Characterization of sialate-*O*-acetylesterases in the gut commensal bacterium *Bacteroides thetaiotamicron*

Fatima Nadat

Masters by Research

University of York

Department of Biology

June 2013

Abstract

The sialic acids are a family of acid sugars that are widely distributed in mammal cells and are found typically linked to the distal end of glycan chains in mammalian glycoproteins, including the mucin proteins found on mucosal surfaces. The most common sialic acid, *N*-acetyl neuraminic acid or Neu5Ac is often found with additional *O*-linked acetyl group s when located on gut epithelial cell surfaces. In *Escherichia coli* the enzyme NanS functions as a sialate *O*-acetylesterase, to remove the additional acetyl groups and allow use of these *O*-acetylated sialic acids for nutrition. Although NanS has a cytoplasmic location in *E. coli*, human faecal extracts have been found to contain extracellular sialate-*O*-acetylesterase activity, which is thought to correlate to the presence of bacteria from the *Bacteroides* genes. We have identified *in silico* the genes encoding potential sialate *O*-acetylesterases across a range of *Bacteroides* species and for *Bacteroides thetaiotamicron*, the most abundant commensal species of *Bacteroides* in the gut, there are 10 potential candidates. Preliminary expression experiments indicated that these proteins are highly insoluble after expression in *E. coli* conventional vectors and increased solubility was dependent on the expression of the target protein with a range of different protein fusion partners. Basic acetylesterase enzyme assays, using p-nitrophenylacetate as a standard substrate, demonstrated that a number of these proteins have detectable acetyl esterase activity. A subset of these proteins are predicted to be secreted and hence could be responsible for the previously reported sialate *O*-acetylesterase activity present in the gut. The identification of this activity is potentially significant as alteration in levels of *O*-acetylation of gut proteins are known to be connected to progression of gastrointestinal cancer.

Declaration

This thesis is a presentation of my own research under the supervision of Dr Gavin Thomas at the University of York. Every effort has been made to clearly indicated wherever contributions were made by others, with reference to literature and acknowledgement of collaborative research and discussions.

Acknowledgements

I would like to thank Dr Gavin Thomas for taking a chance on me and giving me this opportunity. I would also like to thank Dr Jared Cartwright at the Technology Facility for all his patience with the production of soluble proteins. In addition, I would like to thank Dr Marjan Van der Woude and Dr Christoph Baumen for all their input and help and for being fantastic TAP members.

A big thank you to all of L1 for putting up with me for a whole year, especially Dan for his infinite patience with my constant questions and talking and Abbas for being my mentor. Finally I would like to say thank you to Stephen, Gemma, Sara, V-man, Brett and Little Monkey for making York feel like home and all their help with my sanity and research.

Table of Figures

Introduction

Sialic acids (also referred to as neuraminic acids) are nine carbon keto sugars that are widely distributed among mammals and metazoans. They were first identified by two German Biochemists. In 1936 Blix isolated what he termed 'carbohydrate I' in a crystallised form from bovine submaxillary mucin (Blix *et al*, 1936). In 1941 Klenk observed the presence of what he called neuroamine isolated from glangliosides in brain matter (Klenk *et al*, 1941). Following these initial discoveries, many other forms of the substances were identified, but it wasn't until the 1950's that it became evident that these were variant on the compounds described by Blix and Klenk and that the term neuraminic acid was established as the name for the basic, unsubstituted structure common to all the described substances (Blix *et al*, 1936; Klenk *et al*, 1941; Godoy *et al*, 1993; Faillard *et al*, 1989; Vimr *et al*, 2004). Although sialic acid encompasses an entire group of nine carbon sugars, the term itself is used to refer to the most common and widely studied of the sugars, *N*-acetylneuraminic acid (Neu5Ac) (Severi *et al*, 2007). The name sialic acid is consistent with the discovery of the carbohydrate in bovine submaxillary mucin and stems from the Greek word sialon meaning saliva (Blix *et al*, 1936; Vimr *et al*, 2004). Since these initial discoveries over 50 naturally occurring sialic acids have been identified (Angata and Varki, 2002).

Sialic acids were initially thought to be exclusive to higher metazoans but have since been shown to be present in certain prokaryotes (Vimr *et al,* 2004). Given that sialic acids are typically known to be linked to the distal end of glycan chains, they are in a prime position to be utilized by microorganisms. It comes as no surprise then, that many commensal and pathogenic bacteria use environmental and host sialic acids as a source carbon, nitrogen and energy and for assembly into cell wall components (Plumbridge and vimr, 1999; Vimr *et al*, 2004; Severi *et al*, 2005).

Sialic acid structure

Sialic acids originate from a 2-keto-3-deoxy-5-acetamido-D-glycero-D-galacto-nonulosonic acid parent compound. The basic sialic acid backbone is most commonly found in nature as a neuroaminic acid with either an *N*- acetyl or an *N*-glycolyl group at the C-5 position on the carbon ring (Varki *et al*, 2007). Although Neu5Gc is the second most abundant form of sialic acid seen in many vertebrates (mammals, birds, amphibians, reptiles and fish) it is not present in bacteria or archaea (Vimr *et al,* 2004). It is interesting to note that although this form of sialic acid is present in humans, the mammalian structural gene required for the addition of the *N*-glycolyl group is inactive. It is thought that the mutation resulting in a deletion of one of the exons occurred after the

divergence of humans from the great apes (Vimr *et al*, 2004; Angata and Varki, 2002). To date three types of sialic acid have been identified to be produced in bacteria: Neu5Ac, pseudaminic acid (Pse5,7Ac) and legionaminic acid (Leg5,7Ac). All three are derivatives of 4 isomeric 5,7-diamino-3,5,7,9-tetradeoxynon-2-ulosonic acids (Knirel *et al*, 2003). Both Pse and Leg have structural and biosynthetic similarities to sialic acid (Fig 1. Lewis *et al*, 2009). Each of these variations is dependent on the condensation of a 6 carbon sugar intermediate with the 3 carbon phosphoenolpyruvate to generate the final 9 carbon backbone (as described in more detail in below). Pse and Leg modifications both play a role in flagellar assembly and therefore a role in bacterial motility, although in Leg this is less well defined (McNally *et al*, 2006; McNally *et al*, 2007; Schirm *et al*, 2003; Logan *et al*, 2009; Twine *et al*, 2008). Pse and Leg have both been identified in LPS O antigens of various Gram negative bacterium indicating an involvement in biofilm formation, resistance to phage predation and animal association (Almagro and Boyd, 2010; Severi *et al,* 2007; Vimr *et al,* 2007; Lewis *et al,* 2009*)*. Despite being closely related to Neu5Ac the host-pathogen interactions of Pse and Leg are poorly defined (Lewis *et al,* 2009).

Over the years many sialic acid derivatives have been identified and are all biosynthetic derivatives of either Neu5Ac or KDN (3-deoxy-D-glycero-D-galacto-2-nonulosonic acid) (Angata and Varki, 2002). Modifications of sialic acid include the addition of up to three acetyl groups at the C4 position and at hydroxyl positions C-7 to C-9 (Vimr *et al,* 2004). Sialic acids with lactyl, sulfhydryl, phosphate or methyl groups have also been identified but these modifications are less common than those observed through the addition of acetyl groups (Fig. 1. Schauer, 2004).

Although usually found as part of complex structures such as glycolipids and glycoproteins, they can also be bound to structures which are composed entirely of sugars (oligosaccharides and some polysaccharides) (Vimr *et al,* 2004). Positioning of sialic acids at distal ends of glycan chains means they are in the perfect place for interactions with other cells and environmental agents (Angata and Varki, 2002). In eukaryotic cells sialic acids are mostly expressed on glycan chains of glycoproteins and glycolipids. In most cases they are linked via α2-3 linkages to Gal, α2-6 linkages to Gal, GalNAc or via a α2-9 linkage to other sialic acids (Tsuji *et al*, 1996; Angata and Varki, 2002). As well as these commonly seen linkages, there are a few rarer forms of linkages seen in eukaryotes. These include α2-3 linkages to GalNAc, α2-6 linkages to GlcNAc and Glc, α2-4 linkages to Gal and GlcNAc, α2-9 linkages to other sialic acids and Neu5Gc oligomer linked via hydroxyl group of *N*-glycolyl (Iwasaki *et al,* 1984; Kubo *et al*, 1990; Angata and Varki, 2002).
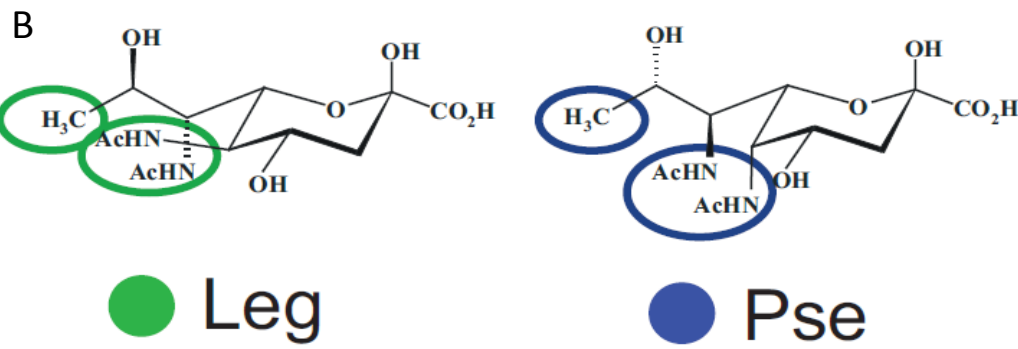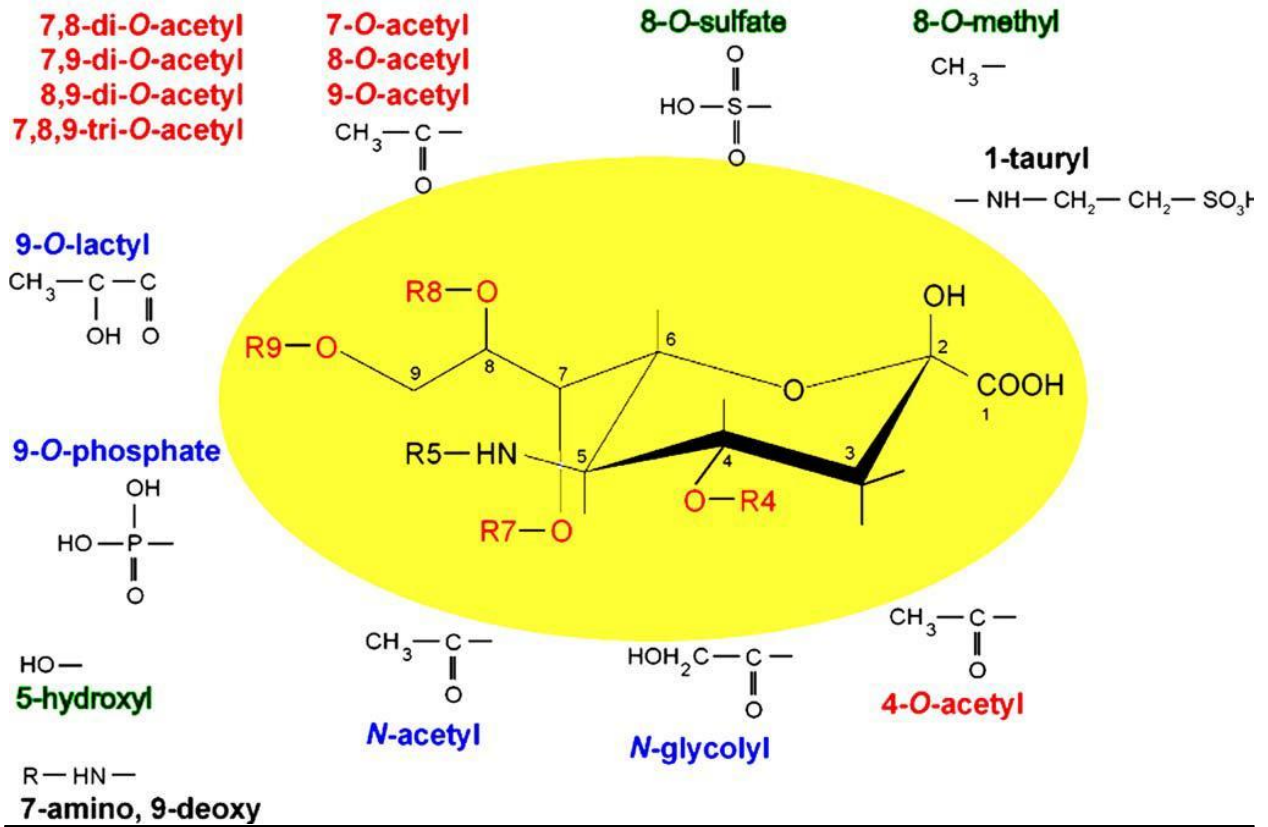
Figure 1 - Naturally occurring sialic acids. A. The various residues indicated in the image can be combined on the basic Neu5Ac backbone depending on the species or tissue (Schauer, 2004). **B**. Variations to the basic 9 carbon sugar backbone seen in legionaminic acid and pseudaminic acid (Lewis *et al*, 2009).

