infect Immun*. Published online 1996.

26. Gaibani P, Caroli F, Nucci C, Sambri V. Major surface protein complex of Treponema denticola induces the production of tumor necrosis factor α, interleukin-1β, interleukin-6 and matrix metalloproteinase 9 by primary human peripheral blood monocytes. *J Periodontal Res*. Published online 2010. doi:10.1111/j.1600-0765.2009.01246.x

27. Fenno JC. Treponema denticola interactions with host proteins. *J Oral Microbiol*. Published online 2012. doi:10.3402/jom.v4i0.9929

28. Chu L, Ebersole JL, Kurzban GP, Holt SC. Cystalysin, a 46-kDa L-Cysteine

Desulfhydrase from Treponema denticola: Biochemical and Biophysical Characterization. *Clin Infect Dis*. Published online 1999. doi:10.1086/515164

29. Rosen G, Genzler T, Sela MN. Coaggregation of Treponema denticola with Porphyromonas gingivalis and Fusobacterium nucleatum is mediated by the major outer sheath protein of Treponema denticola. *FEMS Microbiol Lett*. Published online 2008. doi:10.1111/j.1574-6968.2008.01373.x

30. Tanner ACR, Listgarten MA, Ebersole JL, Strzempko MN. Bacteroides forsythus sp. nov., a slow-growing, fusiform Bacteroides sp. from the human oral cavity. *Int J Syst Bacteriol*. Published online 1986. doi:10.1099/00207713-36-2-213

31. Wyss C. Dependence of proliferation of Bacteroides forsythus on exogenous N-acetylmuramic acid. *Infect Immun*. Published online 1989.

32. Sakamoto M, Suzuki M, Umeda M, Ishikawa I, Benno Y. Reclassification of Bacteroides forsythus (Tanner et al. 1986) as Tannerella forsythensis corrig., gen. nov., comb. nov. *Int J Syst Evol Microbiol*. Published online 2002. doi:10.1099/ijs.0.01945-0

33. Tanner ACR, Izard J. Tannerella forsythia, a periodontal pathogen entering the genomic era. *Periodontol 2000*. Published online 2006. doi:10.1111/j.1600-0757.2006.00184.x

34. Kesavalu L, Sathishkumar S, Bakthavatchalu V, et al. Rat model of polymicrobial infection, immunity, and alveolar bone resorption in periodontal disease. *Infect Immun*. Published online 2007. doi:10.1128/IAI.00733-06

35. Rutherford N, Mourez M. Surface display of proteins by Gram-negative bacterial autotransporters. *Microb Cell Fact*. Published online 2006. doi:10.1186/1475-2859-5-22

36. Sakakibara J, Nagano K, Murakami Y, et al. Loss of adherence ability to human gingival epithelial cells in S-layer protein-deficient mutants of Tannerella forsythensis. *Microbiology*. Published online 2007. doi:10.1099/mic.0.29275-0

37. Bloch S, Thurnheer T, Murakami Y, Belibasakis GN, Schäffer C. Behavior of two Tannerella forsythia strains and their cell surface mutants in multispecies oral biofilms. *Mol Oral Microbiol*. 2017;32(5):404-418. doi:10.1111/omi.12182

38. Lee SW, Sabet M, Um HS, Yang J, Kim HC, Zhu W. Identification and characterization of the genes encoding a unique surface (S-) layer of Tannerella forsythia. *Gene*. Published online 2006. doi:10.1016/j.gene.2005.11.027

39. Tomek MB, Neumann L, Nimeth I, et al. The S-layer proteins of Tannerella forsythia are secreted via a type IX secretion system that is decoupled from protein O - glycosylation. *Mol Oral Microbiol*. 2014;29(6):307-320. doi:10.1111/omi.12062

40. Veith PD, O'Brien-Simpson NM, Tan Y, Djatmiko DC, Dashper SG, Reynolds EC. Outer Membrane Proteome and Antigens of Tannerella forsythia. *J Proteome Res*. Published online 2009. doi:10.1021/pr900372c

41. Sharma A, Sojar HT, Glurich I, Honma K, Kuramitsu HK, Genco RJ. Cloning, expression, and sequencing of a cell surface antigen containing a leucine-rich repeat motif from Bacteroides forsythus ATCC 43037. *Infect Immun*. Published online 1998.

42. Inagaki S, Kuramitsu HK, Sharma A. Contact-dependent regulation of a Tannerella forsythia virulence factor, BspA, in biofilms. *FEMS Microbiol Lett*. Published online 2005. doi:10.1016/j.femsle.2005.06.032

43. Saito Y, Fujii R, Nakagawa KI, Kuramitsu HK, Okuda K, Ishihara K. Stimulation of Fusobacterium nucleatum biofilm formation by Porphyromonas gingivalis. *Oral Microbiol Immunol*. Published online 2008. doi:10.1111/j.1399-302X.2007.00380.x

44. Karim AY, Kulczycka M, Kantyka T, et al. A novel matrix metalloprotease-like enzyme (karilysin) of the periodontal pathogen Tannerella forsythia ATCC 43037. *Biol Chem*. Published online 2010. doi:10.1515/BC.2010.009

45. Ksiazek M, Karim AY, Bryzek D, et al. Mirolase, a novel subtilisin-like serine protease from the periodontopathogen Tannerella forsythia. *Biol Chem*. Published online 2015. doi:10.1515/hsz-2014-0256

46. Honma K, Mishima E, Inagaki S, Sharma A. The OxyR homologue in Tannerella forsythia regulates expression of oxidative stress responses and biofilm formation. *Microbiology*. Published online 2009. doi:10.1099/mic.0.027920-0

47. Braham PH, Moncla BJ. Rapid presumptive identification and further characterization of Bacteroides forsythus. *J Clin Microbiol*. Published online 1992.

48. Loesche WJ, Grossman NS. Periodontal disease as a specific, albeit chronic, infection:

Diagnosis and treatment. *Clin Microbiol Rev*. Published online 2001. doi:10.1128/CMR.14.4.727-752.2001

49. Theilade E. The non-specific theory in microbial etiology of inflammatory periodontal diseases. *J Clin Periodontol*. Published online 1986. doi:10.1111/j.1600-051X.1986.tb01425.x

50. Ximénez-Fyvie LA, Haffajee AD, Socransky SS. Comparison of the microbiota of supra- and subgingival plaque in health and periodontitis. *J Clin Periodontol*. Published online 2000. doi:10.1034/j.1600-051x.2000.027009648.x

51. Marsh PD. Dental plaque as a biofilm and a microbial community - Implications for health and disease. In: *BMC Oral Health*. ; 2006. doi:10.1186/1472-6831-6-S1-S14

52. Nath SG, Raveendran R. Microbial dysbiosis in periodontitis. *J Indian Soc Periodontol*. Published online 2013. doi:10.4103/0972-124X.118334

53. Curtis MA, Zenobia C, Darveau RP. The relationship of the oral microbiotia to periodontal health and disease. *Cell Host Microbe*. Published online 2011. doi:10.1016/j.chom.2011.09.008

54. Hajishengallis G, Liang S, Payne MA, et al. Low-abundance biofilm species orchestrates inflammatory periodontal disease through the commensal microbiota and complement. *Cell Host Microbe*. Published online 2011. doi:10.1016/j.chom.2011.10.006

55. Hojo K, Nagaoka S, Ohshima T, Maeda N. Bacterial Interactions in Dental Biofilm Development. *J Dent Res*. Published online 2009. doi:10.1177/0022034509346811

56. Dunne WM. Bacterial Adhesion: Seen Any Good Biofilms Lately? *Clin Microbiol Rev*. 2002;15(2):155-166. doi:10.1128/CMR.15.2.155-166.2002

57. Garrett TR, Bhakoo M, Zhang Z. Bacterial adhesion and biofilms on surfaces. *Prog Nat Sci*. 2008;18(9):1049-1056. doi:10.1016/j.pnsc.2008.04.001

58. Nesbitt WE, Doyle RJ, Taylor KG, Staat RH, Arnold RR. Positive Cooperativity in the Binding of Streptococcus sanguis to Hydroxylapatite. *Infect Immun*. 1982;35(1):157-165. doi:10.1128/iai.35.1.157-165.1982

59. Hasan A, Palmer RM. A clinical guide to periodontology: Pathology of periodontal

disease. *Br Dent J*. Published online 2014. doi:10.1038/sj.bdj.2014.299

60.   Jakubovics NS. Saliva as the Sole Nutritional Source in the Development of Multispecies Communities in Dental Plaque. Conway T, Cohen P, eds. *Microbiol Spectr*. 2015;3(3). doi:10.1128/microbiolspec.MBP-0013-2014

61.   Mazumdar V, Amar S, Segrè D. Metabolic Proximity in the Order of Colonization of a Microbial Community. Kreth J, ed. *PLoS One*. 2013;8(10):e77617. doi:10.1371/journal.pone.0077617

62.   Marsh PD, Zaura E. Dental biofilm: ecological interactions in health and disease. *J Clin Periodontol*. 2017;44:S12-S22. doi:10.1111/jcpe.12679

63.   Roberts AP, Kreth J. The impact of horizontal gene transfer on the adaptive ability of the human oral microbiome. *Front Cell Infect Microbiol*. 2014;4. doi:10.3389/fcimb.2014.00124

64.   Ali Mohammed MM, Nerland AH, Al-Haroni M, Bakken V. Characterization of extracellular polymeric matrix, and treatment of Fusobacterium nucleatum and Porphyromonas gingivalis biofilms with DNase I and proteinase K. *J Oral Microbiol*. 2013;5(1):20015. doi:10.3402/jom.v5i0.20015

65.   Jang Y-J, Sim J, Jun H-K, Choi B-K. Differential effect of autoinducer 2 of Fusobacterium nucleatum on oral streptococci. *Arch Oral Biol*. 2013;58(11):1594-1602. doi:10.1016/j.archoralbio.2013.08.006

66.   Bao K, Belibasakis GN, Thurnheer T, Aduse-Opoku J, Curtis MA, Bostanci N. Role of Porphyromonas gingivalis gingipains in multi-species biofilm formation. *BMC Microbiol*. 2014;14(1):258. doi:10.1186/s12866-014-0258-7

67.   Pham TK, Roy S, Noirel J, Douglas I, Wright PC, Stafford GP. A quantitative proteomic analysis of biofilm adaptation by the periodontal pathogen Tannerella forsythia. *Proteomics*. 2010;10(17):3130-3141. doi:10.1002/pmic.200900448

68.   Minty M, Canceil T, Serino M, Burcelin R, Tercé F, Blasco-Baque V. Oral microbiota-induced periodontitis: a new risk factor of metabolic diseases. *Rev Endocr Metab Disord*. Published online 2019. doi:10.1007/s11154-019-09526-8

69.   Kilian M, Chapple ILC, Hannig M, et al. The oral microbiome - An update for oral healthcare professionals. *Br Dent J*. Published online 2016.

doi:10.1038/sj.bdj.2016.865

70. Lo AW, Seers CA, Boyce JD, et al. Comparative transcriptomic analysis of Porphyromonas gingivalisbiofilm and planktonic cells. *BMC Microbiol*. 2009;9(1):18. doi:10.1186/1471-2180-9-18

71. Roodman GD. Role of cytokines in the regulation of bone resorption. *Calcif Tissue Int*. 1993;53(S1):S94-S98. doi:10.1007/BF01673412

72. Baggiolini M, Dewald B, Moser B. lnterleukin-8 and Related Chemotactic Cytokines—CXC and CC Chemokines. In: ; 1993:97-179. doi:10.1016/S0065-2776(08)60509-X

73. Sekot G, Posch G, Messner P, et al. Potential of the Tannerella forsythia S-layer to Delay the Immune Response. *J Dent Res*. 2011;90(1):109-114. doi:10.1177/0022034510384622

74. Moffatt CE, Whitmore SE, Griffen AL, Leys EJ, Lamont RJ. Filifactor alocis interactions with gingival epithelial cells. *Mol Oral Microbiol*. Published online 2011. doi:10.1111/j.2041-1014.2011.00624.x

75. Darveau RP. Periodontitis: A polymicrobial disruption of host homeostasis. *Nat Rev Microbiol*. Published online 2010. doi:10.1038/nrmicro2337

76. Ji S, Kim Y, Min BM, Han SH, Choi Y. Innate immune responses of gingival epithelial cells to nonperiodontopathic and periodontopathic bacteria. *J Periodontal Res*. Published online 2007. doi:10.1111/j.1600-0765.2007.00974.x

77. Petersen PE. World Health Organization global policy for improvement of oral health - World Health Assembly 2007. *Int Dent J*. Published online 2008. doi:10.1111/j.1875-595x.2008.tb00185.x

78. Vieira Colombo AP, Magalhães CB, Hartenbach FARR, Martins do Souto R, Maciel da Silva-Boghossian C. Periodontal-disease-associated biofilm: A reservoir for pathogens of medical importance. *Microb Pathog*. Published online 2015. doi:10.1016/j.micpath.2015.09.009

79. Yang J, Zhang Q, Chen M, et al. Association between Helicobacter pylori infection and risk of periodontal diseases in han Chinese: A case-control study. *Med Sci Monit*. Published online 2016. doi:10.12659/MSM.894583

80. Vimr E, Lichtensteiger C. To sialylate, or not to sialylate: That is the question. *Trends Microbiol*. Published online 2002. doi:10.1016/S0966-842X(02)02361-2

81. National Center for Biotechnology Information. PubChem Database. N-Acetyl-Muramic Acid. Accessed November 26, 2019. https://pubchem.ncbi.nlm.nih.gov/compound/N-Acetyl-Muramic-Acid

82. Blix FG, Gottschalk A, Klenk E. Proposed nomenclature in the field of neuraminic and sialic acids. *Nature*. Published online 1957. doi:10.1038/1791088b0

83. Angata T, Varki A. Chemical diversity in the sialic acids and related α-keto acids: An evolutionary perspective. *Chem Rev*. Published online 2002. doi:10.1021/cr000407m

84. Vimr ER, Kalivoda KA, Deszo EL, Steenbergen SM. Diversity of Microbial Sialic Acid Metabolism. *Microbiol Mol Biol Rev*. Published online 2004. doi:10.1128/mmbr.68.1.132-153.2004

85. Martin PT, Kawanishi K, Ashbrook A, et al. Serum Antibodies to N-Glycolylneuraminic Acid Are Elevated in Duchenne Muscular Dystrophy and Correlate with Increased Disease Pathology in Cmahmdx Mice. *Am J Pathol*. 2021;191(8):1474-1486. doi:10.1016/j.ajpath.2021.04.015

86. Ogura H. Development of miracle medicines from sialic acids. *Proc Japan Acad Ser B Phys Biol Sci*. Published online 2011. doi:10.2183/pjab.87.328

87. Severi E, Hood DW, Thomas GH. Sialic acid utilization by bacterial pathogens. *Microbiology*. Published online 2007. doi:10.1099/mic.0.2007/009480-0

88. Roy S, Douglas CWI, Stafford GP. A novel sialic acid utilization and uptake system in the periodontal pathogen Tannerella forsythia. *J Bacteriol*. Published online 2010. doi:10.1128/JB.00079-10

89. Parker D, Soong G, Planet P, Brower J, Ratner AJ, Prince A. The NanA neuraminidase of Streptococcus pneumoniae is involved in biofilm formation. *Infect Immun*. Published online 2009. doi:10.1128/IAI.00228-09

90. Soong G, Muir A, Gomez MI, et al. Bacterial neuraminidase facilitates mucosal infection by participating in biofilm production. *J Clin Invest*. Published online 2006. doi:10.1172/JCI27920

91. Honma K, Ruscitto A, Frey AM, Stafford GP, Sharma A. Sialic acid transporter NanT participates in Tannerella forsythia biofilm formation and survival on epithelial cells. *Microb Pathog*. Published online 2015. doi:10.1016/j.micpath.2015.08.012

92. Byres E, Paton AW, Paton JC, et al. Incorporation of a non-human glycan mediates human susceptibility to a bacterial toxin. *Nature*. Published online 2008. doi:10.1038/nature07428

93. Roggentin P, Schauer R, Hoyer LL, Vimr ER. The sialidase superfamily and its spread by horizontal gene transfer. *Mol Microbiol*. Published online 1993. doi:10.1111/j.1365-2958.1993.tb01221.x

94. Crennell S, Garman E, Laver G, Vimr E, Taylor G. Crystal structure of Vibrio cholerae neuraminidase reveals dual lectin-like domains in addition to the catalytic domain. *Structure*. Published online 1994. doi:10.1016/S0969-2126(00)00053-8

95. Moncla BJ, Braham P, Hillier SL. Sialidase (neuraminidase) activity among gram-negative anaerobic and capnophilic bacteria. *J Clin Microbiol*. Published online 1990.

96. Roggentin P, Rothe B, Kaper JB, et al. Conserved sequences in bacterial and viral sialidases. *Glycoconj J*. Published online 1989. doi:10.1007/BF01047853

97. Kuroiwa A, Hisatsune A, Isohama Y, Katsuki H. Bacterial neuraminidase increases IL-8 production in lung epithelial cells via NF-κB-dependent pathway. *Biochem Biophys Res Commun*. Published online 2009. doi:10.1016/j.bbrc.2008.12.120

98. Corfield T. Bacterial sialidases - roles in pathogenicity and nutrition. *Glycobiology*. Published online 1992. doi:10.1093/glycob/2.6.509

99. Tong HH, Blue LE, James MA, DeMaria TF. Evaluation of the virulence of a Streptococcus pneumoniae neuraminidase- deficient mutant in nasopharyngeal colonization and development of otitis media in the chinchilla model. *Infect Immun*. Published online 2000. doi:10.1128/IAI.68.2.921-924.2000

100. Bradshaw DJ, Marsh PD, Keith Watson G, Allison C. Role of Fusobacterium nucleatum and coaggregation in anaerobe survival in planktonic and biofilm oral microbial communities during aeration. *Infect Immun*. Published online 1998.

101. Vimr ER. Microbial sialidases: does bigger always mean better? *Trends Microbiol*. Published online 1994. doi:10.1016/0966-842X(94)90003-5

102. Honma K, Mishima E, Sharma A. Role of tannerella forsythia nanh sialidase in epithelial cell attachment. *Infect Immun*. Published online 2011. doi:10.1128/IAI.00629-10

103. Cointe D, Leroy Y, Chirat F. Determination of the sialylation level and of the ratio α-(2→3)/α-(2→6) sialyl linkages of N-glycans by methylation and GC/MS analysis. *Carbohydr Res*. Published online 1998. doi:10.1016/S0008-6215(98)00196-7

104. Mizan S, Henk A, Stallings A, Maier M, Lee MD. Cloning and characterization of sialidases with 2-6′ and 2-3′ sialyl lactose specificity from Pasteurella multocida. *J Bacteriol*. Published online 2000. doi:10.1128/JB.182.24.6874-6883.2000

105. Roy S, Honma K, Ian Douglas CW, Sharma A, Stafford GP. Role of sialidase in glycoprotein utilization by Tannerella forsythia. *Microbiology*. Published online 2011. doi:10.1099/mic.0.052498-0

106. Varki A, Diaz S. A neuraminidase from Streptococcus sanguis that can release O-acetylated sialic acids. *J Biol Chem*. Published online 1983.

107. Corfield AP, Wagner SA, Clamp JR, Kriaris MS, Hoskins LC. Mucin degradation in the human colon: Production of sialidase, sialate O- acetylesterase, N-acetylneuraminate lyase, arylesterase, and glycosulfatase activities by strains of fecal bacteria. *Infect Immun*. Published online 1992.

108. Klein A, Roussel P. O-Acetylation of sialic acids. *Biochimie*. Published online 1998. doi:10.1016/S0300-9084(98)80056-4

109. Pillai S, Cariappa A, Pirnie SP. Esterases and autoimmunity: the sialic acid acetylesterase pathway and the regulation of peripheral B cell tolerance. *Trends Immunol*. Published online 2009. doi:10.1016/j.it.2009.07.006

110. Phansopa C, Kozak RP, Liew LP, et al. Characterization of a sialate-O-acetylesterase (NanS) from the oral pathogen Tannerella forsythia that enhances sialic acid release by NanH, its cognate sialidase. *Biochem J*. Published online 2015. doi:10.1042/BJ20150388

111. Frey AM, Satur MJ, Phansopa C, et al. Evidence for a carbohydrate-binding module (CBM) of Tannerella forsythia NanH sialidase, key to interactions at the host–pathogen interface. *Biochem J*. Published online 2018. doi:10.1042/BCJ20170592

112. Noinaj N, Guillier M, Barnard, TJ, Buchanan SK. TonB-Dependent Transporters: Regulation, Structure, and Function. *Annu Rev Microbiol*. Published online 2010. doi:10.1146/annurev.micro.112408.134247

113. Koebnik R. TonB-dependent trans-envelope signalling: The exception or the rule? *Trends Microbiol*. Published online 2005. doi:10.1016/j.tim.2005.06.005

114. Valentin-Hansen P, Eriksen M, Udesen C. The bacterial Sm-like protein Hfq: A key player in RNA transactions. *Mol Microbiol*. Published online 2004. doi:10.1111/j.1365-2958.2003.03935.x

115. Staroń A, Sofia HJ, Dietrich S, Ulrich LE, Liesegang H, Mascher T. The third pillar of bacterial signal transduction: Classification of the extracytoplasmic function (ECF) σ factor protein family. *Mol Microbiol*. Published online 2009. doi:10.1111/j.1365-2958.2009.06870.x

116. Massé E, Gottesman S. A small RNA regulates the expression of genes involved in iron metabolism in Escherichia coli. *Proc Natl Acad Sci U S A*. Published online 2002. doi:10.1073/pnas.032066599

117. Chen Z, Lewis KA, Shultzaberger RK, et al. Discovery of Fur binding site clusters in Escherichia coli by information theory models. *Nucleic Acids Res*. Published online 2007. doi:10.1093/nar/gkm631

118. Chimento DP, Kadner RJ, Wiener MC. The Escherichia coli outer membrane cobalamin transporter BtuB: Structural analysis of calcium and substrate binding, and identification of orthologous transporters by sequence/structure conservation. *J Mol Biol*. Published online 2003. doi:10.1016/j.jmb.2003.07.005

119. Bags A, Neilands JB. Ferric Uptake Regulation Protein Acts as a Repressor, Employing Iron(II) as a Cofactor To Bind the Operator of an Iron Transport Operon in Escherichia coli. *Biochemistry*. Published online 1987. doi:10.1021/bi00391a039

120. Blanvillain S, Meyer D, Boulanger A, et al. Plant carbohydrate scavenging through TonB-dependent receptors: A feature shared by phytopathogenic and aquatic bacteria. *PLoS One*. Published online 2007. doi:10.1371/journal.pone.0000224

121. Gómez J., Criado M., Ferreirós C. Cooperation between the components of the meningococcal transferrin receptor, TbpA and TbpB, in the uptake of transferrin iron

by the 37-kDa ferric-binding protein (FbpA). *Res Microbiol*. 1998;149(6):381-387. doi:10.1016/S0923-2508(98)80320-3

122. Martens EC, Koropatkin NM, Smith TJ, Gordon JI. Complex glycan catabolism by the human gut microbiota: The bacteroidetes sus-like paradigm. *J Biol Chem*. Published online 2009. doi:10.1074/jbc.R109.022848

123. Shipman JA, Berleman JE, Salyers AA. Characterization of four outer membrane proteins involved in binding starch to the cell surface of Bacteroides thetaiotaomicron. *J Bacteriol*. Published online 2000. doi:10.1128/JB.182.19.5365-5372.2000

124. Kitamura M, Okuyama M, Tanzawa F, et al. Structural and functional analysis of a glycoside hydrolase family 97 enzyme from Bacteroides thetaiotaomicron. *J Biol Chem*. Published online 2008. doi:10.1074/jbc.M806115200

125. Xu J, Mahowald MA, Ley RE, et al. Evolution of Symbiotic Bacteria in the Distal Human Intestine. *PLoS Biol*. Published online 2007. doi:10.1371/journal.pbio.0050156

126. Martens EC, Chiang HC, Gordon JI. Mucosal Glycan Foraging Enhances Fitness and Transmission of a Saccharolytic Human Gut Bacterial Symbiont. *Cell Host Microbe*. Published online 2008. doi:10.1016/j.chom.2008.09.007

127. Martens EC, Roth R, Heuser JE, Gordon JI. Coordinate regulation of glycan degradation and polysaccharide capsule biosynthesis by a prominent human gut symbiont. *J Biol Chem*. Published online 2009. doi:10.1074/jbc.M109.008094

128. Phansopa C, Roy S, Rafferty JB, et al. Structural and functional characterization of NanU, a novel high-affinity sialic acid-inducible binding protein of oral and gut-dwelling Bacteroidetes species. *Biochem J*. Published online 2014. doi:10.1042/BJ20131415

129. Braun V. Avoidance of iron toxicity through regulation of bacterial iron transport. *Biol Chem*. Published online 1997.

130. Bradbeer C. The proton motive force drives the outer membrane transport of cobalamin in Escherichia coli. *J Bacteriol*. Published online 1993. doi:10.1128/jb.175.10.3146-3150.1993

131. Reynolds PR, Mottur GP, Bradbeer C. Transport of vitamin B 12 in Escherichia coli. Some observations on the roles of the gene products of BtuC and TonB. *J Biol Chem*.

Published online 1980.

132. Higgs PI, Myers PS, Postle K. Interactions in the TonB-dependent energy transduction complex: ExbB and ExbD form homomultimers. *J Bacteriol*. Published online 1998.

133. Celia H, Noinaj N, Buchanan SK. Structure and Stoichiometry of the Ton Molecular Motor. *Int J Mol Sci*. 2020;21(2):375. doi:10.3390/ijms21020375

134. Maki-Yonekura S, Matsuoka R, Yamashita Y, et al. Hexameric and pentameric complexes of the ExbBD energizer in the Ton system. *Elife*. 2018;7. doi:10.7554/eLife.35419

135. Celia H, Noinaj N, Zakharov SD, et al. Structural insight into the role of the Ton complex in energy transduction. *Nature*. 2016;538(7623):60-65. doi:10.1038/nature19757

136. Higgs PI, Larsen RA, Postle K. Quantification of known components of the Escherichia coli TonB energy transduction system: TonB, ExbB, ExbD and FepA. *Mol Microbiol*. Published online 2002. doi:10.1046/j.1365-2958.2002.02880.x

137. Stork M, Di Lorenzo M, Mouriño S, Osorio CR, Lemos ML, Crosa JH. Two tonB Systems Function in Iron Transport in Vibrio anguillarum , but Only One Is Essential for Virulence. *Infect Immun*. 2004;72(12):7326-7329. doi:10.1128/IAI.72.12.7326-7329.2004

138. Palyada K, Threadgill D, Stintzi A. Iron Acquisition and Regulation in Campylobacter jejuni. *J Bacteriol*. 2004;186(14):4714-4729. doi:10.1128/JB.186.14.4714-4729.2004

139. Foley TL, Simeonov A. Targeting iron assimilation to develop new antibacterials. *Expert Opin Drug Discov*. 2012;7(9):831-847. doi:10.1517/17460441.2012.708335

140. Wang CC, Newton A. Iron transport in Escherichia coli: roles of energy-dependent uptake and 2,3-dihydroxybenzoylserine. *J Bacteriol*. Published online 1969.

141. O'Brien IG, Gibson F. The structure of enterochelin and related 2,3-dihydroxy-N-benzoyne conjugates from Eschericha Coli. *BBA - Gen Subj*. Published online 1970. doi:10.1016/0304-4165(70)90038-3

142. Frost GE, Rosenberg H. Relationship between the tonB locus and iron transport in Escherichia coli. *J Bacteriol*. Published online 1975.

143. Jana B, Manning M, Postle K. Mutations in the ExbB cytoplasmic carboxy terminus prevent energy-dependent interaction between the TonB and ExbD periplasmic domains. *J Bacteriol*. Published online 2011. doi:10.1128/JB.05674-11

144. Postle K, Skare JT. Escherichia coli TonB protein is exported from the cytoplasm without proteolytic cleavage of its amino terminus. *J Biol Chem*. Published online 1988.

145. Manoil C, Beckwith J. A genetic approach to analyzing membrane protein topology. *Science (80-)*. Published online 1986. doi:10.1126/science.3529391

146. Moore KE, Miura S. A small hydrophobic domain anchors leader peptidase to the cytoplasmic membrane of Escherichia coli. *J Biol Chem*. Published online 1987.

147. Hannavy K, Barr GC, Dorman CJ, et al. TonB protein of Salmonella typhimurium. A model for signal transduction between membranes. *J Mol Biol*. Published online 1990. doi:10.1016/S0022-2836(99)80009-6

148. Letain TE, Postle K. TonB protein appears to transduce energy by shuttling between the cytoplasmic membrane and the outer membrane in Escherichia coli. *Mol Microbiol*. Published online 1997. doi:10.1046/j.1365-2958.1997.3331703.x

149. Chu BCH, Peacock RS, Vogel HJ. Bioinformatic analysis of the TonB protein family. In: *BioMetals*. ; 2007. doi:10.1007/s10534-006-9049-4

150. Chang C, Mooser A, Plückthun A, Wlodawer A. Crystal Structure of the Dimeric C-terminal Domain of TonB Reveals a Novel Fold. *J Biol Chem*. 2001;276(29):27535-27540. doi:10.1074/jbc.M102778200

151. Sauter A, Howard SP, Braun V. In Vivo Evidence for TonB Dimerization. *J Bacteriol*. 2003;185(19):5747-5754. doi:10.1128/JB.185.19.5747-5754.2003

152. WIENER M. TonB-dependent outer membrane transport: going for Baroque? *Curr Opin Struct Biol*. 2005;15(4):394-400. doi:10.1016/j.sbi.2005.07.001

153. Khursigara CM, De Crescenzo G, Pawelek PD, Coulton JW. Enhanced Binding of TonB to a Ligand-loaded Outer Membrane Receptor. *J Biol Chem*. 2004;279(9):7405-7412. doi:10.1074/jbc.M311784200

154. Postle K, Kastead KA, Gresock MG, Ghosh J, Swayne CD. The TonB Dimeric Crystal

Structures Do Not Exist In Vivo. Wickner R, ed. *MBio*. 2010;1(5). doi:10.1128/mBio.00307-10

155. Sean Peacock R, Weljie AM, Peter Howard S, Price FD, Vogel HJ. The Solution Structure of the C-terminal Domain of TonB and Interaction Studies with TonB Box Peptides. *J Mol Biol*. 2005;345(5):1185-1197. doi:10.1016/j.jmb.2004.11.026

156. Pawelek PD, Croteau N, Ng-Thow-Hing C, et al. Structure of TonB in Complex with FhuA, E. coli Outer Membrane Receptor. *Science (80-)*. 2006;312(5778):1399-1402. doi:10.1126/science.1128057

157. Shultis DD, Purdy MD, Banchs CN, Wiener MC. Outer Membrane Active Transport: Structure of the BtuB:TonB Complex. *Science (80-)*. 2006;312(5778):1396-1399. doi:10.1126/science.1127694

158. Freed DM, Lukasik SM, Sikora A, Mokdad A, Cafiso DS. Monomeric TonB and the Ton Box Are Required for the Formation of a High-Affinity Transporter–TonB Complex. *Biochemistry*. 2013;52(15):2638-2648. doi:10.1021/bi3016108

159. Oeemig JS, Ollila OHS, Iwaï H. NMR structure of the C-terminal domain of TonB protein from Pseudomonas aeruginosa. *PeerJ*. 2018;6:e5412. doi:10.7717/peerj.5412

160. Gumbart J, Wiener MC, Tajkhorshid E. Mechanics of Force Propagation in TonB-Dependent Outer Membrane Transport. *Biophys J*. 2007;93(2):496-504. doi:10.1529/biophysj.107.104158

161. Josts I, Veith K, Tidow H. Ternary structure of the outer membrane transporter FoxA with resolved signalling domain provides insights into TonB-mediated siderophore uptake. *Elife*. 2019;8. doi:10.7554/eLife.48528

162. Fischer E, Günter K, Braun V. Involvement of ExbB and TonB in transport across the outer membrane of Escherichia coli: phenotypic complementation of exb mutants by overexpressed tonB and physical stabilization of TonB by ExbB. *J Bacteriol*. Published online 1989. doi:10.1128/jb.171.9.5127-5134.1989

163. Ahmer BMM, Thomas MG, Larsen RA, Postle K. Characterization of the exbBD operon of Escherichia coli and the role of ExbB and ExbD in TonB function and stability. *J Bacteriol*. Published online 1995. doi:10.1128/jb.177.16.4742-4747.1995

164. Ollis AA, Manning M, Held KG, Postle K. Cytoplasmic membrane protonmotive force

energizes periplasmic interactions between ExbD and TonB. *Mol Microbiol*. Published online 2009. doi:10.1111/j.1365-2958.2009.06785.x

165. Held KG, Postle K. ExbB and ExbD do not function independently in TonB-dependent energy transduction. *J Bacteriol*. Published online 2002. doi:10.1128/JB.184.18.5170-5173.2002

166. Garcia-Herrero A, Peacock RS, Howard SP, Vogel HJ. The solution structure of the periplasmic domain of the TonB system ExbD protein reveals an unexpected structural homology with siderophore-binding proteins. *Mol Microbiol*. 2007;66(4):872-889. doi:10.1111/j.1365-2958.2007.05957.x

167. Celia H, Botos I, Ni X, et al. Cryo-EM structure of the bacterial Ton motor subcomplex ExbB–ExbD provides information on structure and stoichiometry. *Commun Biol*. 2019;2(1):358. doi:10.1038/s42003-019-0604-2

168. Sverzhinsky A, Chung JW, Deme JC, et al. Membrane Protein Complex ExbB 4 -ExbD 1 -TonB 1 from Escherichia coli Demonstrates Conformational Plasticity. Armitage JP, ed. *J Bacteriol*. 2015;197(11):1873-1885. doi:10.1128/JB.00069-15

169. Chorev DS, Baker LA, Wu D, et al. Protein assemblies ejected directly from native membranes yield complexes for mass spectrometry. *Science (80-)*. 2018;362(6416):829-834. doi:10.1126/science.aau0976

170. James W Coulton. High-resolution electron microscopy of energy-transducing proteins of «Escherichia coli». doi:https://escholarship.mcgill.ca/concern/theses/2f75rd026

171. Pramanik A, Zhang F, Schwarz H, Schreiber F, Braun V. ExbB Protein in the Cytoplasmic Membrane of Escherichia coli Forms a Stable Oligomer. *Biochemistry*. 2010;49(40):8721-8728. doi:10.1021/bi101143y

172. Fanucci GE, Cadieux N, Kadner RJ, Cafiso DS. Competing ligands stabilize alternate conformations of the energy coupling motif of a TonB-dependent outer membrane transporter. *Proc Natl Acad Sci U S A*. Published online 2003. doi:10.1073/pnas.1932486100

173. Fanucci GE, Lee JY, Cafiso DS. Spectroscopic Evidence that Osmolytes Used in Crystallization Buffers Inhibit a Conformation Change in a Membrane Protein. *Biochemistry*. Published online 2003. doi:10.1021/bi035439t

174.	Howard SP, Herrmann C, Stratilo CW, Braun V. In Vivo Synthesis of the Periplasmic Domain of TonB Inhibits Transport through the FecA and FhuA Iron Siderophore Transporters of Escherichia coli. *J Bacteriol*. 2001;183(20):5885-5895. doi:10.1128/JB.183.20.5885-5895.2001

175.	Hickman SJ, Cooper REM, Bellucci L, Paci E, Brockwell DJ. Gating of TonB-dependent transporters by substrate-specific forced remodelling. *Nat Commun*. 2017;8(1):14804. doi:10.1038/ncomms14804

176.	Cadieux N, Bradbeer C, Kadner RJ. Sequence Changes in the Ton Box Region of BtuB Affect Its Transport Activities and Interaction with TonB Protein. *J Bacteriol*. 2000;182(21):5954-5961. doi:10.1128/JB.182.21.5954-5961.2000

177.	Gudmundsdottir A, Bell PE, Lundrigan MD, Bradbeer C, Kadner RJ. Point mutations in a conserved region (TonB box) of Escherichia coli outer membrane protein BtuB affect vitamin B12 transport. *J Bacteriol*. 1989;171(12):6526-6533. doi:10.1128/jb.171.12.6526-6533.1989

178.	Chimento DP, Kadner RJ, Wiener MC. Comparative structural analysis of TonB-dependent outer membrane transporters: Implications for the transport cycle. *Proteins Struct Funct Bioinforma*. 2005;59(2):240-251. doi:10.1002/prot.20416

179.	Brockwell DJ, Paci E, Zinober RC, et al. Pulling geometry defines the mechanical resistance of a β-sheet protein. *Nat Struct Mol Biol*. 2003;10(9):731-737. doi:10.1038/nsb968