9

Sialic Acid and Bacteria

As mentioned previously it was initially thought that Neu5Ac was absent in prokaryotes, however over the years an increasing number of pathogenic bacteria have been found to encode sialic acid metabolism genes. Sialic acid metabolism in bacteria falls into two broad categories: biosynthesis with sialylation and scavenging with catabolism. Bacteria that are able to synthesize sialic acid are pathogenic in general and are able to utilize the sialic acid to coat several structures on their surface. The ability to decorate their surfaces gives these bacteria a competitive advantage in that they are able to mask themselves from the immune system and are able to alter the host cell specificity (Severi *et al*, 2005; Steenbergen *et al*, 2005; Brigham *et al*, 2009; Sharma *et al*, 2011; Almagro- moreno and Boyd, 2010; Takoa *et al*, 2010; Stafford *et al*, 2012; Spinola *et al*, 2012). The terminal position of host sialic acid has also made them the most commonly targeted carbohydrate receptor for pathogen adherence with examples of sialic acid binding agglutinins, adhesions and lectins in mammalian pathogens (Angeta *et al,* 2002).

The use of sialic acid is best characterized in *E. coli*, *Haemophilus influenzae*, *Pasteurella multocida*, *Neisseria* spp, *Campylobacter jejuni* and *Streptococcus agalactiae*. Each of these use the sialic acid in different ways to colonize, persist and cause disease to a mammalian host (Severi *et al,* 2007). Bacteria use sialic acid in at least two different ways; they are able to either use it as a source of nutrients or as part of the host immune evasion mechanism (Severi *et al,* 2007).

Many pathogenic bacteria have evolved to include the sialic acid in a large range of glycoconjugates decorating the cell surface including capsular polysaccharides or CPS (K-antigens), lipopolysacharides or LPS (*O*-antigens), S-layer glycoproteins and peptidoglycans (cell wall component). When sialic acids are present in bacteria they are mainly found in the CPS or LPS and confers an important phenotype involved in innate immune resistance and interaction with host cells (Severi *et al,* 2007).

Following sialic acid uptake bacterial cells are able to incorporate the sialic acid into cell surface macromolecules leading to the ability to modulate the interactions with the host (Fig. 2. Severi *et al*, 2007). The first step in this process is the activation of the sialic acid into CMP-Neu5Ac, which is then added to acceptors by sialyltranferases which are linkage specific (Severi *et al*, 2007). In *E. coli* it is NeuA which activates the Neu5Ac prior to addition of the sialic acid into K1 and K92 capsules. NeuS is the polysialyltranferase which adds the activated Neu5Ac to the oligosialic acid receptor leading to the production of the PSA capsule (Vimr *et al*, 2004; Severi *et al*, 2007). Through the action of NeuO

and NeuD, the sialic acid in the synthesised PSA is modified by *O*-acetylation. This process causes a large variation in the surface of the PSA capsule (Claus *et al*, 2004; Steenbergen *et al*, 2006; Deszo *et al*, 2005; Vimr *et al*, 2004). LPS sialylation is processed in a similar manner but is reliant on the linkage specific sialyltransferase Lst (Severi *et al*, 2007). It is interesting to note that *Neisseria gonorrhoea* relies on a scavenger (outer membrane associated) sialyltransferase to get the CMP-Neu5Ac directly from the secretion of the host (Shell *et al*, 2002). In addition to this *H. influenzae* actually has multiple charachterized sialyltransferase (4 in any given strain), the main one of which (Lic3A) has been shown to be essential for bacterial survival in the middle ear in the animal model for otitis media (Bochet *et al*, 2003; Severi *et al,* 2007).

Although the process of incorporation is relatively well characterized, it is still unclear how bacteria maintain a balance between the catabolism of sialic acid and its incorporation into surface structures (Severi *et al*, 2007). In *E. coli* K1 there is the possibility of competition between sialic acid catabolism and polysialic acid (PSA) synthesis. However, it is predicted a tight gene regulation exists as *neu* genes (involved in sialic acid incorporation into PSA) are continuously expressed whereas genes involved in sialic acid catabolism (*nan* genes) are expressed only during exogenous sialic acid presence (Vimr *et al*¸2004; Severi *et al,* 2007). Evidence does suggest that in *H. influenzae* catabolism of sialic acid is in direct competition with the LPS sialylation pathway as indicated by hypersialylation in *nanA* (gene essential for sialic acid catabolism) mutant (Vimr *et al*, 2000; Severi *et al,* 2007).

The addition of sialic acid into both the LPS and CPS allows bacterial cell immune evasion. A good example is the *E. coli* K1 PSA capsule which is poorly immunogenic due to its structural similarity of the PSA chains in the mammalian neuronal cell adhesion molecule NCAM (Vimr and Lichtensteiger, 2002). Although the K1 capsule does not confer serium resistance it is able to inhibit opsonisation and phagocytosis *in vivo* (Vimr and Lichtensteiger, 2002). *S. agalactiae* CPS has similar properties and is able to inhibit phagocytosis by impairing the C3 deposition onto the cell surface and so leading to the prevention of activation of the complement alternative pathway (Marques *et al*, 1992; Severi *et al,* 2007).

In a similar way, LPS sialylation also inhibits the complement alternative pathway in both *H. influenzae* and *N. gonorrhoeae*. The gonococcal sialyated LPS causes an increase in the binding of factor H (fH) to the bacterial cell surface. The fH is an anti-activator of complement alternative pathway hence allowing the bacterial cell to evade innate immune system of the host cell in a way that mimics the host cells own defences (Ram *et al*, 1998). In contrast, *H. influenzae* inhibits the deposition of C3 in an fH independent manner and it has been postulated that this is due to the C3

11

target being LPS through which sialylation causes the masking of the binding site (Figueira *et al*, 2007). There is also evidence that the sialylated bacterial surfaces of *C. jejjuni*, *Neisseria meningitidis* and *S. agalactiae* allows the bacterial cells to interact directly with sialic acid specific lectins of the siglec family found on the surface of immune cells. This direct interaction allows the bacterial cells to modulate the immune responses of the host (Avril *et al*, 2006; Carlin *et al*, 2007; Jones *et al*, 2003).

Given the high predominance of sialic acid in complex animals, microorganisms that can degrade sialic acid are often close commensals or pathogens of their hosts. Genes required for sialic acid catabolism are mostly confined to the pathogens or commensals that colonise sialic acid rich environments such as the lungs and gut (Almagro-Mareono and Boyd, 2009). Although more recently a soil bacterium has also been shown to grow on Neu5Ac (Grutser *et al*, 2012). The first indication of exogenous sialic acid metabolism (as a potential carbon source) was seen in 1974 through its uptake and degradation by *Clostridium perfingenus* (Ness *et al*, 1976). It wasn't until the 1980's that Vimr and Troy characterized the sialic acid degradation pathway in *E. coli.* While working on polysialic acid (PSA) synthesis in an *E. coli* K-12 strain expressing the *neu* genes from an *E. coli* K1 strain, they observed that only 10% of the radiolabled sialic acid that was added to the growth medium was actually incorporated into the PSA despite the evidence of quantitative uptake. Using mutational analysis, under the presumption that the bacteria expressed an effective transporter and the appropriate degradation system, they identified a sialic acid transporter and aldolase and named the genes *nanT* and *nanA* respectively (Vimr and Troy, 1985).

Genetic analysis alongside the completion of the K-12 genome sequence indicated the presence of a *nan* operon which not only contained the *nanTA* genes but also had several other genes involved in sialic acid metabolism as well as an upstream regulator. The divergently transcribed transcriptional regulator NanR is able to repress the expression of the *nan* operon in the absence of sialic acid. Since the completion of the genome, the genes included in the *nan* operon have been characterized in *E. coli* and a summary of their functions is outlined below.
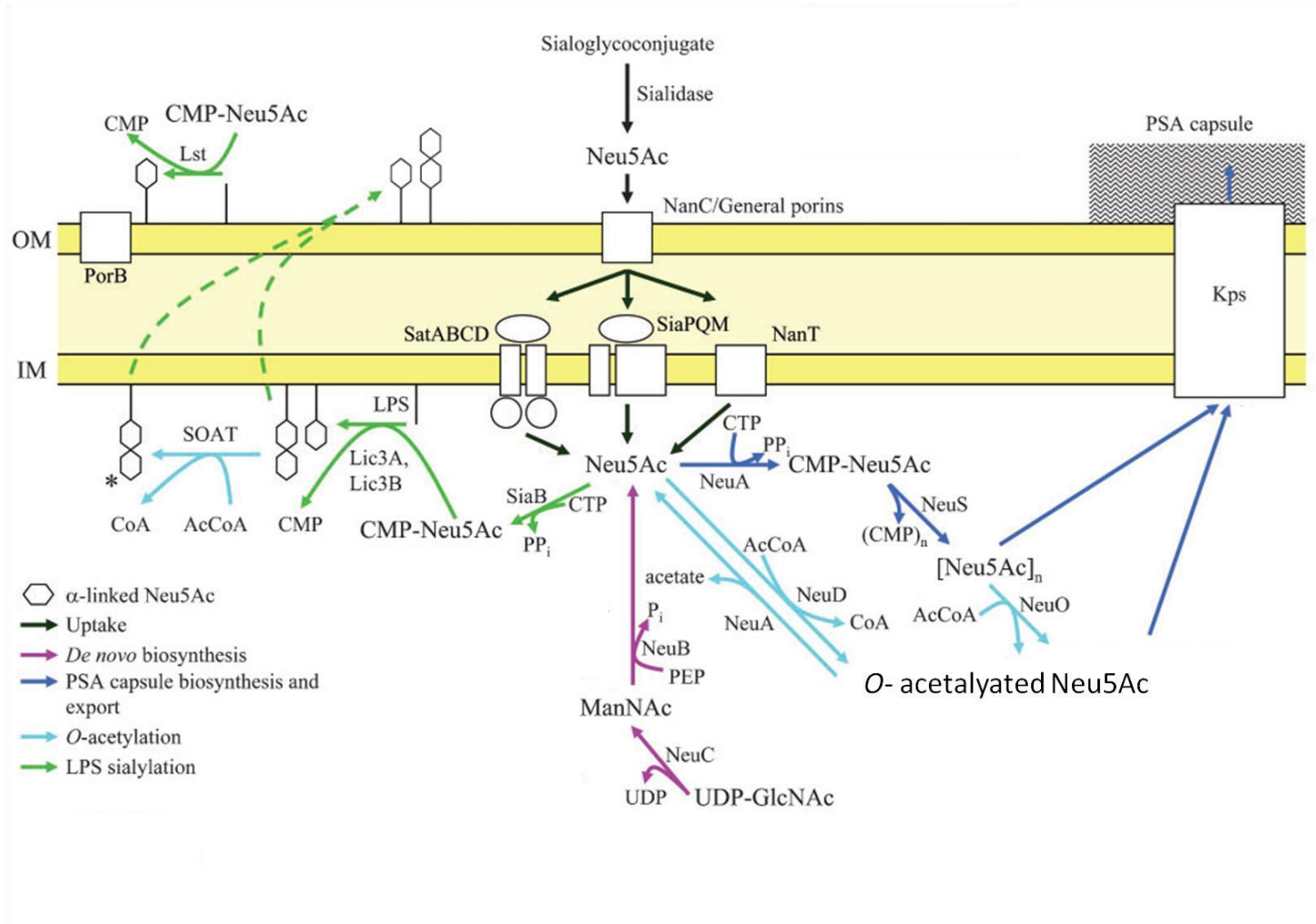
Figure 2 - Bacterial sialic acid biosynthesis and subsequent surface modifications. Schematic outlining the pathways utilized by bacteria for sialic acid biosynthesis and modification previous to incorporation into PSA and LPS (Adapted from Severi *et al*, 2007).

Sialic acid biosynthesis

Bacterial biosynthesis of sialic acid has evolved in two ways with some bacteria using a pathway for biosynthesis that differs to the eukaryotic method and others using truncated pathways which is reliant on host or environment derived sialic acid for sialylation (Fig. 2. Severi *et al*,2007). *E. coli* K1, *N. meningitidis* and other pathogens utilize the *de novo* sialic acid biosynthesis pathway, bacteria such as *N. gonorrhea* use a sialic acid scavenging pathway and *H. influenzae* scavenge precursors to obtain sialic acid for cell surface decoration(Almagro-Moreno and Boyd, 2009; Vimr *et al*, 2004; Severi *et al*, 2007).

The biosynthetic pathway for Neu5Ac was initially established in the 1960's by the Roseman and Warren groups (Rosenberg *et al*, 1976). As reflected by sequence similarities of enzymes involved the process of Neu5Ac biosynthesis is similar in bacteria and vertebrates (Fig. 2). In vertebrates and higher invertebrates sialic acid biosynthesis takes place in the cytosol in four step process involving three enzymes (Li and Chen, 2012). Neu5Ac is synthesised by the condensation of a neutral six carbon unit and the three carbon molecules pyruvate. KOD is the only sugar that is synthesised in this way (Angata and Varki, 2002).

The ultimate source of Neu5Ac is glucose with Neu5Ac synthesis beginning with the epimerization of *N*-acetylglucosamine (GlcNAc) to *N*-acetylmannosamine (ManNAc) (Angata and Varki, 2002). In veterbrates this process in catalysed by a bifunctional enzyme, UDP-GlcNAc/ManNAc kinase. This enzyme converts UDP-GlcNAc to ManNAc which is then immediately phosphorylated at the C6 position to give ManNAc-6-P (Hinderlich *et al,* 1997; Stasche *et al,* 1997; Angata and Varki, 2002). The ManNAc-6-P is then condensed with phosphoenolpyruvate to give rise to Neu5Ac-9-P, a process that is catalysed by Neu5Ac-9-P synthestase (Lawrence *et al,* 2000; Nakata *et al,* 2000; Angata and Varki, 2002). The phosphate group is then removed by a specific phosphatase.

In bacteria GlcNAc-6-P is epimerized to ManNAc-6-P by a GlcNAc-6-P 2-epimerase, in *E. coli* K1 this is encoded by the *neuC* gene and in *N. meningitidis* the *siaA* gene (Zapata *et al,* 1992; Petersen *et al,* 2000; Angata and Varki, 2002). The ManNAc-6-P is then dephosphorylated. The ManNAc and a phosphoenolpyruvate are then condensed through the activity of a synthestase (NeuB in *E. coli* and SiaC in *N. meningitidis*) to yield Neu5Ac (Vann *et al,* 1997; Angata and Varki, 2002).

In both bacteria and vertebrates the Neu5Ac is activated to CMP-Neu5Ac using cytidine 5-triphosphate (CTP). This process is catalysed by a cytosine 5'-monophosphate N acetylneuraminic acid (CMP-Neu5Ac) synthetase which is NeuA in *E. coli* and SiaB in *N. meningitidis* (Zapata *et al,*

1989; Edwards and Frosch, 1992; Angata and Varki, 2002). In vertebrates the synthesis takes place in the nuclei after which it is transported to the Golgi body via a CMP-sialic acid transporter where further modification and formation of glycoconjugates take place (Kean *et al,* 2004; Altheide *et al*, 2006; Eckhardt *et al,* 1996).

The CMP-Neu5Ac is added to Neu5Ac acceptor substrates by sialyltransferases. In mammals many different sialytransferases have been identified including ones for Siaα2-6Gal, Siaα2-3Gal, Siaα2-8Sia and Siaα2-6GalNAc linkages (Tsuji, 1996; Harduin-lepers *et al,* 1995). In bacteria a series of unrelated sialyltransferases have also been identified; α2-8(9) polysialyltransferases of *E. coli* K1/K92 and *N. meningitides,* α2-6 sialyltransferase of *Photobacterium damsel*a, α2-3/6 sialyltransferase of *N. gonorrhoeae,* α2-3 sialyltransferase of *Haemophilus. ducreyi* and *H. influenzae* and α2-3 sialyltransferase of *C. jejuni* (Steenbergen *et al,* 1992; Yamamoto *et al,* 1996; Gilbert *et al,* 1996; Bozue *et al,* 1999).
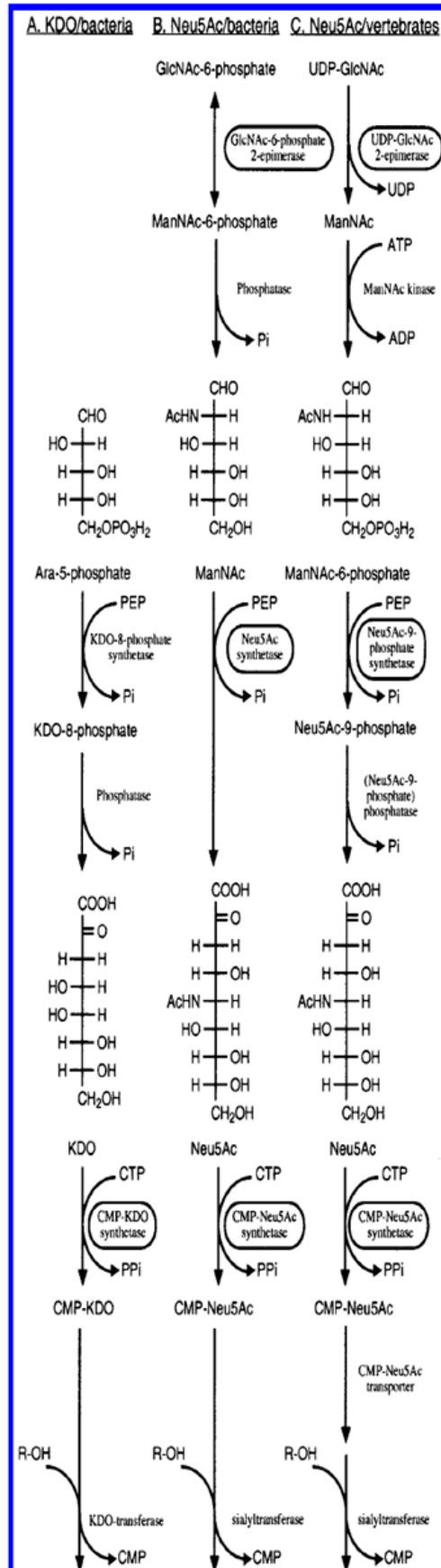
Figure 3 - Sialic acid biosynthesis. The sialic acid biosynthesis differs between bacteria and vertebrates during the initial stages of synthesis. Figure taken from Angata and Varki, 2002.