180.	Carrion-Vazquez M, Li H, Lu H, Marszalek PE, Oberhauser AF, Fernandez JM. The mechanical stability of ubiquitin is linkage dependent. *Nat Struct Mol Biol*. 2003;10(9):738-743. doi:10.1038/nsb965

181.	Lundrigan MD, Kadner RJ. Nucleotide sequence of the gene for the ferrienterochelin receptor FepA in Escherichia coli. Homology among outer membrane receptors that interact with TonB. *J Biol Chem*. 1986;261(23):10797-10801. doi:10.1016/S0021-9258(18)67457-5

182.	Annamalai R, Jin B, Cao Z, Newton SMC, Klebba PE. Recognition of Ferric Catecholates by FepA. *J Bacteriol*. 2004;186(11):3578-3589. doi:10.1128/JB.186.11.3578-3589.2004

183. Braun M, Endriss F, Killmann H, Braun V. In Vivo Reconstitution of the FhuA Transport Protein of Escherichia coli K-12. *J Bacteriol*. 2003;185(18):5508-5518. doi:10.1128/JB.185.18.5508-5518.2003

184. Magnet S, Dubost L, Marie A, Arthur M, Gutmann L. Identification of the <scp>l,d</scp> -Transpeptidases for Peptidoglycan Cross-Linking in *Escherichia coli*. *J Bacteriol*. 2008;190(13):4782-4785. doi:10.1128/JB.00025-08

185. Meroueh SO, Bencze KZ, Hesek D, et al. Three-dimensional structure of the bacterial cell wall peptidoglycan. *Proc Natl Acad Sci*. 2006;103(12):4404-4409. doi:10.1073/pnas.0510182103

186. Park JS, Lee WC, Yeo KJ, et al. Mechanism of anchoring of OmpA protein to the cell wall peptidoglycan of the gram-negative bacterial outer membrane. *FASEB J*. 2012;26(1):219-228. doi:10.1096/fj.11-188425

187. Roujeinikova A. Crystal structure of the cell wall anchor domain of MotB, a stator component of the bacterial flagellar motor: Implications for peptidoglycan recognition. *Proc Natl Acad Sci*. 2008;105(30):10348-10353. doi:10.1073/pnas.0803039105

188. Wojdyla JA, Cutts E, Kaminska R, et al. Structure and Function of the Escherichia coli Tol-Pal Stator Protein TolR. *J Biol Chem*. 2015;290(44):26675-26687. doi:10.1074/jbc.M115.671586

189. Horiyama T, Nishino K. AcrB, AcrD, and MdtABC Multidrug Efflux Systems Are Involved in Enterobactin Export in Escherichia coli. Cloeckaert A, ed. *PLoS One*. 2014;9(9):e108642. doi:10.1371/journal.pone.0108642

190. Vollmer W, Seligman SJ. Architecture of peptidoglycan: more data and more models. *Trends Microbiol*. 2010;18(2):59-66. doi:10.1016/j.tim.2009.12.004

191. Brewer S, Tolley M, Trayer IP, et al. Structure and function of X-Pro dipeptide repeats in the TonB proteins of Salmonella typhimurium and Escherichia coli. *J Mol Biol*. 1990;216(4):883-895. doi:10.1016/S0022-2836(99)80008-4

192. Di Masi DR, White JC, Schnaitman CA, Bradbeer C. Transport of Vitamin B 12 in Escherichia coli : Common Receptor Sites for Vitamin B 12 and the E Colicins on the Outer Membrane of the Cell Envelope. *J Bacteriol*. 1973;115(2):506-513. doi:10.1128/jb.115.2.506-513.1973

193. Bateman A, Bycroft M. The structure of a LysM domain from E. coli membrane-bound lytic murein transglycosylase D (MltD) 1 1Edited by P. E. Wight. *J Mol Biol*. 2000;299(4):1113-1119. doi:10.1006/jmbi.2000.3778

194. Leo JC, Oberhettinger P, Chaubey M, et al. The <scp>I</scp> ntimin periplasmic domain mediates dimerisation and binding to peptidoglycan. *Mol Microbiol*. 2015;95(1):80-100. doi:10.1111/mmi.12840

195. Maxwell KL, Fatehi Hassanabad M, Chang T, et al. Structural and Functional Studies of gpX of Escherichia coli Phage P2 Reveal a Widespread Role for LysM Domains in the Baseplates of Contractile-Tailed Phages. *J Bacteriol*. 2013;195(24):5461-5468. doi:10.1128/JB.00805-13

196. Kaserer WA, Jiang X, Xiao Q, et al. Insight from TonB Hybrid Proteins into the Mechanism of Iron Transport through the Outer Membrane. *J Bacteriol*. 2008;190(11):4001-4016. doi:10.1128/JB.00135-08

197. Wayne R, Neilands JB. Evidence for common binding sites for ferrichrome compounds and bacteriophage phi 80 in the cell envelope of Escherichia coli. *J Bacteriol*. 1975;121(2):497-503. doi:10.1128/jb.121.2.497-503.1975

198. Gresock MG, Kastead KA, Postle K. From Homodimer to Heterodimer and Back: Elucidating the TonB Energy Transduction Cycle. Parkinson JS, ed. *J Bacteriol*. 2015;197(21):3433-3445. doi:10.1128/JB.00484-15

199. Jordan LD, Zhou Y, Smallwood CR, et al. Energy-dependent motion of TonB in the Gram-negative bacterial inner membrane. *Proc Natl Acad Sci*. 2013;110(28):11553-11558. doi:10.1073/pnas.1304243110

200. Jakobs S, Subramaniam V, Schönle A, Jovin TM, Hell SW. EGFP and DsRed expressing cultures of Escherichia coli imaged by confocal, two-photon and fluorescence lifetime microscopy. *FEBS Lett*. 2000;479(3):131-135. doi:10.1016/S0014-5793(00)01896-2

201. Swaminathan R, Hoang CP, Verkman AS. Photobleaching recovery and anisotropy decay of green fluorescent protein GFP-S65T in solution and cells: cytoplasmic viscosity probed by green fluorescent protein translational and rotational diffusion. *Biophys J*. 1997;72(4):1900-1907. doi:10.1016/S0006-3495(97)78835-0

202. Condemine G, Berrier C, Plumbridge J, Ghazi A. Function and expression of an N-acetylneuraminic acid-inducible outer membrane channel in Escherichia coli. *J Bacteriol*. Published online 2005. doi:10.1128/JB.187.6.1959-1965.2005

203. Stafford G, Roy S, Honma K, Sharma A. Sialic acid, periodontal pathogens and Tannerella forsythia: Stick around and enjoy the feast! *Mol Oral Microbiol*. Published online 2012. doi:10.1111/j.2041-1014.2011.00630.x

204. Vimr ER, Troy FA. Identification of an inducible catabolic system for sialic acids (nan) in Escherichia coli. *J Bacteriol*. Published online 1985.

205. Tangvoranuntakul P, Gagneux P, Diaz S, et al. Human uptake and incorporation of an immunogenic nonhuman dietary sialic acid. *Proc Natl Acad Sci U S A*. Published online 2003. doi:10.1073/pnas.2131556100

206. Martin MJ, Rayner JC, Gagneux P, Barnwell JW, Varki A. Evolution of human-chimpanzee differences in malaria susceptibility: Relationship to human genetic loss of N-glycolylneuraminic acid. *Proc Natl Acad Sci U S A*. Published online 2005. doi:10.1073/pnas.0503819102

207. Guzman LM, Belin D, Carson MJ, Beckwith J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose P(BAD) promoter. *J Bacteriol*. Published online 1995. doi:10.1128/jb.177.14.4121-4130.1995

208. Fraser GM, Hirano T, Ferris HU, Devgan LL, Kihara M, Macnab RM. Substrate specificity of type III flagellar protein export in Salmonella is controlled by subdomain interactions in FlhB. *Mol Microbiol*. Published online 2003. doi:10.1046/j.1365-2958.2003.03487.x

209. AMINOFF D. Methods for the quantitative estimation of N -acetylneuraminic acid and their application to hydrolysates of sialomucoids. *Biochem J*. 1961;81(2):384-392. doi:10.1042/bj0810384

210. Ge SX, Son EW, Yao R. iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. *BMC Bioinformatics*. 2018;19(1):534. doi:10.1186/s12859-018-2486-6

211. Fruzangohar M, Ebrahimie E, Ogunniyi AD, Mahdi LK, Paton JC, Adelson DL. Comparative GO: A Web Application for Comparative Gene Ontology and Gene

Ontology-Based Gene Selection in Bacteria. Patterson RL, ed. *PLoS One*. 2013;8(3):e58759. doi:10.1371/journal.pone.0058759

212. Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. 2000;28(1):27-30. doi:10.1093/nar/28.1.27

213. Bu D, Luo H, Huo P, et al. KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acids Res*. 2021;49(W1):W317-W325. doi:10.1093/nar/gkab447

214. Phansopa C, Roy S, Rafferty JB, et al. Structural and functional characterization of NanU, a novel high-affinity sialic acid-inducible binding protein of oral and gut-dwelling Bacteroidetes species. *Biochem J*. 2014;458(3):499-511. doi:10.1042/BJ20131415

215. Friedrich V, Pabinger S, Chen T, Messner P, Dewhirst FE, Schäffer C. Draft genome sequence of Tannerella forsythia type strain ATCC 43037. *Genome Announc*. Published online 2015. doi:10.1128/genomeA.00660-15

216. Lu S, Wang J, Chitsaz F, et al. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res*. 2020;48(D1):D265-D268. doi:10.1093/nar/gkz991

217. Buchanan JM. The Amidotransferases. In: ; 2006:91-183. doi:10.1002/9780470122846.ch2

218. Abe Y, Fujisaki N, Miyoshi T, Watanabe N, Katayama T, Ueda T. Functional analysis of CedA based on its structure: residues important in binding of DNA and RNA polymerase and in the cell division regulation. *J Biochem*. 2016;159(2):217-223. doi:10.1093/jb/mvv096

219. Larsen RA, Postle K. Conserved Residues Ser16 and His20 and Their Relative Positioning Are Essential for TonB Activity, Cross-linking of TonB with ExbB, and the Ability of TonB to Respond to Proton Motive Force. *J Biol Chem*. Published online 2001. doi:10.1074/jbc.M007479200

220. Wei Zheng. No Title. http://zhanglab.ccmb.med.umich.edu/I-TASSER

221. PrediSi. No Title. http://www.predisi.de/index.html

222. SignalP - 6.0. No Title. https://services.healthtech.dtu.dk/service.php?SignalP

223. LipoP - 1.0. No Title. https://services.healthtech.dtu.dk/service.php?LipoP-1.0

224. Shemesh R, Novik A, Cohen Y. Follow the Leader: Preference for Specific Amino Acids Directly Following the Initial Methionine in Proteins of Different Organisms. *Genomics Proteomics Bioinformatics*. 2010;8(3):180-189. doi:10.1016/S1672-0229(10)60020-4

225. Kenney LJ, Anand GS. EnvZ/OmpR Two-Component Signaling: An Archetype System That Can Function Noncanonically. Slauch JM, ed. *EcoSal Plus*. 2020;9(1). doi:10.1128/ecosalplus.ESP-0001-2019

226. Guillier M, Gottesman S. Remodelling of the Escherichia coli outer membrane by two small regulatory RNAs. *Mol Microbiol*. 2006;59(1):231-247. doi:10.1111/j.1365-2958.2005.04929.x

227. Chetri S, Singha M, Bhowmik D, et al. Transcriptional response of OmpC and OmpF in Escherichia coli against differential gradient of carbapenem stress. *BMC Res Notes*. 2019;12(1):138. doi:10.1186/s13104-019-4177-4

228. Mattison K, Oropeza R, Byers N, Kenney LJ. A phosphorylation site mutant of OmpR reveals different binding conformations at ompF and ompC 1 1Edited by R. Ebright. *J Mol Biol*. 2002;315(4):497-511. doi:10.1006/jmbi.2001.5222

229. Chakraborty S, Winardhi RS, Morgan LK, Yan J, Kenney LJ. Non-canonical activation of OmpR drives acid and osmotic stress responses in single bacterial cells. *Nat Commun*. 2017;8(1):1587. doi:10.1038/s41467-017-02030-0

230. Huang L, Tsui P, Freundlich M. Integration host factor is a negative effector of in vivo and in vitro expression of ompC in Escherichia coli. *J Bacteriol*. 1990;172(9):5293-5298. doi:10.1128/jb.172.9.5293-5298.1990

231. Sheridan SD, Benham CJ, Hatfield GW. Activation of Gene Expression by a Novel DNA Structural Transmission Mechanism That Requires Supercoiling-induced DNA Duplex Destabilization in an Upstream Activating Sequence. *J Biol Chem*. 1998;273(33):21298-21308. doi:10.1074/jbc.273.33.21298

232. Efimova VS, Isaeva L V., Makeeva DS, Rubtsov MA, Novikova LA. Expression of Cholesterol Hydroxylase/Lyase System Proteins in Yeast S. cerevisiae Cells as a Self-Processing Polyprotein. *Mol Biotechnol*. 2017;59(9-10):394-406. doi:10.1007/s12033-

017-0028-5

233. Tribble GD, Rigney TW, Dao DH V., et al. Natural competence is a major mechanism for horizontal DNA transfer in the oral pathogen Porphyromonas gingivalis. *MBio*. Published online 2012. doi:10.1128/mBio.00231-11

234. Nishikawa K, Tanaka Y. A simple mutagenesis using natural competence in Tannerella forsythia. *J Microbiol Methods*. Published online 2013. doi:10.1016/j.mimet.2013.07.011

235. Dong Y, Geng J, Liu J, et al. Roles of three TonB systems in the iron utilization and virulence of the Aeromonas hydrophila Chinese epidemic strain NJ-35. *Appl Microbiol Biotechnol*. 2019;103(10):4203-4215. doi:10.1007/s00253-019-09757-4

236. Noinaj N, Guillier M, Barnard, TJ, Buchanan SK. TonB-Dependent Transporters: Regulation, Structure, and Function. *Annu Rev Microbiol*. 2010;64(1):43-60. doi:10.1146/annurev.micro.112408.134247

237. Pradenas G, Myers J, Torres A. Characterization of the Burkholderia cenocepacia TonB Mutant as a Potential Live Attenuated Vaccine. *Vaccines*. 2017;5(4):33. doi:10.3390/vaccines5040033

238. Larsen RA, Wood GE, Postle K. The conserved proline-rich Motif is not essential for energy transduction by Escherichia coliTonB protein. *Mol Microbiol*. Published online 1993. doi:10.1111/j.1365-2958.1993.tb00966.x

239. Seliger SS, Mey AR, Valle AM, Payne SM. The two TonB systems of Vibrio cholerae: Redundant and specific functions. *Mol Microbiol*. Published online 2001. doi:10.1046/j.1365-2958.2001.02273.x

240. Khursigara CM, De Crescenzo G, Pawelek PD, Coulton JW. Deletion of the proline-rich region of TonB disrupts formation of a 2:1 complex with FhuA, an outer membrane receptor of Escherichia coli. *Protein Sci*. 2005;14(5):1266-1273. doi:10.1110/ps.051342505

241. Carter DM, Miousse IR, Gagnon J-N, et al. Interactions between TonB from Escherichia coli and the Periplasmic Protein FhuD. *J Biol Chem*. 2006;281(46):35413-35424. doi:10.1074/jbc.M607611200

242. Zhao Q, Poole K. Mutational analysis of the TonB1 energy coupler of Pseudomonas

aeruginosa. *J Bacteriol*. Published online 2002. doi:10.1128/JB.184.6.1503-1513.2002

243. Hanique S, Colombo ML, Goormaghtigh E, Soumillion P, Frère JM, Joris B. Evidence of an Intramolecular Interaction between the Two Domains of the BlaR1 Penicillin Receptor during the Signal Transduction. *J Biol Chem*. Published online 2004. doi:10.1074/jbc.M313488200

244. Río SJ, Osorio CR, Lemos ML. Heme uptake genes in human and fish isolates of Photobacterium damselae: existence of hutA pseudogenes. *Arch Microbiol*. 2005;183(5):347-358. doi:10.1007/s00203-005-0779-4

245. Parker AC, Seals NL, Baccanale CL, Rocha ER. Analysis of Six tonB Gene Homologs in Bacteroides fragilis Revealed That tonB3 is Essential for Survival in Experimental Intestinal Colonization and Intra-Abdominal Infection. Raffatellu M, ed. *Infect Immun*. 2022;90(1). doi:10.1128/IAI.00469-21

246. Liao H, Cheng X, Zhu D, et al. TonB Energy Transduction Systems of Riemerella anatipestifer Are Required for Iron and Hemin Utilization. *PLoS One*. 2015;10(5):e0127506. doi:10.1371/journal.pone.0127506

247. Zimbler DL, Arivett BA, Beckett AC, Menke SM, Actis LA. Functional Features of TonB Energy Transduction Systems of Acinetobacter baumannii. Payne SM, ed. *Infect Immun*. 2013;81(9):3382-3394. doi:10.1128/IAI.00540-13

248. Álvarez B, Álvarez J, Menéndez A, Guijarro JA. A mutant in one of two exbD loci of a TonB system in Flavobacterium psychrophilum shows attenuated virulence and confers protection against cold water disease. *Microbiology*. 2008;154(4):1144-1151. doi:10.1099/mic.0.2007/010900-0

249. Wiggerich H-G, Pühler A. The exbD2 gene as well as the iron-uptake genes tonB, exbB and exbD1 of Xanthomonas campestris pv. campestris are essential for the induction of a hypersensitive response on pepper (Capsicum annuum). *Microbiology*. 2000;146(5):1053-1060. doi:10.1099/00221287-146-5-1053

250. Derrien M, van Passel MWJ, van de Bovenkamp JHB, Schipper R, de Vos W, Dekker J. Mucin-bacterial interactions in the human oral cavity and digestive tract. *Gut Microbes*. 2010;1(4):254-268. doi:10.4161/gmic.1.4.12778

251. Hottmann I, Borisova M, Schäffer C, Mayer C. Peptidoglycan Salvage Enables the

Periodontal Pathogen Tannerella forsythia to Survive within the Oral Microbial Community. *Microb Physiol*. 2021;31(2):123-134. doi:10.1159/000516751

252.   Chatterjee A, O'Brian MR. Rapid evolution of a bacterial iron acquisition system. *Mol Microbiol*. 2018;108(1):90-100. doi:10.1111/mmi.13918

253.   Miao S, Xing L, Qi J, et al. Roles of the TonB1 and TonB2 proteins in haemin iron acquisition and virulence in Riemerella anatipestifer. *Microbiology*. 2015;161(8):1592-1599. doi:10.1099/mic.0.000123

254.   Naikare H, Butcher J, Flint A, Xu J, Raymond KN, Stintzi A. Campylobacter jejuni ferric–enterobactin receptor CfrA is TonB3 dependent and mediates iron acquisition from structurally different catechol siderophores. *Metallomics*. 2013;5(8):988. doi:10.1039/c3mt20254b

255.   Huang B, Ru K, Yuan Z, Whitchurch CB, Mattick JS. tonB3 Is Required for Normal Twitching Motility and Extracellular Assembly of Type IV Pili. *J Bacteriol*. 2004;186(13):4387-4389. doi:10.1128/JB.186.13.4387-4389.2004

256.   Hatcher CL, Muruato LA, Torres AG. Recent Advances in Burkholderia mallei and B. pseudomallei Research. *Curr Trop Med Reports*. 2015;2(2):62-69. doi:10.1007/s40475-015-0042-2

257.   Nierman WC, Feldblyum T V., Laub MT, et al. Complete genome sequence of Caulobacter crescentus. *Proc Natl Acad Sci*. 2001;98(7):4136-4141. doi:10.1073/pnas.061029298

258.   Roy S, Douglas CWI, Stafford GP. A Novel Sialic Acid Utilization and Uptake System in the Periodontal Pathogen Tannerella forsythia. *J Bacteriol*. 2010;192(9):2285-2293. doi:10.1128/JB.00079-10

259.   Bloch S, Tomek MB, Friedrich V, Messner P, Schäffer C. Nonulosonic acids contribute to the pathogenicity of the oral bacterium Tannerella forsythia. *Interface Focus*. 2019;9(2):20180064. doi:10.1098/rsfs.2018.0064

260.   Honma K, Inagaki S, Okuda K, Kuramitsu HK, Sharma A. Role of a Tannerella forsythia exopolysaccharide synthesis operon in biofilm development. *Microb Pathog*. Published online 2007. doi:10.1016/j.micpath.2007.01.003

261.   Honma K, Ruscitto A, Frey AM, Stafford GP, Sharma A. Sialic acid transporter NanT

participates in Tannerella forsythia biofilm formation and survival on epithelial cells. *Microb Pathog*. 2016;94:12-20. doi:10.1016/j.micpath.2015.08.012

262. Pollet RM, Martin LM, Koropatkin NM. TonB-dependent transporters in the Bacteroidetes: Unique domain structures and potential functions. *Mol Microbiol*. 2021;115(3):490-501. doi:10.1111/mmi.14683

263. Cheng Q, Yu MC, Reeves AR, Salyers AA. Identification and characterization of a Bacteroides gene, csuF, which encodes an outer membrane protein that is essential for growth on chondroitin sulfate. *J Bacteriol*. 1995;177(13):3721-3727. doi:10.1128/jb.177.13.3721-3727.1995

264. Lapébie P, Lombard V, Drula E, Terrapon N, Henrissat B. Bacteroidetes use thousands of enzyme combinations to break down glycans. *Nat Commun*. 2019;10(1):2043. doi:10.1038/s41467-019-10068-5

265. Martens EC, Kelly AG, Tauzin AS, Brumer H. The Devil Lies in the Details: How Variations in Polysaccharide Fine-Structure Impact the Physiology and Evolution of Gut Microbes. *J Mol Biol*. 2014;426(23):3851-3865. doi:10.1016/j.jmb.2014.06.022

266. Déjean G, Tamura K, Cabrera A, et al. Synergy between Cell Surface Glycosidases and Glycan-Binding Proteins Dictates the Utilization of Specific Beta(1,3)-Glucans by Human Gut Bacteroides. Comstock LE, ed. *MBio*. 2020;11(2). doi:10.1128/mBio.00095-20

267. Rogowski A, Briggs JA, Mortimer JC, et al. Glycan complexity dictates microbial resource allocation in the large intestine. *Nat Commun*. 2015;6(1):7481. doi:10.1038/ncomms8481

268. Porter NT, Hryckowian AJ, Merrill BD, et al. Phase-variable capsular polysaccharides and lipoproteins modify bacteriophage susceptibility in Bacteroides thetaiotaomicron. *Nat Microbiol*. 2020;5(9):1170-1181. doi:10.1038/s41564-020-0746-5

269. Tuckman M, Osburne MS. In vivo inhibition of TonB-dependent processes by a TonB box consensus pentapeptide. *J Bacteriol*. Published online 1992. doi:10.1128/jb.174.1.320-323.1992

270. Martens EC, Lowe EC, Chiang H, et al. Recognition and Degradation of Plant Cell Wall Polysaccharides by Two Human Gut Symbionts. Eisen JA, ed. *PLoS Biol*.

2011;9(12):e1001221. doi:10.1371/journal.pbio.1001221

271.   El-Gebali S, Mistry J, Bateman A, et al. The Pfam protein families database in 2019. *Nucleic Acids Res*. 2019;47(D1):D427-D432. doi:10.1093/nar/gky995

272.   Friedrich V, Janesch B, Windwarder M, et al. Tannerella forsythia strains display different cell-surface nonulosonic acids: biosynthetic pathway characterization and first insight into biological implications. *Glycobiology*. Published online December 16, 2016. doi:10.1093/glycob/cww129

273.   Tomek MB, Maresch D, Windwarder M, et al. A General Protein O-Glycosylation Gene Cluster Encodes the Species-Specific Glycan of the Oral Pathogen Tannerella forsythia: O-Glycan Biosynthesis and Immunological Implications. *Front Microbiol*. 2018;9. doi:10.3389/fmicb.2018.02008

274.   Hanley SA, Aduse-Opoku J, Curtis MA. A 55-kilodalton immunodominant antigen of Porphyromonas gingivalis W50 has arisen via horizontal gene transfer. *Infect Immun*. Published online 1999.

275.   Glenwright AJ, Pothula KR, Bhamidimarri SP, et al. Structural basis for nutrient acquisition by dominant members of the human gut microbiota. *Nature*. 2017;541(7637):407-411. doi:10.1038/nature20828

276.   Mohammad MM, Howard KR, Movileanu L. Redesign of a Plugged β-Barrel Membrane Protein. *J Biol Chem*. 2011;286(10):8000-8013. doi:10.1074/jbc.M110.197723

277.   Locher KP, Rees B, Koebnik R, et al. Transmembrane Signaling across the Ligand-Gated FhuA Receptor. *Cell*. 1998;95(6):771-778. doi:10.1016/S0092-8674(00)81700-6

278.   Costliow ZA, Degnan PH. Thiamine Acquisition Strategies Impact Metabolism and Competition in the Gut Microbe Bacteroides thetaiotaomicron. Gilbert JA, ed. *mSystems*. 2017;2(5). doi:10.1128/mSystems.00116-17

279.   Bolam DN, van den Berg B. TonB-dependent transport by the gut microbiota: novel aspects of an old problem. *Curr Opin Struct Biol*. 2018;51:35-43. doi:10.1016/j.sbi.2018.03.001

280.   Moore GWK, Howell SEL, Brady M, Xu X, McNeil K. Anomalous collapses of Nares

Strait ice arches leads to enhanced export of Arctic sea ice. *Nat Commun*. 2021;12(1):1. doi:10.1038/s41467-020-20314-w

281. Garcia-Herrero A, Vogel HJ. Nuclear magnetic resonance solution structure of the periplasmic signalling domain of the TonB-dependent outer membrane transporter FecA from Escherichia coli. *Mol Microbiol*. 2005;58(5):1226-1237. doi:10.1111/j.1365-2958.2005.04889.x

282. Malki I, Simenel C, Wojtowicz H, et al. Interaction of a Partially Disordered Antisigma Factor with Its Partner, the Signaling Domain of the TonB-Dependent Transporter HasR. Cascales E, ed. *PLoS One*. 2014;9(4):e89502. doi:10.1371/journal.pone.0089502

283. Benjdia A, Martens EC, Gordon JI, Berteau O. Sulfatases and a Radical S-Adenosyl-l-methionine (AdoMet) Enzyme Are Key for Mucosal Foraging and Fitness of the Prominent Human Gut Symbiont, Bacteroides thetaiotaomicron. *J Biol Chem*. 2011;286(29):25973-25982. doi:10.1074/jbc.M111.228841

284. Briliūtė J, Urbanowicz PA, Luis AS, et al. Complex N-glycan breakdown by gut Bacteroides involves an extensive enzymatic apparatus encoded by multiple co-regulated genetic loci. *Nat Microbiol*. 2019;4(9):1571-1581. doi:10.1038/s41564-019-0466-x

285. Pudlo NA, Urs K, Kumar SS, German JB, Mills DA, Martens EC. Symbiotic Human Gut Bacteria with Variable Metabolic Priorities for Host Mucosal Glycans. Sperandio V, ed. *MBio*. 2015;6(6). doi:10.1128/mBio.01282-15

286. Rogers TE, Pudlo NA, Koropatkin NM, et al. Dynamic responses of B acteroides thetaiotaomicron during growth on glycan mixtures. *Mol Microbiol*. 2013;88(5):876-890. doi:10.1111/mmi.12228

287. Bolam DN, Koropatkin NM. Glycan recognition by the Bacteroidetes Sus-like systems. *Curr Opin Struct Biol*. 2012;22(5):563-569. doi:10.1016/j.sbi.2012.06.006

288. Cadieux N, Kadner RJ. Site-directed disulfide bonding reveals an interaction site between energy-coupling protein TonB and BtuB, the outer membrane cobalamin transporter. *Proc Natl Acad Sci*. 1999;96(19):10673-10678. doi:10.1073/pnas.96.19.10673

289. Coggshall KA, Cadieux N, Piedmont C, Kadner RJ, Cafiso DS. Transport-Defective Mutations Alter the Conformation of the Energy-Coupling Motif of an Outer Membrane Transporter. *Biochemistry*. 2001;40(46):13964-13971. doi:10.1021/bi015602p

290. Fujita M, Mori K, Hara H, Hishiyama S, Kamimura N, Masai E. A TonB-dependent receptor constitutes the outer membrane transport system for a lignin-derived aromatic compound. *Commun Biol*. 2019;2(1):432. doi:10.1038/s42003-019-0676-z

291. Ferreira D, Seca AML, C.G.A. D, Silva AMS. Targeting human pathogenic bacteria by siderophores: A proteomics review. *J Proteomics*. 2016;145:153-166. doi:10.1016/j.jprot.2016.04.006

292. Antunes LCS, Imperi F, Towner KJ, Visca P. Genome-assisted identification of putative iron-utilization genes in Acinetobacter baumannii and their distribution among a genotypically diverse collection of clinical isolates. *Res Microbiol*. 2011;162(3):279-284. doi:10.1016/j.resmic.2010.10.010

293. Nairn BL, Eliasson OS, Hyder DR, et al. Fluorescence High-Throughput Screening for Inhibitors of TonB Action. DiRita VJ, ed. *J Bacteriol*. 2017;199(10). doi:10.1128/JB.00889-16

294. Yep A, McQuade T, Kirchhoff P, Larsen M, Mobley HLT. Inhibitors of TonB Function Identified by a High-Throughput Screen for Inhibitors of Iron Acquisition in Uropathogenic Escherichia coli CFT073. Kolter R, ed. *MBio*. 2014;5(2). doi:10.1128/mBio.01089-13

295. Martinez J, Steenbergen S, Vimr E. Derived structure of the putative sialic acid transporter from Escherichia coli predicts a novel sugar permease domain. *J Bacteriol*. 1995;177(20):6005-6010. doi:10.1128/jb.177.20.6005-6010.1995

296. Shakhnovich EA, King SJ, Weiser JN. Neuraminidase Expressed by Streptococcus pneumoniae Desialylates the Lipopolysaccharide of Neisseria meningitidis and Haemophilus influenzae : a Paradigm for Interbacterial Competition among Pathogens of the Human Respiratory Tract. *Infect Immun*. 2002;70(12):7161-7164. doi:10.1128/IAI.70.12.7161-7164.2002

297. Post DMB, Mungur R, Gibson BW, Munson RS. Identification of a novel sialic acid

transporter in Haemophilus ducreyi. *Infect Immun*. 2005;73(10):6727-6735. doi:10.1128/IAI.73.10.6727-6735.2005

298. Almagro-Moreno S, Boyd EF. Insights into the evolution of sialic acid catabolism among bacteria. *BMC Evol Biol*. 2009;9:118. doi:10.1186/1471-2148-9-118

299. Fletcher CM, Coyne MJ, Comstock LE. Theoretical and Experimental Characterization of the Scope of Protein O-Glycosylation in Bacteroides fragilis. *J Biol Chem*. 2011;286(5):3219-3226. doi:10.1074/jbc.M110.194506

300. Chen T, Abbey K, Deng W -j., Cheng M -c. The bioinformatics resource for oral pathogens. *Nucleic Acids Res*. 2005;33(Web Server):W734-W740. doi:10.1093/nar/gki361

301. Friedrich V, Pabinger S, Chen T, Messner P, Dewhirst FE, Schäffer C. Draft Genome Sequence of Tannerella forsythia Type Strain ATCC 43037. *Genome Announc*. 2015;3(3). doi:10.1128/genomeA.00660-15

302. Finn RD, Bateman A, Clements J, et al. Pfam: the protein families database. *Nucleic Acids Res*. 2014;42(Database issue):D222-30. doi:10.1093/nar/gkt1223

303. National Center for Biotechnology Information (NCBI)[Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information. Published 1988. https://www.ncbi.nlm.nih.gov/

304. N. Terrapon N, Lombard V, Drula É, Lapébie P, Al-Masaudi S, Gilbert HJ HB. PULDB: the expanded database of Polysaccharide Utilization Loci; Nucleic Acids Research.