Bacterial sialic acid catabolism

Over recent years an increasing number of bacterial pathogens and human commensals are being identified which are able to utilize sialic acid as a growth factor or sole carbon source (Corfield, 1992; Vimr *et* al, 2004; Severi *et al,* 2007). The pathway for sialic acid catabolism has been established in *E. coli* (Fig. 3).

One of the first genes identified in the *E. coli* sialic acid metabolism pathway was the *nanT.* This has been established to code a secondary transporter of the major facilitator superfamily that can transport exogenous sialic acid. Once the sialic acid is successfully transported into the bacterial cell it is degraded intracellularly by NanA. NanA acts as an aldolase which breaks down the sialic acid into pyruvate and ManNAc. Evolutionary analysis shows that NanA has significant horizontal gene transfer between eukaryotes and prokaryotes. Of particular interest is that all members of *Bacteroides* have NanA proteins that branch closely to members of the eukarya (Almegro-Moreono and Boyd, 2009; De Koning *et al,* 2000). A kinase, NanK, then phosphorylates the ManNAc at the C-6 position producing the substrate for an epimerase (NanE). NanE converts the ManNAc-6-P to GlcNAc-6-P which is then fed into the NagA/NagB pathway, eventually giving rise to fructose-6-P that can be used as a carbon source. NagA (N-acetylglucosamine-6-P deacetylase) removes an acetyle group from GlcNAc-6-P giving rise to Glucosamine-6-P (Glc-6-P). NagB (N-glucosamine-6-P deaminase) then removes the amino group from Glc-6-P leading to the production of fructose-6-P (Plumbridge and Vimr, 1999)

Orthologues of the genes encoding NanAEK were initially regarded as essential requirements for a complete *nan* operon in those bacteria that are able to catabolise sialic acid. Although evidence suggests this is the case there is the potential that alternative methods exists for sialic acid catabolism, Although the final product of the sialic acid metabolism pathway is fed into the NagAB system, in *E. coli* the genes encoding these are not in close association with the *nan* operon. This holds true for most sialic acid catabolising bacteria with only a few exceptions.

As well as the need for the enzymes involved in sialic acid catabolism, it is essential for bacterial cells utilizing sialic acid to have transporters. To date there are three main functionally characterized types of bacterial sialic acid transporters. Although there is a general consensus of genes required for catabolism of sialic acid, the method by which the sialic acid is transported from an extracellular location to an intracellular varies between bacterial species.

Transport across the inner membrane has been studied in detail and four major classes of inner membrane transporters have been identified. NanT falls into the category of Major Facilitator Permeases whereas SiaPQM is a tripartite ATP-independent periplasmic (TRAP) transporter (Vimr *et al*,

2004; Steenbergen *et al,* 2005; Mulligan *et al,* 2012). More recently a third type of transporter has been identified. The Sodium Solute Symporter (SSS) family of transporters were first identified in *Salmonella tymphimirium* and is predicted to be present in many pathogenic bacteria (Severi *et al,* 2010). In *H. ducreyi* there appears to be a novel transporter which falls into the ABC family of transporters (SatABCD), which are common in Gram positives including *Corynebacterium glutamicum* (Post *et al,* 2005; Grutser *et al*, 2012; Chowdhury *et al*, 2012).

Outer membrane transporters for sialic acid fall into two main categories; NanC and NanOU. NanC was first identified in *E. coli K12* and is a sialic acid specific porin. It was shown to be essential for bacterial growth on sialic acid in the absence of the more general porins OmpF and OmpC (Condemine *et al,* 2005). The NanOU transporter is coded by two individual genes (*nanO* and *nanU*) and is a TonB dependent transporter found in *Tanerella forsythia* (Roy *et al,* 2010). In association with this transporter there is also NanO which is part of the TonB dependent receptor family involved in small molecule transport and energized by a TomB-ExbB-ExbD complex (Xu *et al*, 2003; Xu *et al*, 2007; Blanvillain *et al*, 2007).

Alongside these sialic acid catabolic genes, accessory genes have also been identified. These aid the bacteria in scavenging and utilizing exogenous sialic acids. Neuraminidase or sialidase (NanH) is the first accessory gene that was identified to be involved in Neu5Ac scavenging. The accessory genes aid in utilizing exogenous Neu5Ac but are not present in all sialic acid catabolising bacteria (Almagro-moreno and Boyd, 2010). The majority of Neu5Ac is bound as terminal residues on glycolipids and glycoproteins thus making them inaccessible to the bacteria (Sillanaukee et al, 1999). To compensate for this, certain bacteria (excluding *E. coli)* are able to secrete a sialidase to promote cleavage of the glycoketosidic linkages producing free Neu5Ac (Vimr and Troy, 1985). These enzymes act as specific glycosidases that use a double displacement mechanism and a covalent glycosyl-enzyme intermediate (Newstead *et* al, 2008). Along with host sialidases, these secreted bacterial sialidases increase the Neu5Ac pool available for utilization by all bacteria in their niche. Bacteria encoding the *nanATEK* genes then use the free Neu5Ac as a carbon source or as decoration for their cell surface. Sialidases cleave terminal sialic acid residues linked in an alpha configuration to oligosaccharides, but underlying carbohydrate residues and modification on side chains of the sialic acid (such as addition acetyl groups) affect the activity of the sialidase (Achyuthan *et al,* 2001; Roggentin *et al,* 1993; Lewis and Lewis, 2012). For example in *T. forsythia*, NanH is only able to cleave α-2,3-sialyl and α-2,6-sialyl bonds (Thompson *et al,* 2009; Stafford *et al,* 2011). As well as releasing sialic acid, bacterial sialidases are also involved in unmasking toxin receptors (exemplified by *V. cholera*) and biofilm formation (as

seen in *P. aruginosa*) (Holmgren *et al,* 1975; Soong *et al,* 2006). These enzymes are also able act as adhesins in their own right as well as exposing adhesive epitopes (Stafford *et al,* 2011).

In many pathogenic bacteria it has been shown that sialidases not only contain a catalytic domain for the release of sialic acid, but also contain carbohydrate binding domains (CBM) at either the C- or N-terminal of the sialidase. These CBM domains have also been identified incorporated within the β-propeller domain of the sialidase. It is postulated that these additional domains increase sialidase activity, especially in the presence of polysaccharides (Varghese *et al,* 1983; Crennell *et al*, 1993; Crennell *et al*, 1994; Park *et al,* 2013). It is interesting to note that in *B. thetaiotamicron*, a gram negative anerobic commensal of the gut, the N-terminal CBM found on its sialidase has a lectin like fold that differs from those previously described (Achyuthan *et al,* 2001; Xu *et al*, 2008; Boraston *et al*, 2007; Park *et al*, 2013).

A sialic acid mutarotase (NanM) has also been identified in *E. coli* and is associated with the successful utilization of Neu5Ac by bacteria. When bound to sialoglycoconjugates the Neu5Ac is found in its α-anomer form, bacteria however utilize Neu5Ac in the β-anomer form (Severi *et al*, 2008; Amegro-Moreono and Boyd, 2010). The alpha and beta configuration refers to the C7 and C1 carboxyl groups in either axial or equatorial orientations respectively (Angata and Varki. 2002). The pool of free Neu5Ac exists in a state of equilibrium where 95% of the Neu5Ac is found as a β-anomer. Spontaneous mutarotation of free α-Neu5Ac to β-Nue5Ac occurs at a slow rate. NanM is able to increase the rate at which this mutarotation occurs to reach equilibrium levels (Severi et al, 2007).

In the gastrointestinal tract (GI) many of the Neu5Ac molecules on the epithelial cell surfaces and mucin proteins have additional *O*-acetyl groups. Recently a novel enzyme has been identified in *E. coli* with sialate-*O*-acetylesterase activity (Rangarajan *et al.*, 2011).This enzyme, now called NanS, is able to cleave the *O*-acetyl group and is found at a cytoplasmic location in *E. coli* (Thomas *et al,* unpublished). The molecular characterization of a sialate O-acetylesterase is of particular interest as changes in levels of gut epithelial acetylated sialic acids are associated with GI cancers. Mutations in the human sialate-O-acetyltransferase, that lead to a decrease in sialic acid acetylation, are thought to be an important factor in the progression of GI cancer (Corfield *et al.,* 1999). Human faecal extracts have been found to contain extracellular sialate-*O*-acetylesterase activity which is thought to correlate to the presence of bacteria from the *Bacteroides* genus, where the NanS is potentially secreted and converts the acetylated sialic acid back to Neu5Ac, which can then be cleaved from the glycan via the action of the bacterial sialidase NanH (Corfied, 1993).

19
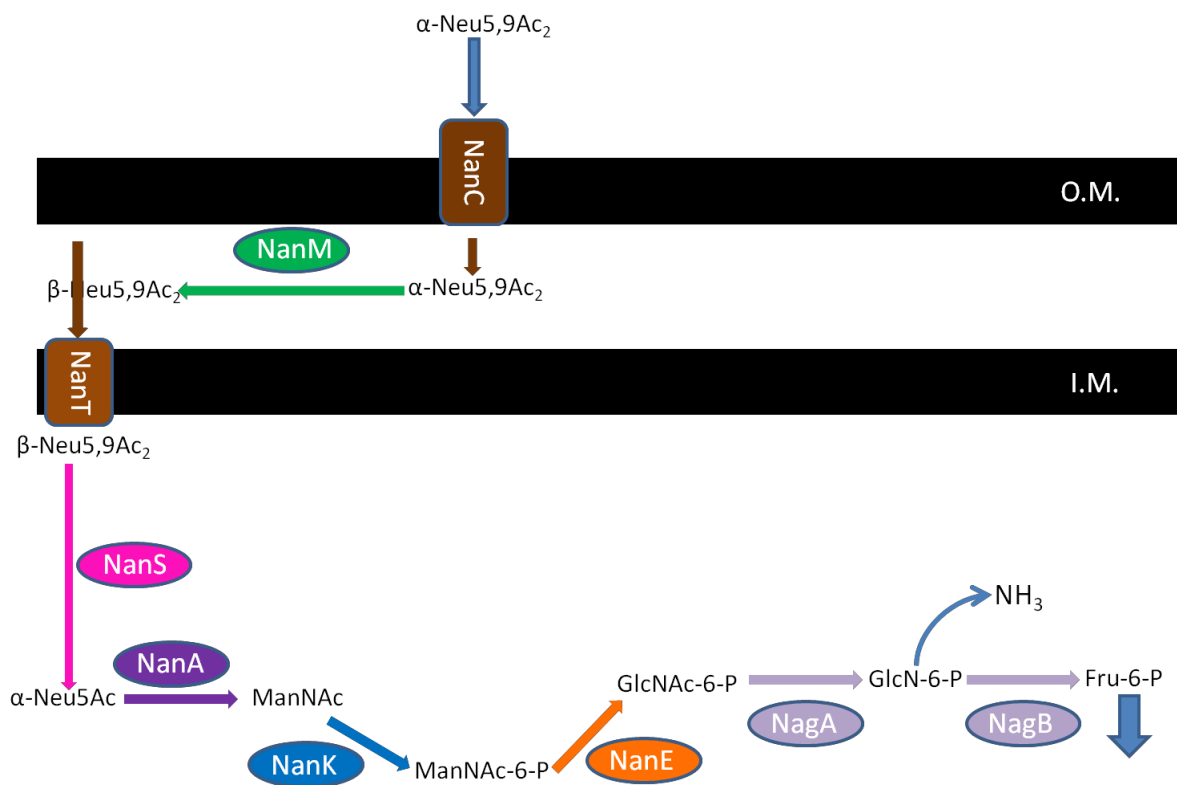
Figure 4 – Catabolism of Neu5,9Ac$_2$ by *Escherichia coli*. Di-acetylated sialic acid is de-acetylated to allow α-Neu5Ac release from its glycan bound state by the action of NanH. NanM increases rotation rate of α-anomer to the β-anomer which can then be transported through the outer membrane by NanO/U and through the inner membrane by NanT. Once inside the cell an unknown mechanism rotates the β-Neu5Ac back to the α-form in preparation for NanA action. Once NanA cleaves the Neu5Ac into ManNAc and Pyruvate the pathway can take one of two routes. In *E. coli* NanK phosphorylates the ManNAc as the NanE requires a phosphorylated substrate for conversion to GlcNAc.

Sialate-*O*-acetylesterase

Acetylation and deacetylation of both *N*- and O-acetylated carbohydrates is a common occurrence in nature. Enzymes involved in the deacetylation of carbohydrates fall into 16 groups of carbohydrate hydrolases and are further separated into two functional classes of enzymes: those that remove acetyl groups from carbohydrate units of polysaccharides (including pectin, xylan, conjugated sialic acids and others) as part of a catabolic pathway and those that are associated with regulation or virulence functions (Sundberg and Poutanen, 1991; Corfield *et al*, 1992; Steenbergen *et al*, 2009; Vimr and Steenbergen, 2006). Many of these enzymes have a broad substrate specificity including a range of naturally occurring and synthetic substrates (Rangarajan *et al*, 2010).

Sialate-*O*-acetylesterases known as NanS in bacterial cells belong to the SGNH superfamily of hydrolases (Molgaard *et al*, 2000). This superfamily is a distinct class of α/β hydrolases which are based on the presence of a catalytic serine (S), an oxyanion hole forming residue glycine (G), an asparagine (N) and an invariant histidine (H) (Ollis *et al*, 1992; Lo *et al*, 2003). Each of these residues is found within blocks of conserved residues in all SGNH hydrolases regardless of catalytic function with each member of the group utilizing a nucleophilic-His-acid catalytic triad for hydrolysis of its given substrate (Rangarajan *et al*, 2010). The conserved residues were shown to be incorporated into four distinct blocks within the SGNH hydrolases. Block I contains a GDS motif with the nucleophilc Ser residue and a hydrogen bond donor for the oxyanion hole. Block II contains the conserved Gly while Block III contains the conserved Asn which is involved in the formation of the oxyanion hole within a conserved GXN motif. Finally Block IV has two catalytic residues, aspartic acid and histidine, within a DXXH motif (Rangarajan *et al*, 2010).

Structural analysis shows the SGNH hydrolases to contain a central, five stranded parallel β-sheet which is flanked by an α-helix on either side. The nucleophile is shown to sit on the sharp turn between the β-strand and the leading α-helix. The His sits within the loop preceding the C-terminal α-helix with the aspartate (part of the catalytic triad) found two residues previous to the His. This makes the SGNH hydrolase family distinct from other hydrolases where it is the aspartate that follows the His (Rangarajan *et al*, 2010).

In *E. coli* the sialate-*O*-acetylesterase was actually initially thought to be an acetylxylan esterase but has since been shown to specific for 9-*O*-acetylated Neu5Ac and is essential for growth of *E. coli* on sialic acid (Steenbergen *et al*, 2009). The *nanS* gene in *E. coli* is found within the operon also containing *nanC* and *nanM* (Houliston *et al*, 2006). The *E. coli* NanS protein is found as a monomer in solution and has a single compact domain that is similar to the conical α/β hydrolase fold and in

specific the SGNH hydrolyse fold (Molgaard *et al*, 2000; Rangarajan *et al*, 2010). It is comprised of a seven stranded mixed β-sheet and two β-hairpins following the β-2 and β-6 strands. Both C- and N-terminal α-helices are situated on the concave side of the β-sheet with the rest sitting on the convex side (Rangarajan *et al*, 2010).

Although the backbone of NanS is similar to the SGNH hydrolases previously characterized, it became the founding member of a sub-group of the family (now known as Group II SGNH hydrolases) due to differences observed in comparison to those previously described. These differences are observed within the conserved blocks described above. Block I of Group II hysrolases contains a GXSN conserved region as appose to the GDS motif seen in Group I SGNH hydrolases. The glutamate in this conserved motif (Gln18 in NanS), which is adjacent to the catalytic serine (Ser19), functions as a hydrogen bond donor for the formation of the oxyanion hole and so functionally replaces the asparagine found within the Block III GXN motif in Group I SGNH hydrolases. In addition to this the Block III motif is replaced by a QGEX conserved sequence, where the residues act to stabilise the Gln18. Interestingly the Asp found in Block IV of the Group I SGNH hydrolases is not present within the Group II hydrolases (Rangarajan *et al*, 2010). Although the *E. coli* NanS falls into a distinct sub-group of the SGNH hydrolase family, not all prokaryotic esterases do (Lewis *et al*, 2007; Steenbergen *et al*, 2006). The mammalian sialate *O*-acetylesterase, interestingly, does fall into the group II sub-group of SGNH hydrolases.

The presence of a mammalian sialate *O*-acetylesterase was first predicted in 1985 when it was suggested that the esterase existed in a membrane bound form (Varki *et al*, 1985). At this time little was known about the catabolism or turnover of the *O*-acetyl group found on sialic acids. In 1989 experimental data showed the presence of the acetylesterase within the rat liver. Golgi vesicles containing [3]H- labelled acetyl were shown to incorporate a large portion of the labelled compound into acetylated sialic acid. When this was incubated with neurominidase a significant amount of free labelled acetate was detected, indicating the presence of an esterase within the Golgi vesicle (Higa *et al*, 1989).

Since their initial identification, it has been shown that two isoforms of the mammalian sialate *O*-acetylesterases exist. Both forms of the acetylesterase in mammals are coded by an individual gene but expressed differently in various tissues and have different subcellular localization (Takematsu *et al*, 1999). Of the two forms one is found localized in the cytoplasm and the other in lysosomal compartments. The lysosomal acetylesterase contains a predicted signal peptide and is through to participate in the deacetylation of 9-*O*-acetylated sialoglycoconjugates. The cytosolic isoform, however, contains no signal peptide and is thought to salvage any acetylated molecules that escape

the initial action of the lysosomal esterase (Higa *et al*, 1989; Butor *et al*, 1992; Guimaraes *et al*, 1996; Takematsu *et al*, 1999; Schauer *et al*, 1989). Following identification of these acetylesterases in rats and mice, the first indication of the presence of lysosomal sialate acetylesterase in human tissue was shown in 2003. The human sialate acetylesterase (SIAE) was shown to structural similarities to those previously identified in mice and rats (Zhu *et al*, 2003; Yangin *et al*, 2004).

Project objectives

An intimate relationship has been established between sialic acid catabolism and bacterial fitness in the gut due to this prevalence of sialic acid in this environment (Almagro-Moreo and Boyd, 2009; Jong *et al*, 2009; Chang *et al,* 2008). The human intestine is dominated by two phyla of bacteria, the Bacteroidetes and the Firmicutes. These account for 95 to 98% of the microbiota but the relative abundance of each varies. *Bacteroides* (Gram-negative, nonsporing, rod shaped bacteria) are the most abundant in the gut microbiota and corresponding to 25% of all intestinal bacteria (Karlsson *et al,* 2010; Eckburg *et al*, 2005).

The aim of the project is to determine if gut commensals have the potential to alter levels of acetylated sialic acid on the host cells surface, through identification and characterisation of *Bacteroides* sialate *O*-acetylesterase activities.

Materials and Methods

Media

Preparation of liquid and agar media was achieved by dissolving the specific quantities in distilled water and autoclaved at 121°C and 15 Psi. Approximately 20 ml of liquid agar was poured into Petri dishes for the preparation of agar plates. These were dried in a category 2 microbiology cabinet before storage at 4°C.

For the preparation of Lauria Broth (LB) 20 g of tryptone, 10 g of yeast extract and 10 g of NaCl was dissolved in a total volume of 1 litre of dH_2O. For agar media 5 g agar was added to the above mix.

Media was infused with appropriate antibiotics at the following concentrations. Ampicillin 100 µg/ml and Kanamycin 30 µg/ml

Polymerase Chain Reaction (PCR)

PCR was performed as previously described (Mullis *et al*, 1986). Primers were designed and purchased from MWG Biotech (Germany). Each primer was diluted as per manufactures instructions to give a final concentration of 100 pm/µl. Primers sequences are described in Table 1. dNTP stock was prepared to a final concentration of 2.5mM.

For Taq polymerase PCR the following 50 µl reaction mix was prepared in a 0.5 ml PCR tubes; 5 µl of 10x Thermo. Polymerase buffer (NEB), 1 µl dNTPs (of a 10 mM stock), 5 µl forward and reverse primers (from a 10 pm/µl stock), 1 µl of genomic DNA, and dH_2O to make up the final volume. For whole cell PCR, a colony from an agar plate was resuspended in the above mix in place of the genomic DNA.

For KOD hot start polymerase PCR the following 50 µl reaction mix was prepared; 5 µl of 10x conc. KOD hotstart buffer, dNTPs (10mM final concentration), forward and reverse primers (10 pMol/µl final concentration), $MgSO_4$ (20mM final concentration), 1 µl of genomic DNA, and dH_2O to make up the final volume.