305. Magrane M, UniProt Consortium. UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford)*. 2011;2011:bar009. doi:10.1093/database/bar009

306. Jones P, Binns D, Chang H-Y, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 2014;30(9):1236-1240. doi:10.1093/bioinformatics/btu031

307. Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25(17):3389-3402. doi:10.1093/nar/25.17.3389

308. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 1994;22(22):4673-4680. doi:10.1093/nar/22.22.4673

309. Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics*. 2007;23(1):127-128. doi:10.1093/bioinformatics/btl529

310. Alikhan N-F, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics*. 2011;12(1):402. doi:10.1186/1471-2164-12-402

311. Hughes C V., Malki G, Loo CY, Tanner ACR, Ganeshkumar N. Cloning and expression of α- <scp>d</scp> -glucosidase and N-acetyl-β-glucosaminidase from the periodontal pathogen, *Tannerella forsythensis* (*Bacteroides forsythus*). *Oral Microbiol Immunol*. 2003;18(5):309-312. doi:10.1034/j.1399-302X.2003.00091.x

312. Jung Y-J, Choi Y-J, An S-J, Lee H-R, Jun H-K, Choi B-K. Tannerella forsythia GroEL induces inflammatory bone resorption and synergizes with interleukin-17. *Mol Oral Microbiol*. 2017;32(4):301-313. doi:10.1111/omi.12172

313. Lee J-Y, Jung Y-J, Jun H-K, Choi B-K. Pathogenic potential of Tannerella forsythia enolase. *Mol Oral Microbiol*. 2016;31(2):189-203. doi:10.1111/omi.12115

314. Settem RP, Honma K, Shankar M, et al. Tannerella forsythia -produced methylglyoxal causes accumulation of advanced glycation endproducts to trigger cytokine secretion in human monocytes. *Mol Oral Microbiol*. 2018;33(4):292-299. doi:10.1111/omi.12224

315. Yoo JY, Kim HC, Zhu W, et al. Identification of Tannerella forsythia antigens specifically expressed in patients with periodontal disease. *FEMS Microbiol Lett*. 2007;275(2):344-352. doi:10.1111/j.1574-6968.2007.00906.x

316. Sekot G, Posch G, Jin Oh Y, et al. Analysis of the cell surface layer ultrastructure of the oral pathogen Tannerella forsythia. *Arch Microbiol*. Published online 2012. doi:10.1007/s00203-012-0792-3

317. Megson Z, Koerdt A, Schuster H, et al. Characterization of an α- <scp>l</scp> -

fucosidase from the periodontal pathogen *Tannerella forsythia*. *Virulence*.
2015;6(3):282-292. doi:10.1080/21505594.2015.1010982

318. Ksiazek M, Mizgalska D, Eick S, Thøgersen IB, Enghild JJ, Potempa J. KLIKK
proteases of Tannerella forsythia: putative virulence factors with a unique domain
structure. *Front Microbiol*. 2015;6. doi:10.3389/fmicb.2015.00312

319. Sharma A. Virulence mechanisms of Tannerella forsythia. *Periodontol 2000*.
2010;54(1):106-116. doi:10.1111/j.1600-0757.2009.00332.x

320. Severi E, Randle G, Kivlin P, et al. Sialic acid transport in Haemophilus influenzae is
essential for lipopolysaccharide sialylation and serum resistance and is dependent on a
novel tripartite ATP-independent periplasmic transporter. *Mol Microbiol*.
2005;58(4):1173-1185. doi:10.1111/j.1365-2958.2005.04901.x

321. Severi E, Hosie AHF, Hawkhead JA, Thomas GH. Characterization of a novel sialic
acid transporter of the sodium solute symporter (SSS) family and in vivo comparison
with known bacterial sialic acid transporters. *FEMS Microbiol Lett*. 2010;304(1):47-
54. doi:10.1111/j.1574-6968.2009.01881.x

322. Brigham CJ, Malamy MH. Characterization of the RokA and HexA broad-substrate-
specificity hexokinases from Bacteroides fragilis and their role in hexose and N-
acetylglucosamine utilization. *J Bacteriol*. 2005;187(3):890-901.
doi:10.1128/JB.187.3.890-901.2005

323. Potempa J, Madej M, Scott DA. The RagA and RagB proteins of Porphyromonas
gingivalis. *Mol Oral Microbiol*. 2021;36(4):225-232. doi:10.1111/omi.12345

324. Roy S, Phansopa C, Stafford P, et al. Beta-hexosaminidase activity of the oral
pathogen Tannerella forsythia influences biofilm formation on glycoprotein substrates.
FEMS Immunol Med Microbiol. Published online 2012. doi:10.1111/j.1574-
695X.2012.00933.x

325. Severi E, Müller A, Potts JR, et al. Sialic acid mutarotation is catalyzed by the
Escherichia coli beta-propeller protein YjhT. *J Biol Chem*. 2008;283(8):4841-4849.
doi:10.1074/jbc.M707822200

326. Berkhout MD, Plugge CM, Belzer C. How microbial glycosyl hydrolase activity in the
gut mucosa initiates microbial cross-feeding. *Glycobiology*. Published online October

18, 2021. doi:10.1093/glycob/cwab105

327. Pacheco AR, Curtis MM, Ritchie JM, et al. Fucose sensing regulates bacterial intestinal colonization. *Nature*. 2012;492(7427):113-117. doi:10.1038/nature11623

328. Ravcheev DA, Thiele I. Comparative Genomic Analysis of the Human Gut Microbiome Reveals a Broad Distribution of Metabolic Pathways for the Degradation of Host-Synthetized Mucin Glycans and Utilization of Mucin-Derived Monosaccharides. *Front Genet*. 2017;8:111. doi:10.3389/fgene.2017.00111

329. Zwickl NF, Stralis-Pavese N, Schäffer C, Dohm JC, Himmelbauer H. Comparative genome characterization of the periodontal pathogen Tannerella forsythia. *BMC Genomics*. 2020;21(1):150. doi:10.1186/s12864-020-6535-y

330. Rosa LT, Bianconi ME, Thomas GH, Kelly DJ. Tripartite ATP-Independent Periplasmic (TRAP) Transporters and Tripartite Tricarboxylate Transporters (TTT): From Uptake to Pathogenicity. *Front Cell Infect Microbiol*. 2018;8:33. doi:10.3389/fcimb.2018.00033

331. Desai MS, Seekatz AM, Koropatkin NM, et al. A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. *Cell*. 2016;167(5):1339-1353.e21. doi:10.1016/j.cell.2016.10.043

332. Brigham C, Caughlan R, Gallegos R, Dallas MB, Godoy VG, Malamy MH. Sialic acid (N-acetyl neuraminic acid) utilization by Bacteroides fragilis requires a novel N-acetyl mannosamine epimerase. *J Bacteriol*. 2009;191(11):3629-3638. doi:10.1128/JB.00811-08

333. Kuwahara T, Yamashita A, Hirakawa H, et al. Genomic analysis of Bacteroides fragilis reveals extensive DNA inversions regulating cell surface adaptation. *Proc Natl Acad Sci U S A*. 2004;101(41):14919-14924. doi:10.1073/pnas.0404172101

334. Settem RP, El-Hassan AT, Honma K, Stafford GP, Sharma A. Fusobacterium nucleatum and Tannerella forsythia induce synergistic alveolar bone loss in a mouse periodontitis model. *Infect Immun*. 2012;80(7):2436-2443. doi:10.1128/IAI.06276-11

335. Ng KM, Ferreyra JA, Higginbottom SK, et al. Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature*. 2013;502(7469):96-99. doi:10.1038/nature12503

336. Dalia AB, Standish AJ, Weiser JN. Three surface exoglycosidases from Streptococcus pneumoniae, NanA, BgaA, and StrH, promote resistance to opsonophagocytic killing by human neutrophils. *Infect Immun*. 2010;78(5):2108-2116. doi:10.1128/IAI.01125-09

337. Mahowald MA, Rey FE, Seedorf H, et al. Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla. *Proc Natl Acad Sci U S A*. 2009;106(14):5859-5864. doi:10.1073/pnas.0901529106

338. Rey FE, Gonzalez MD, Cheng J, Wu M, Ahern PP, Gordon JI. Metabolic niche of a prominent sulfate-reducing human gut bacterium. *Proc Natl Acad Sci U S A*. 2013;110(33):13582-13587. doi:10.1073/pnas.1312524110

339. Praharaj AB, Dehury B, Mahapatra N, Kar SK, Behera SK. Molecular dynamics insights into the structure, function, and substrate binding mechanism of mucin desulfating sulfatase of gut microbe Bacteroides fragilis. *J Cell Biochem*. 2018;119(4):3618-3631. doi:10.1002/jcb.26569

340. Hoskins LC, Agustines M, McKee WB, Boulding ET, Kriaris M, Niedermeyer G. Mucin degradation in human colon ecosystems. Isolation and properties of fecal strains that degrade ABH blood group antigens and oligosaccharides from mucin glycoproteins. *J Clin Invest*. 1985;75(3):944-953. doi:10.1172/JCI111795

341. Masahiro Sato, Kanta Kajikawa, Tomoya Kumon, Daisuke Watanabe, Ryuichi Takase WH. Mutually Beneficial Symbiosis Between Human and Gut-Dominant Bacteroides Species Through Bacterial Assimilation of Host Mucosubstances. BioRxiv. doi:https://doi.org/10.1101/2020.08.21.262261

342. Kilcoyne M, Gerlach JQ, Gough R, et al. Construction of a natural mucin microarray and interrogation for biologically relevant glyco-epitopes. *Anal Chem*. 2012;84(7):3330-3338. doi:10.1021/ac203404n

343. Vimr ER. Unified theory of bacterial sialometabolism: how and why bacteria metabolize host sialic acids. *ISRN Microbiol*. 2013;2013:816713. doi:10.1155/2013/816713

344. Lawrence JG, Roth JR. Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics*. 1996;143(4):1843-1860. doi:10.1093/genetics/143.4.1843

345. Woese C. The universal ancestor. *Proc Natl Acad Sci U S A*. 1998;95(12):6854-6859. doi:10.1073/pnas.95.12.6854

346. Omelchenko M V, Makarova KS, Wolf YI, Rogozin IB, Koonin E V. Evolution of mosaic operons by horizontal gene transfer and gene displacement in situ. *Genome Biol*. 2003;4(9):R55. doi:10.1186/gb-2003-4-9-r55

347. Itoh T, Takemoto K, Mori H, Gojobori T. Evolutionary instability of operon structures disclosed by sequence comparisons of complete microbial genomes. *Mol Biol Evol*. 1999;16(3):332-346. doi:10.1093/oxfordjournals.molbev.a026114

348. Dandekar T, Snel B, Huynen M, Bork P. Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem Sci*. 1998;23(9):324-328. doi:10.1016/s0968-0004(98)01274-2

349. Lipničanová S, Chmelová D, Ondrejovič M, Frecer V, Miertuš S. Diversity of sialidases found in the human body - A review. *Int J Biol Macromol*. 2020;148:857-868. doi:10.1016/j.ijbiomac.2020.01.123

350. Tailford LE, Owen CD, Walshaw J, et al. Discovery of intramolecular trans-sialidases in human gut microbiota suggests novel mechanisms of mucosal adaptation. *Nat Commun*. 2015;6:7624. doi:10.1038/ncomms8624

351. Juge N, Tailford L, Owen CD. Sialidases from gut bacteria: a mini-review. *Biochem Soc Trans*. 2016;44(1):166-175. doi:10.1042/BST20150226

352. Kim S-Y, Volsky DJ. PAGE: parametric analysis of gene set enrichment. *BMC Bioinformatics*. 2005;6:144. doi:10.1186/1471-2105-6-144

353. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9(1):559. doi:10.1186/1471-2105-9-559

354. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(D1):D607-D613. doi:10.1093/nar/gky1131

355. Lagedroste M, Reiners J, Smits SHJ, Schmitt L. Impact of the nisin modification machinery on the transport kinetics of NisT. *Sci Rep*. 2020;10(1):12295. doi:10.1038/s41598-020-69225-2

356. Kumar P, Satyanarayana T. Microbial glucoamylases: characteristics and applications. *Crit Rev Biotechnol*. 2009;29(3):225-255. doi:10.1080/07388550903136076

357. Rigal A, Bouveret E, Lloubes R, Lazdunski C, Benedetti H. The TolB protein interacts with the porins of Escherichia coli. *J Bacteriol*. 1997;179(23):7274-7279. doi:10.1128/jb.179.23.7274-7279.1997

358. Hagan CL, Kahne D. The Reconstituted Escherichia coli Bam Complex Catalyzes Multiple Rounds of β-Barrel Assembly. *Biochemistry*. 2011;50(35):7444-7446. doi:10.1021/bi2010784

359. Sperandeo P, Lau FK, Carpentieri A, et al. Functional Analysis of the Protein Machinery Required for Transport of Lipopolysaccharide to the Outer Membrane of Escherichia coli. *J Bacteriol*. 2008;190(13):4460-4469. doi:10.1128/JB.00270-08

360. Thong S, Ercan B, Torta F, et al. Defining key roles for auxiliary proteins in an ABC transporter that maintains bacterial outer membrane lipid asymmetry. *Elife*. 2016;5. doi:10.7554/eLife.19042

361. Evans TG. Considerations for the use of transcriptomics in identifying the 'genes that matter' for environmental adaptation. Podrabsky JE, Stillman JH, Tomanek L, eds. *J Exp Biol*. 2015;218(12):1925-1935. doi:10.1242/jeb.114306

362. Wu Z, Liu W, Jin X, et al. NormExpression: An R Package to Normalize Gene Expression Data Using Evaluated Methods. *Front Genet*. 2019;10. doi:10.3389/fgene.2019.00400

363. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*. 2010;11(10):R106. doi:10.1186/gb-2010-11-10-r106

364. Koch CM, Chiu SF, Akbarpour M, et al. A Beginner's Guide to Analysis of RNA Sequencing Data. *Am J Respir Cell Mol Biol*. 2018;59(2):145-157. doi:10.1165/rcmb.2017-0430TR

365. Eschen-Lippold L, Jiang X, Elmore JM, et al. Bacterial AvrRpt2-Like Cysteine Proteases Block Activation of the Arabidopsis Mitogen-Activated Protein Kinases, MPK4 and MPK11. *Plant Physiol*. 2016;171(3):2223-2238. doi:10.1104/pp.16.00336

366. Doyle MT, Bernstein HD. Bacterial outer membrane proteins assemble via asymmetric interactions with the BamA β-barrel. *Nat Commun*. 2019;10(1):3358.

doi:10.1038/s41467-019-11230-9

367. Kauko A, Illergård K, Elofsson A. Coils in the membrane core are conserved and functionally important. *J Mol Biol*. 2008;380(1):170-180. doi:10.1016/j.jmb.2008.04.052

368. Way M, Sanders M, Chafel M, Tu YH, Knight A, Matsudaira P. beta-Scruin, a homologue of the actin crosslinking protein scruin, is localized to the acrosomal vesicle of Limulus sperm. *J Cell Sci*. 1995;108(10):3155-3162. doi:10.1242/jcs.108.10.3155

369. Sanchez-Mejias A, Tay Y. Competing endogenous RNA networks: tying the essential knots for cancer biology and therapeutics. *J Hematol Oncol*. 2015;8(1):30. doi:10.1186/s13045-015-0129-1

370. Sztukowska MN, Dutton LC, Delaney C, et al. Community Development between Porphyromonas gingivalis and Candida albicans Mediated by InlJ and Als3. Biswas I, ed. *MBio*. 2018;9(2). doi:10.1128/mBio.00202-18

371. Lee SA, Kim T-W, Sang M-K, Song J, Kwon S-W, Weon H-Y. Arachidicoccus soli sp. nov., a bacterium isolated from soil. *Int J Syst Evol Microbiol*. 2021;71(1). doi:10.1099/ijsem.0.004566

372. Chen M, Li N, Zhang X-F, et al. Sphingobacterium faecale sp. nov., a 1-aminocyclopropane-1-carboxylate deaminase producing bacterium isolated from camel faeces. *Int J Syst Evol Microbiol*. 2022;72(1). doi:10.1099/ijsem.0.005215

Chapter VIII

Appendices

Appendix VIII: 8.1.

License Number	4698290015625
License date	Oct 29, 2019
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Journal of Clinical Periodontology
Licensed Content Title	Microbial complexes in subgingival plaque
Licensed Content Author	R. L. Kent, C. Smith, M.A. Cugini, et al
Licensed Content Date	Dec 21, 2005
Licensed Content Volume	25
Licensed Content Issue	2
Licensed Content Pages	11
Type of use	Dissertation/Thesis
Requestor type	University/Academic
Format	Print and electronic

Appendix VIII: 8.2.

Copyright Clearance Center

RightsLink®

Home | Account Info | Help | LIVE CHAT

Publisher: American Chemical Society

Copyright © 1969, American Chemical Society

Logged in as:
Ahmed Almuntashiri
Account #:
3001421294

LOGOUT

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

If credit is given to another source for the material you requested, permission must be obtained from that source.

BACK | CLOSE WINDOW

Appendix VIII: 8.3

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS

Sep 10, 2019

This Agreement between Ahmed Almuntashiri ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	4665490093158
License date	Sep 10, 2019
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Molecular Oral Microbiology
Licensed Content Title	Sialic acid, periodontal pathogens and Tannerella forsythia: stick around and enjoy the feast!
Licensed Content Author	G. Stafford, S. Roy, K. Honma, et al
Licensed Content Date	Nov 12, 2011
Licensed Content Volume	27
Licensed Content Issue	1
Licensed Content Pages	12
Type of use	Dissertation/Thesis
Requestor type	University/Academic
Format	Print and electronic
Portion	Figure/table
Number of figures/tables	1
Original Wiley figure/table number(s)	Figure 1
Will you be translating?	No
Title of your thesis / dissertation	The tale of tonB in TF
Expected completion date	Dec 2023
Expected size (number of pages)	200
Requestor Location	Ahmed Almuntashiri Communal Areas Flat 2, 278 Glossop Road Sheffield, other United Kingdom Attn:
Publisher Tax ID	EU826007151
Total	0.00 GBP
Terms and Conditions	

TERMS AND CONDITIONS

Appendix VIII: 8.4.

Copyright Clearance Center

RightsLink®

Home | Account Info | Help | LIVE CHAT

AMERICAN SOCIETY FOR MICROBIOLOGY

Title: A Novel Sialic Acid Utilization and Uptake System in the Periodontal Pathogen *Tannerella forsythia*

Author: Sumita Roy, C. W. Ian Douglas, Graham P. Stafford

Publication: Journal of Bacteriology

Publisher: American Society for Microbiology

Date: Apr 14, 2010

Copyright © 2010, American Society for Microbiology

Permissions Request

ASM authorizes an advanced degree candidate to republish the requested material in his/her doctoral thesis or dissertation. If your thesis, or dissertation, is to be published commercially, then you must reapply for permission.

BACK | CLOSE WINDOW

Appendix VIII: 8.5.

Annual Reviews, Inc. - License Terms and Conditions

Order Date		06-Nov-2019	
Order license ID		1002450-1	
ISSN		1545-3251	
Type of Use		Republish in a thesis/dissertation	
Publisher		ANNUAL REVIEWS	
Portion		Image/photo/illustration	

LICENSED CONTENT

Publication Title	Annual review of microbiology	Rightsholder	Annual Reviews, Inc.
Date	01/01/1947	Publication Type	e-Journal
Language	English	URL	http://arjournals.annualreviews.org/loi/micro
Country	United States of America		

REQUEST DETAILS

Portion Type	Image/photo/illustration	Distribution	U.K. and Commonwealth (excluding Canada)
Number of images / photos / illustrations	1	Translation	Original language of publication
Format (select all that apply)	Print, Electronic	Copies for the disabled?	No
Who will republish the content?	Not-for-profit entity	Minor editing privileges?	No
Duration of Use	Life of current edition	Incidental promotional use?	No
Lifetime Unit Quantity	Up to 499	Currency	GBP
Rights Requested	Main product		

NEW WORK DETAILS

Title	Tale tonB in Tannerella	Institution name	University of Sheffield
Instructor name	Ahmed Almuntashiri	Expected presentation date	2023-11-01

ADDITIONAL DETAILS

Order reference number	N/A	The requesting person / organization to appear on the license	Ahmed Almuntashiri

REUSE CONTENT DETAILS

Title, description or numeric reference of the portion(s)	Transport and regulation of siderophores	Title of the article/chapter the portion is from	N/A

Appendix VIII: 8.6.

Microbiology Society - License Terms and Conditions

Order Date   26-Nov-2019
Order license ID   1005843-1
ISSN   1350-0872
Type of Use   Republish in a thesis/dissertation
Publisher   Microbiology Society
Portion   Chart/graph/table/figure

LICENSED CONTENT

Publication Title	Microbiology	Country	United Kingdom of Great Britain and Northern Ireland
Author/Editor	SOCIETY FOR GENERAL MICROBIOLOGY.		
Date	01/01/1994	Rightsholder	Microbiology Society
Language	English	Publication Type	Journal

REQUEST DETAILS

Portion Type	Chart/graph/table/figure	Distribution	Worldwide
Number of charts / graphs / tables / figures requested	1	Translation	Original language of publication
Format (select all that apply)	Electronic	Copies for the disabled?	No
		Minor editing privileges?	No
Who will republish the content?	Not-for-profit entity	Incidental promotional use?	No
Duration of Use	Current edition and up to 10 years	Currency	GBP
Lifetime Unit Quantity	Up to 499		
Rights Requested	Main product		

NEW WORK DETAILS

Title	A tole of TonB	Institution name	University of Sheffield
Instructor name	Ahmed	Expected presentation date	2023-11-01

ADDITIONAL DETAILS

Order reference number	N/A	The requesting person / organization to appear on the license	NA

REUSE CONTENT DETAILS

Appendix VIII: 8.7.

MDPI Open Access Information and Policy

All articles published by MDPI are made immediately available worldwide under an open access license. This means:

- everyone has free and unlimited access to the full-text of *all* articles published in MDPI journals;

- everyone is free to re-use the published material if proper accreditation/citation of the original publication is given;

- open access publication is supported by the authors' institutes or research funding agencies by payment of a comparatively low **Article Processing Charge (APC) (/about/apc)** for accepted articles.

Permissions

No special permission is required to reuse all or part of article published by MDPI, including figures and tables. For articles published under an open access Creative Common CC BY license, any part of the article may be reused without permission provided that the original article is clearly cited. Reuse of an article does not imply endorsement by the authors or MDPI.

External Open Access Resources

Appendix VIII: 8.8.

Open Access Articles

Articles that are "open access" in PMC are made available under a Creative Commons or similar license that allows more liberal redistribution and reuse than a traditional copyrighted work. Many articles in PMC are protected by U.S. and/or foreign copyright laws, even though PMC provides free access to it. Some journals use the label "open access" for an article that is available free at time of publication, but is still subject to traditional copyright restrictions.

Open access articles are included in the PMC Open Access Subset and can be found in PMC by searching open access[filter]. This subset also includes a number of U.S. government publications, e.g.,

- Addiction Science & Clinical Practice (vol. 1 through vol. 6)
- Alcohol Research: Current Reviews
- Emerging Infectious Diseases
- Environmental Health Perspectives
- Journal of Research of NIST
- Morbidity and Mortality Weekly Report
- Preventing Chronic Disease

Articles published in these journals are in the public domain and may be used and reproduced without special permission. However, anyone using the material is requested to properly cite and acknowledge the source.

Please note open access and public domain articles may still contain photographs or illustrations copyrighted by other commercial organizations or individuals that may not be used without obtaining prior approval from the holder of the copyright.

Appendix VIII: 8.9.

Attribution 4.0
International (CC BY 4.0)

This is a human-readable summary of (and not a substitute for) the license.

You are free to:

Appendix VIII: 8.10.

Copyright Clearance Center — RightsLink®

	Home	Account Info	Help	LIVE CHAT

AMERICAN SOCIETY FOR MICROBIOLOGY

Title: Diversity of Microbial Sialic Acid Metabolism

Author: Eric R. Vimr,Kathryn A. Kalivoda,Eric L. Deszo,Susan M. Steenbergen

Publication: Microbiology and Molecular Biology Reviews

Publisher: American Society for Microbiology

Date: Mar 8, 2004

Copyright © 2004, American Society for Microbiology

Logged in as:
Ahmed Almuntashiri

Account #:
3001421294

LOGOUT

Permissions Request

BACK   CLOSE WINDOW

Appendix VIII: 8.10.

TonB sequences

TonB BFO_0333

METW**TYLLKVNAAIAVFYMVYRW**CYRNDTFFTLR**RYLLQGILFLSVVYPFTDFSKWFV**RS

HTLTEAAFTYTQYMPELTVTPSVEAPATYSLGDYALWLYLAVTGFLLLRLMVRLTQLVWL

RWRSPQVEMEGARVSRLQRPATPFSFFNWIFIHPEMHRKEELKEILAHEQIHVRQHHSFD

ILLAELSCAVCWFNPMAWMLKKEIHSNLEFLVDHRVVKDGADARSYQYHLLRMAHQPVQA

ALANQFNTSPLKKRIRMLNAKQSPKVKLVAYTLVLPLALLLLVANNAGAMTERMSGLLDA

SLPMLQGHTENLSDTKKEDTLVVSGTVVDGQGALQGVSILISGTYSGTISDENGRFQIRM

QAGDTLNFSYIGMAPVHLVPKASGDVGKIEMKRKKENLDEIVVVGYTSSEKTTRNDRKPS

TGSEGEPIFTVVEEMPQFPGGQAALMQYIARNIKYPVIAQENGIQGRVVCQFIIDERGNT

ADVQVVQGVDPLLDKEAVRLIEEMPRWKPGKQRGKPVSVEYTVPVNFRLTKGAVCSIQKD

TTRQTQSRQAPIRYYLNGKEIFSERMKSIPPAEIENIHVLNTENRQMEHKDKTASKIIII

TTLDATADERARNEEIKKKFN

- **Underline text represents transmembrane helices. Red font colour represents strand motifs. Text highlight pink colour presents predictable periplasmic domain. Text highlight yellow colour presents C-terminal domain**

TonB BFO_0233

MGAF**FSYSLQSALCLAVFYLFYTVVL**SRETFHRLNRCM**LLFIMALSWIIPLFAGALSF**RTGAEVVP**EMLVQ**

PLATLTA**GGAAASLYLSV**WQSVFV**PVFTLYLVGIAFCVLYQALSYLR**LWRLLSKGKHAALDNDIRLILHTD

TSMAPFSWMKYIVLSEDDYTEDGAAIIAHETAHIRLHHS**WDLMAAHVFVILQWFNPFAWL**LYRDLRNTHEY

EADKSVLDEGFEAQHYQLLLIKKAVGTRLYSMANGYNHSNNLKKRITMMLQK**KSNAWARLKYALVLPLAAT**

TVVAFARPEISRPFDEISSVKVSYFAPTAQPVEAKSVEKVVVPPRQEVAVPVETVAETPPAPRETKAPAAK

PKPAPTTPEDTVVFMVVEQMPEFPGGQGELMKYIAQNIKYPVKAIEKKMEGRVVVSFVVSCDGSLRDFQVV

RSVDPDLDAEAIRVLKLMPNWKPGKQRGKNVSVKYTLPINFSLSAKKDSLQTK

- **Underline text represents transmembrane helices. Red font colour represents strand motifs. Text highlight pink colour presents predictable periplasmic domain. Text highlight yellow colour presents C-terminal domain.**

TonB BFO_0953

MEIKKSRKADLEGGKGLSILMGIVVGLAVLFVGFEWGTQEKTIQKDEGIADIIAEEEIDI
TRQEETPPPPVEQVAEVLNVVEDDVEVENTDLLSSEDNQAEAQTQTYVPPVVKVEEEEES
SQQIFMVVEEMPEFPGGQAALMSFIAKSIKYPVVAQENGIQGRVTCSFVVNKDGSIVDAEVIRGIDPSLDK
EALRVINTMPKWKPGKQRGKPVRVKFTVPINFRLNQ

- **Underline text represents transmembrane helices. Red font colour represents strand motifs. Text highlight pink colour presents predictable periplasmic domain. Text highlight yellow colour presents C-terminal domain**

TonB BFO_3116

MKFEKNDIIALIGSIIFHLSILLVLYFTVLKTIVPEEDSGIMVNFGNVNLAAGQFEPRGE
PASSDMTAAPPPQPAHPTPREEMITQDLEESVSLTNKKNEEERKKKEKKEQEEARRKETE
QRKVEAERKKQTDEQRRKEQAIRDRVAGAFGSTGGSENNNQGDAASGSGNQGSPFGNSDS
GANEGVGGIGSFNLNGRSIGAGGLPKPAYTIQEEGRIVINITVDPRGNVIMAEIGKGTNI
DHATMRKSALEAARKAKFNSIQGTNNQSGTITYRYSLR

- **Underline text represents transmembrane helices. Red font colour represents strand motifs. Text highlight pink colour presents predictable periplasmic domain. Text highlight yellow colour presents C-terminal domain**

Appendix VIII: 8.11.

NanO-Mutagenesis

tgatgggcttctgtacgactttcgaacaggtctattggttgcgtgcattgatgggagtcagcgaagcattatatatccctgcgggactgtcgct
gatcaccgactatcatcgggagaaaacacgttcgctggctgtcggcatccacatgacgggattgtatgtcggacaggccttaggcggctttggg
gcaacggtttccagtattttcagttggcatgttactttccagagctttggctggataggtattatatacagtctcatcctgatccttttcctga
aagataaaaagagtgagacaccggtaagtactccacagcccacttcgttgggcgaaacggtaaaatcggcattgaaaggtatcggcatgttggc
cggtaacatttcattctggatcatcctgctctattttgcgactccgagtcttccgggatgggcgacgaagaactggttaccgacgctgttctcc
gaaaacctgtcgatcgatatgacgaaagccggaccgttatcgacgattgtcattgccgtgtcatcattcatcggagtgattgccggtggcatcc
tttccgaccggtggatacaacgtaaccttcgaggacgtatttataccggagcgatcgggttgtcactgatgatcccttccttgctgctgctcgg
cttcggacatagtctgacgatgattatcggcggaggactgtgcttcggtatcggttacggcatgttcgatgcgaataacatgccgatcctgtgt
cagtttatttcgtcgagataccgtgctacagcctacggactcatgaacatgacgggggtatttgccggcgccatgattacgaatttcctcggac
ggtcgaccgatgccggcaaccttgggaaagacttcgcgatgctggcagtcgtagtcgctggtgctctcgtcgtacaattactctttctgcgtcc
ccaaaccgcagacaagaaagagaaataaaggtcttcattaatcaataactatttaaatttatgacaaaaaagaaattgcccgttcgtttacgg
gtcagcactttactattgataaagtgctaataaaagatgcaataagacaagcaaatataagtaatcaggatacggttttagatattggggcagg
caaggggtttcttactgttcatttattaaaaatcgccaacaatgttgttgctattgaaaacgacacagctttggttgaacatttacgaaaatta
ttttctgatgcccgaaatgttcaagttgtcggttgtgattttaggaattttgcagttccgaaatttcctttcaaagtggtgtcaaatattcctt
atggcattacttccgatattttcaaaatcctgatgtttgagagtcttggaaattttctgggaggttccattgtccttcaattagaacctacaca
aaagttattttcgaggaagctttacaatccatataccgtttctatcatacttttttttgatttgaaacttgtctatgaggtaggtcctgaaagt
ttcttgccaccgccaactgtcaaatcagccctgttaaacattaaaagaaaacacttatttttttgattttaagtttaaagccaaatacttagcat
ttatttcctatctgttagagaaacctgatttatctgtaaaaacagctttaaagtcgattttcaggaaaagtcaggtcaggtcaatttcggaaaa
attcggtttaaaccttaatgctcaaattgtttgtttgtctccaagtcaatggttaaactgtttttttggaaatgctggaagttgtccctgaaaaa
tttcatccttcgtagttgattaacgaatataaaaatacacagtatatgaaaaagatagcaaatcatattgtcctatggatcgctttcatggttg
ttctcatgtcgtgcgactcattgaatctggctcccgaagattatttcggaagcggaaacttctggaaaaatgaagcgcaggtgaacggtgcact
catcggattgcatgccgacttaagaagaagttatttcatgttctatcatttgggagaaatacggggaggcacccagcgtgtaggctcctcgtcg
ttgaacaccagtcttcactatgccaacgagcgttccaacctgtttgataaggatcgcacaggcatccgaaactggtatgggctttacgacaatc
tctttcaggtaaaccattttatcgatcaggtagaaaacaaatgtacgtttttgagcgacgagagccgtaaaaatctgctggcacaggcttacgg
gttcagagcgttgtactattttatgctctataagacgtacggcggcgtgccgatcgtgacggaaaccgaaatcctcaacgggaaggcaactgcc
gagaaattttatgtcgaacgggctacaccgcaggctacgatggaattcatcaagaaagacatcttgaaatcggaagagtattatgccggtaaga
cagtaagtaacaactatccgaagatgatgtggtcgaaggcggctaccttgatgctcaaggccgaaatctacttgtgggcggccaaagtatccat
ccagggatatacggcgaccggtacggacgatctgaaagtggcgaaaacagccctgaatgaagtagtcggtaagttcgaattgctcaaagacttc
ggaagcatttttcagcacaagtaatcgcaataacaaggagcttgtttttacactgcattttgccgatcgcgaagccacgaataatgcttcaagtt
tcctttaccaacaagcgctgtttgtgggtcaggtatatggccgcaacgggaaaaagatcgagaacgacacgctga

BFO_0333 Mutagenesis

aacgctcgaaagacttttttggagacactcaccggccaatgcgtcggaggcttccgcatgccccgcatggccaacgtctcggacgctgcgcttg
tcgaggccggatacacctacaactcttccctcaaccccaccttcctgcccggacgatacaatcattggcacgaaccgcgcacttggttccggca
gaacggactgcttcaactgcccgcctcaacaactccgaacatacgtttcccgctcttctggctgtcttttcatcacctgcccggagccctctac
cgacgactggcactccgaacatacaagcacgacggctacctgctcacctactttcatccgtgggagttcatgcccatacagcaaatcaaaggac
taccatatacggtcacgcacaatgccggaaataatgccgaaatacgtttggataagttgatcggtttcttccgaaaacgaaacgtaccattcgg
aacattccgcgaatttattgaaacattacccccccacgtcataacttcttcacggcgatacgaaaaaaatatcggccaataacatcttttaaca
tcactctcgcatcaacattacgaaacattcgtacattcgcgttcgattagtaacataaaacgattgtaatcatggaaaaattaacacaacaaga
agaagaagccatgttgattgtatggcaacaggggggagggcacgattcgtagctatcacgaacaaagcggaaagccgctcatgccttacacgacc
tttgcgtccatcattaaaaatctggatcgcaagaaatatttgaagatgcggcgtgtagggaacgacgaacttgtgtaaaccgctgatttcggaga
aagaatacaaacgtaaattcatgtcgggattttattcgcgattattttcagaactcgtacaaagaaatggtctcgttcttcgcccgcgaagagaa
gatttcggccgatgatttgaaagagattatccgtattatcgaaaacaaaaatagaaaacgatgacaaaaaagaaattgcccgttcgtttacg
ggtcagcactttactattgataaagtgctaataaaagatgcaataagacaagcaaatataagtaatcaggatacggttttagatattggggcag
gcaaggggtttcttactgttcatttattaaaaatcgccaacaatgttgttgctattgaaaacgacacagctttggttgaacatttacgaaaatt
attttctgatgcccgaaatgttcaagttgtcggttgtgattttaggaattttgcagttccgaaatttcctttcaaagtggtgtcaaatattcct
tatggcattacttccgatattttcaaaatcctgatgtttgagagtcttggaaattttctgggaggttccattgtccttcaattagaacctacac
aaaagttattttcgaggaagctttacaatccatataccgtttctatcatacttttttttgatttgaaacttgtctatgaggtaggtcctgaaag
ttcttgccaccgccaactgtcaaatcagccctgttaaacattaaaagaaaacacttatttttttgattttaagtttaaagccaaatacttagca
tttatttcctatctgttagagaaacctgatttatctgtaaaaacagctttaaagtcgattttcaggaaaagtcaggtcaggtcaatttcggaaa
aattcggtttaaaccttaatgctcaaattgtttgtttgtctccaagtcaatggttaaactgtttttttggaaatgctggaagttgtccctgaaaaa
atttcatccttcgtagtctcagattactaattactaagcgatcgatttcaaaagggtgtgcccgaaacgtttttttcggcacaccctttcgatat
gtagaagaagtatagaaaaacaacgcgctaaagataagagcaagagcccctaccgacaaggagagcgagtatcataggcaagcgacggaata
atatccccatagcgatgcacgaaacatttttgatacgtatctttgacgcgccgaaagtaagcggtacagggcacaacccctacaacaaaacata
tacgcgtaaataaaacgaaatgaaacagttctgtttaatccttgccggcgtgttgatgatactcgcatgttcgcacgaaaagaaagtgaccgag
aaaaaaaaaaagcccggaaaccatcatcggaatggaaaaagatacattgtctatcgacctcgacgatacgtcattgccccactatcatacccta
gcggcattcagcttgaatacgaacgattatttaatcggttataacgatcaaacgcatgctctagacgtttttgattttacaaacggggttgtcc
gcgatataaagctctaccgcgaaggaagacaggcggttgccggcgaaataaacggattatacgccacatcgccggattcgatctggattttggc
ggacaaaatcgtttacttgctcgacacgaccggcatcatacacgaaaaacacgaattaacggtcgacgaaggagaatacatcatcaatacatcg
aactattcgaacgcttcgatccaattatatgtcaacacacacagccaatcaatgtattacgtcacccttaaaatgggagatgaacctatatttt
acattaacgaggccgatttacggcaggagaaacacaaaaaaaacagtccttccgtataccggcgccgagcaaggagtcggacatctatacggatg
gatgcaacatcctaacgtcacctaccattttccgctggtgatctacaattttccttttaactcgaatgtgtacacc

BFO_0233 Mutagenesis

tgataccggctgtcagttccatatatttatgctcgatcagaccgcccggacccaagatgaatccgtacaggcgggaatctttattccccgattt
cagtccgtcaaacagtccggcaatcacgttgtgaccgcccggcgcctgtcctcccgacaatatcacgccgacgttgatcggcggcatcgctttc
tttcgttgcttttctcaaaagtcacaaccggcatgccgtacgtattcgggaacagtttcttgatctcttcctgattggctaccgattcggtgt
actcgccgtctacggcctttacggctcctcgcagggcggtgggcatcttgggaacatagtgtgcccgctcaatttgtaatgcgctctttgtcat
gatgtattgttttttttataaatcttttttgtctcttgttttttgcgtccgcaaagataatattttttccgcaaacctcataccgcaacgcccggcct
ttttcgcaccgtttatttgtgtatgtcctcattttttaatggggttaaaactgctttctttagttatatttaactataaaaattagttttgaaac
taaattatctcgttatttgtagtcgtaatttgaacaggaggatagacgtatgcataagttaacagcaaaagaagaagaattaatgctctttttt
tgggaaagaggagcgttgtttgtcagggagatactcaattttttatccggatccgaagccccattttaataccctttcgaccattgtccgggggt
tggaagaaaaagggtttctggcgcataagacgttcgggaatacctatcagtactacgccgtcatcagtaaggaggagcaccggcgcggtacgtt
acggggtgtcatccgtaaatatttcgataattcgtatttgcacgccgtttcgtcgcttgtgcagagcgaaaatatctctgtcgaagaattgaag
gctttgatccgcgaagtggaaaaaaactcaccgcccgcaggcgataaatcttgatgtgtcatgacaaaaagaaattgcccgttcgttttacgg
gtcagcactttactattgataaagtgctaataaaagatgcaataagacaagcaaatataagtaatcaggatacggtttttagatattggggcagg
caaggggtttcttactgttcatttattaaaaatcgccaacaatgttgttgctattgaaaacgacacagctttggttgaacatttacgaaaatta
ttttctgatgcccgaaatgttcaagttgtcggttgtgattttaggaattttgcagttccgaaatttcctttcaaagtggtgtcaaatattcctt
atggcattacttccgatattttcaaaatcctgatgtttgagagtcttggaaattttctgggaggttccattgtccttcaattagaacctacaca
aaagttattttcgaggaagctttacaatccatataccgtttttctatcatactttttttgatttgaaacttgtctatgaggtaggtcctgaaagt
ttcttgccaccgccaactgtcaaatcagccctgttaaacattaaaagaaaacacttatttttttgattttaagtttaaagccaaatacttagcat
ttatttcctatctgttagagaaacctgatttatctgtaaaaacagctttaaagtcgattttcaggaaaagtcaggtcaggtcaatttcggaaaa
attcggtttaaaccttaatgctcaaattgtttgtttgtctccaagtcaatggttaaactgtttttttggaaatgctggaagttgtccctgaaaaa
tttcatccttcgtagtaaggtttcgtttgtttcaagaagtgataggagccgcctcgcaacgttgggattgcggggcggcttttttctttgcggat
actgccaatatcattcgtttggttgatgatcacctggatcggagtgaactacctgcctgcggcgcaaggaagcattcacgtgtattaaataaga
tagattatttttttaattttcaaaaataagtaataaatttgccccggataataaataaaataaaacgaaagaacgtgtgagagtttcgatgaaa
tcaagcatatcggtcgtgcaatgctacatgcttgtacgccgtttcggtgtgcggataggtgaaagtaatcttaccgtaacggaactcaattata
aagataatacgcgtacggaaaaacagcaagccggatattccgtatatcttgtgaataatactattaccgatagaaatgaactataaaataaagt
caaacaattaaattttttagcttatgaaaagaaaagtaacaagtagtgtatcaaaacttatgctcgcagcttttctccttgctgcaagtagcgga
gggatagcgtggggcaggcgtatattagtggccctgctggagtgcaaaatactttcggagcaggcgagagtgccaggctaaataacggtgcag
tcatttttcccggaggaggaacgtggactggaagcgaaatagagatcgctaccggtagtggtaatattgtcataaattcatataacgcaggagg
acgttatacggattatacgaataagtggactggaaactggagcggtagtggaaatggccgatgtaattctccttggcctgattggacaggagt
gttaactggaacgggccgacagctcagataactcctgattatacggacccgtgtaattcaagtcatacacattctcatccgagtgatggaacga
gtgatccatatacttccaatcagagggagcaacagatgatggcgttgctttggtacagacaaacggatcacaga