Using a thermocycler, PCRs were performed with an initial denaturing at 95 °C for 5 minutes and subsequent denaturing at 94 °C for 30 seconds. Annealing was carried out to an appropriate temperature and time for the given primer set. Extension and final extension were both done at 72 °C for 3 minutes and 5 minutes respectively, and the completed reaction held at 10 °C.

Agarose gel electrophoresis

The product of PCR reactions and restriction enzyme digests were analysed through electrophoresis of the DNA samples using a 1% agarose gel. A stock solution of 10x TBE was made as follows: 2.5 L dH2O, 405g Tris, 68.75 g Boric acid and 23.75 g EDTA. The 1% agarose gel was made up in 100 ml of 1x TBE. 10 µl of Sybersafe or 10 µl ethidium bromide was added to molten agarose before 50 ml of the agarose was allowed to set in a gel tank containing well combs. 10x conc. Loading buffer was added to each sample before samples were loaded into the wells. The gel was run in 1x TBE buffer at 80V for approximately 40 minutes. DNA was visualized using a Transilluminator.

Restriction enzyme digests

Both pBAD and pET vectors were digested using either Swa1 or BseR1 respectively for cloning for subsequent ligation independent cloning. The purified plasmids were digested in a total volume of 60 µl containing the following; 50 µl purified plasmid (pBAD vector was purified to (4.4 ng/µl and the pET vector was purified to 16.3 ng/µl) 1 µl restriction enzyme, 6 µl 10x buffer and 4 µl dH$_2$O. For the digestion of pET vectors buffer 4 (NEB) was used and samples were incubated at 37 °C for 2 hours before enzyme inactivation at 65 °C for 20 minutes. Digestion of pBAD vectors used buffer 3 (NEB) and incubation was carried out at 25 °C for 2 hours before enzyme inactivation.

For determination of insert orientation and presence in traditional expression vectors (pET and pBAD), restriction enzyme digests was carried out using two enzymes (as outlined in Table 2). The total reaction volume was 20 µl and consisted of 5 µl of the purified construct, 1 µl of each enzyme, 2 µl of the appropriate 10x buffer and dH2O to make up the final volume. Reaction conditions are outlined in Table 2.

For the screening of insert presence in novel vectors the following components were included in a total reaction volume of 20 µl; 16 µl of the purified plasmid, 1 µl Xho1, 1 µl Nco1, 2 µl buffer 4 and 0.2 µl BSA. The reactions were incubated at 37 °C for 2 hours before DNA was visualised on an agarose gel.

SDS-PAGE

SDS-PAGE was performed for analysis of protein expression, solubility and purification. Whole cell samples were resuspended in 25 x b µl 5x loading dye (12.6 % stacking buffer [20 mM TRIS and 0.55 mM SDS, pH 6.8], 10.5 % glycerol, 2.1 % SDS, 0.005 % bromophenol blue and 5 % 2- mercaptoethanol) where b is the OD of the cell culture at the point of harvest. In the case of non- whole cell samples used in the SDS assay, 10 µl of the sample was added to 10 µl of 5x loading dye. Samples were boiled for 5 minutes and loaded onto a 12 % SDS polyacrylamide gel. Samples loaded onto the gel were separated at 200 V in running buffer (25 mM TRIS, 250 mM Glycine and 0.1 % SDS) until the gel front was between 1 and 5 mm from the bottom. Gels were stained in bromophenol blue to visualise proteins.

Bioinformatics

Analysis of gene and protein sequences was achieved using Blast searches, Xbase and Uniprot. Genes previously established as being involved in sialic acid metabolism were used as Blast queries for the identification of homolgous genes within *Bacteroides.* Gene maps were constructed using Xbase and Uniprot for size and orientation of the *Bacteroides* genes. Sequence alignments were constructed using Clustal X which was also utilized to construct trees analysing relationships between various NanS proteins from different bacterial species (Fig. 16).

Gene	Vector	Forward Primer	Reverse Primer
BT_0457	*pETYSBLIC*-3C	**CCAGGGACCA**GCAATGCAGGAAAATCGTATTAAAG	GAGGAGAAGGCGCGTTATCATTTCATCGAAAATTCG
BT_0447	*pETYSBLIC*-3C	**CCAGGGACCA**GCAATGTTCAGAAACAGATACT	GAGGAGAAGGCGCGTTATCATAAGGAGCCTCCTCTT
BT_0457	*pBADcLIC*	ATGGGTGGTGGATTTGCTATGATGAAAAATGTATTAT	TTGGAAGTATAAATTTTCTTTCATCGAAAATTCGGCA
BT_0447	*pBADcLIC*	ATGGGTGGTGGATTTGCTATGTTCAGAAACAGATACT	TTGGAAGTATAAATTTTCTAAGGAGCCTCCTCTTTTC
BT_0457 Full length	pETFPP_2/3/4/5	**TCCAGGGACCAGCAATG**CAGGAAAATCGTATTAAAGTTGC	**TGAGGAGAAGGCGCGTTA**TTTCATCGAAAATTCGGCACG
BT_0457 DUF303 domain	pETFPP_2/3/4/5	**TCCAGGGACCAGCAATG**TTCATATTGACGGACGCCGTACACC	**TGAGGAGAAGGCGCGTTA**TTTCATCGAAAATTCGGCACG
BT_0457 N-terminal	pETFPP_2/3/4/5	**TCCAGGGACCAGCAATG**CAGGAAAATCGTATTAAAGTTGC	**TGAGGAGAAGGCGCGTTA**CATATTGTCTGTATAAAGC
BT_0447 Full length	pETFPP_2/3/4/5	**TCCAGGGACCAGCAATG**TTCAGAAACAGATACTTTTGGC	**TGAGGAGAAGGCGCGTTA**TAAGGAGCCTCCTCTTTTCATTCC
BT_0447 DUF303 domain	pETFPP_2/3/4/5	**TCCAGGGACCAGCAATG**TTCAGAAACAGATACTTTTGGC	**TGAGGAGAAGGCGCGTTA**AAGGCCGGTTTCTTTTTGATCACG
BT_0447 C-terminal	pETFPP_2/3/4/5	**TCCAGGGACCAGCAATG**AAAAAGAAACCGGCCTTAATGTGG	**TGAGGAGAAGGCGCGTTA**TAAGGAGCCTCCTCTTTTCATTCC
BT_4091	pETFPP_2/3/4/5	**TCCAGGGACCAGCAATG**CAAGTGAAATTGCTTCCG	**TGAGGAGAAGGCGCGTTA**TTTATTGACATAAGTAATACCTTTCC
BT_0895	pETFPP_2/3/4/5	**TCCAGGGACCAGCAATG**GAAGTGAAACTGCCCGCTATTTC	**TGAGGAGAAGGCGCGTTA**CCAATCGTCCGAACGGAAAGACC
BT_3171	pETFPP_2/3/4/5	**TGAGGAGAAGGCGCGTTA**CCAATCGTCCGAACGGAAAGACC	**TGAGGAGAAGGCGCGTTA**CTTTACATCATCCCATGTATCC
BT_4180	pETFPP_2/3/4/5	**TCCAGGGACCAGCAATG**GAGAAACCGCTGAAGACTCTGG	**TGAGGAGAAGGCGCGTTA**TTTCTTTATTAGTTTCTTAAC

Table 1- Primer list. List of primers utilized during the course of this study. Primers for cloning into conventional vectors were designed to incorporate the Ligation Independent Cloning (LIC) cassette (**Bold**). Primers designed for utilization with the novel cloning vectors were designed to include a vector complementary sequence in the final PCR product for downstream InFusion cloning.

Preparation of pBADcLIC and pETYSBLIC3C based vectors

Genes of interest (BT_0447 and BT_0457) were initially cloned for insertion into the conventional vectors used in the Thomas Lab. These are the pETYSBLIC3C vector and the pBADcLIC vector. In all cases the primers were designed to include the LIC cassette as well as cloning the gene of interest (Table 1).

Both pET and pBAD vectors were prepared for cloning through restriction enzyme digestion with enzymes BSer1 and Swa1 respectively. Following vector digestions, products were run on a 1% agarose gel and bands migrating to the correct molecular weight were extracted and purified using a gel extraction kit (Macherey – Nagel). Purified digested vectors were then prepared for Ligation Independent cloning (LIC). This was achieved by treating 45 μl of the vector with 9 μl dTTP at a final concentration of 25 mM (for pET vectors) or 9 μl dCTP at a final concentration of 25 mM (for pBAD vectors). The reaction was carried out in a total volume of 62.5 μl and also contained 6 μl of 10x T4 DNAse polymerase buffer. The reaction mix was incubated at room temperature with 2.5 μl T4 DNA polymerase for 30 minutes before inactivation of the polymerase by incubation at 75°C for 20 minutes. To determine if the digest and LIC preparation was successful, the resulting DNA was transformed into Solopack Gold supercompetent *E. coli* (Invitrogen) cells (as per manufacturer's instructions). Addition of the LIC cassette interferes with antibiotic resistance and so a lack of colonies on a transformation plate containing the appropriate antibiotic correlates to the plasmids with the LIC cassette.

BT_0447 and BT_0457 were both amplified from *B. thetaiotamicron* genomic DNA by PCR using a high fidelity KOD polymerase, with an annealing temperature of 55°C and using primers described in Table 1. The PCR products were seperated on a 1% agarose before the bands representing the PCR products were extracted and purified using the a gel extraction kit. A volume of 14 μl purified PCR products were treated in the same way as the vectors for LIC annealing by the addition of 2 μl dATP at a final concentration of 25 mM (for genes being cloned into the pET vector) or 2 μl dGTP at a final concentration of 25 mM (for genes being cloned into the pBAD vector). In this case the reaction mix also contained 1 μl 100 mM DTT as well as 0.5 μl T4 DNAse polymerase and 2 μl T4 polymerase buffer.

Insertion of the genes of interests into LIC ready vectors was achieved through LIC annealing at a 1:2 molar ratio of vector to insert. Reactions were incubated at room temperature for 20 minutes

before the addition of 1.5 µl 25 mM EDTA and a further incubation for 15 minutes. Resultant products were then transformed into Solopack Gold supercompetent *E. coli* cells.