BFO_3116 Mutagenesis

atggggggttctcctcatcctttctttagtcgctatttacttttttttattcaacggattttggtgattcgtcaggccgccaagagcgacgaaacgt
ttatgaaccgaatccgggattatgtgtatgaagggaaagtcgattctgctcttaatctctgtcgtacgacgaaaacaccggcagcccgcatgat
tgaaaagggaattatccgtttgggacgtcccacgaacgatgtgttggcggcaatcgagaatgcgggcaatattgaaattgcaaaactggagaaa
ggctttccggtacttgccactacagcagccggcgctccgatgatcggctttctgggcacggtgaccggaatggtacgtgcttttttacgatatgg
ctaatgccggaacgaatgtggatgtatccttactgtcgggaggtatctacgaagctttggtgacaaccgtcggcggtctggtagtaggcatcat
tgctcttttcgcgtacaattacttggtcgcgcgtgtcgacaatgttgtcaatcgattagaggccaatacgatggaatttatggacatactgaac
gaaccggctaactgaacatcatggcgttgaaacgaagaaccaaaatcaatgcagcattcagcatggcttcgatgacagacgtcatctttctctt
gctgatctttttttatggtcacgtcgaccgtggtcgtgccgaatgcgatcaaactgaccttgcctcaagctcagaagcaaacggcagcgaaaccg
cttacacgagttacgatcgatgcggccggaaattattatgttgcggccgggaagcaacgcgaacggatggtgacattcgacgagattactcctt
tccttcaggagcagtacgttcaggagccggagatgttcgttgcattatatgccgatgaaaccgtgccatataaagaaatcgtgaaagtattgaa
tatcgcaaacgaaaataaattcaaggtggtattggctactcgtccaccgtctcacaagccatgacaaaaagaaattgcccgttcgttttacgg
gtcagcactttactattgataaagtgctaataaaagatgcaataagacaagcaaatataagtaatcaggatacggtttttagatattggggcagg
caaggggtttcttactgttcatttattaaaaatcgccaacaatgttgttgctattgaaaacgacacagctttggttgaacatttacgaaaatta
ttttctgatgcccgaaatgttcaagttgtcggttgtgattttaggaattttgcagttccgaaatttcctttcaaagtggtgtcaaatattcctt
atggcattacttccgatattttcaaaatcctgatgtttgagagtcttggaaattttctgggaggttccattgtccttcaattagaacctacaca
aaagttattttcgaggaagctttacaatccatataccgtttttctatcatactttttttgatttgaaacttgtctatgaggtaggtcctgaaagt
ttcttgccaccgccaactgtcaaatcagccctgttaaacattaaaagaaaacacttatttttttgattttaagtttaaagccaaatacttagcat
ttatttcctatctgttagagaaacctgatttatctgtaaaaacagctttaaagtcgattttcaggaaaagtcaggtcaggtcaatttcggaaaa
attcggtttaaaccttaatgctcaaattgtttgtttgtctccaagtcaatggttaaactgtttttttggaaatgctggaagttgtccctgaaaaa
tttcatccttcgtagaacgcatgtaataaccattcaaaaaacttataaacgatgaaaaaggcttatgtattttttagcaaccggctttgaagaag
tcgaagcgatcggtacaatcgatgtactacgccgtggaggaatgcacgttgtaacggtttcaatgaccggtgccgcaaccgttaagggagggca
tgatatctctgttgttgccgatgcgctatttgaagacctcagctatgccgatgccgatgtgctcgtactgcccggaggaggtccaatgcttaac
gattacgaggcattgaagaaggaaatcgtgcgccatcacgatgccggcaagccgctggcagctatttgcgctgcgccgctcgtttttcggcggat
tgggactgcttgaaggcaagcgcgcgacatgttatccggggatttgaagagtacctcaaaggtgcaactgttgttaatgtaccgacggttaccga
tggttctatcattaccggtcggggaccgggactggttttcgatttcgggttggctatactggagcatctcgaagggctcgagtcggcggaacaa
gtcgccaaagatttattattgaaataaaaaagccgctctacgcggctttttttatgtatgacaggtgaaagtactattctttcacttcccgatc
tctccgataaacagaatgacgccggtgctctttttcgcgtatggcaaaaacaaacggcttgtttatgtggaagttgataggagtggtagaggggg
atgacgtttttttaccatttctacggctgttacggcagcagcttccgtaccttcttcatcgacctgcacaaacgtcttatggatcacacgtgaaat
aaatatcgcagcgtcggtgatgccgggaaaatcagccgtctccgtaaacggcacgttcattccatctccggcagtatctttttctcgagcccg
tatttacactctgtcttgaagcgcggcatgcgcagcgatacttgtgttgggcgaatgcctttttataatcatgctc

BFO_0953 Mutagenesis

gaacaaaactatcctcgagattcttttccgccagcgtctcgtctacatatttgaacttccattgcttcagatgcatggccgttacggatacacg
gatcggagcatgcgccattcttttggaaattcccatctgaatgtcccaaggcaactgcctgcgttccgaatcgtatgctttcagttgtgccccg
atattttttcaaggccaccccgaaagagagaccctttttcttcgtcgtaatagctcagcccggcatcaaccccgatcccgaacgacgaatattcag
ccagatgagagtataacatcttcatcgagagcccccccctccatttatccgaaagatcgtacgaataaaaagcctgtacactcatgtctttagc

cgagaattctcctttgatcatccgtgacgcatcgacttctttaaagttgccatagctgatgaaggaccctccgattccccaagcacctcgctcg
cggtgggctttggtatagatcgcactgcctacatgaatatcgccaatataactcatgtaattcaggttcaccatgccgtccatctctccaccta
acagggccggactgtgaaaagcaagcgacgggtcagcttcgatcagtgatacggtatgccctcccaaagcattggctcgtgccgaggtcggaaa
acgtaaaaacgtaaacacttcgccaccttcctgcgccgtcagaaaccagggtaagaggcacagtagccctccgattaatcgtatatttctttt
ctcattctttcaaaagcgtgtgcaaaagtacacaattttaccactatggtaatcttttcataccacgataatcataacggataaaaacaaaaaa
aggtatttgaaggctaacttttttttccgggaatcaaaaaaaaagcaaaaaagcaagtttttttttccgctttcacatgcaatctgtaatttttt
gtctatctttgcctccacgagaatcaaattaagagtaatcattacttaaagaagatagctatgacaaaaaagaaattgcccgttcgtttacgg
gtcagcactttactattgataaagtgctaataaaagatgcaataagacaagcaaatataagtaatcaggatacggttttagatattggggcagg
caaggggtttcttactgttcatttattaaaaatcgccaacaatgttgttgctattgaaaacgacacagctttggttgaacatttacgaaaatta
ttttctgatgcccgaaatgttcaagttgtcggttgtgattttaggaattttgcagttccgaaatttcctttcaaagtggtgtcaaatattcctt
atggcattacttccgatattttcaaaatcctgatgtttgagagtcttggaaattttctgggaggttccattgtccttcaattagaacctacaca
aaagttattttcgaggaagctttacaatccatataccgttttctatcatacttttttttgatttgaaacttgtctatgaggtaggtcctgaaagt
ttcttgccaccgccaactgtcaaatcagccctgttaaacattaaaagaaaacacttattttttgattttaagtttaaagccaaatacttagcat
ttatttcctatctgttagagaaacctgatttatctgtaaaaacagctttaaagtcgattttcaggaaaagtcaggtcaggtcaatttcggaaaa
attcggtttaaaccttaatgctcaaattgtttgtttgtctccaagtcaatggttaaactgttttttggaaatgctggaagttgtccctgaaaaa
tttcatccttcgtaggaagcacatatacaagcagtaaaattctttttttcatttcacaaaaaaggtcgccctctgaaaaagcggcctttttttttc
atgtctaaaatatgctatcatttgaagatacgctgccgatcattgaacatgccgttcgagagctgcggtttgatgacccaccacagagcctttt
tgatccgattgtttatacgctttcgttgggaggtaagcgcattcgtcctgctttcaccttgatggcgtgtaatctttacagctctgaagtagag
caggcaatcaaacccgctctcggaatagaagtgtttcataatttacgttgctgcacgacgatctgatggatgaagctcgaaagcgtcgggggc
aaccgacggtgcatgtcaagtggaaccggaatacggcggtgctctcgggagatgccatgctcatcacggcgtatcgtcttatcggggagacatc
atccgaactgctaaaaaaaatactggatttgtttacccggacagccttggagatatgtgaagggcaacagtatgatatggagtttgaatcgagg
ccggacgtgacagaagcggagtatcttgaaatgattcgtttaaaaacggctgtactgctggcttgcagcctgaaaacagggcgcttatcggcg
gcgcttccgagcatgaagcggatattctctatcgcttcggcatctatatcgggttggcctttcagttacaggatgatctgctcgatgtgtatgg
cagtccggaaactttcggaaagaatatcggcggagatattttgtgcaacaagaagacatttctattgattcatactttggcaagcgcctcgccc
gaacaaaaagaacagattgtctattatcagcatccggataaagcggcatgctatacggccgacgaaaaaattaaggcgattacttctatttatg
atcaattaaacataagggagctgacacaagagcagatgaagcattattataatcttgcgatggatgaattagcgc

Appendix VIII: 8.12.

Appendix VIII: Figure 8.13.

List	ATCC43037	92A.2	UB4	UB20	UB22	k16		3313	9610 WW10960	WW11663	TF1426
groEL	812	BFO_1437	SCQ20282.1	SCQ20998.1	SCQ21617.1	BAR51102.1	BAR48841.1	OLQ20225.1	PDP71731.1	PDP44099.1	RRD77829.1
enolase	1369	BFO_2068	SCQ21702.1	SCQ22643.1	SCQ23820.1	BAR52073.1	BAR49380.1	OLQ21459.1	PDP70557.1	PDP44032.1	RRD75651.1
methylglyoxal	1073	BFO_1726	SCQ20885.1	SCQ22751.1	SCQ22615.1	BAR53179.1	BAR49091.1	OLQ20716.1	PDP70264.1	PDP45217.1	RRD77289.1
TF2926	1241	BFO_1924	SCQ21316.1	SCQ22257.1	SCQ23302.1	BAR51947.1	BAR49263.1	OLQ21599.1	PDP71666.1	PDP43856.1	RRD70118.1
TF1717	198	BFO_0705	SCQ18824.1	SCQ19259.1	SCQ19341.1	BAR50948.1	BAR48250.1	OLQ20398.1	PDP69853.1	PDP42802.1	RRD73294.1
TF1012	2333	BFO_3356	SCQ23988.1	SCQ24893.1	SCQ25461.1	BAR53185.1	BAR47703.1	OLQ21024.1	PDP69978.1	PDP43318.1	RRD77775.1
TF0015	1472	BFO_2183	SCQ21919.1	SCQ22909.1	SCQ24094.1	BAR52172.1	BAR49468.1	OLQ19791.1	PDP70937.1	PDP42631.1	RRD79561.1
TF2930	1246	BFO_1930	SCQ21326.1	SCQ22267.1	SCQ23313.1	BAR51952.1	BAR49268.1	OLQ21603.1	PDP71669.1	PDP43859.1	RRD72771.1
TF0751	2140	BFO_3080	SCQ23527.1	SCQ24544.1	SCQ25268.1	BAR52964.1	BAR50132.1	OLQ19952.1	PDP70944.1	PDP44773.1	RRD71009.1
TF1711	192	BFO_0699	SCQ23527.1	SCQ19236.1	SCQ19321.1	BAR50942.1	BAR48244.1	OLQ21515.1	PDP70315.1	PDP42072.1	RRD79650.1
tfsA	1004	BFO_1651	SCQ20752.1	SCQ21558.1	SCQ22347.1	BAR51709.1	BAR49026.1	OLQ21954.1	PDP71942.1	PDP44596.1	RRD77757.1
tfsB	1005	BFO_1652	SCQ20755.1	SCQ21562.1	SCQ22351.1	BAR51710.1	BAR49027.1	none	none	none	RRD77802.1
fucO	1838	BFO_2737	SCQ22791.1	SCQ23999.1	SCQ24954.1	BAR52652.1	BAR49830.1	OLQ21900.1	PDP71996.1	PDP42886.1	RRD79220.1
karilysin	1793	BFO_2683	SCQ22676.1	SCQ23876.1	SCQ24909.1	*none*	BAR48584.1	OLQ20954.1	*none*	PDP43308.1	*none*
DPP IV_A	438	BFO_1007	SCQ19505.1	SCQ20211.1	SCQ20350.1	BAR51420.1	BAR48496.1	OLQ19518.1	PDP69945.1	PDP44409.1	RRD77362.1
DPP IV_B	1008	BFO_1659	SCQ20763.1	SCQ21581.1	SCQ22368.1	BAR51714.1	BAR49030.1	OLQ20581.1	PDP71079.1	PDP44506.1	RRD75440.1
mirolase	1785	BFO_2665	SCQ19818.1	SCQ23869.1	SCQ24873.1	BAR51284.1	BAR49778.1	OLQ19794.1	PDP69033.1	PDP41746.1	RRD76625.1
miropsin-1	579	BFO_1179	SCQ19822.1	SCQ19451.1	SCQ20845.1	BAR50819.1	*none*	OLQ20753.1	PDP72086.1	PDP44898.1	RRD79377.1
mirolysin	1783	BFO_2661	SCQ22625.1	SCQ23757.1	SCQ24836.1	BAR52584.1	*none*	OLQ19796.1	*none*	PDP41744.1	*none*
miropsin-2	1791	BFO_2679	SCQ22673.1	SCQ23873.1	SCQ24907.1	BAR51279.1	BAR48585.1	OLQ20320.1	*none*	PDP45134.1	*none*
forsilysin	575	BFO_1168	SCQ19814.1	SCQ19442.1	SCQ20791.1	BAR51288.1	BAR48635.1	OLQ21518.1	PDP72093.1	PDP41275.1	*none*
TF1320	2625	BFO_0300	SCQ18572.1	SCQ18188.1	SCQ18191.1	BAR50620.1	BAR47851.1	OLQ19664.1	PDP71395.1	PDP43405.1	RRD73476.1
abfA-TF0780	2166	BFO_3112	SCQ23593.1	SCQ24593.1	SCQ25299.1	BAR52991.1	*none*	OLQ19978.1	PDP70967.1	PDP44749.1	RRD72658.1
susB	1542	BFO_2272	SCQ22082.1	SCQ23106.1	SCQ24311.1	BAR52277.1	BAR49539.1	OLQ19859.1	PDP71184.1	PDP44305.1	RRD75044.1
bglX_β-glucosida	1471	BFO_2182	SCQ21917.1	SCQ22907.1	SCQ24092.1	BAR52171.1	BAR49467.1	OLQ19790.1	PDP70936.1	PDP42632.1	RRD79650.1
hexA	1240	BFO_1923	SCQ21313.1	SCQ22254.1	SCQ23300.1	BAR51946.1	BAR49262.1	OLQ21598.1	PDP71665.1	PDP43855.1	RRD70119.1
nanH	1491	BFO_2207	SCQ21971.1	SCQ22945.1	SCQ24155.1	BAR52188.1	BAR49484.1	OLQ19823.1	PDP70034.1	PDP43204.1	RRD74006.1
siaHI	606	BFO_1210	SCQ19914.1	SCQ19539.1	SCQ20957.1	BAR51252.1	BAR48673.1	OLQ20726.1	PDP72064.1	PDP42607.1	RRD76208.1
bspA-TF1589	93	BFO_0575	SCQ18633.1	SCQ18925.1	SCQ19038.1	BAR50687.1	BAR48111.1	OLQ21538.1	PDP71315.1	PDP45013.1	RRD69398.1
bspA_2	307	BFO_2000	SCQ19179.1	SCQ19835.1	SCQ19833.1	BAR52016.1	BAR48361.1	OLQ20465.1	PDP69940.1	PDP42880.1	RRD69779.1
bspA_4	1305	BFO_0833	SCQ21478.1	SCQ22455.1	SCQ23285.1	BAR51559.1	BAR48468.1	OLQ20898.1	PDP69579.1	PDP42880.1	RRD69518.1
prtH	796	BFO_1417	SCQ20228.1	SCQ20963.1	SCQ21571.1	BAR52595.1	BAR48825.1	OLQ20239.1	PDP71748.1	PDP44085.1	*none*
TF0371	1795	BFO_2686	SCQ22693.1	SCQ23879.1	SCQ24911.1	BAR52607.1	BAR49784.1	OLQ20957.1	PDP69971.1	PDP43311.1	RRD78374.1
TF1502	22	BFO_0490	SCQ18295.1	SCQ18748.1	SCQ18841.1	BAR50762.1	BAR48035.1	OLQ19569.1	PDP71254.1	PDP44955.1	RRD79434.1
TF1705	187	BFO_1055	SCQ19604.1	SCQ19223.1	SCQ20464.1	BAR51377.1	BAR48534.1	OLQ19538.1	PDP70313.1	PDP44182.1	RRD72046.1
hemagglutinin/T	204	BFO_0711	SCQ18840.1	SCQ19275.1	SCQ19362.1	BAR50954.1	*none*	OLQ20411.1	PDP69857.1	PDP42806.1	*none*
TF2392	769	BFO_1388	SCQ20170.1	SCQ20895.1	SCQ21474.1	BAR51143.1	BAR48799.1	OLQ20287.1	PDP71815.1	PDP44132.1	*none*
TF0636	2039	BFO_2968	SCQ23286.1	SCQ24326.1	SCQ25170.1	BAR52865.1	BAR50037.1	OLQ21295.1	PDP71826.1	PDP43252.1	RRD75627.1
TF0960	2282	BFO_3298	SCQ23840.1	SCQ24802.1	SCQ25411.1	BAR53131.1	BAR50306.1	OLQ21698.1	PDP70100.1	PDP42690.1	RRD73877.1
TF0229	1685	BFO_2553	SCQ22396.1	SCQ23506.1	SCQ24691.1	BAR52418.1	BAR49681.1	OLQ20917.1	PDP71525.1	PDP43649.1	RRD73877.1
TF1468	2756	BFO_0450	SCQ18123.1	SCQ18604.1	SCQ18690.1	BAR50799.1	BAR47976.1	OLQ19614.1	PDP71227.1	PDP44928.1	RRD73466.1
TF2386	764	BFO_1381	SCQ20158.1	SCQ20880.1	SCQ21453.1	BAR51149.1	BAR48793.1	OLQ20288.1	PDP71816.1	PDP44131.1	RRD72931.1
TF0617	2021	BFO_2947	SCQ23248.1	SCQ24297.1	SCQ25153.1	BAR50763.1	BAR50019.1	OLQ21279.1	PDP71841.1	PDP43501.1	RRD77009.1
TF1335	2638	BFO_0315	SCQ18599.1	SCQ18230.1	SCQ18231.1	BAR50635.1	BAR47866.1	OLQ19675.1	PDP71408.1	PDP43416.1	RRD75924.1
β-hexosaminida	1492	BFO_2208	SCQ21974.1	SCQ22947.1	SCQ24158.1	BAR52189.1	BAR49262.1	OLQ19824.1	PDP70026.1	PDP43205.1	RRD70119.1

Appendix VIII: 8.1.

Table 8.1: Literature reported bacteria with sialic acid inner membrane transporter

Species	Family	Locus	Presence of *nan* operon	Ref
Tannerella forsythia ATCC 43037	MFS	TF0032/ Tanf_13715	Yes	Stafford G, Roy S, Honma K, Sharma A
Salmonella enterica	SSS		No	Severi E, Hosie AHF, Hawkhead JA,
Clostridium botulinum C str. Eklund	SSS	CBC_A1223	No	Almagro-Moreno S, Boyd EF
Clostridium perfringens SM101	SSS	CPR_0176	No	Almagro-Moreno S, Boyd EF
Dorea formicigenerans ATCC 27755	SSS	DORFOR_00337	No	Almagro-Moreno S, Boyd EF
Dorea longicatena DSM 13814	SSS	DORLON_02532	No	Almagro-Moreno S, Boyd EF
Lactobacillus plantarum WCFS1	SSS	lp_3563	No	Almagro-Moreno S, Boyd EF
Lactobacillus sakei subsp. 23K	SSS	LSA1642	No	Almagro-Moreno S, Boyd EF
Mycoplasma capricolum ATCC 27343	SSS	MCAP_0416	No	Almagro-Moreno S, Boyd EF
Mycoplasma mycoides SC str. PG1	SSS	MSC_0558	No	Almagro-Moreno S, Boyd EF
Mycoplasma synoviae 53	SSS	MS_0191	No	Almagro-Moreno S, Boyd EF
Pseudoalteromonas haloplanktis TAC125	SSS	PSHAb0151	No	Almagro-Moreno S, Boyd EF
Pseudoalteromonas translucida	SSS	CAI89198.1	Yes	Almagro-Moreno S, Boyd EF
Shewanella pealeana ATCC 700345	SSS	Spea_1518	Yes	Almagro-Moreno S, Boyd EF
Vibrio fischeri ES114	SSS	VF0668	No	Almagro-Moreno S, Boyd EF
Streptococcus pneumoniae	SSS	SP_1328	No	Almagro-Moreno S, Boyd EF

Fusobacterium nucleatum ATCC 25586	TRAP	FN1473	No	Almagro-Moreno S, Boyd EF
Haemophilus influenzae Rd	TRAP	HI0147	No	Aeveri E, Randle G, Kivlin P, et al.
Haemophilus somnus 129PT	TRAP	HS_0702	No	Almagro-Moreno S, Boyd EF
Pasteurella multocida subsp.Multocida Pm70	TRAP	PM1708	No	Almagro-Moreno S, Boyd EF
Photobacterium profundum SS9	TRAP	PBPRA2279	No	Almagro-Moreno S, Boyd EF
Vibrio cholerae N16961	TRAP	VC1777	No	Almagro-Moreno S, Boyd EF
Vibrio vulnificus YJ016	TRAP	VVA1200	No	Almagro-Moreno S, Boyd EF
Haemophilus ducreyi	ABC	HD_1669	No	Post DMB, Mungur R, Gibson BW, Munson RS.
Arachidicoccus sp. 5GH13-10[371]	ABC	K9M52_08165	Yes	Lee SA, Kim T-W, Sang M-K, Song J, Kwon S-W, Weon H-Y.
Pedobacter sp. BS3 BS3-1	ABC	WP_149348386.1	Yes	Hankyong National University-NCBI
Sphingobacterium sp. C459-1T[372]	ABC	WP_202103915.1	Yes	Almagro-Moreno S, Boyd EF
Streptococcus gordonii str. Challis substr. CH1	ABC	SAG0035	No	Almagro-Moreno S, Boyd EF
Streptococcus pneumoniae TIGR4	ABC	SP_1688/SP_1682	No	Almagro-Moreno S, Boyd EF
Streptococcus pyogenes M1 GAS	ABC	Spy_254	No	Almagro-Moreno S, Boyd EF
Streptococcus sanguinis SK36	ABC	SSA_0076	No	Almagro-Moreno S, Boyd EF
Streptococcus agalactiae 2603V/R	ABC	SAG0035	No	Almagro-Moreno S, Boyd EF
Corynebacterium glutamicum	ABC	BAC00046.1	No	Almagro-Moreno S, Boyd EF
Bifidobacterium breve UCC2003	ABC	Bbr_0156	No	Almagro-Moreno S, Boyd EF

369
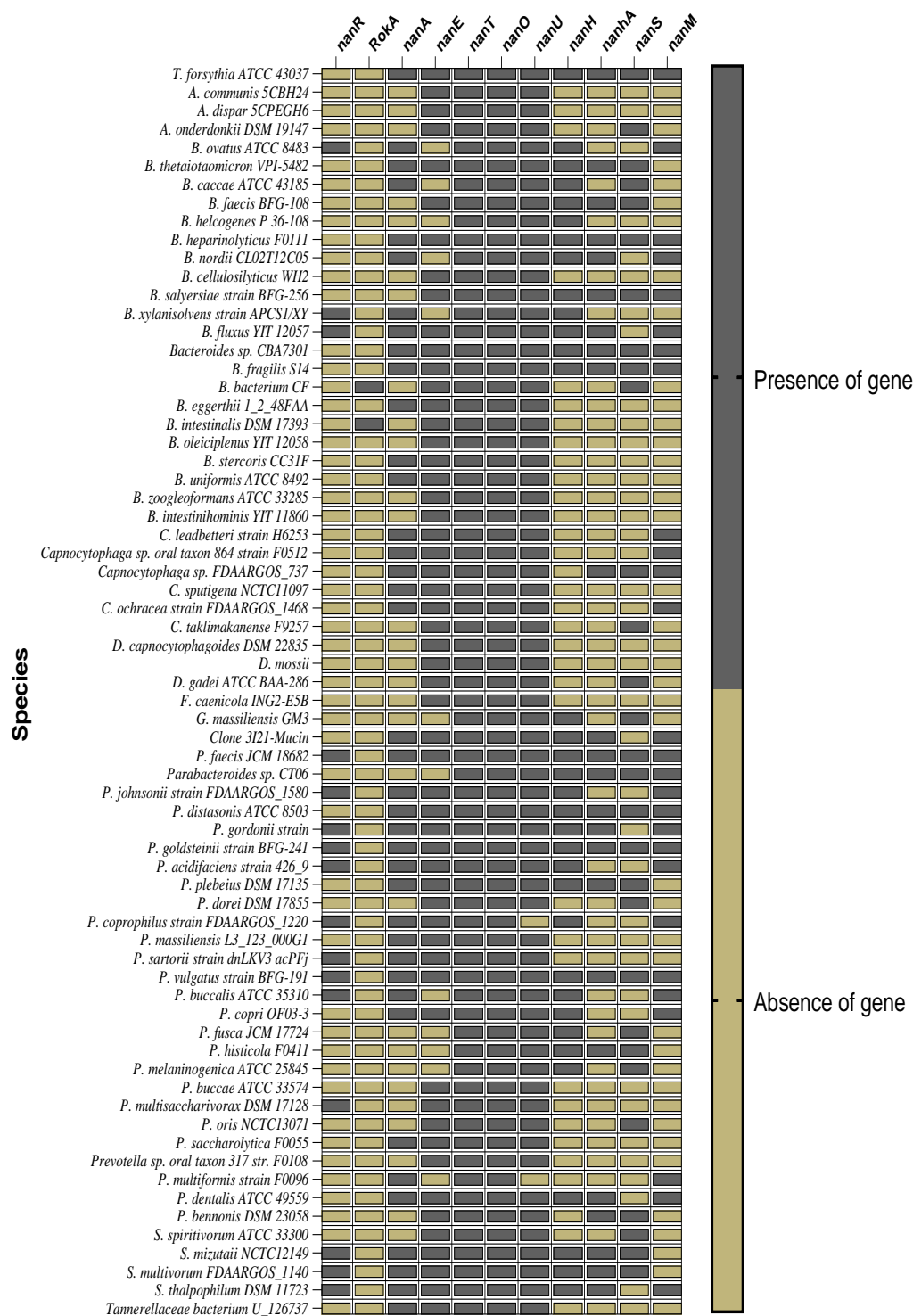
Appendix VIII: Figure 8.15 (A). Sialic acid utilization, catabolism, and transport clusters from non-human isolated microbes. First-half predicted sialic acid *nan* operon based on the inner membrane sialic acid (nanT). All predicted microbes were structured to the order of *nan* operon of *T. forsythia*, although many bacterial genomes showed reverse start and end of *nan* operons. Key: dark gray means presence of a gene within the *nan* operon structures.
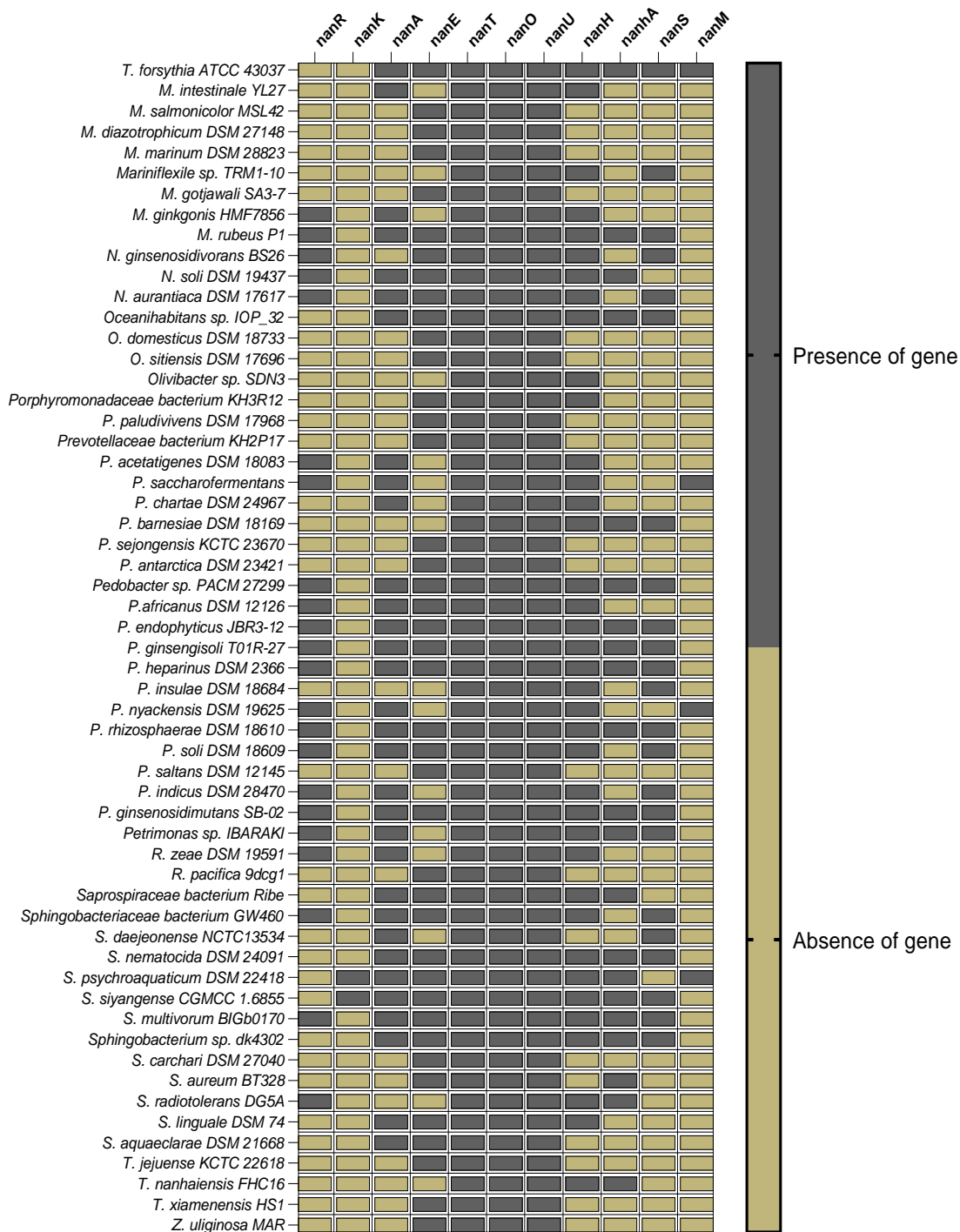
Appendix VIII: Figure 8.15. (B). Sialic acid utilization, catabolism, and transport clusters from non-human isolated microbes. Second-half predicted sialic acid *nan* operon based on the inner membrane (nanT). All predicted microbes were structured to the order of *nan* operon of *T. forsythia*, although many bacterial genomes showed reverse start and end of nan operons. Key: dark gray colour means presence of a gene within the *nan* operon structures.
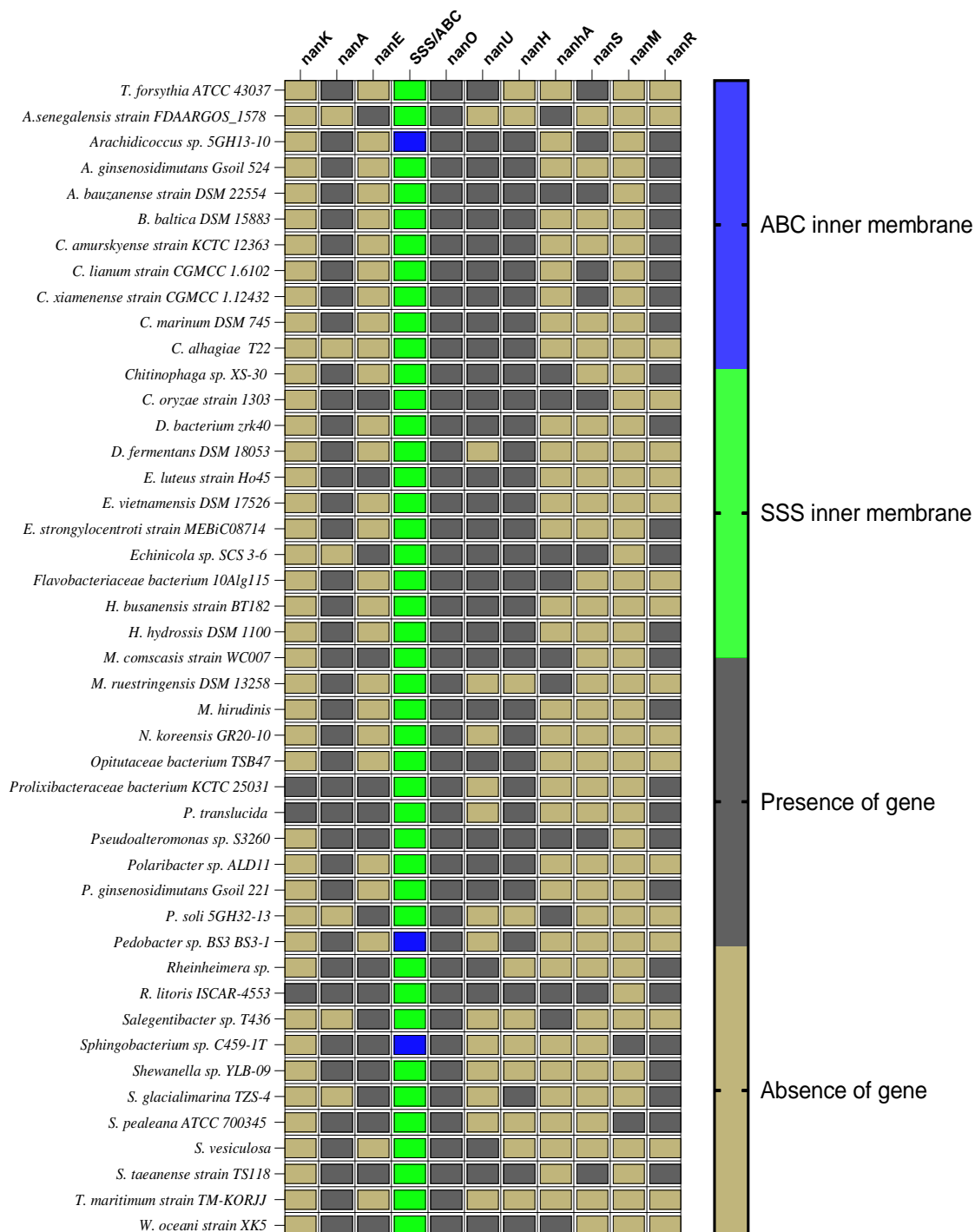
Appendix VIII: Figure 8.15 (C). Sialic acid utilization, catabolism, and transport clusters from non-human isolated microbes. Predicted sialic acid *nan* operon based on the inner membranes SSS (green colour) and ABC (blue colour). All predicted microbes were structured to the order of *nan* operon of *T. forsythia*, although many bacterial genomes showed reverse start and end of nan operons. Key: dark gray colour means presence of a gene within the *nan* operon structures.

Appendix VIII: Figure 8.16. Phylogenetic tree using nanT amino acid sequences of all spices with *nan* operon of genes. The tree is viewed and built by Interactive Tree of Life, version 6.5.2. Bootstrap analysis was performed and calculated based on alignment of the amino acid sequences by ClustalW.

373
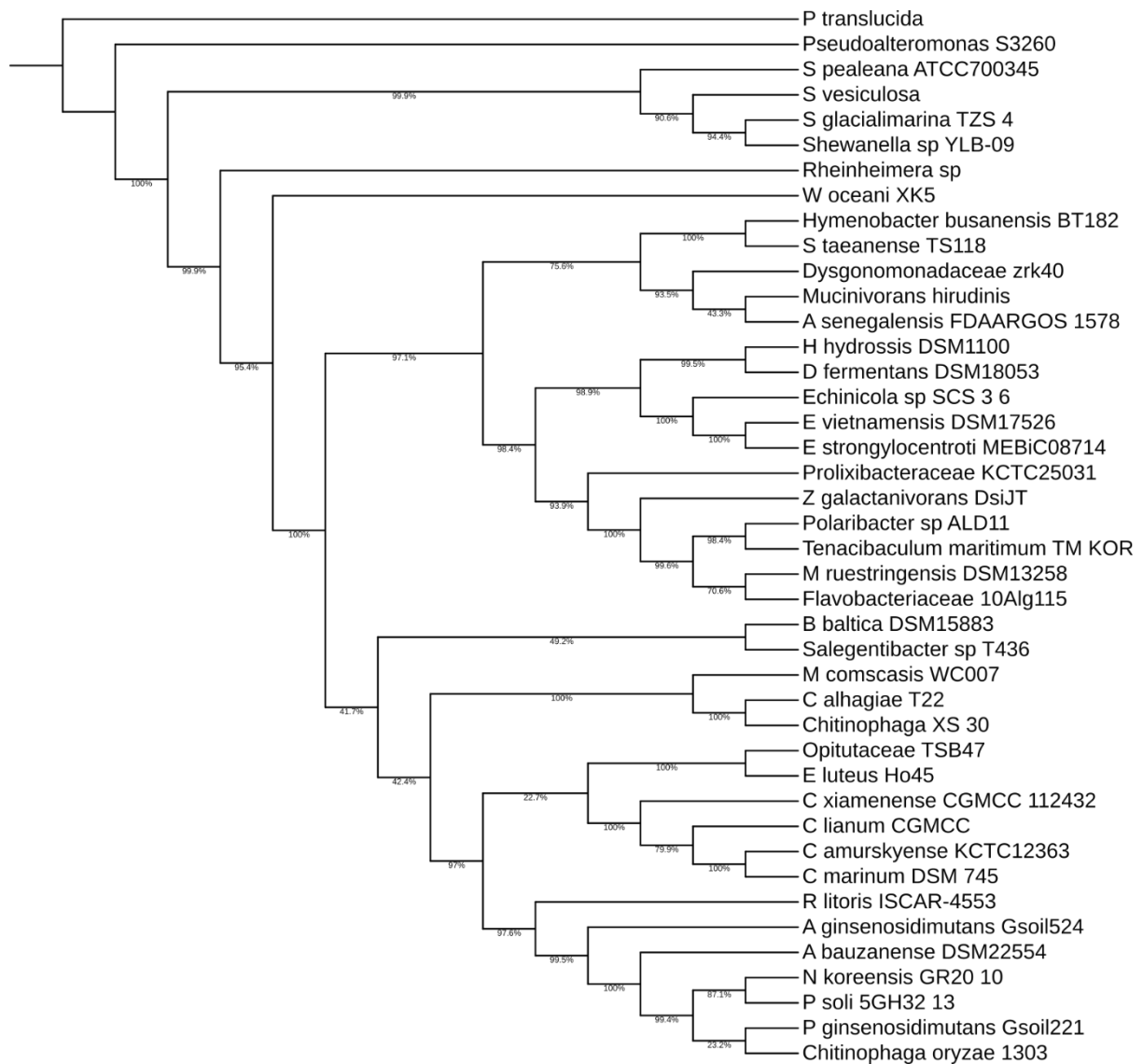
Appendix VIII: Figure 8.17. Phylogenetic tree using SSS amnio acid sequences of all bacterial spices with *nan* operons. The tree is viewed and built by Interactive Tree of Life, version 6.5.2. Bootstrap analysis was performed and calculated based on alignment of the amino acid sequences by ClustalW.
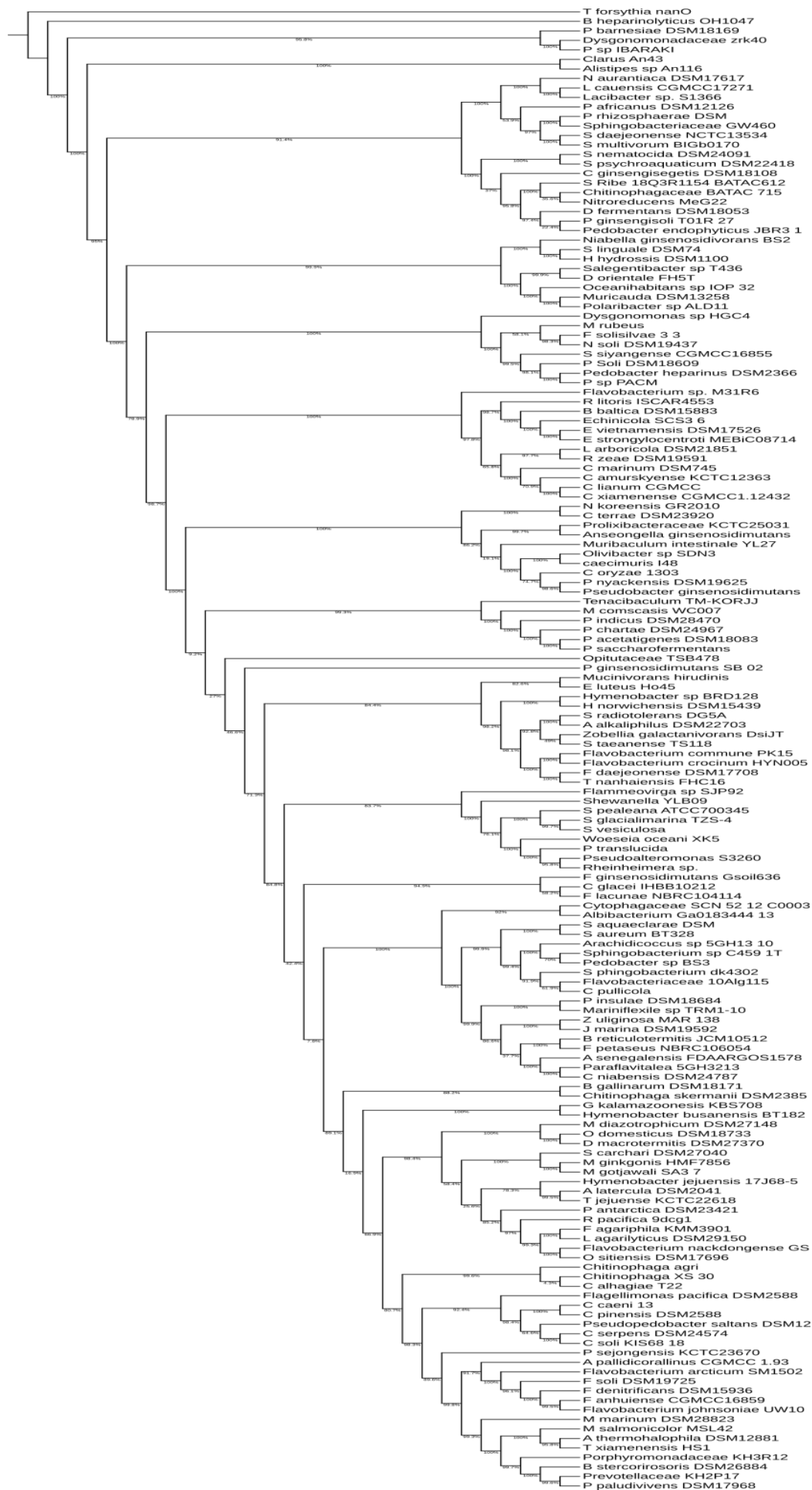
Appendix VIII: Figure 8.18. Phylogenetic tree using NanO amino acid sequences of all non-human microbes with *nan* operons. The tree is viewed and built by Interactive Tree of Life, version 6.5.2. Bootstrap analysis was performed and calculated based on alignment of the amino acid sequences by ClustalW.

375
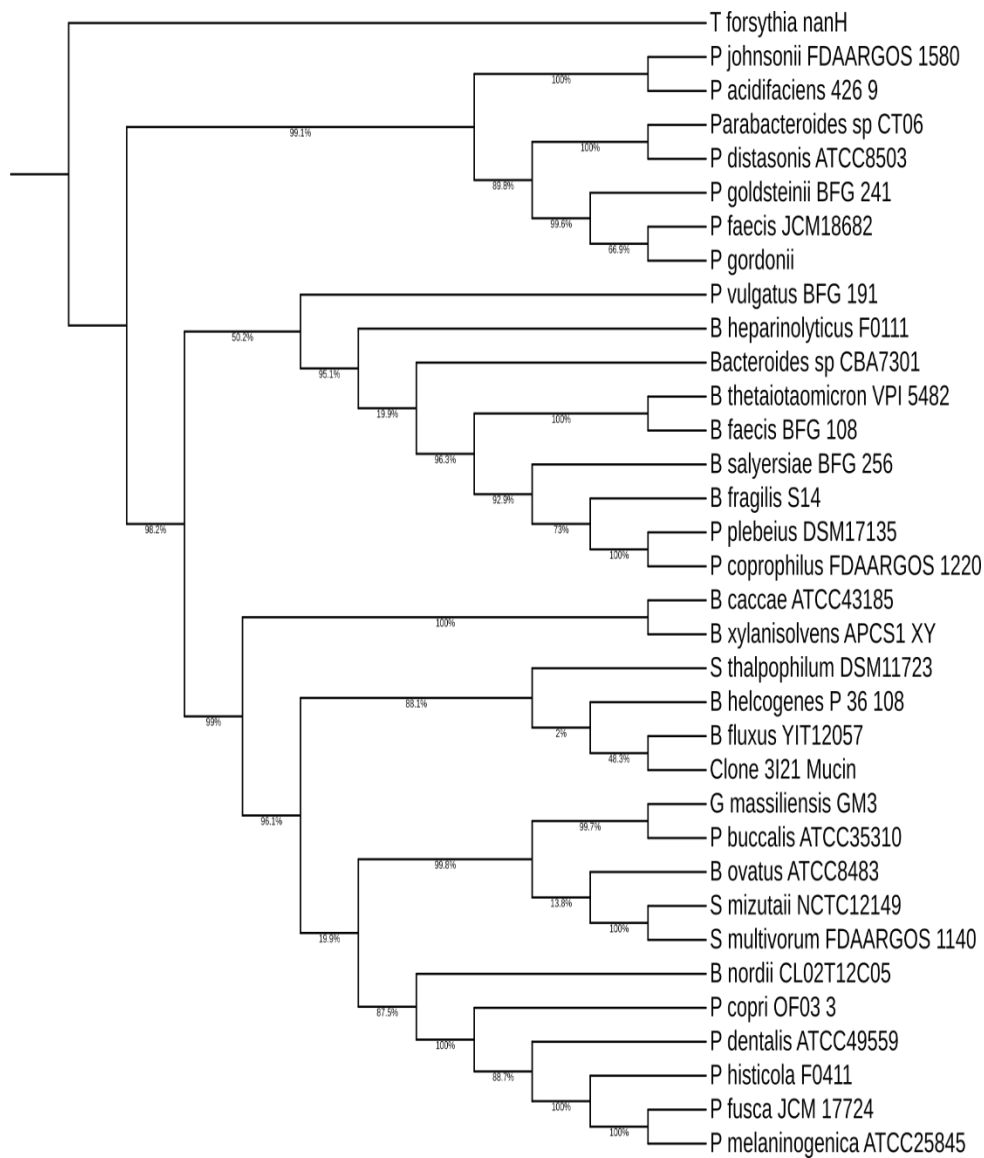
Appendix VIII: Figure 8.19. Phylogenetic tree using NanH amino acid sequences of all human microbes. The tree is viewed and built by Interactive Tree of Life, version 6.5.2. Bootstrap analysis was performed and calculated based on alignment of the amino acid sequences by ClustalW.
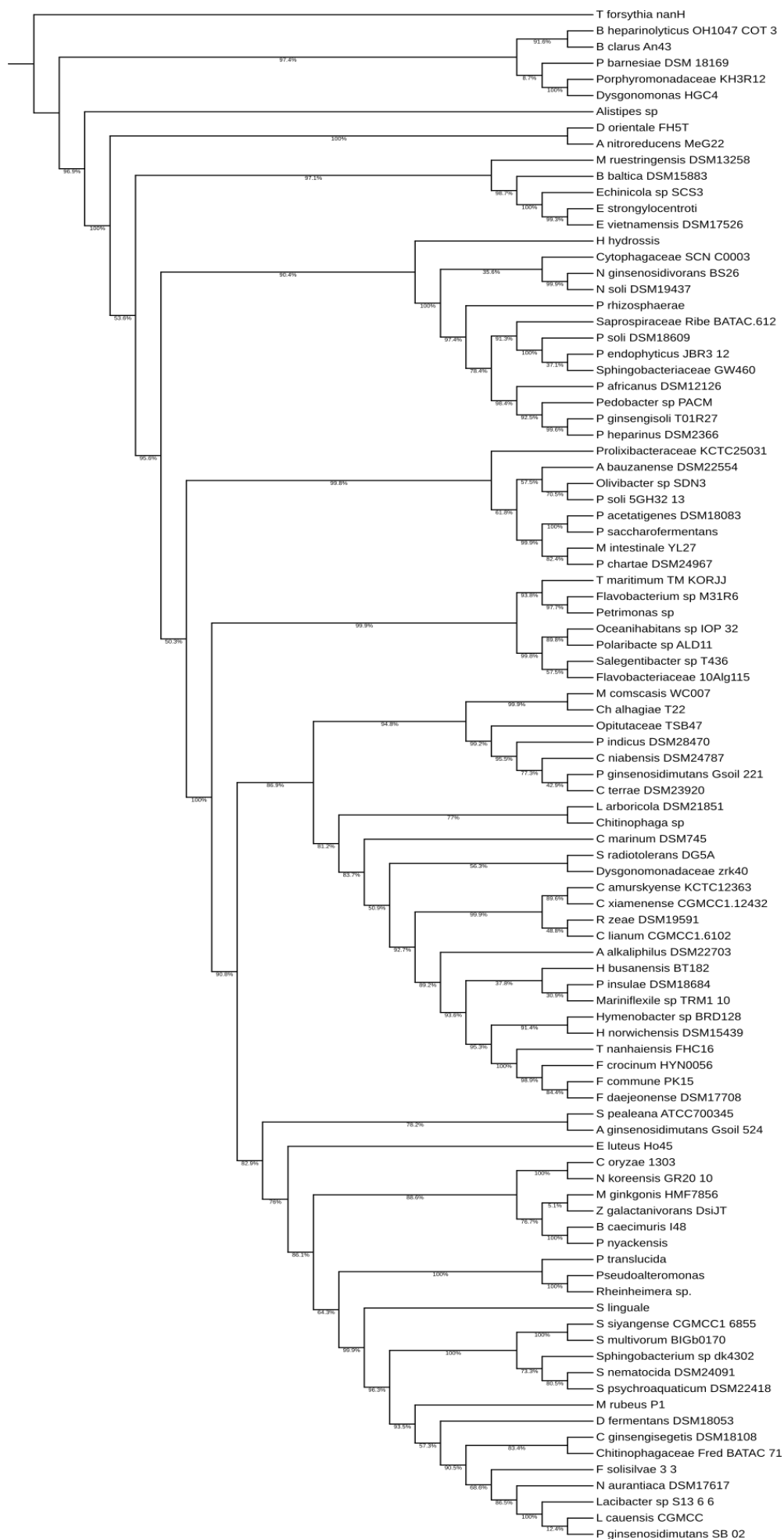
Appendix VIII: Figure 8.20. Phylogenetic tree using nanH amino acid sequences. This tree compared NanH from *T. forsythia* to non-human microbes with sialidases. The tree is viewed and built by Interactive Tree of Life, version 6.5.2. Bootstrap analysis was performed and calculated based on alignment of the amino acid sequences by ClustalW.

Appendix VIII: Table 8.2: Biological process for the significant upregulated genes in Neu5Ac

Gene Ontology for Biological Process	sample 3 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
phosphorelay signal transduction system	8.48			BFO_0146,BFO_1220,BFO_0269,BFO_0562,BFO_0895, G8UQ79,	BFO_0146,BFO_1220,BFO_0269,BFO_0562,BFO_0895, G8UQ79,
polysaccharide biosynthetic process	1.03			BFO_2943,	BFO_2943,
enzyme-directed rRNA pseudouridine synthesis	1.58			BFO_0137,	BFO_0137,
transcription-coupled nucleotide-excision repair, DNA damage recognition	2.27			G8UJ21,	G8UJ21,
telomere maintenance	1.55			BFO_2003,	BFO_2003,
tRNA threonylcarbamoyladenosine modification	1.27			G8UIN7,	G8UIN7,
carbohydrate metabolic process	15.86			BFO_2203,BFO_1193,G8UKW6,G8UL71,G8ULC7,BFO_0515, BFO_0699,BFO_2806,BFO_2947,BFO_0973,	BFO_2203,BFO_1193,G8UKW6,G8UL71,G8ULC7,BFO_0515, BFO_0699,BFO_2806,BFO_2947,BFO_0973,
glycogen biosynthetic process	1.88			BFO_0974,	BFO_0974,
glucose catabolic process	1.09			G8UMY8,	G8UMY8,
galactose metabolic process	1.42			G8UJL0,	G8UJL0,
N-acetylglucosamine metabolic process	1.02			G8UL71,	G8UL71,
acetate metabolic process	1.52			G8UL43,	G8UL43,

Gene Ontology for Biological Process	sample 3 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
glycolytic process	1.09			G8UMY8,	G8UMY8,
pentose-phosphate shunt	5.53			G8UKI0,G8UKW6,G8ULC7,	G8UKI0,G8UKW6,G8ULC7,
malate metabolic process	1.63			BFO_0515,	BFO_0515,
pyrimidine nucleotide biosynthetic process	1.18			BFO_0019,	BFO_0019,
DNA replication	7.6			BFO_2503,BFO_1723,BFO_1849,G8UQ18,	BFO_2503,BFO_1723,BFO_1849,G8UQ18,
DNA repair	5.58			G8UKU5,G8ULI3,G8UPN6,BFO_2003,	G8UKU5,G8ULI3,G8UPN6,BFO_2003,
DNA catabolic process	2.09			BFO_2314,	BFO_2314,
DNA recombination	10.91			G8UKU5,G8ULI3,BFO_0635,BFO_1731,G8UPN6,G8UQ18,	G8UKU5,G8ULI3,BFO_0635,BFO_1731,G8UPN6,G8UQ18,
transposition, DNA-mediated	8.67			BFO_0385,BFO_0831,G8UPP4,BFO_0056,	BFO_0385,BFO_0831,G8UPP4,BFO_0056,
DNA-templated transcription, initiation	2.74			BFO_0199,BFO_1778,	BFO_0199,BFO_1778,
DNA-templated transcription, termination	1.02			G8UPP8,	G8UPP8,
DNA-templated transcription, elongation	1.02			G8UPP8,	G8UPP8,
regulation of transcription, DNA-templated	12.75			G8UJ21,BFO_1220,BFO_0269,BFO_0562,BFO_2823,BFO_0791,BFO_0895,G8UQ79,	G8UJ21,BFO_1220,BFO_0269,BFO_0562,BFO_2823,BFO_0791,BFO_0895,G8UQ79,
RNA processing	1.82			BFO_0828,	BFO_0828,
tRNA modification	2.01			G8UMF5,	G8UMF5,
RNA catabolic process	1.34			G8UMP6,	G8UMP6,
translation	1.25			G8UN42,	G8UN42,
methionyl-tRNA aminoacylation	1.45			G8UPI8,	G8UPI8,

Gene Ontology for Biological Process	sample 3 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
protein folding	3.22			BFO_0138,G8UKP6,	BFO_0138,G8UKP6,
proteolysis	1.59			G8UL67,	G8UL67,
glutamine metabolic process	1.6			G8UQG4,	G8UQG4,
SRP-dependent cotranslational protein targeting to membrane	1.04			G8UNT6,	G8UNT6,
glycerol ether metabolic process	2.35			G8UN43,G8UN67,	G8UN43,G8UN67,
one-carbon metabolic process	1.22			G8ULQ7,	G8ULQ7,
NADP biosynthetic process	1.68			G8UQZ4,	G8UQZ4,
FAD biosynthetic process	1.19			G8UM71,	G8UM71,
chloride transport	1.64			BFO_1802,	BFO_1802,
amino acid transport	1.83			BFO_0948,	BFO_0948,
response to stress	1.6			BFO_1778,	BFO_1778,
cell cycle	1.27			G8UPP4,	G8UPP4,
chromosome segregation	1.27			G8UPP4,	G8UPP4,
isoprenoid biosynthetic process	1.03			BFO_0954,	BFO_0954,
regulation of cell shape	3.51			G8UJG8,G8UL10,	G8UJG8,G8UL10,
pyridoxine biosynthetic process	1.7			G8UQZ3,	G8UQZ3,
queuosine biosynthetic process	1.8			G8UJV2,	G8UJV2,
biosynthetic process	2.44			G8UMM0,BFO_1777,	G8UMM0,BFO_1777,

Gene Ontology for Biological Process	sample 3 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
lysine biosynthetic process via diaminopimelate	2.04			G8UIY5,	G8UIY5,
nucleoside metabolic process	2.81			G8UM36,	G8UM36,
nucleotide metabolic process	1.04			G8UQC1,	G8UQC1,
ribonucleoside monophosphate biosynthetic process	2.81			G8UM36,	G8UM36,
nucleotide biosynthetic process	2.81			G8UM36,	G8UM36,
riboflavin biosynthetic process	1.19			G8UM71,	G8UM71,
cobalamin biosynthetic process	5.25			BFO_2779,G8UMU8,G8UMU9,G8UQG4,	BFO_2779,G8UMU8,G8UMU9,G8UQG4,
lipid A biosynthetic process	1.08			BFO_1428,	BFO_1428,
glycolipid biosynthetic process	1.46			BFO_0644,	BFO_0644,
peptidoglycan biosynthetic process	3.72			G8UL10,BFO_0523,	G8UL10,BFO_0523,
DNA restriction-modification system	1.16			BFO_1962,	BFO_1962,
FMN biosynthetic process	1.19			G8UM71,	G8UM71,
nitrogen fixation	4.97			BFO_0086,BFO_1424,	BFO_0086,BFO_1424,
NAD biosynthetic process	1.01			G8UQ66,	G8UQ66,

Gene Ontology for Biological Process	sample 3 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
protein transport	4.04			G8UPY7,BFO_3117,BFO_3118,	G8UPY7,BFO_3117,BFO_3118,
DNA integration	1.71			BFO_0635,	BFO_0635,
coenzyme A biosynthetic process	1.2			G8UJC7,	G8UJC7,
carbohydrate biosynthetic process	1			G8UHP9,	G8UHP9,
carbohydrate catabolic process	5.73			BFO_0450,BFO_2686,	BFO_0450,BFO_2686,
terpenoid biosynthetic process	3.94			G8UQS9,	G8UQS9,
iron-sulfur cluster assembly	1.97			BFO_1485,	BFO_1485,
protein flavinylation	1.25			BFO_2219,	BFO_2219,
methylglyoxal biosynthetic process	1.1			G8UN40,	G8UN40,
isopentenyl diphosphate biosynthetic process, methylerythritol 4-phosphate pathway	3.94			G8UQS9,	G8UQS9,
pentose catabolic process	3.1			G8UKI0,	G8UKI0,
glycine decarboxylation via glycine cleavage system	1.44			G8UP98,	G8UP98,
lactate oxidation	1.09			BFO_2715,	BFO_2715,
histidine catabolic process to glutamate and formamide	1.7			G8UNG1,	G8UNG1,

Gene Ontology for Biological Process	sample 3 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
histidine catabolic process to glutamate and formate	1.7			G8UNG1,	G8UNG1,
aerobic electron transport chain	1.39			G8UNM1,	G8UNM1,
NAD metabolic process	1.68			G8UQZ4,	G8UQZ4,
propionate metabolic process, methylcitrate cycle	1.52			G8UL43,	G8UL43,
carboxylic acid metabolic process	1.32			BFO_2283,	BFO_2283,
diaminopimelate biosynthetic process	2.04			G8UIY5,	G8UIY5,
protein catabolic process	1.23			G8UKS1,	G8UKS1,
negative regulation of Wnt signaling pathway	1.21			G8UHK8,	G8UHK8,
tRNA pseudouridine synthesis	1.16			G8UQG2,	G8UQG2,
transcription antitermination	1.02			G8UPP8,	G8UPP8,
methylation	10.36			BFO_2273,G8UMU8,G8UNZ5, G8UP98,BFO_0946,	BFO_2273,G8UMU8,G8UNZ5, G8UP98,BFO_0946,
regulation of DNA-templated transcription, elongation	2.36			G8UPP8,G8UQN3,	G8UPP8,G8UQN3,
7,8-dihydroneopterin 3'-triphosphate biosynthetic process	1.22			G8ULQ7,	G8ULQ7,
protein refolding	1.04			G8UKP7,	G8UKP7,

Gene Ontology for Biological Process	sample 3 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
alginic acid biosynthetic process	1.27			BFO_0673,	BFO_0673,
ribosome biogenesis	1.28			G8UIK2,	G8UIK2,
'de novo' GDP-L-fucose biosynthetic process	1.26			G8UQC2,	G8UQC2,
pteridine-containing compound metabolic process	2.72			G8UNZ5,	G8UNZ5,
D-xylose metabolic process	1.13			G8UN89,	G8UN89,
'de novo' CTP biosynthetic process	1.43			G8UPZ9,	G8UPZ9,
nicotinamide nucleotide metabolic process	2.16			G8UHZ1,	G8UHZ1,
tetrahydrofolate biosynthetic process	1.22			G8ULQ7,	G8ULQ7,
cell division	1.27			G8UPP4,	G8UPP4,
transmembrane transport	1.15			BFO_1493,	BFO_1493,
trans-translation	2			G8UPS3,	G8UPS3,
cell wall organization	4.15			G8UL10,BFO_1449,	G8UL10,BFO_1449,
membrane assembly	1.17			BFO_1445,	BFO_1445,
DNA biosynthetic process	1.04			BFO_2503,	BFO_2503,
tRNA-guanine transglycosylation	1.8			G8UJV2,	G8UJV2,

Appendix VIII: Table 8.3: Molecular Function for the significant upregulated genes in Neu5Ac

Gene Ontology for Molecular Function	sample3 Enrichment Score	Log2 Average Fold Change		Common Genes Involved	All Genes Involved
tRNA binding	1.45			G8UPI8,	G8UPI8,
phosphorelay sensor kinase activity	7.51			BFO_1304,BFO_0561,BFO_1922,	BFO_1304,BFO_0561, BFO_1922,
nucleotide binding	6.47			BFO_2503,BFO_0722,BFO_2928,BFO_2003,	BFO_2503,BFO_0722, BFO_2928,BFO_2003,
magnesium ion binding	5.22			G8UIK2,G8UM36,G8UN89,	G8UIK2,G8UM36,G8UN89,
nucleic acid binding	6.5			G8UKY5,G8ULI3,BFO_1731,BFO_2003,	G8UKY5,G8ULI3,BFO_1731,BFO_2003,
DNA binding	34.1			BFO_0146,BFO_1220,BFO_2253,BFO_0269,BFO_0385, BFO_0562,BFO_0635,BFO_0791,BFO_1778,G8UNT4, BFO_0831,BFO_1849,BFO_0895,BFO_0905,BFO_1900,G8UPP4, BFO_1962,G8UQ79,BFO_0056,G8UQN3,BFO_2043,	BFO_0146,BFO_1220,BFO_2253,BFO_0269,BFO_0385, BFO_0562,BFO_0635,BFO_0791,BFO_1778,G8UNT4, BFO_0831,BFO_1849,BFO_0895,BFO_0905,BFO_1900,G8UPP4,BFO_1962,G8UQ79,BFO_0056,G8UQN3,BFO_2043,
DNA helicase activity	7.81			G8ULI3,BFO_1731,BFO_1900,BFO_2003,	G8ULI3,BFO_1731,BFO_1900,BFO_2003,
damaged DNA binding	2.27			G8UJ21,	G8UJ21,
DNA-binding transcription factor activity	12.59			BFO_0121,BFO_0199,BFO_1304,BFO_2280,BFO_2793, BFO_1721,BFO_1922,	BFO_0121,BFO_0199,BFO_1304,BFO_2280,BFO_2793,BFO_1721,BFO_1922,
RNA binding	8.97			BFO_0137,G8UMP6,G8UMY6,BFO_0828,G8UPS3,G8UQG2,	BFO_0137,G8UMP6,G8UMY6,BFO_0828,G8UPS3,G8UQG2,
structural constituent of ribosome	1.25			G8UN42,	G8UN42,
translation initiation	1.65			BFO_1899,	BFO_1899,

Gene Ontology for Molecular Function	sample3 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
factor activity					
translation elongation factor activity	1.34			G8UQN3,	G8UQN3,
peptidyl-prolyl cis-trans isomerase activity	2.8			BFO_1216,BFO_1497,	BFO_1216,BFO_1497,
catalytic activity	15.04			BFO_0086,BFO_2272,BFO_1424,BFO_0545,G8UPI7, BFO_0054,BFO_1032,	BFO_0086,BFO_2272, BFO_1424,BFO_0545, G8UPI7, BFO_0054,BFO_1032,
1,4-alpha-glucan branching enzyme activity	1.88			BFO_0974,	BFO_0974,
DNA-directed DNA polymerase activity	1.58			BFO_1723,	BFO_1723,
FMN adenylyltransferase activity	1.19			G8UM71,	G8UM71,
GTPase activity	2.31			G8UIK2,G8UNT6,	G8UIK2,G8UNT6,
GTP cyclohydrolase I activity	1.22			G8ULQ7,	G8ULQ7,
NAD+ kinase activity	1.68			G8UQZ4,	G8UQZ4,
UDP-glucose 4-epimerase activity	1.42			G8UJL0,	G8UJL0,
acetyl-CoA hydrolase activity	1.52			G8UL43,	G8UL43,
alcohol dehydrogena	1.49			BFO_1487,	BFO_1487,

Gene Ontology for Molecular Function	sample3 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
se (NAD+) activity					
alkaline phosphatase activity	2.55			BFO_3041,	BFO_3041,
aminomethyltransferase activity	1.44			G8UP98,	G8UP98,
endopeptidase activity	1.21			G8UHK8,	G8UHK8,
ATP-dependent peptidase activity	1.23			G8UKS1,	G8UKS1,
aminopeptidase activity	4.8			BFO_1179,G8UKI4,BFO_2679,	BFO_1179,G8UKI4,BFO_2679,
aspartic-type endopeptidase activity	2.94			G8UPA1,	G8UPA1,
metalloendopeptidase activity	3.62			BFO_0225,G8UKS1,BFO_2027,	BFO_0225,G8UKS1,BFO_2027,
serine-type endopeptidase activity	3.56			BFO_1179,G8UKI4,	BFO_1179,G8UKI4,
helicase activity	2.27			G8UJ21,	G8UJ21,
histidine ammonia-lyase activity	1.7			G8UNG1,	G8UNG1,
mannose-6-phosphate isomerase activity	1.42			BFO_0973,	BFO_0973,
methylmalonyl-CoA mutase activity	1.02			BFO_2747,	BFO_2747,
nicotinate-nucleotide diphosphorylase (carboxylating) activity	1.01			G8UQ66,	G8UQ66,

Gene Ontology for Molecular Function	sample3 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
endonuclease activity	3.04			G8UQ18,	G8UQ18,
RNA-DNA hybrid ribonuclease activity	2.77			G8UKY5,G8UMP6,	G8UKY5,G8UMP6,
deoxyribonuclease activity	1.6			BFO_0955,	BFO_0955,
hydrolase activity, hydrolyzing O-glycosyl compounds	8.18			BFO_1193,BFO_2686,BFO_0699,BFO_0974,	BFO_1193,BFO_2686, BFO_0699,BFO_0974,
beta-galactosidase activity	3.5			BFO_0450,	BFO_0450,
pantothenate kinase activity	1.2			G8UJC7,	G8UJC7,
peroxidase activity	2.82			BFO_0505,BFO_1720,	BFO_0505,BFO_1720,
phosphoglycerate dehydrogenase activity	1.45			BFO_0195,	BFO_0195,
protein serine/threonine kinase activity	1.78			BFO_2981,	BFO_2981,
protein serine/threonine/tyrosine kinase activity	1.78			BFO_2981,	BFO_2981,
ribonucleoside-diphosphate reductase activity, thioredoxin disulfide as acceptor	1.04			BFO_2503,	BFO_2503,
ribose phosphate diphosphokinase activity	2.81			G8UM36,	G8UM36,

Gene Ontology for Molecular Function	sample3 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
D-ribulose-phosphate 3-epimerase activity	3.1			G8UKI0,	G8UKI0,
transaldolase activity	1.18			G8ULC7,	G8ULC7,
transposase activity	7.4			BFO_0385,BFO_0831,BFO_0056,	BFO_0385,BFO_0831, BFO_0056,
methionine-tRNA ligase activity	1.45			G8UPI8,	G8UPI8,
tryptophan synthase activity	1.1			G8UNI9,	G8UNI9,
voltage-gated chloride channel activity	1.64			BFO_1802,	BFO_1802,
iron ion binding	1.97			BFO_1485,	BFO_1485,
calcium ion binding	1.53			BFO_0021,	BFO_0021,
ATP binding	48.06			BFO_0082,BFO_0088,BFO_0102,G8UHZ1,BFO_0138, G8UJ21,G8UJC7,BFO_2263,G8UKP6,G8UKP7,G8UKS1,G8UKU5, BFO_1493,G8ULI3,G8UM36,G8UM71,BFO_2779,BFO_1731, BFO_1849,G8UPI8,BFO_1900, BFO_2981,G8UPZ9,G8UQG4, ⊻	BFO_0082,BFO_0088, BFO_0102,G8UHZ1,BFO_0138, G8UJ21,G8UJC7,BFO_2263,G8UKP6,G8UKP7,G8UKS1,G8UKU5, BFO_1493,G8ULI3,G8UM36,G8UM71,BFO_2779,BFO_1731, BFO_1849,G8UPI8,BFO_1900,BFO_2981,G8UPZ9,G8UQG4, ⊻
GTP binding	3.53			G8UIK2,G8ULQ7,G8UNT6,	G8UIK2,G8ULQ7,G8UNT6,
methyltransferase activity	4.78			BFO_2273,BFO_0946,	BFO_2273,BFO_0946,
N-methyltransferase activity	1.16			BFO_1962,	BFO_1962,
RNA methyltransf	1.82			BFO_0828,	BFO_0828,

Gene Ontology for Molecular Function	sample3 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
erase activity					
peptidase activity	11.03			BFO_0082,BFO_0088,BFO_0102,	BFO_0082,BFO_0088, BFO_0102,
serine-type peptidase activity	6.5			BFO_0087,BFO_0097,BFO_0742,BFO_3080,BFO_1007,	BFO_0087,BFO_0097, BFO_0742,BFO_3080, BFO_1007,
metallopeptidase activity	1.7			BFO_1449,	BFO_1449,
zinc ion binding	9.81			G8UHU0,G8UKS1,BFO_1449, G8ULQ7,G8UNZ5,BFO_0973,	G8UHU0,G8UKS1,BFO_1449,G8ULQ7,G8UNZ5,BFO_0973,
protein methyltransferase activity	1.19			G8UMU9,	G8UMU9,
3'-5' exonuclease activity	4.62			BFO_1723,G8UQ18,	BFO_1723,G8UQ18,
N-acetylglucosamine-6-phosphate deacetylase activity	1.02			G8UL71,	G8UL71,
queuine tRNA-ribosyltransferase activity	1.8			G8UJV2,	G8UJV2,
transaminase activity	2.7			G8UMM0,G8UP98,	G8UMM0,G8UP98,
sulfuric ester hydrolase activity	3.06			G8UJW8,G8UNI7,	G8UJW8,G8UNI7,
riboflavin kinase activity	1.19			G8UM71,	G8UM71,
ubiquitin-like modifier activating enzyme activity	1.32			BFO_2283,	BFO_2283,
methionine synthase activity	2.72			G8UNZ5,	G8UNZ5,