Colonies present following transformation were analyzed using Taq polymerase PCR and primers used for cloning. Successful transformation was established by presence of a band on an agarose gel with a molecular weight representing the gene of interest. Colonies producing positive results for each transformation were then used to inoculate 5 ml of LB (containing appropriate antibiotics) followed by overnight incubation at 37°C. Overnight cultures were then used for purification of plasmid constructs using a Nucleo-spin plasmid mini prep kit (Macherey - Nagel). Correct plasmid orientation and insertion was checked through restriction enzyme digests of the purified plasmid constructs (Table 2).

Preparation of plasmids based on novel pET vectors

Due to insolubility of proteins using the standard cloning method an alternative approach was taken to produce soluble proteins. This approach relied on the InFusion cloning method and a range of vectors engineered in the Protein production lab of the York Technology Facility (TF) by Jared Cartwright to facilitate rapid cloning into a pET-based vector to create fusions with MBP (pETFPP_2), GST(pETFPP_3), IM9 (pETFPP_4) or GFP (pETFPP_5) and improve solubility.

Achieving In-Fusion cloning involved designing primers which not only amplified the gene of interest during the PCR process, but also incorporated vector complimentary regions flanking the gene in the final PCR product. The vectors were treated by the TF with BseR1 to produce these overhangs for which the adaptors on the PCR products are complementary. During the In-Fusion ligation independent cloning process, the In-Fusion enzyme (Clontech) then anneals the complementary sequences and inserts the gene of interest at the specific position within the vector.

Each target was amplified from genomic DNA using a high fidelity PCR process with KOD polymerase at an annealing temperature of 52 °C. Successful PCR was determined through analysis on an agarose gel and the total PCR reaction mix was treated with 1 µl Dpn1 for 1 hour at a temperature of 37 °C. Treated PCR products were analysed on an agarose gel and correct bands were gel extracted and purified using a gel extraction kit.

1 µl of the PCR product was then incubated with 1 µl of the digested and treated vector at 37 °C in the presence of 2 µl of the InFusion enzyme. The reaction was incubated for 15 minutes before the enzyme was heat inactivated by incubation at 50 °C for 15 minutes. Following InFusion cloning the

resultant plasmids were transformed into XL1-blue supercompetant cells and insert presence was determined through colony screening.

Transformation into XL1-blue supercompetant cells was achieved as follows: 10 µl of XL1-blue cells were thawed on ice for 10 minutes before 1 µl of the purified plasmid construct was added to the cells. The cells were gently mixed before being left on ice for a further 10 minutes. The samples were then heat shocked at 42 °C for a total of 30 seconds and then allowed to recover on ice for 2 minutes. 100 µl of LB was then added to the cells and samples were incubated with shaking at 37 °C for 1 hour. The resultant samples were then plated on kanamycin selective LB plates which were incubated at 37 °C overnight.

Any colonies present of antibiotic selective plates were used to inoculate 5 ml LB and grown overnight at 37 °C. Plasmids were then purified from the cultures using a plasmid mini prep kit and a restriction enzyme digest with Xho1 and Nco1 was carried out to determine correct insertion of gene of interest into the vectors. This was analyzed on a 1% agarose gel. Each of the 10 targets were cloned into 4 different vectors and a series of 40 constructs created.

Cell growth for protein expression trials

Plasmids were initially transformed into either Solopackgold competent cells (during the conventional cloning process) or XL1-blue super competent cells (for the alternative cloning methods). Following transformation colonies grown on antibiotic selective plates were picked and screened (through purification and restriction enzyme digests or whole cell PCR) for correct insert presence. These purified plasmids were then transformed into MC1061 cells (for pBAD based plasmids) or BL21 cells (for pET based plasmids).

Following successful transformation into *E. coli* expression strains, colonies were picked and used to inoculate 5 ml LB. These starter cultures were allowed to grow overnight at 37°C before being used to inoculate 25 ml LB to an OD_{650} of 0.05. 25 ml cultures were then incubated at either 20, 30 or 37 °C with shaking until they reached an OD_{650} between 0.4 and 0.6 at which point expression of the recombinant genes were induced with 0.2 mM isopropylthiogalactoside (IPTG) (pET vectors) or 0.005% arabinose (pBAD vectors).

During expression from conventional vectors 1 ml samples for protein expression analysis were taken every hour for 3 or 4 hours and after 24 hours. When proteins were expressed from the pET vectors containing the fusion partners, cultures were grown at 20 °C before and after induction and

were harvested after overnight growth for expression trials. Protein expression was analysed using SDS PAGE as described above.

Solubility tests

Once proteins were expressing correctly from the conventional vectors, three hour cell pellet samples were re-suspended in 200 µl of Bugbuster (Novagen). The samples were incubated at room temperature for 20 minutes before the addition of 100 µl of a phosphate buffer. After a further 10 minute incubation samples were centrifuged for 1 minute and 10 µl of clear lysate was added to SDS-PAGE loading dye. These samples were boiled at 95 °C and analysed using SDS-PAGE.

In the case of expression and solubility from novel vectors, cells were initially lysed using Bugbuster as described above but lysis was also achieved through the use of a detergent based lysis buffer. The cell pellets were resuspended in lysis buffer (50 mM HEPES, 100 mM KCl, 15 mM NaCl, 10 mM $MgCl_2$, 10 mM $CaCl_2$, 0.5 % [W/V] glycerol, 1 mg/ml lysozyme, 20 µg/ml DNAse and 0.5 % Triton X100) and incubated with agitation for 1 hour. These samples were then analyzed as described above.

Protein purification

Starter cultures were grown for 6 hours before being used to inoculate 25ml cultures. These were allowed to grow at the appropriate temperature overnight with shaking and were then used to inoculate large volumes (600ml or 800ml) to an OD_{650} of 0.05. Cultures were grown to mid log phase and induced as described previously. Cells from these cultures were harvested by centrifugation at 5000 rpm and 4°C for 30 minutes.

Following harvest cells were lysed in lysis buffer, centrifuged at 5000 rpm for 20 minutes at 4 °C and proteins purified from the clear lysate using nickel affinity chromatography via a HisTrap column. The 5 ml HisTrap column was initially equilibrated by running approximately 25 ml equilibration buffer (50 mM Tris, 200 mM NaCl, 20% Glycerol, 5 mM Imidazole) down the column using a peristaltic pump. The clear lysate sample was then run down the HisTrap column to allow binding of the protein His tag to the nickel within the column. The resultant wash through was collected and stored for analysis. Following this, non-specifically bound protein was removed using approximately 100 ml (amounting to 20 column volumes) wash buffer (50 mM Tris, 200 mM NaCl, 20% Glycerol, 20 mM Imidazole) down the column. The wash through was also collected and stored for analysis. Once the non-specific protein binding was eliminated an elution buffer (50 mM Tris, 200 mM NaCl, 20%

Glycerol, 500 mM Imidazole) was run through the column and 10 samples at volumes of 4 ml were collected. Each of the elution fractions were analysed using SDS-PAGE and fractions containing the protein of interests were dialysed over three hours in 50 mM Tris HCL buffer (pH6.8).

Acetylesterase assay

All enzyme assays were carried out in 1 ml cuvette. 0.5 µM of the protein (See Table 2 for volumes) was added to sodium phosphate buffer at pH 6.7 (at a final concentration of 20 mM). These samples were incubated at 40 °C for 10 minutes. Following this pre-incubation 0.5 mM of p-nitrophenyl acetate (pNPA) was added to each cuvette. The absorbance at OD_{405} was measured immediately after addition of substrate and then every 5 minutes for 30 minutes. In between readings samples were kept at 40 °C.

In addition to assaying the purified proteins of interests a control assay was also carried out using various concentrations of an acetylxylan esterase from *Orpinomyces sp.* (Megazyme). 1 mM stock of PNPA was made by diluting 100 mM PNPA (dissolved in methanol) in sodium phosphate buffer (pH 6.7).

Protein	Molecular Weight	Extinction co-efficient	A_{280} of 1 ml purified protein	Concentration in μM	Volume required for 0.5 μM final concentration
BT1full-MBP	119695.8	215480	4.64	21.53	23.2 μl
BT1ext-MBP	72506.3	98335	0.288	2.93	170.6 μl
BT2ext-GST	757773.6	148435	0.275	1.85	270.3 μl
BT4-Im9	65144.6	121155	0.170	1.40	357.1 μl
BT6-Im9	41796.4	52620	0.748	14.22	35.2 μl

Table 2 - Concentrations of proteins for enzyme assays was determined using a calculated extinction co-efficient.

Results

Comparative analysis of sialic acid catabolism genes across the *Bacteroides* genus

Through the sequencing of many hundreds of bacterial genomes and the analysis of their gene content and layout, it has been shown that the *nan* operon has a complex genetic organisation in different organisms. The observation that the operons have very limited synteny is observed in both Gram positive and Gram negative bacteria, suggesting independent evolution of the operon structure prompted by the need to adapt to various environments (Vimr *et al*, 2004). It is logical to assume commensals of the human GI, an environment rich in sialic acids, might contain the appropriate genes for sialic acid utilization.

Using this assumption and the available *Bacteroides* genome sequences in the public databases, the presence of the *nan* operon across the genus as well as positioning of *nanS* in relation to the other genes was established. To achieve this a series of Blast searches, using *T. forsythia, E. coli, H. influenzae* and *B. fragilis nan* genes as search queries, were used to establish the presence of potential sialic acid catabolism genes within the *Bacteroides*. Blast results were then used to construct gene maps of all the species investigated using Xbase and Uniprot for gene orientation and positioning (Fig. 5).

From the analysis it was evident that very little consistency existed in the gene layout between the different species investigated. As expected, most of the species investigated contained the *nanAET* genes together when all three were present in the genome. This appeared to be the only consistency observed between the different species. An exception to this general layout was observed in *B. uniformis* which contained a *nanA* within one cluster and the *nanE* within a second. Interestingly in *B. uniformis* there appeared to be no direct homologue of the *nanT* transporter observed in other species. There was, however, a gene encoding a member of the MFS transporter family adjacent to the *nanE* gene. When looking specifically within *Bacteroides* the closest homologue for the protein encoded by this gene was annotated as a sugar:proton symporter. In many *Bacteroides* these homologues are annotated as arabinose:proton symporter. A wider BLAST search showed that homologues of the protein observed in *B. uniformis* was seen across many different bacterial genus with the closest match seen in *Dysgonomonas gadei* (a member of the Bacteroidetes phylum) with a maximum sequence identity of 68% and *Prevotella oralis* (also belonging to the Bacteroidetes phylum) with a maximum sequence identity of 63%.