Gene Ontology for Molecular Function	sample3 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
acetate CoA-transferase activity	1.52			G8UL43,	G8UL43,
cob(I)yrinic acid a,c-diamide adenosyltransferase activity	1.03			BFO_2779,	BFO_2779,
4-hydroxy-tetrahydrodipicolinate synthase activity	2.04			G8UIY5,	G8UIY5,
exodeoxyribonuclease VII activity	2.09			BFO_2314,	BFO_2314,
exoribonuclease II activity	1.09			G8UMY6,	G8UMY6,
glutamate racemase activity	2.44			G8UL10,	G8UL10,
methylglyoxal synthase activity	1.1			G8UN40,	G8UN40,
site-specific DNA-methyltransferase (adenine-specific) activity	1.16			BFO_1962,	BFO_1962,
tyrosine-based site-specific recombinase activity	1.27			G8UPP4,	G8UPP4,
xylose isomerase activity	1.13			G8UN89,	G8UN89,
electron transfer activity	4.18			G8UL47,G8UL48,G8UNM1,	G8UL47,G8UL48,G8UNM1,
protein-disulfide	2.35			G8UN43,G8UN67,	G8UN43,G8UN67,

Gene Ontology for Molecular Function	sample3 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
reductase activity					
antiporter activity	1.07			BFO_0228,	BFO_0228,
solute:proton antiporter activity	2.64			BFO_0837,	BFO_0837,
siderophore transmembrane transporter activity	1.16			G8UPY7,	G8UPY7,
efflux transmembrane transporter activity	1.04			BFO_0388,	BFO_0388,
kinase activity	4.17			G8UM36,BFO_2806,	G8UM36,BFO_2806,
pyrophosphatase activity	1.08			BFO_1428,	BFO_1428,
oxidoreductase activity	11.41			BFO_0079,BFO_0136,BFO_3326,BFO_1635,BFO_0712, BFO_0019,BFO_3174,	BFO_0079,BFO_0136, BFO_3326,BFO_1635, BFO_0712, BFO_0019,BFO_3174,
malate dehydrogenase activity	1.63			BFO_0515,	BFO_0515,
oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	3.98			BFO_2283,BFO_0515,BFO_2943,	BFO_2283,BFO_0515, BFO_2943,
oxidoreductase activity, acting on the CH-CH group of donors	2.3			G8UL41,	G8UL41,
oxidoreductase activity, acting on the CH-CH	1.03			BFO_2943,	BFO_2943,

Gene Ontology for Molecular Function	sample3 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
group of donors, NAD or NADP as acceptor					
transferase activity	6.64			BFO_2219,BFO_0523,BFO_0954,BFO_1987,BFO_1995,	BFO_2219,BFO_0523, BFO_0954,BFO_1987, BFO_1995,
acyltransferase activity	2.72			BFO_0644,BFO_0673,	BFO_0644,BFO_0673,
glycosyltransferase activity	5.38			BFO_0104,BFO_2575,BFO_0900,	BFO_0104,BFO_2575, BFO_0900,
transferase activity, transferring alkyl or aryl (other than methyl) groups	1.3			BFO_0914,	BFO_0914,
transferase activity, transferring phosphorus-containing groups	1.52			G8UNI7,	G8UNI7,
phosphotransferase activity, alcohol group as acceptor	1.36			BFO_2806,	BFO_2806,
hydrolase activity	15.97			BFO_0197,G8UJ21,BFO_1194, BFO_3347,G8UL67, BFO_2554,BFO_2719,BFO_1731,BFO_2805,BFO_1799,	BFO_0197,G8UJ21,BFO_1194,BFO_3347,G8UL67, BFO_2554,BFO_2719, BFO_1731,BFO_2805, BFO_1799,
dipeptidase activity	1.7			BFO_1449,	BFO_1449,
lyase activity	6.42			BFO_0133,BFO_0136,G8UL41, BFO_1777,	BFO_0133,BFO_0136, G8UL41,BFO_1777,
aldehyde-lyase activity	1.18			G8ULC7,	G8ULC7,
isomerase activity	6.26			BFO_2919,G8UQC2,G8UQS9,	BFO_2919,G8UQC2,G8UQS9,

Gene Ontology for Molecular Function	sample3 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
ATP hydrolysis activity	11.57			BFO_0138,G8UKP6,G8UKP7,G8UKS1,G8ULI3,BFO_1849, BFO_1900,	BFO_0138,G8UKP6,G8UKP7,G8UKS1,G8ULI3,BFO_1849, BFO_1900,
sigma factor activity	1.6			BFO_1778,	BFO_1778,
precorrin-8X methylmutase activity	1.42			G8UMU8,	G8UMU8,
6-phosphogluconolactonase activity	1.25			G8UKW6,	G8UKW6,
transmembrane transporter activity	9.93			BFO_0135,BFO_0387,BFO_1586,BFO_2791,BFO_3117,	BFO_0135,BFO_0387, BFO_1586,BFO_2791, BFO_3117,
manganese ion binding	2.43			G8UMP6,G8UMY8,	G8UMP6,G8UMY8,
pyridoxal phosphate binding	3.54			G8UMM0,G8UNI9,BFO_1777,	G8UMM0,G8UNI9,BFO_1777,
carbohydrate binding	7.21			BFO_2272,BFO_0450,BFO_2947,	BFO_2272,BFO_0450, BFO_2947,
1-deoxy-D-xylulose-5-phosphate reductoisomerase activity	3.94			G8UQS9,	G8UQS9,
precorrin-3B C17-methyltransferase activity	1.42			G8UMU8,	G8UMU8,
cobalamin binding	4.77			BFO_2503,BFO_2747,G8UNZ5,	BFO_2503,BFO_2747, G8UNZ5,
energy transducer activity	1.16			G8UPY7,	G8UPY7,
pyridoxine 5'-phosphate synthase activity	1.7			G8UQZ3,	G8UQZ3,
UMP kinase activity	1.43			G8UPZ9,	G8UPZ9,

Gene Ontology for Molecular Function	sample3 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
NADH pyrophosphatase activity	1.04			G8UQC1,	G8UQC1,
methylthiotransferase activity	2.01			G8UMF5,	G8UMF5,
dTTP diphosphatase activity	1.04			G8UQC1,	G8UQC1,
UTP diphosphatase activity	1.04			G8UQC1,	G8UQC1,
cobyrinic acid a,c-diamide synthase activity	1.6			G8UQG4,	G8UQG4,
3'-tRNA processing endoribonuclease activity	1.52			G8UHU0,	G8UHU0,
xenobiotic transmembrane transporter activity	1.07			BFO_0228,	BFO_0228,
cation binding	1.88			BFO_0974,	BFO_0974,
sequence-specific DNA binding	8.1			BFO_0121,BFO_1304,BFO_2793,BFO_1922,	BFO_0121,BFO_1304, BFO_2793,BFO_1922,
precorrin-6Y C5,15-methyltransferase (decarboxylating) activity	1.19			G8UMU9,	G8UMU9,
2,3-bisphosphoglycerate-independent phosphoglycerate mutase activity	1.09			G8UMY8,	G8UMY8,

Gene Ontology for Molecular Function	sample3 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
metal ion binding	48.47			BFO_0079,BFO_0086,BFO_1079,G8UHZ1,BFO_0136, BFO_0225,BFO_2219,G8UKI0, BFO_1424,BFO_1428,BFO_0474, G8UKY5,BFO_1487,G8UL71,BFO_0545,BFO_1635, BFO_2715,G8UMD6,G8UMF5, BFO_2747,G8UNM1,G8UPI8,BFO_0019,⊗	BFO_0079,BFO_0086, BFO_1079,G8UHZ1,BFO_0136, BFO_0225,BFO_2219, G8UKI0,BFO_1424,BFO_1428,BFO_0474, G8UKY5,BFO_1487,G8UL71,BFO_0545,BFO_1635, BFO_2715,G8UMD6, G8UMF5,BFO_2747,G8UNM1,G8UPI8,BFO_0019,⊗
protein dimerization activity	2			BFO_0561,	BFO_0561,
N-acetylgalactosamine-6-phosphate deacetylase activity	1.02			G8UL71,	G8UL71,
glyoxylate reductase (NAD+) activity	1.03			BFO_0938,	BFO_0938,
N-acylglucosamine 2-epimerase activity	1.68			BFO_2203,	BFO_2203,
S-adenosylhomocysteine deaminase activity	3.37			G8UMD6,	G8UMD6,
vinylacetyl-CoA delta-isomerase activity	2.3			G8UL41,	G8UL41,
N-acetylneuraminate synthase activity	1			G8UHP9,	G8UHP9,
GDP-L-fucose synthase activity	1.26			G8UQC2,	G8UQC2,

Gene Ontology for Molecular Function	sample3 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
flavin adenine dinucleotide binding	2.61			G8UL48,BFO_0019,	G8UL48,BFO_0019,
unfolded protein binding	2.27			G8UKP7,BFO_2546,	G8UKP7,BFO_2546,
NAD binding	5.19			BFO_0195,BFO_2943,BFO_0938,G8UQZ4,	BFO_0195,BFO_2943,BFO_0938,G8UQZ4,
iron-sulfur cluster binding	14.43			BFO_0086,BFO_1424,BFO_1485,BFO_0545,BFO_1635,BFO_0054,	BFO_0086,BFO_1424,BFO_1485,BFO_0545,BFO_1635,BFO_0054,
2 iron, 2 sulfur cluster binding	2.2			BFO_0474,BFO_0019,	BFO_0474,BFO_0019,
4 iron, 4 sulfur cluster binding	5.85			BFO_0079,BFO_0136,BFO_2715,G8UMF5,	BFO_0079,BFO_0136,BFO_2715,G8UMF5,
ADP-dependent NAD(P)H-hydrate dehydratase activity	2.16			G8UHZ1,	G8UHZ1,
NADHX epimerase activity	2.16			G8UHZ1,	G8UHZ1,
NADPHX epimerase activity	2.16			G8UHZ1,	G8UHZ1,
RNA polymerase binding	1.34			G8UQN3,	G8UQN3,
NADP+ binding	1.26			G8UQC2,	G8UQC2,
NADPH binding	3.94			G8UQS9,	G8UQS9,
5'-methylthioadenosine deaminase activity	3.37			G8UMD6,	G8UMD6,

Gene Ontology for Molecular Function	sample3 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
1,4-alpha-glucan branching enzyme activity (using a glucosylated glycogenin as primer for glycogen synthesis)	1.88			BFO_0974,	BFO_0974,
tRNA pseudouridine synthase activity	1.16			G8UQG2,	G8UQG2,
protein serine kinase activity	1.78			BFO_2981,	BFO_2981,
rRNA pseudouridine synthase activity	1.58			BFO_0137,	BFO_0137,
ABC-type transporter activity	11.03			BFO_0082,BFO_0088,BFO_0102,	BFO_0082,BFO_0088, BFO_0102,

Appendix VIII: Table 8.4. Cellular Component for the significant upregulated genes in Neu5Ac

Gene Ontology for Cellular Component	sample3 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
tRNA binding	1.45			G8UPI8,	G8UPI8,
phosphorelay sensor kinase activity	7.51			BFO_1304,BFO_0561,BFO_1922,	BFO_1304,BFO_0561,BFO_1922,
nucleotide binding	6.47			BFO_2503,BFO_0722,BFO_2928,BFO_2003,	BFO_2503,BFO_0722,BFO_2928,BFO_2003,
magnesium ion binding	5.22			G8UIK2,G8UM36,G8UN89,	G8UIK2,G8UM36,G8UN89,
nucleic acid binding	6.5			G8UKY5,G8ULI3,BFO_1731,BFO_2003,	G8UKY5,G8ULI3,BFO_1731,BFO_2003,
DNA binding	34.1			BFO_0146,BFO_1220,BFO_2253,BFO_0269,BFO_0385, BFO_0562,BFO_0635,BFO_0791,BFO_1778,G8UNT4, BFO_0831,BFO_1849,BFO_0895,BFO_0905,BFO_1900,G8UPP4, BFO_1962,G8UQ79,BFO_0056,G8UQN3,BFO_2043,	BFO_0146,BFO_1220,BFO_2253,BFO_0269,BFO_0385, BFO_0562,BFO_0635,BFO_0791,BFO_1778,G8UNT4, BFO_0831,BFO_1849,BFO_0895,BFO_0905,BFO_1900,G8UPP4, BFO_1962,G8UQ79,BFO_0056,G8UQN3,BFO_2043,
DNA helicase activity	7.81			G8ULI3,BFO_1731,BFO_1900,BFO_2003,	G8ULI3,BFO_1731,BFO_1900,BFO_2003,
damaged DNA binding	2.27			G8UJ21,	G8UJ21,
DNA-binding transcription factor activity	12.59			BFO_0121,BFO_0199,BFO_1304,BFO_2280,BFO_2793, BFO_1721,BFO_1922,	BFO_0121,BFO_0199,BFO_1304,BFO_2280,BFO_2793, BFO_1721,BFO_1922,
RNA binding	8.97			BFO_0137,G8UMP6,G8UMY6,BFO_0828,G8UPS3,G8UQG2,	BFO_0137,G8UMP6,G8UMY6,BFO_0828,G8UPS3,G8UQG2,
structural constituent of ribosome	1.25			G8UN42,	G8UN42,
translation initiation factor activity	1.65			BFO_1899,	BFO_1899,
translation elongation factor activity	1.34			G8UQN3,	G8UQN3,
peptidyl-prolyl cis-trans isomerase activity	2.8			BFO_1216,BFO_1497,	BFO_1216,BFO_1497,
catalytic activity	15.04			BFO_0086,BFO_2272,BFO_1424,BFO_0545,G8UPI7, BFO_0054,BFO_1032,	BFO_0086,BFO_2272,BFO_1424,BFO_0545,G8UPI7, BFO_0054,BFO_1032,
1,4-alpha-glucan branching enzyme activity	1.88			BFO_0974,	BFO_0974,
DNA-directed DNA polymerase activity	1.58			BFO_1723,	BFO_1723,

Gene Ontology for Cellular Component	sample3 Enrichment Score	Log2 Average Fold Change	Common Genes Involved	All Genes Involved
FMN adenylyltransferase activity	1.19		G8UM71,	G8UM71,
GTPase activity	2.31		G8UIK2,G8UNT6,	G8UIK2,G8UNT6,
GTP cyclohydrolase I activity	1.22		G8ULQ7,	G8ULQ7,
NAD+ kinase activity	1.68		G8UQZ4,	G8UQZ4,
UDP-glucose 4-epimerase activity	1.42		G8UJL0,	G8UJL0,
acetyl-CoA hydrolase activity	1.52		G8UL43,	G8UL43,
alcohol dehydrogenase (NAD+) activity	1.49		BFO_1487,	BFO_1487,
alkaline phosphatase activity	2.55		BFO_3041,	BFO_3041,
aminomethyltransferase activity	1.44		G8UP98,	G8UP98,
endopeptidase activity	1.21		G8UHK8,	G8UHK8,
ATP-dependent peptidase activity	1.23		G8UKS1,	G8UKS1,
aminopeptidase activity	4.8		BFO_1179,G8UKI4,BFO_2679,	BFO_1179,G8UKI4,BFO_2679,
aspartic-type endopeptidase activity	2.94		G8UPA1,	G8UPA1,
metalloendopeptidase activity	3.62		BFO_0225,G8UKS1,BFO_2027,	BFO_0225,G8UKS1,BFO_2027,
serine-type endopeptidase activity	3.56		BFO_1179,G8UKI4,	BFO_1179,G8UKI4,
helicase activity	2.27		G8UJ21,	G8UJ21,
histidine ammonia-lyase activity	1.7		G8UNG1,	G8UNG1,
mannose-6-phosphate isomerase activity	1.42		BFO_0973,	BFO_0973,
methylmalonyl-CoA mutase activity	1.02		BFO_2747,	BFO_2747,
nicotinate-nucleotide diphosphorylase (carboxylating) activity	1.01		G8UQ66,	G8UQ66,
endonuclease activity	3.04		G8UQ18,	G8UQ18,
RNA-DNA hybrid ribonuclease activity	2.77		G8UKY5,G8UMP6,	G8UKY5,G8UMP6,
deoxyribonuclease activity	1.6		BFO_0955,	BFO_0955,
hydrolase activity, hydrolyzing O-glycosyl compounds	8.18		BFO_1193,BFO_2686,BFO_0699,BFO_0974,	BFO_1193,BFO_2686,BFO_0699,BFO_0974,

Gene Ontology for Cellular Component	sample3 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
beta-galactosidase activity	3.5			BFO_0450,	BFO_0450,
pantothenate kinase activity	1.2			G8UJC7,	G8UJC7,
peroxidase activity	2.82			BFO_0505,BFO_1720,	BFO_0505,BFO_1720,
phosphoglycerate dehydrogenase activity	1.45			BFO_0195,	BFO_0195,
protein serine/threonine kinase activity	1.78			BFO_2981,	BFO_2981,
protein serine/threonine/tyrosine kinase activity	1.78			BFO_2981,	BFO_2981,
ribonucleoside-diphosphate reductase activity, thioredoxin disulfide as acceptor	1.04			BFO_2503,	BFO_2503,
ribose phosphate diphosphokinase activity	2.81			G8UM36,	G8UM36,
D-ribulose-phosphate 3-epimerase activity	3.1			G8UKI0,	G8UKI0,
transaldolase activity	1.18			G8ULC7,	G8ULC7,
transposase activity	7.4			BFO_0385,BFO_0831,BFO_0056,	BFO_0385,BFO_0831,BFO_0056,
methionine-tRNA ligase activity	1.45			G8UPI8,	G8UPI8,
tryptophan synthase activity	1.1			G8UNI9,	G8UNI9,
voltage-gated chloride channel activity	1.64			BFO_1802,	BFO_1802,
iron ion binding	1.97			BFO_1485,	BFO_1485,
calcium ion binding	1.53			BFO_0021,	BFO_0021,
ATP binding	48.06			BFO_0082,BFO_0088,BFO_0102,G8UHZ1,BFO_0138, G8UJ21,G8UJC7,BFO_2263,G8UKP6,G8UKP7,G8UKS1,G8UKU5, BFO_1493,G8ULI3,G8UM36,G8UM71,BFO_2779,BFO_1731, BFO_1849,G8UPI8,BFO_1900,BFO_2981,G8UPZ9,G8UQG4, ⧩	BFO_0082,BFO_0088,BFO_0102,G8UHZ1,BFO_0138, G8UJ21,G8UJC7,BFO_2263,G8UKP6,G8UKP7,G8UKS1,G8UKU5, BFO_1493,G8ULI3,G8UM36,G8UM71,BFO_2779,BFO_1731, BFO_1849,G8UPI8,BFO_1900,BFO_2981,G8UPZ9,G8UQG4, ⧩
GTP binding	3.53			G8UIK2,G8ULQ7,G8UNT6,	G8UIK2,G8ULQ7,G8UNT6,
methyltransferase activity	4.78			BFO_2273,BFO_0946,	BFO_2273,BFO_0946,
N-methyltransferase activity	1.16			BFO_1962,	BFO_1962,

Gene Ontology for Cellular Component	sample3 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
RNA methyltransferase activity	1.82			BFO_0828,	BFO_0828,
peptidase activity	11.03			BFO_0082,BFO_0088,BFO_0102,	BFO_0082,BFO_0088,BFO_0102,
serine-type peptidase activity	6.5			BFO_0087,BFO_0097,BFO_0742,BFO_3080,BFO_1007,	BFO_0087,BFO_0097,BFO_0742,BFO_3080,BFO_1007,
metallopeptidase activity	1.7			BFO_1449,	BFO_1449,
zinc ion binding	9.81			G8UHU0,G8UKS1,BFO_1449,G8ULQ7,G8UNZ5,BFO_0973,	G8UHU0,G8UKS1,BFO_1449,G8ULQ7,G8UNZ5,BFO_0973,
protein methyltransferase activity	1.19			G8UMU9,	G8UMU9,
3'-5' exonuclease activity	4.62			BFO_1723,G8UQ18,	BFO_1723,G8UQ18,
N-acetylglucosamine-6-phosphate deacetylase activity	1.02			G8UL71,	G8UL71,
queuine tRNA-ribosyltransferase activity	1.8			G8UJV2,	G8UJV2,
transaminase activity	2.7			G8UMM0,G8UP98,	G8UMM0,G8UP98,
sulfuric ester hydrolase activity	3.06			G8UJW8,G8UNI7,	G8UJW8,G8UNI7,
riboflavin kinase activity	1.19			G8UM71,	G8UM71,
ubiquitin-like modifier activating enzyme activity	1.32			BFO_2283,	BFO_2283,
methionine synthase activity	2.72			G8UNZ5,	G8UNZ5,
acetate CoA-transferase activity	1.52			G8UL43,	G8UL43,
cob(I)yrinic acid a,c-diamide adenosyltransferase activity	1.03			BFO_2779,	BFO_2779,
4-hydroxy-tetrahydrodipicolinate synthase activity	2.04			G8UIY5,	G8UIY5,
exodeoxyribonuclease VII activity	2.09			BFO_2314,	BFO_2314,
exoribonuclease II activity	1.09			G8UMY6,	G8UMY6,
glutamate racemase activity	2.44			G8UL10,	G8UL10,
methylglyoxal synthase activity	1.1			G8UN40,	G8UN40,
site-specific DNA-methyltransferase	1.16			BFO_1962,	BFO_1962,

Gene Ontology for Cellular Component	sample3 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
(adenine-specific) activity					
tyrosine-based site-specific recombinase activity	1.27			G8UPP4,	G8UPP4,
xylose isomerase activity	1.13			G8UN89,	G8UN89,
electron transfer activity	4.18			G8UL47,G8UL48,G8UNM1,	G8UL47,G8UL48,G8UNM1,
protein-disulfide reductase activity	2.35			G8UN43,G8UN67,	G8UN43,G8UN67,
antiporter activity	1.07			BFO_0228,	BFO_0228,
solute:proton antiporter activity	2.64			BFO_0837,	BFO_0837,
siderophore transmembrane transporter activity	1.16			G8UPY7,	G8UPY7,
efflux transmembrane transporter activity	1.04			BFO_0388,	BFO_0388,
kinase activity	4.17			G8UM36,BFO_2806,	G8UM36,BFO_2806,
pyrophosphatase activity	1.08			BFO_1428,	BFO_1428,
oxidoreductase activity	11.41			BFO_0079,BFO_0136,BFO_3326,BFO_1635,BFO_0712, BFO_0019,BFO_3174,	BFO_0079,BFO_0136,BFO_3326,BFO_1635,BFO_0712, BFO_0019,BFO_3174,
malate dehydrogenase activity	1.63			BFO_0515,	BFO_0515,
oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	3.98			BFO_2283,BFO_0515,BFO_2943,	BFO_2283,BFO_0515,BFO_2943,
oxidoreductase activity, acting on the CH-CH group of donors	2.3			G8UL41,	G8UL41,
oxidoreductase activity, acting on the CH-CH group of donors, NAD or NADP as acceptor	1.03			BFO_2943,	BFO_2943,
transferase activity	6.64			BFO_2219,BFO_0523,BFO_0954,BFO_1987,BFO_1995,	BFO_2219,BFO_0523,BFO_0954,BFO_1987,BFO_1995,
acyltransferase activity	2.72			BFO_0644,BFO_0673,	BFO_0644,BFO_0673,
glycosyltransferase activity	5.38			BFO_0104,BFO_2575,BFO_0900,	BFO_0104,BFO_2575,BFO_0900,
transferase activity, transferring alkyl or aryl (other than methyl) groups	1.3			BFO_0914,	BFO_0914,

Gene Ontology for Cellular Component	sample3 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
transferase activity, transferring phosphorus-containing groups	1.52			G8UNI7,	G8UNI7,
phosphotransferase activity, alcohol group as acceptor	1.36			BFO_2806,	BFO_2806,
hydrolase activity	15.97			BFO_0197,G8UJ21,BFO_1194,BFO_3347,G8UL67, BFO_2554,BFO_2719,BFO_1731,BFO_2805,BFO_1799,	BFO_0197,G8UJ21,BFO_1194,BFO_3347,G8UL67, BFO_2554,BFO_2719,BFO_1731,BFO_2805,BFO_1799,
dipeptidase activity	1.7			BFO_1449,	BFO_1449,
lyase activity	6.42			BFO_0133,BFO_0136,G8UL41,BFO_1777,	BFO_0133,BFO_0136,G8UL41,BFO_1777,
aldehyde-lyase activity	1.18			G8ULC7,	G8ULC7,
isomerase activity	6.26			BFO_2919,G8UQC2,G8UQS9,	BFO_2919,G8UQC2,G8UQS9,
ATP hydrolysis activity	11.57			BFO_0138,G8UKP6,G8UKP7,G8UKS1,G8ULI3,BFO_1849, BFO_1900,	BFO_0138,G8UKP6,G8UKP7,G8UKS1,G8ULI3,BFO_1849, BFO_1900,
sigma factor activity	1.6			BFO_1778,	BFO_1778,
precorrin-8X methylmutase activity	1.42			G8UMU8,	G8UMU8,
6-phosphogluconolactonase activity	1.25			G8UKW6,	G8UKW6,
transmembrane transporter activity	9.93			BFO_0135,BFO_0387,BFO_1586,BFO_2791,BFO_3117,	BFO_0135,BFO_0387,BFO_1586,BFO_2791,BFO_3117,
manganese ion binding	2.43			G8UMP6,G8UMY8,	G8UMP6,G8UMY8,
pyridoxal phosphate binding	3.54			G8UMM0,G8UNI9,BFO_1777,	G8UMM0,G8UNI9,BFO_1777,
carbohydrate binding	7.21			BFO_2272,BFO_0450,BFO_2947,	BFO_2272,BFO_0450,BFO_2947,
1-deoxy-D-xylulose-5-phosphate reductoisomerase activity	3.94			G8UQS9,	G8UQS9,
precorrin-3B C17-methyltransferase activity	1.42			G8UMU8,	G8UMU8,
cobalamin binding	4.77			BFO_2503,BFO_2747,G8UNZ5,	BFO_2503,BFO_2747,G8UNZ5,
energy transducer activity	1.16			G8UPY7,	G8UPY7,
pyridoxine 5'-phosphate synthase activity	1.7			G8UQZ3,	G8UQZ3,
UMP kinase activity	1.43			G8UPZ9,	G8UPZ9,

Gene Ontology for Cellular Component	sample3 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
NADH pyrophosphatase activity	1.04			G8UQC1,	G8UQC1,
methylthiotransferase activity	2.01			G8UMF5,	G8UMF5,
dTTP diphosphatase activity	1.04			G8UQC1,	G8UQC1,
UTP diphosphatase activity	1.04			G8UQC1,	G8UQC1,
cobyrinic acid a,c-diamide synthase activity	1.6			G8UQG4,	G8UQG4,
3'-tRNA processing endoribonuclease activity	1.52			G8UHU0,	G8UHU0,
xenobiotic transmembrane transporter activity	1.07			BFO_0228,	BFO_0228,
cation binding	1.88			BFO_0974,	BFO_0974,
sequence-specific DNA binding	8.1			BFO_0121,BFO_1304,BFO_2793,BFO_1922,	BFO_0121,BFO_1304,BFO_2793,BFO_1922,
precorrin-6Y C5,15-methyltransferase (decarboxylating) activity	1.19			G8UMU9,	G8UMU9,
2,3-bisphosphoglycerate-independent phosphoglycerate mutase activity	1.09			G8UMY8,	G8UMY8,
metal ion binding	48.47			BFO_0079,BFO_0086,BFO_1079,G8UHZ1,BFO_0136, BFO_0225,BFO_2219,G8UKI0,BFO_1424,BFO_1428,BFO_0474, G8UKY5,BFO_1487,G8UL71,BFO_0545,BFO_1635, BFO_2715,G8UMD6,G8UMF5,BFO_2747,G8UNM1,G8UPI8,BFO_0019, ⨅	BFO_0079,BFO_0086,BFO_1079,G8UHZ1,BFO_0136, BFO_0225,BFO_2219,G8UKI0,BFO_1424,BFO_1428,BFO_0474, G8UKY5,BFO_1487,G8UL71,BFO_0545,BFO_1635, BFO_2715,G8UMD6,G8UMF5,BFO_2747,G8UNM1,G8UPI8,BFO_0019, ⨅
protein dimerization activity	2			BFO_0561,	BFO_0561,
N-acetylgalactosamine-6-phosphate deacetylase activity	1.02			G8UL71,	G8UL71,
glyoxylate reductase (NAD+) activity	1.03			BFO_0938,	BFO_0938,
N-acylglucosamine 2-epimerase activity	1.68			BFO_2203,	BFO_2203,
S-adenosylhomocysteine deaminase activity	3.37			G8UMD6,	G8UMD6,

Gene Ontology for Cellular Component	sample3 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
vinylacetyl-CoA delta-isomerase activity	2.3			G8UL41,	G8UL41,
N-acetylneuraminate synthase activity	1			G8UHP9,	G8UHP9,
GDP-L-fucose synthase activity	1.26			G8UQC2,	G8UQC2,
flavin adenine dinucleotide binding	2.61			G8UL48,BFO_0019,	G8UL48,BFO_0019,
unfolded protein binding	2.27			G8UKP7,BFO_2546,	G8UKP7,BFO_2546,
NAD binding	5.19			BFO_0195,BFO_2943,BFO_0938,G8UQZ4,	BFO_0195,BFO_2943,BFO_0938,G8UQZ4,
iron-sulfur cluster binding	14.43			BFO_0086,BFO_1424,BFO_1485,BFO_0545,BFO_1635,BFO_0054,	BFO_0086,BFO_1424,BFO_1485,BFO_0545,BFO_1635,BFO_0054,
2 iron, 2 sulfur cluster binding	2.2			BFO_0474,BFO_0019,	BFO_0474,BFO_0019,
4 iron, 4 sulfur cluster binding	5.85			BFO_0079,BFO_0136,BFO_2715,G8UMF5,	BFO_0079,BFO_0136,BFO_2715,G8UMF5,
ADP-dependent NAD(P)H-hydrate dehydratase activity	2.16			G8UHZ1,	G8UHZ1,
NADHX epimerase activity	2.16			G8UHZ1,	G8UHZ1,
NADPHX epimerase activity	2.16			G8UHZ1,	G8UHZ1,
RNA polymerase binding	1.34			G8UQN3,	G8UQN3,
NADP+ binding	1.26			G8UQC2,	G8UQC2,
NADPH binding	3.94			G8UQS9,	G8UQS9,
5'-methylthioadenosine deaminase activity	3.37			G8UMD6,	G8UMD6,
1,4-alpha-glucan branching enzyme activity (using a glucosylated glycogenin as primer for glycogen synthesis)	1.88			BFO_0974,	BFO_0974,
tRNA pseudouridine synthase activity	1.16			G8UQG2,	G8UQG2,
protein serine kinase activity	1.78			BFO_2981,	BFO_2981,
rRNA pseudouridine synthase activity	1.58			BFO_0137,	BFO_0137,
ABC-type transporter activity	11.03			BFO_0082,BFO_0088,BFO_0102,	BFO_0082,BFO_0088,BFO_0102,

Appendix VIII: Table 8.5 Biological process for the significant downregulated genes in Neu5Ac.

Gene Ontology	sample4 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
tyrosyl-tRNA aminoacylation	1			G8UKY9,	G8UKY9,
tyrosine biosynthetic process	1			BFO_1679,	BFO_1679,
tRNA wobble uridine modification	1			G8UHT1,	G8UHT1,
tRNA threonylcarbam oyladenosine modification	1			G8UNZ9,	G8UNZ9,
tRNA modification	1			G8UMA1,	G8UMA1,
transposition, DNA-mediated	1			BFO_3048,	BFO_3048,
transmembrane transport	1			G8UMS8,	G8UMS8,
translation	37			G8UHW7,G8UHW9,G8UJE4,G8UJH7,G8UJL3,G8ULL4, G8ULL5,G8ULL6,G8ULL7,G8ULL8,G8ULL9,G8ULM1,G8ULM2, G8ULM3,G8ULM4,G8ULM5,G8ULM6,G8ULM7,G8ULM8,G8ULM9, G8ULN0,G8ULN1,G8ULN2,G8ULN3,G8ULN5,G8ULN9,G8ULP0, ⯆	G8UHW7,G8UHW9,G8UJE4,G8UJH7, G8UJL3,G8ULL4, G8ULL5,G8ULL6,G8ULL7,G8ULL8,G8ULL9,G8ULM1,G8ULM2, G8ULM3,G8ULM4,G8ULM5,G8ULM6, G8ULM7,G8ULM8,G8ULM9, G8ULN0,G8ULN1,G8ULN2,G8ULN3, G8ULN5,G8ULN9,G8ULP0, ⯆
transcription, DNA-templated	2			G8ULP3,G8UPQ3,	G8ULP3,G8UPQ3,
threonine biosynthetic process	1			BFO_0717,	BFO_0717,
tetrahydrofolate biosynthetic process	1			G8UNJ6,	G8UNJ6,
sodium ion transport	1			G8UI14,	G8UI14,
signal peptide processing	2			G8UK66,G8UNC2,	G8UK66,G8UNC2,
rRNA base methylation	1			G8UKG3,	G8UKG3,

Gene Ontology	sample4 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
rRNA 5'-end processing	1			BFO_3046,	BFO_3046,
ribosome biogenesis	1			G8UKZ0,	G8UKZ0,
ribosomal small subunit biogenesis	1			G8UKZ2,	G8UKZ2,
riboflavin biosynthetic process	1			G8ULZ9,	G8ULZ9,
rhamnose catabolic process	1			BFO_0217,	BFO_0217,
response to stimulus	2			BFO_0424,BFO_2618,	BFO_0424,BFO_2618,
regulation of translation	2			G8UPP5,G8UPQ0,	G8UPP5,G8UPQ0,
regulation of transcription, DNA-templated	1			BFO_0753,	BFO_0753,
regulation of DNA replication	1			G8UPW7,	G8UPW7,
regulation of DNA repair	1			G8UKP9,	G8UKP9,
queuosine biosynthetic process	1			BFO_1303,	BFO_1303,
putrescine biosynthetic process	1			BFO_2324,	BFO_2324,
purine nucleoside triphosphate catabolic process	1			G8UPX8,	G8UPX8,
protein transport by the Sec complex	1			G8ULN6,	G8ULN6,
protein transport	1			G8UMS0,	G8UMS0,
protein targeting	1			G8ULN6,	G8ULN6,
protein secretion	2			BFO_0843,G8UPV1,	BFO_0843,G8UPV1,
protein initiator methionine removal	1			G8ULN7,	G8ULN7,

Gene Ontology	sample4 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
protein folding	3			G8ULR6,G8UMI8,G8UN53,	G8ULR6,G8UMI8,G8UN53,
protein complex oligomerization	1			G8UNP5,	G8UNP5,
proline catabolic process	1			BFO_2738,	BFO_2738,
primary metabolic process	1			G8UPP5,	G8UPP5,
porphyrin-containing compound biosynthetic process	1			BFO_1580,	BFO_1580,
polyphosphate biosynthetic process	1			G8UQR9,	G8UQR9,
phosphorelay signal transduction system	1			BFO_2048,	BFO_2048,
phospholipid biosynthetic process	3			BFO_3357,BFO_0006,BFO_3139,	BFO_3357,BFO_0006,BFO_3139,
phosphate-containing compound metabolic process	1			G8UPS7,	G8UPS7,
phenylalanyl-tRNA aminoacylation	2			G8UN57,G8UP75,	G8UN57,G8UP75,
pentose-phosphate shunt	1			G8UKV1,	G8UKV1,
organic substance metabolic process	1			G8UK96,	G8UK96,
organic acid metabolic process	1			G8UK90,	G8UK90,
organic acid catabolic process	1			G8UKV1,	G8UKV1,
nucleotide metabolic process	1			G8UPX8,	G8UPX8,
nucleotide catabolic process	1			BFO_1096,	BFO_1096,

Gene Ontology	sample4 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
nucleoside metabolic process	2			BFO_2153,BFO_0294,	BFO_2153,BFO_0294,
nucleobase-containing small molecule biosynthetic process	1			G8UPX8,	G8UPX8,
nitrogen compound metabolic process	1			G8UHY5,	G8UHY5,
negative regulation of DNA recombination	1			G8UJ12,	G8UJ12,
NAD biosynthetic process	1			BFO_1025,	BFO_1025,
N-acetylmuramic acid catabolic process	1			BFO_0044,	BFO_0044,
N-acetylglucosamine metabolic process	1			BFO_3156,	BFO_3156,
mismatch repair	1			G8UJ12,	G8UJ12,
menaquinone biosynthetic process	1			G8UQ30,	G8UQ30,
lysyl-tRNA aminoacylation	1			G8UJH2,	G8UJH2,
lysine biosynthetic process via diaminopimelate	1			BFO_0717,	BFO_0717,
lipid metabolic process	1			BFO_2191,	BFO_2191,
lipid A biosynthetic process	2			G8UKA8,G8UKA9,	G8UKA8,G8UKA9,
leucine biosynthetic process	4			G8UNX8,G8UNY0,BFO_0823,G8UNY2,	G8UNX8,G8UNY0,BFO_0823,G8UNY2,
L-phenylalanine biosynthetic process	1			BFO_1682,	BFO_1682,

Gene Ontology	sample4 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
L-arabinose metabolic process	1			BFO_3112,	BFO_3112,
iron-sulfur cluster assembly	1			G8UMN1,	G8UMN1,
intracellular protein transmembrane transport	1			G8ULN6,	G8ULN6,
hexose metabolic process	1			G8UNH5,	G8UNH5,
glycolytic process	1			G8UJH4,	G8UJH4,
glycerol-3-phosphate catabolic process	1			BFO_3357,	BFO_3357,
glutamine metabolic process	2			BFO_0129,G8UMS1,	BFO_0129,G8UMS1,
gluconeogenesis	1			G8UJH4,	G8UJH4,
fucose metabolic process	1			BFO_2737,	BFO_2737,
FtsZ-dependent cytokinesis	1			G8UKF4,	G8UKF4,
fructose 6-phosphate metabolic process	2			G8UHS7,G8UIG9,	G8UHS7,G8UIG9,
folic acid biosynthetic process	1			G8UNJ6,	G8UNJ6,
fatty acid biosynthetic process	2			G8UIM6,G8UKA9,	G8UIM6,G8UKA9,
electron transport chain	3			G8UMS8,G8UMT0,G8UMT1,	G8UMS8,G8UMT0,G8UMT1,
dUTP catabolic process	1			G8ULI7,	G8ULI7,
dUMP biosynthetic process	1			G8ULI7,	G8ULI7,
dTDP-rhamnose biosynthetic process	1			G8UHZ9,	G8UHZ9,

Gene Ontology	sample4 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
DNA-templated transcription, initiation	3			BFO_0424,BFO_2618, BFO_2011,	BFO_0424,BFO_2618,BFO_2011,
DNA unwinding involved in DNA replication	1			G8UR29,	G8UR29,
DNA replication, synthesis of RNA primer	1			G8UR29,	G8UR29,
DNA replication initiation	1			G8UPW7,	G8UPW7,
DNA replication	1			BFO_2081,	BFO_2081,
DNA recombination	2			BFO_0679,BFO_1958,	BFO_0679,BFO_1958,
DNA integration	2			BFO_0679,BFO_1958,	BFO_0679,BFO_1958,
defense response to virus	1			BFO_0208,	BFO_0208,
D-gluconate metabolic process	1			G8UKV1,	G8UKV1,
D-amino acid catabolic process	1			G8UHT3,	G8UHT3,
D-alanine biosynthetic process	1			G8ULP9,	G8ULP9,
cytokinin biosynthetic process	1			BFO_2938,	BFO_2938,
cobalamin biosynthetic process	1			BFO_0479,	BFO_0479,
chorismate biosynthetic process	1			G8UN73,	G8UN73,
cellular response to heat	1			BFO_3089,	BFO_3089,
cellular biosynthetic process	1			BFO_3201,	BFO_3201,
cellular amino acid metabolic process	1			BFO_0184,	BFO_0184,

Gene Ontology	sample4 Enrichment Score		Log2 Average Fold Change	Common Genes Involved	All Genes Involved
cellular amino acid catabolic process	1			BFO_1954,	BFO_1954,
cellular amino acid biosynthetic process	1			G8UN73,	G8UN73,
cell division	3			BFO_2745,BFO_0719, BFO_1762,	BFO_2745,BFO_0719,BFO_1762,
cell cycle	1			BFO_1762,	BFO_1762,
carbohydrate metabolic process	5			BFO_2208,BFO_3357, BFO_0676,BFO_0044, BFO_3156,	BFO_2208,BFO_3357,BFO_0676,BFO_ 0044,BFO_3156,
carbohydrate catabolic process	1			BFO_0300,	BFO_0300,
biosynthetic process	3			BFO_1681,BFO_3031, BFO_2075,	BFO_1681,BFO_3031,BFO_2075,
base-excision repair	1			G8UNV4,	G8UNV4,
aromatic amino acid family biosynthetic process	1			G8UN73,	G8UN73,
arginine catabolic process	1			BFO_2923,	BFO_2923,
arginine biosynthetic process	2			BFO_1814,G8UNP5,	BFO_1814,G8UNP5,
amino sugar catabolic process	1			BFO_0044,	BFO_0044,
acetyl-CoA biosynthetic process	1			G8UK90,	G8UK90,
'de novo' IMP biosynthetic process	1			G8UPB5,	G8UPB5,
'de novo' CTP biosynthetic process	1			G8UMS1,	G8UMS1,

Appendix VIII: Table 8.6 Molecular function for the significant downregulated genes in Neu5Ac.

Gene Ontology	sample4 Enrichment Score	Log2 Average Fold Change	Common Genes Involved	All Genes Involved
tRNA binding	8		G8UHT3,G8ULL4,G8ULM3, G8ULM8,G8ULP0,G8UN57, G8UP75,G8UPQ0,	G8UHT3,G8ULL4,G8ULM3,G8UL M8,G8ULP0,G8UN57, G8UP75,G8UPQ0,
nucleotide binding	2		BFO_1096,G8UPX8,	BFO_1096,G8UPX8,
magnesium ion binding	10		BFO_1954,G8UJH2,G8UK90 ,G8ULI7,G8UN57,G8UN73, G8UNY2,G8UP75,G8UPS7, G8UQ30,	BFO_1954,G8UJH2,G8UK90,G8UL I7,G8UN57,G8UN73, G8UNY2,G8UP75,G8UPS7,G8UQ3 0,
purine-specific mismatch base pair DNA N-glycosylase activity	1		G8UNV4,	G8UNV4,
6,7-dimethyl-8-ribityllumazine synthase activity	1		G8ULZ9,	G8ULZ9,
nucleic acid binding	4		BFO_0209,BFO_3340,G8UJ H2,BFO_0272,	BFO_0209,BFO_3340,G8UJH2,BF O_0272,
DNA binding	17		BFO_2133,BFO_0424,BFO_ 0531,G8ULP3,BFO_2618, BFO_0679,BFO_0753,G8UN P5,G8UNV2,G8UNV4,G8UP Q3, BFO_1958,BFO_3048,BFO_ 3089,BFO_2011,BFO_2048, G8UR29,	BFO_2133,BFO_0424,BFO_0531,G 8ULP3,BFO_2618, BFO_0679,BFO_0753,G8UNP5,G8 UNV2,G8UNV4,G8UPQ3, BFO_1958,BFO_3048,BFO_3089,B FO_2011,BFO_2048,G8UR29,
DNA helicase activity	1		G8UR29,	G8UR29,
DNA replication origin binding	1		G8UPW7,	G8UPW7,
DNA-binding transcription factor activity	3		BFO_1752,G8UNP5,BFO_19 03,	BFO_1752,G8UNP5,BFO_1903,
RNA binding	2		G8UKK1,G8UKY9,	G8UKK1,G8UKY9,
double-stranded RNA binding	1		BFO_1804,	BFO_1804,
single-stranded RNA binding	1		BFO_3089,	BFO_3089,
mRNA binding	1		G8ULM2,	G8ULM2,
structural constituent of ribosome	34		G8UHW7,G8UHW9,G8UJE4 ,G8UJH7,G8UJL3,G8ULL4, G8ULL5,G8ULL6,G8ULL7, G8ULL8,G8ULL9,G8ULM1, G8ULM2, G8ULM3,G8ULM4,G8ULM 5,G8ULM6,G8ULM7,G8UL M8,G8ULM9, G8ULN0,G8ULN1,G8ULN2, G8ULN3,G8ULN5,G8ULN9, G8ULP0,⊻	G8UHW7,G8UHW9,G8UJE4,G8UJ H7,G8UJL3,G8ULL4, G8ULL5,G8ULL6,G8ULL7,G8ULL 8,G8ULL9,G8ULM1,G8ULM2, G8ULM3,G8ULM4,G8ULM5,G8U LM6,G8ULM7,G8ULM8,G8ULM9, G8ULN0,G8ULN1,G8ULN2,G8UL N3,G8ULN5,G8ULN9,G8ULP0,⊻

Gene Ontology	sample4 Enrichment Score	Log2 Average Fold Change	Common Genes Involved	All Genes Involved
translation initiation factor activity	2		G8ULN8,G8UMM7,	G8ULN8,G8UMM7,
translation elongation factor activity	2		G8UHX0,G8ULL3,	G8UHX0,G8ULL3,
peptidyl-prolyl cis-trans isomerase activity	1		G8ULR6,	G8ULR6,
1-acylglycerol-3-phosphate O-acyltransferase activity	1		BFO_0006,	BFO_0006,
2-amino-4-hydroxy-6-hydroxymethyldihydropteridine diphosphokinase activity	1		G8UNJ6,	G8UNJ6,
2-isopropylmalate synthase activity	1		BFO_0823,	BFO_0823,
3-isopropylmalate dehydratase activity	2		G8UNX8,G8UNY0,	G8UNX8,G8UNY0,
3-isopropylmalate dehydrogenase activity	1		G8UNY2,	G8UNY2,
6-phosphofructokinase activity	2		G8UHS7,G8UIG9,	G8UHS7,G8UIG9,
CTP synthase activity	1		G8UMS1,	G8UMS1,
DNA-directed 5'-3' RNA polymerase activity	2		G8ULP3,G8UPQ3,	G8ULP3,G8UPQ3,
GTPase activity	4		G8UKZ2,G8ULL3,G8UMM7,G8UPL4,	G8UKZ2,G8ULL3,G8UMM7,G8UPL4,
aldose 1-epimerase activity	1		G8UNH5,	G8UNH5,
amidase activity	1		BFO_1375,	BFO_1375,
aminoacyl-tRNA hydrolase activity	1		G8UQC8,	G8UQC8,
argininosuccinate synthase activity	1		BFO_1814,	BFO_1814,
asparaginase activity	1		BFO_0184,	BFO_0184,
aspartate kinase activity	1		BFO_0717,	BFO_0717,
carbamoyl-phosphate synthase (glutamine-hydrolyzing) activity	1		G8UHY5,	G8UHY5,
chorismate mutase activity	1		BFO_1682,	BFO_1682,
coproporphyrinogen oxidase activity	1		BFO_1580,	BFO_1580,
dUTP diphosphatase activity	1		G8ULI7,	G8ULI7,
ATP-dependent peptidase activity	1		G8UNC2,	G8UNC2,

Gene Ontology	sample4 Enrichment Score	Log2 Average Fold Change	Common Genes Involved	All Genes Involved
aminopeptidase activity	1		BFO_1818,	BFO_1818,
metalloendopeptidase activity	3		BFO_1168,BFO_0441,BFO_0011,	BFO_1168,BFO_0441,BFO_0011,
serine-type endopeptidase activity	4		G8UK66,BFO_2770,BFO_2771,G8UNC2,	G8UK66,BFO_2770,BFO_2771,G8UNC2,
exo-alpha-sialidase activity	1		BFO_2207,	BFO_2207,
3-oxoacyl-[acyl-carrier-protein] synthase activity	1		G8UIM6,	G8UIM6,
3-hydroxypalmitoyl-[acyl-carrier-protein] dehydratase activity	1		G8UKA9,	G8UKA9,
glucosamine-6-phosphate deaminase activity	1		BFO_3156,	BFO_3156,
glucose-6-phosphate isomerase activity	1		G8UJH4,	G8UJH4,
glycerol-3-phosphate dehydrogenase [NAD+] activity	1		BFO_3357,	BFO_3357,
inorganic diphosphatase activity	1		G8UPS7,	G8UPS7,
nuclease activity	3		BFO_0209,BFO_0272,BFO_3046,	BFO_0209,BFO_0272,BFO_3046,
endonuclease activity	2		G8UJ12,BFO_0433,	G8UJ12,BFO_0433,
ribonuclease activity	1		BFO_3332,	BFO_3332,
alpha-L-fucosidase activity	1		BFO_2737,	BFO_2737,
beta-N-acetylhexosaminidase activity	1		BFO_2208,	BFO_2208,
beta-galactosidase activity	1		BFO_0300,	BFO_0300,
peroxidase activity	1		BFO_0478,	BFO_0478,
phosphogluconate dehydrogenase (decarboxylating) activity	1		G8UKV1,	G8UKV1,
phospholipase A2 activity	1		BFO_2191,	BFO_2191,
phosphoribosylaminoimidazolesuccinocarboxamide synthase activity	1		G8UPB5,	G8UPB5,
proline dehydrogenase activity	1		BFO_2738,	BFO_2738,
prephenate dehydratase activity	1		BFO_1682,	BFO_1682,

Gene Ontology	sample4 Enrichment Score	Log2 Average Fold Change	Common Genes Involved	All Genes Involved
prephenate dehydrogenase (NADP+) activity	1		BFO_1679,	BFO_1679,
protein-arginine deiminase activity	1		BFO_2324,	BFO_2324,
purine-nucleoside phosphorylase activity	1		BFO_0294,	BFO_0294,
ribonucleoside-diphosphate reductase activity, thioredoxin disulfide as acceptor	1		BFO_2081,	BFO_2081,
shikimate kinase activity	1		G8UN73,	G8UN73,
transposase activity	1		BFO_3048,	BFO_3048,
lysine-tRNA ligase activity	1		G8UJH2,	G8UJH2,
phenylalanine-tRNA ligase activity	2		G8UN57,G8UP75,	G8UN57,G8UP75,
tyrosine-tRNA ligase activity	1		G8UKY9,	G8UKY9,
endopeptidase inhibitor activity	1		BFO_0469,	BFO_0469,
iron ion binding	1		G8UNZ9,	G8UNZ9,
ATP binding	30		G8UHS7,BFO_2081,BFO_2098,G8UHY5,G8UIG9,BFO_1166, G8UJ12,G8UJH2,G8UK90,G8UKM4,BFO_0479,G8UKY9, G8ULP9,BFO_1575,G8UMA1,G8UMI8,G8UMN0,G8UMS1, BFO_0717,G8UN53,G8UN57,G8UN73,G8UNJ6,BFO_1814,G8UP75, ⊻	G8UHS7,BFO_2081,BFO_2098,G8UHY5,G8UIG9,BFO_1166, G8UJ12,G8UJH2,G8UK90,G8UKM4,BFO_0479,G8UKY9, G8ULP9,BFO_1575,G8UMA1,G8UMI8,G8UMN0,G8UMS1, BFO_0717,G8UN53,G8UN57,G8UN73,G8UNJ6,BFO_1814,G8UP75, ⊻
GTP binding	7		G8UKM4,BFO_0479,G8UKZ0,G8UKZ2,G8ULL3,G8UMM7, G8UPL4,	G8UKM4,BFO_0479,G8UKZ0,G8UKZ2,G8ULL3,G8UMM7, G8UPL4,
5S rRNA binding	1		G8UQC7,	G8UQC7,
N-methyltransferase activity	1		BFO_2133,	BFO_2133,
peptidase activity	1		G8UNZ9,	G8UNZ9,
serine-type peptidase activity	1		BFO_0222,	BFO_0222,
zinc ion binding	2		BFO_2935,G8UR29,	BFO_2935,G8UR29,
cation transmembrane transporter activity	1		BFO_2524,	BFO_2524,
3'-5' exonuclease activity	1		BFO_3340,	BFO_3340,

Gene Ontology	sample4 Enrichment Score	Log2 Average Fold Change	Common Genes Involved	All Genes Involved
transaminase activity	1		BFO_1681,	BFO_1681,
sulfuric ester hydrolase activity	1		BFO_0762,	BFO_0762,
penicillin binding	1		BFO_0406,	BFO_0406,
(3R)-hydroxymyristoyl-[acyl-carrier-protein] dehydratase activity	1		G8UKA9,	G8UKA9,
3-hydroxydecanoyl-[acyl-carrier-protein] dehydratase activity	1		G8UKA9,	G8UKA9,
L-aspartate oxidase activity	1		BFO_1025,	BFO_1025,
UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase activity	1		G8UKA9,	G8UKA9,
UDP-N-acetylglucosamine 2-epimerase activity	1		BFO_2566,	BFO_2566,
acetate kinase activity	1		G8UK90,	G8UK90,
acyl-[acyl-carrier-protein]-UDP-N-acetylglucosamine O-acyltransferase activity	1		G8UKA8,	G8UKA8,
alanine racemase activity	1		G8ULP9,	G8ULP9,
arginine decarboxylase activity	1		BFO_2923,	BFO_2923,
cobinamide phosphate guanylyltransferase activity	1		BFO_0479,	BFO_0479,
dTDP-4-dehydrorhamnose 3,5-epimerase activity	1		G8UHZ9,	G8UHZ9,
isochorismate synthase activity	1		BFO_3201,	BFO_3201,
phospholipase A1 activity	1		BFO_2191,	BFO_2191,
polyphosphate kinase activity	1		G8UQR9,	G8UQR9,
prephenate dehydrogenase (NAD+) activity	1		BFO_1679,	BFO_1679,
rhamnulokinase activity	1		BFO_0217,	BFO_0217,
electron transfer activity	4		G8UJI1,BFO_1638,G8UMS7,G8UMS9,	G8UJI1,BFO_1638,G8UMS7,G8UMS9,
FMN binding	1		G8UMS9,	G8UMS9,
proton transmembrane transporter activity	1		BFO_0554,	BFO_0554,

Gene Ontology	sample4 Enrichment Score	Log2 Average Fold Change	Common Genes Involved	All Genes Involved
potassium ion transmembrane transporter activity	1		BFO_0518,	BFO_0518,
lipopolysaccharide transmembrane transporter activity	1		BFO_3037,	BFO_3037,
antiporter activity	2		BFO_0443,BFO_2848,	BFO_0443,BFO_2848,
protein-transporting ATPase activity	1		G8UPV1,	G8UPV1,
efflux transmembrane transporter activity	1		BFO_0459,	BFO_0459,
site-specific DNA-methyltransferase (cytosine-N4-specific) activity	1		BFO_2133,	BFO_2133,
antioxidant activity	1		BFO_1034,	BFO_1034,
kinase activity	2		G8UNJ6,BFO_2059,	G8UNJ6,BFO_2059,
oxidoreductase activity	2		BFO_3030,BFO_1034,	BFO_3030,BFO_1034,
oxidoreductase activity, acting on NAD(P)H, quinone or similar compound as acceptor	1		G8UI14,	G8UI14,
transferase activity	5		BFO_0129,BFO_1303,G8ULL8,BFO_1803,BFO_1933,	BFO_0129,BFO_1303,G8ULL8,BFO_1803,BFO_1933,
acyltransferase activity, transferring groups other than amino-acyl groups	1		BFO_1314,	BFO_1314,
glycosyltransferase activity	3		BFO_2153,BFO_0320,BFO_0036,	BFO_2153,BFO_0320,BFO_0036,
hexosyltransferase activity	1		BFO_0330,	BFO_0330,
phosphotransferase activity, alcohol group as acceptor	1		BFO_2059,	BFO_2059,
nucleotidyltransferase activity	1		BFO_3031,	BFO_3031,
phosphotransferase activity, for other substituted phosphate groups	1		BFO_3139,	BFO_3139,
hydrolase activity	8		BFO_1096,BFO_1313,G8UK96,BFO_1470,BFO_0721,BFO_2938,BFO_3140,G8UR29,	BFO_1096,BFO_1313,G8UK96,BFO_1470,BFO_0721,BFO_2938,BFO_3140,G8UR29,
lyase activity	2		BFO_2202,BFO_0721,	BFO_2202,BFO_0721,
carbon-oxygen lyase activity	1		BFO_0044,	BFO_0044,
ammonia-lyase activity	1		BFO_0733,	BFO_0733,

419

Gene Ontology	sample4 Enrichment Score	Log2 Average Fold Change	Common Genes Involved	All Genes Involved
isomerase activity	1		BFO_1303,	BFO_1303,
racemase and epimerase activity, acting on carbohydrates and derivatives	1		BFO_2526,	BFO_2526,
ligase activity	1		G8ULZ9,	G8ULZ9,
ligase activity, forming carbon-nitrogen bonds	2		G8ULP9,G8UMA1,	G8ULP9,G8UMA1,
ATP hydrolysis activity	5		BFO_1166,G8UJ12,G8UKM4,G8UMI8,G8UN53,	BFO_1166,G8UJ12,G8UKM4,G8UMI8,G8UN53,
sigma factor activity	3		BFO_0424,BFO_2618,BFO_2011,	BFO_0424,BFO_2618,BFO_2011,
nucleoside-triphosphatase activity	1		G8UPX8,	G8UPX8,
nitronate monooxygenase activity	1		BFO_2519,	BFO_2519,
rRNA binding	24		G8UJH7,G8ULL5,G8ULL6,G8ULL7,G8ULL8,G8ULL9,G8ULM1,G8ULM2,G8ULM3,G8ULM5,G8ULM6,G8ULM7,G8ULM8,G8ULM9,G8ULN0,G8ULN1,G8ULN2,G8ULN3,G8ULN5,G8ULN8,G8ULP0,G8ULP1,G8ULP2,G8UPQ0,	G8UJH7,G8ULL5,G8ULL6,G8ULL7,G8ULL8,G8ULL9,G8ULM1,G8ULM2,G8ULM3,G8ULM5,G8ULM6,G8ULM7,G8ULM8,G8ULM9,G8ULN0,G8ULN1,G8ULN2,G8ULN3,G8ULN5,G8ULN8,G8ULP0,G8ULP1,G8ULP2,G8UPQ0,
transmembrane transporter activity	4		BFO_2520,BFO_2790,BFO_2040,BFO_3142,	BFO_2520,BFO_2790,BFO_2040,BFO_3142,
manganese ion binding	1		G8UQ30,	G8UQ30,
pyridoxal phosphate binding	2		G8ULP9,BFO_1681,	G8ULP9,BFO_1681,
carbohydrate binding	2		BFO_0300,G8UNH5,	BFO_0300,G8UNH5,
thiamine pyrophosphate binding	1		G8UQ30,	G8UQ30,
mismatched DNA binding	1		G8UJ12,	G8UJ12,
ribonucleoside binding	1		G8UPQ3,	G8UPQ3,
membrane insertase activity	1		G8UMS0,	G8UMS0,
beta-ketoacyl-acyl-carrier-protein synthase III activity	1		G8UIM6,	G8UIM6,
arginine binding	1		G8UNP5,	G8UNP5,
dITP diphosphatase activity	1		G8UPX8,	G8UPX8,
ITP diphosphatase activity	1		G8UPX8,	G8UPX8,

Gene Ontology	sample4 Enrichment Score	Log2 Average Fold Change	Common Genes Involved	All Genes Involved
XTP diphosphatase activity	1		G8UPX8,	G8UPX8,
adenosylcobinamide kinase (GTP-specific) activity	1		BFO_0479,	BFO_0479,
adenosylcobinamide kinase (ATP-specific) activity	1		BFO_0479,	BFO_0479,
peptide deformylase activity	1		G8UQ87,	G8UQ87,
peptidoglycan binding	2		BFO_2745,BFO_0719,	BFO_2745,BFO_0719,
xenobiotic transmembrane transporter activity	2		BFO_0443,BFO_2848,	BFO_0443,BFO_2848,
ribosome binding	2		G8UKM4,G8ULN8,	G8UKM4,G8ULN8,
ribosomal large subunit binding	2		G8UKM4,BFO_3089,	G8UKM4,BFO_3089,
[formate-C-acetyltransferase]-activating enzyme activity	1		G8UQN0,	G8UQN0,
sequence-specific DNA binding	1		BFO_1903,	BFO_1903,
Ser(Gly)-tRNA(Ala) hydrolase activity	1		G8UHT3,	G8UHT3,
alpha-L-arabinofuranosidase activity	1		BFO_3112,	BFO_3112,
metal ion binding	20		G8UHS7,G8UHY5,G8UIG9, BFO_1168,BFO_2191,G8UKA9, BFO_0441,BFO_1580,BFO_2695,G8UMS1,G8UMS7,G8UNV4, G8UNX8,G8UPL4,G8UPW0, G8UPX8,BFO_0985,G8UQ87, BFO_0011,G8UQR9,	G8UHS7,G8UHY5,G8UIG9,BFO_1168,BFO_2191,G8UKA9, BFO_0441,BFO_1580,BFO_2695,G8UMS1,G8UMS7,G8UNV4, G8UNX8,G8UPL4,G8UPW0,G8UPX8,BFO_0985,G8UQ87, BFO_0011,G8UQR9,
transition metal ion binding	1		G8ULN7,	G8ULN7,
protein dimerization activity	1		G8ULP3,	G8ULP3,
diphosphate-fructose-6-phosphate 1-phosphotransferase activity	1		G8UHS7,	G8UHS7,
3-hydroxyoctanoyl-[acyl-carrier-protein] dehydratase activity	1		G8UKA9,	G8UKA9,

421

Gene Ontology	sample4 Enrichment Score	Log2 Average Fold Change	Common Genes Involved	All Genes Involved
4-hydroxythreonine-4-phosphate dehydrogenase activity	1		G8UPW0,	G8UPW0,
flavin adenine dinucleotide binding	2		G8UHT1,BFO_1741,	G8UHT1,BFO_1741,
NADP binding	1		G8UKV1,	G8UKV1,
unfolded protein binding	2		G8UMI8,G8UN53,	G8UMI8,G8UN53,
NAD binding	3		BFO_3357,G8UNY2,G8UPW0,	BFO_3357,G8UNY2,G8UPW0,
D-aminoacyl-tRNA deacylase activity	1		G8UHT3,	G8UHT3,
2 iron, 2 sulfur cluster binding	1		BFO_1638,	BFO_1638,
4 iron, 4 sulfur cluster binding	5		BFO_1580,BFO_1638,G8UMS7,G8UNV4,G8UNX8,	BFO_1580,BFO_1638,G8UMS7,G8UNV4,G8UNX8,
raffinose alpha-galactosidase activity	1		BFO_2083,	BFO_2083,
phosphatidylserine 1-acylhydrolase activity	1		BFO_2191,	BFO_2191,
1-acyl-2-lysophosphatidylserine acylhydrolase activity	1		BFO_2191,	BFO_2191,
16S rRNA (adenine(1518)-N(6)/adenine(1519)-N(6))-dimethyltransferase activity	1		G8UKK1,	G8UKK1,
N(6)-L-threonylcarbamoyladenine synthase activity	1		G8UNZ9,	G8UNZ9,
metalloaminopeptidase activity	1		G8ULN7,	G8ULN7,
large ribosomal subunit rRNA binding	1		G8UPQ1,	G8UPQ1,
small ribosomal subunit rRNA binding	1		G8UKZ2,	G8UKZ2,
2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylic-acid synthase activity	1		G8UQ30,	G8UQ30,
rRNA (cytosine-N4-)-methyltransferase activity	1		G8UKG3,	G8UKG3,
carbohydrate derivative binding	1		BFO_0044,	BFO_0044,

Gene Ontology	sample4 Enrichment Score	Log2 Average Fold Change	Common Genes Involved	All Genes Involved
N-acetyl-beta-D-galactosaminidase activity	1		BFO_2208,	BFO_2208,
UDP-3-O-acyl-N-acetylglucosamine deacetylase activity	1		G8UKA9,	G8UKA9,
Gly-tRNA(Ala) hydrolase activity	1		G8UHT3,	G8UHT3,

Appendix VIII: Table 8.7. Cellular component for the significant downregulated genes in Neu5Ac.

Gene Ontology	sample 4 Enrichment Score	Log2 Average Fold Change	Common Genes Involved	All Genes Involved
extracellular region	2		BFO_1168,BFO_1853,	BFO_1168,BFO_1853,
cytoplasm	48		G8UHS7,G8UHT1,G8UHT3,G8UHX0,G8UIG9,G8UJH2, G8UJH4,G8UJH7,G8UJL3,BFO_1303,G8UK90,G8UKA9,G8UKG3, G8UKK1,G8UKP9,G8UKY9,G8UKZ2,G8ULL3,G8ULL4,G8ULL6, G8ULL7,G8ULM2,G8ULM3,G8ULM6,G8ULM8,G8ULM9,G8ULN2,⮟	G8UHS7,G8UHT1,G8UHT3,G8UHX0,G8UIG9,G8UJH2, G8UJH4,G8UJH7,G8UJL3,BFO_1303,G8UK90,G8UKA9,G8UKG3, G8UKK1,G8UKP9,G8UKY9,G8UKZ2,G8ULL3,G8ULL4,G8ULL6, G8ULL7,G8ULM2,G8ULM3,G8ULM6,G8ULM8,G8ULM9,G8ULN2,⮟
ribosome	26		G8UHW7,G8UJE4,G8UJH7,G8UJL3,G8ULL4,G8ULL5, G8ULL6,G8ULL7,G8ULM2,G8ULM3,G8ULM4,G8ULM5,G8ULM7, G8ULM8,G8ULM9,G8ULN0,G8ULN1,G8ULN2,G8ULN9,G8ULP0, G8ULP1,G8UM17,G8UP83,G8UPQ1,G8UPQ2,G8UQC7,	G8UHW7,G8UJE4,G8UJH7,G8UJL3,G8ULL4,G8ULL5, G8ULL6,G8ULL7,G8ULM2,G8ULM3,G8ULM4,G8ULM5,G8ULM7, G8ULM8,G8ULM9,G8ULN0,G8ULN1,G8ULN2,G8ULN9,G8ULP0, G8ULP1,G8UM17,G8UP83,G8UPQ1,G8UPQ2,G8UQC7,
plasma membrane	16		G8UI14,BFO_0151,BFO_0330,G8UKZ2,BFO_0518,G8ULN6, BFO_0649,G8UMS0,G8UMS7,G8UMS8,G8UMS9,G8UMT0, G8UMT1,BFO_1762,G8UPV1,BFO_0023,	G8UI14,BFO_0151,BFO_0330,G8UKZ2,BFO_0518,G8ULN6, BFO_0649,G8UMS0,G8UMS7,G8UMS8,G8UMS9,G8UMT0, G8UMT1,BFO_1762,G8UPV1,BFO_0023,
integral component of plasma membrane	2		BFO_0843,BFO_3037,	BFO_0843,BFO_3037,
cell outer membrane	16		BFO_2191,BFO_2205,BFO_2206,BFO_1202,BFO_1299, BFO_0395,BFO_1406,BFO_1407,BFO_1412,BFO_0521,	BFO_2191,BFO_2205,BFO_2206,BFO_1202,BFO_1299, BFO_0395,BFO_1406,BFO_1407,BFO_1412,BFO_0521,

Gene Ontology	sample 4 Enrichment Score	Log2 Average Fold Change	Common Genes Involved	All Genes Involved
			BFO_2639,BFO_2812,BFO_2813,BFO_2914,BFO_2076,BFO_2077,	BFO_2639,BFO_2812,BFO_2813,BFO_2914,BFO_2076,BFO_2077,
3-isopropylmalate dehydratase complex	1		G8UNY0,	G8UNY0,
glycerol-3-phosphate dehydrogenase complex	1		BFO_3357,	BFO_3357,
beta-galactosidase complex	1		BFO_0300,	BFO_0300,
riboflavin synthase complex	1		G8ULZ9,	G8ULZ9,
polyphosphate kinase complex	1		G8UQR9,	G8UQR9,
cytoplasmic side of plasma membrane	1		G8UKF4,	G8UKF4,
large ribosomal subunit	5		G8ULL8,G8ULM1,G8ULM6,G8ULN5,G8UPQ0,	G8ULL8,G8ULM1,G8ULM6,G8ULN5,G8UPQ0,
small ribosomal subunit	4		G8UHW9,G8ULL9,G8ULN3,G8ULP2,	G8UHW9,G8ULL9,G8ULN3,G8ULP2,
membrane	1		BFO_3142,	BFO_3142,
integral component of membrane	53		BFO_2099,BFO_2113,G8UI14,BFO_0151,BFO_1162, BFO_1163,BFO_1167,BFO_2164,BFO_2191,BFO_3288, BFO_2211,BFO_0330,G8UK66,BFO_0406,BFO_1396,BFO_1415, BFO_2520,BFO_2524,BFO_0443,BFO_0542,BFO_0554,G8ULN6,⌄	BFO_2099,BFO_2113,G8UI14,BFO_0151,BFO_1162, BFO_1163,BFO_1167,BFO_2164,BFO_2191,BFO_3288, BFO_2211,BFO_0330,G8UK66,BFO_0406,BFO_1396,BFO_1415, BFO_2520,BFO_2524,BFO_0443,BFO_0542,BFO_0554,G8ULN6,⌄
cell division site	1		G8UKF4,	G8UKF4,
proton-transporting V-type	1		BFO_0554,	BFO_0554,

Gene Ontology	sample 4 Enrichment Score	Log2 Average Fold Change	Common Genes Involved	All Genes Involved
ATPase, V0 domain				
ATP-binding cassette (ABC) transporter complex	1		BFO_1576,	BFO_1576,
cell periphery	1		BFO_2790,	BFO_2790,
toxin-antitoxin complex	1		BFO_3332,	BFO_3332,
cellular anatomical entity	1		BFO_3112,	BFO_3112,
primosome complex	1		G8UR29,	G8UR29,

Appendix VIII: Table 8.8. Biological process for the significant upregulated genes in Mucin.

Gene Ontology	sample5 Enrichment Score	Log 2 Average Fold Change	Common Genes Involved	All Genes Involved
polysaccharide biosynthetic process	1		BFO_1051,	BFO_1051,
tRNA threonylcarbamoyladenosine modification	1		G8UNZ9,	G8UNZ9,
carbohydrate metabolic process	3		BFO_2208,BFO_0676,BFO_0044,	BFO_2208,BFO_0676,BFO_0044,
fucose metabolic process	1		BFO_2737,	BFO_2737,
DNA replication	1		G8UL57,	G8UL57,
DNA topological change	1		BFO_1082,	BFO_1082,
DNA repair	1		G8UL57,	G8UL57,
base-excision repair	1		G8UN58,	G8UN58,
mismatch repair	1		G8UJ12,	G8UJ12,
DNA recombination	1		G8UL57,	G8UL57,
transcription, DNA-templated	1		G8ULP3,	G8ULP3,
tRNA modification	1		G8UMA1,	G8UMA1,
translation	27		G8UJH8,G8ULL4,G8ULL5,G8ULL6,G8ULL7,G8ULL8, G8ULL9,G8ULM1,G8ULM2,G8ULM3, G8ULM4,G8ULM5,G8ULM6, G8ULM7,G8ULM8,G8ULM9,G8ULN0, G8ULN1,G8ULN2,G8ULN3, G8ULN5,G8ULN9,G8ULP0,G8ULP1,G8ULP2,G8UM17,G8UQ87,	G8UJH8,G8ULL4,G8ULL5,G8ULL6,G8ULL7,G8ULL8, G8ULL9,G8ULM1,G8ULM2,G8ULM3, G8ULM4,G8ULM5,G8ULM6, G8ULM7,G8ULM8,G8ULM9,G8ULN0, G8ULN1,G8ULN2,G8ULN3, G8ULN5,G8ULN9,G8ULP0,G8ULP1,G8ULP2,G8UM17,G8UQ87,
phenylalanyl-tRNA aminoacylation	1		G8UN57,	G8UN57,
protein folding	2		G8ULR6,G8UMI8,	G8ULR6,G8UMI8,
signal peptide processing	1		G8UNC2,	G8UNC2,

Gene Ontology	sample5 Enrichment Score	Log 2 Average Fold Change	Common Genes Involved	All Genes Involved
glutamine metabolic process	1		G8UMS1,	G8UMS1,
proline catabolic process	1		BFO_2738,	BFO_2738,
protein targeting	1		G8ULN6,	G8ULN6,
lipid metabolic process	3		BFO_2321,BFO_0550,BFO_1591,	BFO_2321,BFO_0550,BFO_1591,
fatty acid biosynthetic process	1		BFO_1782,	BFO_1782,
ion transport	1		BFO_0555,	BFO_0555,
sodium ion transport	4		G8UI14,G8UI15,G8UI16,G8UI17,	G8UI14,G8UI15,G8UI16,G8UI17,
queuosine biosynthetic process	1		BFO_1303,	BFO_1303,
biosynthetic process	2		BFO_1681,G8UPR7,	BFO_1681,G8UPR7,
L-phenylalanine biosynthetic process	1		BFO_1682,	BFO_1682,
leucine biosynthetic process	3		G8UNY0,BFO_0823,G8UNY2,	G8UNY0,BFO_0823,G8UNY2,
nucleoside metabolic process	1		BFO_2784,	BFO_2784,
menaquinone biosynthetic process	2		G8UQ30,G8UQY3,	G8UQ30,G8UQY3,
protein secretion	1		BFO_0843,	BFO_0843,
DNA restriction-modification system	1		BFO_2134,	BFO_2134,
SOS response	1		G8UL57,	G8UL57,
NAD biosynthetic process	1		G8UNH7,	G8UNH7,
cytokinin biosynthetic process	1		BFO_2938,	BFO_2938,
coenzyme A biosynthetic process	1		G8UHR4,	G8UHR4,

Gene Ontology	sample5 Enrichment Score	Log 2 Average Fold Change	Common Genes Involved	All Genes Involved
D-amino acid catabolic process	1		G8UHT3,	G8UHT3,
L-threonine catabolic process to glycine	1		G8UPR7,	G8UPR7,
electron transport chain	3		G8UMS8,G8UMT0,G8UMT1,	G8UMS8,G8UMT0,G8UMT1,
respiratory electron transport chain	1		G8UI16,	G8UI16,
ribosome biogenesis	1		G8UKZ0,	G8UKZ0,
ribosomal small subunit biogenesis	1		G8UKZ2,	G8UKZ2,
arginine biosynthetic process via ornithine	1		G8UKY2,	G8UKY2,
FtsZ-dependent cytokinesis	1		G8UKF4,	G8UKF4,
protein transport by the Sec complex	1		G8ULN6,	G8ULN6,
protein transport by the Tat complex	1		G8UJG0,	G8UJG0,
'de novo' CTP biosynthetic process	1		G8UMS1,	G8UMS1,
cellular biosynthetic process	1		BFO_3201,	BFO_3201,
negative regulation of DNA recombination	1		G8UJ12,	G8UJ12,
ATP metabolic process	1		BFO_0556,	BFO_0556,
amino sugar catabolic process	1		BFO_0044,	BFO_0044,
cell division	1		BFO_0719,	BFO_0719,
transmembrane transport	2		G8UI16,G8UMS8,	G8UI16,G8UMS8,

Gene Ontology	sample5 Enrichment Score	Log 2 Average Fold Change	Common Genes Involved	All Genes Involved
intracellular protein transmembrane transport	1		G8ULN6,	G8ULN6,
protein initiator methionine removal	1		G8ULN7,	G8ULN7,
rRNA base methylation	1		G8UKG3,	G8UKG3,
N-acetylmuramic acid catabolic process	1		BFO_0044,	BFO_0044,
proton transmembrane transport	1		BFO_0556,	BFO_0556,

Appendix VIII; Table 8.9. Molecular function for the significant upregulated genes in mucin.

Gene Ontology	sample5 Enrichment Score	Log 2 Average Fold Change	Common Genes Involved	All Genes Involved
tRNA binding	6		G8UHT3,G8ULL4,G8ULM3,G8ULM8,G8ULP0,G8UN57,	G8UHT3,G8ULL4,G8ULM3,G8ULM8,G8ULP0,G8UN57,
magnesium ion binding	4		BFO_2134,G8UN57,G8UNY2,G8UQ30,	BFO_2134,G8UN57,G8UNY2,G8UQ30,
nucleic acid binding	2		BFO_2829,G8UNK9,	BFO_2829,G8UNK9,
DNA binding	7		BFO_0093,BFO_1082,BFO_2133,BFO_2134,G8UL57, G8ULP3,G8UN58,	BFO_0093,BFO_1082,BFO_2133,BFO_2134,G8UL57, G8ULP3,G8UN58,
RNA binding	1		BFO_1137,	BFO_1137,
RNA helicase activity	1		BFO_2829,	BFO_2829,
mRNA binding	1		G8ULM2,	G8ULM2,
structural constituent of ribosome	26		G8UJH8,G8ULL4,G8ULL5,G8ULL6,G8ULL7,G8ULL8, G8ULL9,G8ULM1,G8ULM2,G8ULM3, G8ULM4,G8ULM5,G8ULM6, G8ULM7,G8ULM8,G8ULM9,G8ULN0, G8ULN1,G8ULN2,G8ULN3, G8ULN5,G8ULN9,G8ULP0,G8ULP1,G8ULP2,G8UM17,	G8UJH8,G8ULL4,G8ULL5,G8ULL6,G8ULL7,G8ULL8, G8ULL9,G8ULM1,G8ULM2,G8ULM3, G8ULM4,G8ULM5,G8ULM6, G8ULM7,G8ULM8,G8ULM9,G8ULN0, G8ULN1,G8ULN2,G8ULN3, G8ULN5,G8ULN9,G8ULP0,G8ULP1,G8ULP2,G8UM17,
translation initiation factor activity	1		G8ULN8,	G8ULN8,
peptidyl-prolyl cis-trans isomerase activity	1		G8ULR6,	G8ULR6,
2-isopropylmalate synthase activity	1		BFO_0823,	BFO_0823,
3-isopropylmalate dehydratase activity	1		G8UNY0,	G8UNY0,
3-isopropylmalate dehydrogenase activity	1		G8UNY2,	G8UNY2,
CTP synthase activity	1		G8UMS1,	G8UMS1,
DNA-directed 5'-3' RNA	1		G8ULP3,	G8ULP3,

Gene Ontology	sample5 Enrichment Score	Log 2 Average Fold Change	Common Genes Involved	All Genes Involved
polymerase activity				
DNA topoisomerase type II (double strand cut, ATP-hydrolyzing) activity	1		BFO_1082,	BFO_1082,
GTPase activity	1		G8UKZ2,	G8UKZ2,
argininosuccinate lyase activity	1		G8UKY2,	G8UKY2,
chorismate mutase activity	1		BFO_1682,	BFO_1682,
ATP-dependent peptidase activity	1		G8UNC2,	G8UNC2,
metalloendopeptidase activity	3		BFO_1168,BFO_1745,BFO_0011,	BFO_1168,BFO_1745,BFO_0011,
serine-type endopeptidase activity	1		G8UNC2,	G8UNC2,
exo-alpha-sialidase activity	1		BFO_2207,	BFO_2207,
enoyl-[acyl-carrier-protein] reductase (NADH) activity	1		BFO_1782,	BFO_1782,
endonuclease activity	1		G8UJ12,	G8UJ12,
alpha-L-fucosidase activity	1		BFO_2737,	BFO_2737,
beta-N-acetylhexosaminidase activity	1		BFO_2208,	BFO_2208,
pantetheine-phosphate adenylyltransferase activity	1		G8UHR4,	G8UHR4,
proline dehydrogenase activity	1		BFO_2738,	BFO_2738,

Gene Ontology	sample5 Enrichment Score	Log 2 Average Fold Change	Common Genes Involved	All Genes Involved
prephenate dehydratase activity	1		BFO_1682,	BFO_1682,
shikimate 3-dehydrogenase (NADP+) activity	1		BFO_2948,	BFO_2948,
phenylalanine-tRNA ligase activity	1		G8UN57,	G8UN57,
iron ion binding	1		G8UNZ9,	G8UNZ9,
ATP binding	13		G8UHR4,BFO_1082,G8UJ12,G8UKM4, G8UL57,BFO_0556, BFO_1575,G8UMA1,G8UMH5,G8UMI8 ,G8UMS1,G8UN57, BFO_2829,	G8UHR4,BFO_1082,G8UJ12,G8UKM4, G8UL57,BFO_0556, BFO_1575,G8UMA1,G8UMH5,G8UMI8 ,G8UMS1,G8UN57, BFO_2829,
GTP binding	3		G8UKM4,G8UKZ0,G8UKZ2,	G8UKM4,G8UKZ0,G8UKZ2,
N-acetyltransferase activity	1		BFO_2785,	BFO_2785,
phosphoric diester hydrolase activity	1		BFO_0550,	BFO_0550,
ATP-dependent activity, acting on DNA	1		BFO_2829,	BFO_2829,
N-methyltransferase activity	1		BFO_2133,	BFO_2133,
peptidase activity	1		G8UNZ9,	G8UNZ9,
serine-type peptidase activity	1		BFO_0046,	BFO_0046,
lipid binding	1		G8UMH1,	G8UMH1,
protein transmembrane transporter activity	1		G8UJG0,	G8UJG0,
transaminase activity	1		BFO_1681,	BFO_1681,
sulfuric ester hydrolase activity	1		BFO_0762,	BFO_0762,

Gene Ontology	sample5 Enrichment Score	Log 2 Average Fold Change	Common Genes Involved	All Genes Involved
penicillin binding	1		BFO_0406,	BFO_0406,
UDP-N-acetylglucosamine 2-epimerase activity	1		BFO_1050,	BFO_1050,
glycine C-acetyltransferase activity	1		G8UPR7,	G8UPR7,
isochorismate synthase activity	1		BFO_3201,	BFO_3201,
quinolinate synthetase A activity	1		G8UNH7,	G8UNH7,
serine-type D-Ala-D-Ala carboxypeptidase activity	1		G8UHR7,	G8UHR7,
type II site-specific deoxyribonuclease activity	1		BFO_2134,	BFO_2134,
electron transfer activity	3		G8UJI1,G8UMS7,G8UMS9,	G8UJI1,G8UMS7,G8UMS9,
FMN binding	3		G8UI15,G8UI16,G8UMS9,	G8UI15,G8UI16,G8UMS9,
proton transmembrane transporter activity	1		BFO_0554,	BFO_0554,
antiporter activity	1		BFO_2848,	BFO_2848,
site-specific DNA-methyltransferase (cytosine-N4-specific) activity	1		BFO_2133,	BFO_2133,
oxidoreductase activity	1		BFO_1683,	BFO_1683,
oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	1		BFO_1051,	BFO_1051,

Gene Ontology	sample5 Enrichment Score	Log 2 Average Fold Change	Common Genes Involved	All Genes Involved
oxidoreductase activity, acting on the CH-CH group of donors, NAD or NADP as acceptor	1		BFO_1051,	BFO_1051,
enoyl-[acyl-carrier-protein] reductase activity	1		BFO_1782,	BFO_1782,
oxidoreductase activity, acting on NAD(P)H, quinone or similar compound as acceptor	4		G8UI14,G8UI15,G8UI16,G8UI17,	G8UI14,G8UI15,G8UI16,G8UI17,
transferase activity	3		BFO_1303,G8ULL8,BFO_1933,	BFO_1303,G8ULL8,BFO_1933,
acyltransferase activity	1		BFO_0409,	BFO_0409,
glycosyltransferase activity	2		BFO_1989,BFO_1049,	BFO_1989,BFO_1049,
hydrolase activity	8		BFO_2321,BFO_0555,G8UMG9,G8UMH0,G8UMH1,G8UMH2, G8UMH5,BFO_2938,	BFO_2321,BFO_0555,G8UMG9,G8UMH0,G8UMH1,G8UMH2, G8UMH5,BFO_2938,
carbon-oxygen lyase activity	1		BFO_0044,	BFO_0044,
isomerase activity	1		BFO_1303,	BFO_1303,
ligase activity, forming carbon-nitrogen bonds	1		G8UMA1,	G8UMA1,
ATP hydrolysis activity	5		G8UJ12,G8UKM4,G8UL57,G8UMI8,BFO_2829,	G8UJ12,G8UKM4,G8UL57,G8UMI8,BFO_2829,
oxidoreductase activity, acting on the aldehyde or oxo group of donors	1		BFO_2832,	BFO_2832,
tRNA dihydrouridine synthase activity	1		BFO_1137,	BFO_1137,

Gene Ontology	sample5 Enrichment Score	Log 2 Average Fold Change		Common Genes Involved	All Genes Involved
DNA N-glycosylase activity	1			G8UN58,	G8UN58,
rRNA binding	23			G8UJH8,G8ULL5,G8ULL6,G8ULL7,G8ULL8,G8ULL9, G8ULM1,G8ULM2,G8ULM3,G8ULM5, G8ULM6,G8ULM7,G8ULM8, G8ULM9,G8ULN0,G8ULN1,G8ULN2,G8ULN3,G8ULN5,G8ULN8, G8ULP0,G8ULP1,G8ULP2,	G8UJH8,G8ULL5,G8ULL6,G8ULL7,G8ULL8,G8ULL9, G8ULM1,G8ULM2,G8ULM3,G8ULM5, G8ULM6,G8ULM7,G8ULM8, G8ULM9,G8ULN0,G8ULN1,G8ULN2,G8ULN3,G8ULN5,G8ULN8, G8ULP0,G8ULP1,G8ULP2,
transmembrane transporter activity	1			BFO_0382,	BFO_0382,
manganese ion binding	1			G8UQ30,	G8UQ30,
pyridoxal phosphate binding	2			BFO_1681,G8UPR7,	BFO_1681,G8UPR7,
thiamine pyrophosphate binding	1			G8UQ30,	G8UQ30,
mismatched DNA binding	1			G8UJ12,	G8UJ12,
peptide deformylase activity	1			G8UQ87,	G8UQ87,
ATPase-coupled transmembrane transporter activity	1			BFO_0555,	BFO_0555,
peptidoglycan binding	1			BFO_0719,	BFO_0719,
xenobiotic transmembrane transporter activity	1			BFO_2848,	BFO_2848,
ribosome binding	2			G8UKM4,G8ULN8,	G8UKM4,G8ULN8,
ribosomal large subunit binding	1			G8UKM4,	G8UKM4,
3'-5' DNA helicase activity	1			G8UL57,	G8UL57,
Ser(Gly)-tRNA(Ala) hydrolase activity	1			G8UHT3,	G8UHT3,

Gene Ontology	sample5 Enrichment Score	Log 2 Average Fold Change	Common Genes Involved	All Genes Involved
1,4-dihydroxy-2-naphthoate octaprenyltransferase activity	1		G8UQY3,	G8UQY3,
metal ion binding	8		BFO_1168,G8UMS1,G8UMS7,G8UN58, BFO_1745,G8UNH7, G8UQ87,BFO_0011,	BFO_1168,G8UMS1,G8UMS7,G8UN58, BFO_1745,G8UNH7, G8UQ87,BFO_0011,
transition metal ion binding	1		G8ULN7,	G8ULN7,
proton-transporting ATP synthase activity, rotational mechanism	6		G8UMG9,G8UMH0,G8UMH1,G8UMH2 ,BFO_0663,G8UMH5,	G8UMG9,G8UMH0,G8UMH1,G8UMH2 ,BFO_0663,G8UMH5,
proton-transporting ATPase activity, rotational mechanism	1		G8UMH5,	G8UMH5,
protein dimerization activity	1		G8ULP3,	G8ULP3,
flavin adenine dinucleotide binding	1		BFO_1137,	BFO_1137,
unfolded protein binding	1		G8UMI8,	G8UMI8,
NAD binding	2		G8UNY2,BFO_1051,	G8UNY2,BFO_1051,
D-aminoacyl-tRNA deacylase activity	1		G8UHT3,	G8UHT3,
4 iron, 4 sulfur cluster binding	3		G8UMS7,G8UN58,G8UNH7,	G8UMS7,G8UN58,G8UNH7,
raffinose alpha-galactosidase activity	1		BFO_2083,	BFO_2083,
N(6)-L-threonylcarbamoyladenine synthase activity	1		G8UNZ9,	G8UNZ9,
metalloaminopeptidase activity	1		G8ULN7,	G8ULN7,

Gene Ontology	sample5 Enrichment Score	Log 2 Average Fold Change	Common Genes Involved	All Genes Involved
small ribosomal subunit rRNA binding	1		G8UKZ2,	G8UKZ2,
2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylic-acid synthase activity	1		G8UQ30,	G8UQ30,
rRNA (cytosine-N4-)-methyltransferase activity	1		G8UKG3,	G8UKG3,
carbohydrate derivative binding	1		BFO_0044,	BFO_0044,
N-acetyl-beta-D-galactosaminidase activity	1		BFO_2208,	BFO_2208,
Gly-tRNA(Ala) hydrolase activity	1		G8UHT3,	G8UHT3,
class I DNA-(apurinic or apyrimidinic site) endonuclease activity	1		G8UN58,	G8UN58,

Appendix VIII: Table 8.10. Cellular component for the significant upregulated genes in mucin.

Gene Ontology	sample 5 Enrichment Score	Log 2 Average Fold Change	Common Genes Involved	All Genes Involved
extracellular region	1		BFO_1168,	BFO_1168,
cytoplasm	25		G8UHR4,G8UHT3,BFO_1303,G8UKG3,G8UKY2,G8UKZ2, G8ULL4,G8ULL6,G8ULL7,G8ULM2,G8ULM3,G8ULM6,G8ULM8, G8ULM9,G8ULN2,G8ULN3,G8ULN5,G8ULN8,G8ULP1,G8ULP2, G8UMA1,G8UN57,G8UNH7,G8UNY2,G8UNZ9,	G8UHR4,G8UHT3,BFO_1303,G8UKG3,G8UKY2,G8UKZ2, G8ULL4,G8ULL6,G8ULL7,G8ULM2,G8ULM3,G8ULM6,G8ULM8, G8ULM9,G8ULN2,G8ULN3,G8ULN5,G8ULN8,G8ULP1,G8ULP2, G8UMA1,G8UN57,G8UNH7,G8UNY2,G8UNZ9,
ribosome	20		G8UJH8,G8ULL4,G8ULL5,G8ULL6,G8ULL7,G8ULM2, G8ULM3,G8ULM4,G8ULM5,G8ULM7,G8ULM8,G8ULM9,G8ULN0, G8ULN1,G8ULN2,G8ULN9,G8ULP0,G8ULP1,G8UM17,G8UNK9,	G8UJH8,G8ULL4,G8ULL5,G8ULL6,G8ULL7,G8ULM2, G8ULM3,G8ULM4,G8ULM5,G8ULM7,G8ULM8,G8ULM9,G8ULN0, G8ULN1,G8ULN2,G8ULN9,G8ULP0,G8ULP1,G8UM17,G8UNK9,
plasma membrane	16		G8UI14,G8UI15,G8UI16,G8UKZ2,G8ULN6,G8UMG9, G8UMH0,G8UMH1,G8UMH2,G8UMH5,G8UMS7,G8UMS8,G8UMS9, G8UMT0,G8UMT1,BFO_0023,	G8UI14,G8UI15,G8UI16,G8UKZ2,G8ULN6,G8UMG9, G8UMH0,G8UMH1,G8UMH2,G8UMH5,G8UMS7,G8UMS8,G8UMS9, G8UMT0,G8UMT1,BFO_0023,
integral component of plasma membrane	3		G8UJG0,BFO_0843,G8UQY3,	G8UJG0,BFO_0843,G8UQY3,
cell outer membrane	8		BFO_2206,BFO_1198,BFO_1299,BFO_1406,BFO_1407, BFO_1412,BFO_3110,BFO_3111,	BFO_2206,BFO_1198,BFO_1299,BFO_1406,BFO_1407, BFO_1412,BFO_3110,BFO_3111,
3-isopropylmalate dehydratase complex	1		G8UNY0,	G8UNY0,
cytoplasmic side of plasma membrane	1		G8UKF4,	G8UKF4,
large ribosomal subunit	4		G8ULL8,G8ULM1,G8ULM6,G8ULN5,	G8ULL8,G8ULM1,G8ULM6,G8ULN5,
small ribosomal subunit	3		G8ULL9,G8ULN3,G8ULP2,	G8ULL9,G8ULN3,G8ULP2,

Gene Ontology	sample 5 Enrichment Score	Log 2 Average Fold Change	Common Genes Involved	All Genes Involved
membrane	1		BFO_0382,	BFO_0382,
integral component of membrane	27		G8UI14,G8UI15,G8UI16,BFO_3288,BFO_2211,BFO_0406, BFO_0407,BFO_0554,G8ULN6,BFO_2602, BFO_1601, G8UMH0,G8UMH1,G8UMH2,BFO_0662, BFO_1678,G8UMS8,G8UMS9, G8UMT0,G8UMT1,BFO_2830,G8UNC2,BFO_2848,BFO_0761, ⊻	G8UI14,G8UI15,G8UI16,BFO_3288,BFO_2211,BFO_0406, BFO_0407,BFO_0554,G8ULN6,BFO_2602, BFO_1601, G8UMH0,G8UMH1,G8UMH2,BFO_0662, BFO_1678,G8UMS8,G8UMS9, G8UMT0,G8UMT1,BFO_2830,G8UNC2,BFO_2848,BFO_0761, ⊻
cell division site	1		G8UKF4,	G8UKF4,
proton-transporting V-type ATPase, V0 domain	1		BFO_0554,	BFO_0554,
TAT protein transport complex	1		G8UJG0,	G8UJG0,
ATP-binding cassette (ABC) transporter complex	1		BFO_1576,	BFO_1576,
proton-transporting ATP synthase complex, catalytic core F(1)	3		G8UMG9,BFO_0663,G8UMH5,	G8UMG9,BFO_0663,G8UMH5,
proton-transporting ATP synthase complex, coupling factor F(o)	3		G8UMH0,G8UMH1,G8UMH2,	G8UMH0,G8UMH1,G8UMH2,

Appendix VIII: Table 8.11. Biological process for the significant downregulated genes in Mucin.

Gene Ontology	sample6 Enrichment Score	Log2 Average Fold Change	Common Genes Involved	All Genes Involved
transcription-coupled nucleotide-excision repair, DNA damage recognition	1		G8UJ21,	G8UJ21,
carbohydrate metabolic process	4		BFO_1193,G8UL71,BFO_0699,BFO_2947,	BFO_1193,G8UL71,BFO_0699,BFO_2947,
glycogen biosynthetic process	1		BFO_0974,	BFO_0974,
N-acetylglucosamine metabolic process	1		G8UL71,	G8UL71,
acetate metabolic process	1		BFO_0413,	BFO_0413,
acetyl-CoA metabolic process	1		BFO_0413,	BFO_0413,
pentose-phosphate shunt	1		G8UKI0,	G8UKI0,
DNA repair	1		G8UNI6,	G8UNI6,
transposition, DNA-mediated	2		BFO_0385,BFO_0831,	BFO_0385,BFO_0831,
DNA-templated transcription, initiation	2		BFO_0428,BFO_1509,	BFO_0428,BFO_1509,
regulation of transcription, DNA-templated	3		G8UJ21,BFO_2823,BFO_1789,	G8UJ21,BFO_2823,BFO_1789,
lipid metabolic process	1		BFO_2191,	BFO_2191,
chloride transport	1		BFO_1802,	BFO_1802,
iron ion transport	1		G8UKQ8,	G8UKQ8,
oligopeptide transport	1		BFO_2795,	BFO_2795,
cellular iron ion homeostasis	1		G8UKQ8,	G8UKQ8,
metabolic process	1		BFO_1263,	BFO_1263,

Gene Ontology	sample6 Enrichment Score	Log2 Average Fold Change	Common Genes Involved	All Genes Involved
regulation of cell shape	1		G8UL10,	G8UL10,
pyridoxine biosynthetic process	1		G8UKV3,	G8UKV3,
queuosine biosynthetic process	1		G8UJV2,	G8UJV2,
lysine biosynthetic process via diaminopimelate	1		G8UIY5,	G8UIY5,
nucleoside metabolic process	1		BFO_1438,	BFO_1438,
nucleotide-sugar metabolic process	1		G8UI00,	G8UI00,
peptidoglycan biosynthetic process	2		G8UL10,G8UNU1,	G8UL10,G8UNU1,
deoxyribonucleotide catabolic process	1		G8UIG5,	G8UIG5,
DNA-mediated transformation	1		G8UNI6,	G8UNI6,
terpenoid biosynthetic process	1		G8UQS9,	G8UQS9,
iron-sulfur cluster assembly	1		BFO_1485,	BFO_1485,
isopentenyl diphosphate biosynthetic process, methylerythritol 4-phosphate pathway	1		G8UQS9,	G8UQS9,
pentose catabolic process	1		G8UKI0,	G8UKI0,
propionate metabolic process, methylcitrate cycle	1		BFO_0413,	BFO_0413,
cytolysis	1		BFO_1263,	BFO_1263,
diaminopimelate biosynthetic process	1		G8UIY5,	G8UIY5,

Gene Ontology	sample6 Enrichment Score	Log2 Average Fold Change	Common Genes Involved	All Genes Involved
protein catabolic process	1		G8UKS1,	G8UKS1,
methylation	2		BFO_2273,G8UNZ5,	BFO_2273,G8UNZ5,
arginine biosynthetic process via ornithine	1		G8UNA4,	G8UNA4,
pteridine-containing compound metabolic process	1		G8UNZ5,	G8UNZ5,
defense response to bacterium	1		BFO_1263,	BFO_1263,
tRNA aminoacylation	1		BFO_0226,	BFO_0226,
extracellular polysaccharide biosynthetic process	1		G8UJN1,	G8UJN1,
nicotinamide nucleotide metabolic process	1		G8UHZ1,	G8UHZ1,
response to stimulus	1		BFO_1509,	BFO_1509,
trans-translation	1		G8UPS3,	G8UPS3,
cell wall organization	3		G8UL10,BFO_1449,G8UNU1,	G8UL10,BFO_1449,G8UNU1,
tRNA-guanine transglycosylation	1		G8UJV2,	G8UJV2,
organonitrogen compound metabolic process	1		BFO_2642,	BFO_2642,

Appendix VIII: Table 8.12. Molecular function for the significant downregulated genes in mucin.

Gene Ontology	sample 6 Enrichment Score	Log2 Average Fold Change	Common Genes Involved	All Genes Involved
tRNA binding	1		BFO_0226,	BFO_0226,
phosphorelay sensor kinase activity	2		BFO_1304,BFO_1922,	BFO_1304,BFO_1922,
DNA binding	9		BFO_2253,BFO_2260,BFO_0385,BFO_0428,BFO_1509, G8UNI6,BFO_1789,BFO_0831,BFO_1900,	BFO_2253,BFO_2260,BFO_0385,BFO_0428,BFO_1509, G8UNI6,BFO_1789,BFO_0831,BFO_1900,
DNA helicase activity	1		BFO_1900,	BFO_1900,
damaged DNA binding	1		G8UJ21,	G8UJ21,
DNA-binding transcription factor activity	2		BFO_1304,BFO_1922,	BFO_1304,BFO_1922,
RNA binding	1		G8UPS3,	G8UPS3,
lysozyme activity	1		BFO_1263,	BFO_1263,
catalytic activity	4		BFO_2272,BFO_0545,G8UPI7,BFO_0054,	BFO_2272,BFO_0545,G8UPI7,BFO_0054,
1,4-alpha-glucan branching enzyme activity	1		BFO_0974,	BFO_0974,
acetyl-CoA hydrolase activity	1		BFO_0413,	BFO_0413,
acetylglutamate kinase activity	1		G8UNA4,	G8UNA4,
acyl-CoA dehydrogenase activity	1		BFO_0252,	BFO_0252,
alkaline phosphatase activity	1		BFO_3041,	BFO_3041,
deoxyribose-phosphate aldolase activity	1		G8UIG5,	G8UIG5,
ATP-dependent peptidase activity	1		G8UKS1,	G8UKS1,
aminopeptidase activity	1		G8UKI4,	G8UKI4,
aspartic-type endopeptidase activity	1		G8UPA1,	G8UPA1,
metalloendopeptidase activity	2		G8UKS1,BFO_0703,	G8UKS1,BFO_0703,

Gene Ontology	sample 6 Enrichment Score	Log2 Average Fold Change	Common Genes Involved	All Genes Involved
serine-type endopeptidase activity	1		G8UKI4,	G8UKI4,
helicase activity	1		G8UJ21,	G8UJ21,
hydrolase activity, hydrolyzing O-glycosyl compounds	3		BFO_1193,BFO_0699,BFO_0974,	BFO_1193,BFO_0699,BFO_0974,
phosphoglycerate dehydrogenase activity	1		BFO_0195,	BFO_0195,
phospholipase A2 activity	1		BFO_2191,	BFO_2191,
protein serine/threonine kinase activity	1		BFO_2981,	BFO_2981,
protein serine/threonine/tyrosine kinase activity	1		BFO_2981,	BFO_2981,
D-ribulose-phosphate 3-epimerase activity	1		G8UKI0,	G8UKI0,
transposase activity	2		BFO_0385,BFO_0831,	BFO_0385,BFO_0831,
aminoacyl-tRNA ligase activity	1		BFO_0226,	BFO_0226,
voltage-gated chloride channel activity	1		BFO_1802,	BFO_1802,
iron ion binding	1		BFO_1485,	BFO_1485,
ATP binding	10		BFO_0082,BFO_0102,G8UHZ1,BFO_0226,G8UJ21,G8UKS1, BFO_1451,G8UNA4,BFO_1900,BFO_2981,	BFO_0082,BFO_0102,G8UHZ1,BFO_0226,G8UJ21,G8UKS1, BFO_1451,G8UNA4,BFO_1900,BFO_2981,
methyltransferase activity	1		BFO_2273,	BFO_2273,
ferric iron binding	1		G8UKQ8,	G8UKQ8,
peptidase activity	2		BFO_0082,BFO_0102,	BFO_0082,BFO_0102,
metalloexopeptidase activity	1		BFO_3171,	BFO_3171,
metallopeptidase activity	1		BFO_1449,	BFO_1449,
zinc ion binding	4		G8UKS1,BFO_1449,BFO_2596,G8UNZ5,	G8UKS1,BFO_1449,BFO_2596,G8UNZ5,
N-acetylglucosamine-	1		G8UL71,	G8UL71,

444

Gene Ontology	sample 6 Enrichment Score	Log2 Average Fold Change	Common Genes Involved	All Genes Involved
6-phosphate deacetylase activity				
dTDP-glucose 4,6-dehydratase activity	1		G8UI00,	G8UI00,
queuine tRNA-ribosyltransferase activity	1		G8UJV2,	G8UJV2,
methionine synthase activity	1		G8UNZ5,	G8UNZ5,
acetate CoA-transferase activity	1		BFO_0413,	BFO_0413,
4-hydroxy-tetrahydrodipicolinate synthase activity	1		G8UIY5,	G8UIY5,
glucose-1-phosphate thymidylyltransferase activity	1		G8UJN1,	G8UJN1,
glutamate racemase activity	1		G8UL10,	G8UL10,
lytic endotransglycosylase activity	1		G8UNU1,	G8UNU1,
phospholipase A1 activity	1		BFO_2191,	BFO_2191,
lactate transmembrane transporter activity	1		BFO_2713,	BFO_2713,
solute:proton antiporter activity	1		BFO_0837,	BFO_0837,
kinase activity	1		BFO_2254,	BFO_2254,
oxidoreductase activity, acting on the CH-CH group of donors	1		G8UL41,	G8UL41,
glycosyltransferase activity	1		BFO_0475,	BFO_0475,
hydrolase activity	6		BFO_3195,G8UJ21,BFO_1194,BFO_0721,BFO_0798, BFO_2902,	BFO_3195,G8UJ21,BFO_1194,BFO_0721,BFO_0798, BFO_2902,
dipeptidase activity	1		BFO_1449,	BFO_1449,
hydrolase activity, acting on carbon-nitrogen (but not	1		BFO_2642,	BFO_2642,

Gene Ontology	sample 6 Enrichment Score	Log2 Average Fold Change	Common Genes Involved	All Genes Involved
peptide) bonds, in cyclic amides				
lyase activity	3		G8UL41,BFO_0721,G8UNU1,	G8UL41,BFO_0721,G8UNU1,
isomerase activity	1		G8UQS9,	G8UQS9,
ATP hydrolysis activity	2		G8UKS1,BFO_1900,	G8UKS1,BFO_1900,
sigma factor activity	2		BFO_0428,BFO_1509,	BFO_0428,BFO_1509,
transmembrane transporter activity	1		BFO_0387,	BFO_0387,
pyridoxal phosphate binding	1		BFO_0243,	BFO_0243,
carbohydrate binding	2		BFO_2272,BFO_2947,	BFO_2272,BFO_2947,
1-deoxy-D-xylulose-5-phosphate reductoisomerase activity	1		G8UQS9,	G8UQS9,
cobalamin binding	1		G8UNZ5,	G8UNZ5,
4-phosphoerythronate dehydrogenase activity	1		G8UKV3,	G8UKV3,
cation binding	1		BFO_0974,	BFO_0974,
sequence-specific DNA binding	2		BFO_1304,BFO_1922,	BFO_1304,BFO_1922,
metal ion binding	11		G8UHZ1,BFO_2191,G8UJN1,G8UKI0,BFO_0474,G8UL71, BFO_0545,BFO_2642,G8UMD6,BFO_0054,G8UQS9,	G8UHZ1,BFO_2191,G8UJN1,G8UKI0,BFO_0474,G8UL71, BFO_0545,BFO_2642,G8UMD6,BFO_0054,G8UQS9,
N-acetylgalactosamine-6-phosphate deacetylase activity	1		G8UL71,	G8UL71,
S-adenosylhomocysteine deaminase activity	1		G8UMD6,	G8UMD6,
vinylacetyl-CoA delta-isomerase activity	1		G8UL41,	G8UL41,
flavin adenine dinucleotide binding	1		BFO_0252,	BFO_0252,
NAD binding	2		BFO_0195,G8UKV3,	BFO_0195,G8UKV3,

Gene Ontology	sample 6 Enrichment Score	Log2 Average Fold Change	Common Genes Involved	All Genes Involved
iron-sulfur cluster binding	3		BFO_1485,BFO_0545,BFO_0054,	BFO_1485,BFO_0545,BFO_0054,
2 iron, 2 sulfur cluster binding	1		BFO_0474,	BFO_0474,
raffinose alpha-galactosidase activity	1		BFO_1585,	BFO_1585,
phosphatidylserine 1-acylhydrolase activity	1		BFO_2191,	BFO_2191,
1-acyl-2-lysophosphatidylserine acylhydrolase activity	1		BFO_2191,	BFO_2191,
ADP-dependent NAD(P)H-hydrate dehydratase activity	1		G8UHZ1,	G8UHZ1,
NADHX epimerase activity	1		G8UHZ1,	G8UHZ1,
NADPHX epimerase activity	1		G8UHZ1,	G8UHZ1,
N-acylphosphatidylethanolamine-specific phospholipase D activity	1		BFO_2596,	BFO_2596,
NADPH binding	1		G8UQS9,	G8UQS9,
5'-methylthioadenosine deaminase activity	1		G8UMD6,	G8UMD6,
1,4-alpha-glucan branching enzyme activity (using a glucosylated glycogenin as primer for glycogen synthesis)	1		BFO_0974,	BFO_0974,
protein serine kinase activity	1		BFO_2981,	BFO_2981,
ABC-type transporter activity	2		BFO_0082,BFO_0102,	BFO_0082,BFO_0102,
peptide transmembrane transporter activity	1		BFO_2795,	BFO_2795,

Appendix VIII: Table 8.13. Cellular component for the significant downregulated genes in mucin.

Gene Ontology	sample 6 Enrichment Score	Log2 Average Fold Change	Common Genes Involved	All Genes Involved
cytoplasm	8		G8UIG5,G8UIY5,G8UJ21,G8UKQ8,G8UKV3,BFO_1485, G8UNA4,G8UPS3,	G8UIG5,G8UIY5,G8UJ21,G8UKQ8,G8UKV3,BFO_1485, G8UNA4,G8UPS3,
plasma membrane	5		BFO_2299,BFO_0427,G8UKS1,BFO_2795,G8UPA1,	BFO_2299,BFO_0427,G8UKS1,BFO_2795,G8UPA1,
integral component of plasma membrane	2		BFO_2713,G8UNU1,	BFO_2713,G8UNU1,
cell outer membrane	9		BFO_2107,BFO_2191,BFO_1510,BFO_0588,BFO_2801, BFO_2972,BFO_0981,BFO_0982,BFO_2013,	BFO_2107,BFO_2191,BFO_1510,BFO_0588,BFO_2801, BFO_2972,BFO_0981,BFO_0982,BFO_2013,
integral component of membrane	25		BFO_0082,BFO_0102,BFO_2107,BFO_3212,BFO_0189, BFO_2191,G8UIY5,BFO_1196,BFO_2255,BFO_2299, BFO_0387,BFO_0427,G8UKS1,BFO_2596,BFO_0587,BFO_2795, BFO_1714,BFO_1802,BFO_0837,G8UPA1,BFO_1922,BFO_2981, ⨄	BFO_0082,BFO_0102,BFO_2107,BFO_3212,BFO_0189, BFO_2191,G8UIY5,BFO_1196,BFO_2255,BFO_2299, BFO_0387,BFO_0427,G8UKS1,BFO_2596,BFO_0587,BFO_2795, BFO_1714,BFO_1802,BFO_0837,G8UPA1,BFO_1922,BFO_2981, ⨄
cell periphery	1		BFO_0387,	BFO_0387,