Like that seen previously in *B. fragilis,* a select few of the species had the genes involved in sialic acid utilization spread across two clusters. Similar to the gene organization observed in *B. fragilis*, when the genes were seen to be spread across two clusters, one cluster contain the essential catabolic genes *nanAET* while the other cluster included the accessory genes (such as *nanH* and *nanS*) as well as the outer membrane transporter, *nanO* and *nanU*. As described previously the exception to this was observed in *B. uniformis* where the catabolic genes were spread across the two clusters.

In some instances where genes are spread across two clusters, multiple copies of certain *nan* genes was observed. Multiple examples were observed where two copies if the *nanH* genes were spread across two of the clusters. *Parabacteroides merdae* not only contained two *nanH* genes but is the only species investigated that has two *nanA* genes within each of its sialic acid metabolism clusters.

During the analysis only a single example arose, in *B. thetaiotamicron*, where two *nanS* genes were observed within the same sialic acid catabolism gene cluster. This is of particular interest as the *B. thetaiotamicron nan* operon although containing accessory genes for sialic acid scavenging, lacks one of the essential catabolism genes. The cluster contains the *nanE* gene but lacks both the *nanA* and *nanT* genes, meaning *B. thetaiotamicron* is unable to transport the sialic acid into the cell and unable to convert the Neu5Ac to ManNAc.

Many of the *nan* genes are also separated by genes involved in sugar utilization on a more general scale. This is exemplified in *B. fluxus* where the main catabolic genes are separated from the accessory genes by genes mainly denoted as glycosyl hydrolases. In addition to the presence of the glycosyl hydrolases, the genes involved in sialic acid catabolism in *Bacteroides* were also separated by acetylhydrolases, β-mannosidases, β-acetylhexosaminidases, fucosidases and mannosidases.

The most interesting observation was seen when analyzing exceptionally long protein sequences, which not only resulted from NanS blast queries but also showed homology to *E. coli* NanM. Analysis of these sequences showed them to potentially be a fusion protein which could function as both a NanS and a NanM. The protein size correlated with a fusion between the two and the conserved sequences required for functionality of both proteins was easily identified following alignment studies.
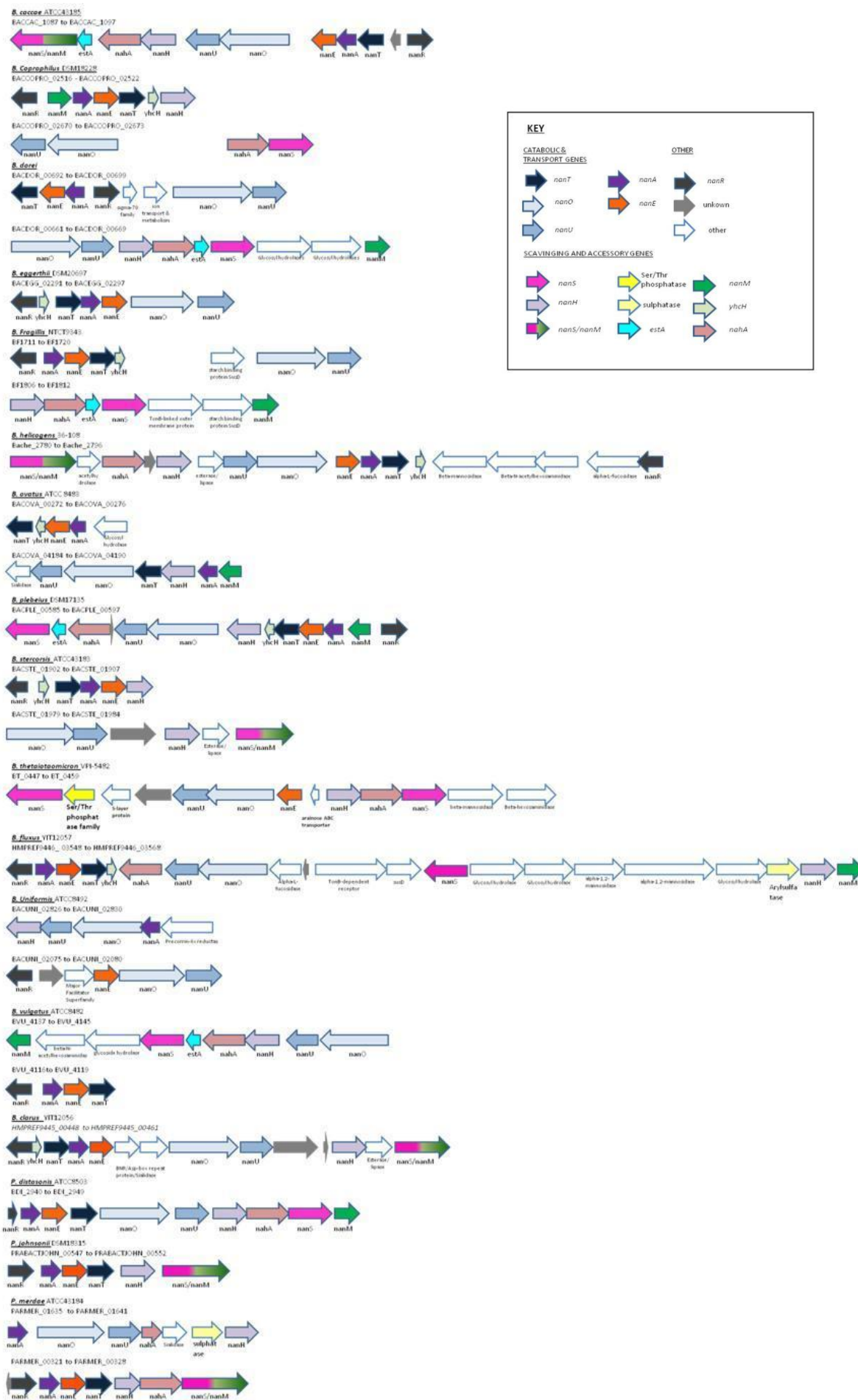
Figure 5 - Sialic acid utilization genes in various Bacteroides species (to scale). The colouring of individual genes is indicated in the key.

36

Analysis of *nanS* genes in the *Bacteroides* clusters

The presence of putative sialate-*O*-acetylesterase in *Bacteroides* was investigated further following the construction of the gene maps. Analysis showed in most cases that a candidate *nanS* gene was present adjacent to other accessory genes, in particular *nanH* and *nahA*.

The NanS proteins were also aligned from various sources to investigate sequence identity between the different NanS proteins. When aligned with *E. coli* NanS there appeared to be very limited sequence identity with a large number of gaps in the *E. coli* protein sequence and very little conservation between the two species (Fig. 5A). During Blast searches using the *E. coli* NanS as a query the maximum sequence identity observed between a hypothetical *Bacteroides* protein and the *E. coli* NanS was a maximum of 34%.

However, when comparing the *Bacteroides* NanS sequences with that of *T. forsythia*, the sequence alignments were almost identical and the expected conserved sequences were evident in all species investigated. Comparing the *T. forsythia* NanS protein with those seen in *Bacteroides* a maximum sequence identity of 65% was observed. In contrast, other genes involved in sialic acid utilization have a higher maximum sequence identity when comparing *T. forsythia* and Bacteroides, for example comparison of NanA proteins shows a maximum identity of 88% between *T. forsythia* and *B. fragilis*.

The expected conserved sequences of group II SGNH hydrolases (the family which NanS is predicted to belong to) are a GQSN sequence in Block I of the protein and QGEX in Block II (Fig. 6). Interestingly there was an exception observed in this alignment, where BT_0447 (a NanS protein of *B. thetaiotamicron*) appeared to align poorly with the other species investigated. On further analysis it was seen that this particular NanS contain an extension at the C-terminal (Fig. 7). Analysis of this extension was seen to contain a domain of unknown function (DUF187). Using this domain as a Blast query against a wide range of bacteria produced significant hits to potential NanS sequences in *Bacteroides* species not previously investigated during the initial studies. Upon further analysis of these protein sequences, it was confirmed these also contained the expected conserved sequences as well as a predicted signal peptide.

Analysis using public databases indicated the DUF187 domain was present in many bacterial species often as part of a protein which also contained a CE4_SF domain (a polysaccharide deacetylase). In the *Bacteroides* species specifically, other than the finding this domain alongside the NanS DUF303 domain, the DUF187 domain is also part of a protein containing a DOMON_like domain which

interacts with sugars in the type 9 carbohydrate binding modules present in a range of glycosyl hydrolases.

Another striking observation was the presence of an N-terminal extension observed in most *Bacteroides* species as well as within the *T. forsythia* NanS protein, which is lacking in the *E. coli* NanS (Fig. 6). A Blast search of this extension shows no conclusive results other then detecting the sequence in various *Bacteroides* NanS. This portion of the protein however is denoted as being part of the SGNH hydrolase superfamily but is different to the DUF303 domain associated with the activity of NanS.

Considering *Bacteroides* NanS could potentially be secreted into the lumen of the GI, it was predicted that these extensions could be involved in aiding the NanS in adhering to gut cell surface sugars in a similar manner to lectin like domain in *V. cholera*. To test this, alignments of the *Bacteroides* NanS containing this extension and *V. cholera* lectin like domains were done with no conclusive results. However, due to the diversity observed in lectin like domains, it is possible these extensions could still be involved in anchoring the enzyme to the surface sugars.

With NanS being the protein of interest, SignalP was used to determine if the various NanS proteins contained a predicted signal peptide. In many cases the NanS found within the *Bacteroides* species were found to have a potential signal peptide. Taking this into consideration alongside the information previously outlined in regards to *Bacteroides* catabolism of sialic acid (Stafford *et al,* 2011) an alternative pathway is predicted to be present in *Bacteroides* when compared to that seen in *E.coli* (Fig. 9).

A

Block I

```
BACDOR_00666       YENVLAGEVWLCSGQSNMEFYLNWSATAAQDVP---QAANSNIRFYDMKARWRTDAVEWD 351
BVU_4140           YENVLAGEVWLCSGQSNMEFYLNWSATAAQDVP---QAANSNIRFYDMKARWRTDAVEWD 351
PRABACTJOHN_00552  YTNVLIGEVWLCSGQSNMAFQVSSAVDSQRKAFLEFAARKPEIRLFDMKPRWATSAVEWD 352
PARMER_00328       YTNILIGEVWLCSGQSNMAFQVSSAVDSQRKAFLEFAARKPQIRLFDMKPRWLTNAVEWD 352
BDI_2948           YRDVVAGEVWLCSGQSNMAFRVNESVKEEQQQQLDYAKQHSQIRLFDLKPRWETYAVEWD 355
BACCAC_01087       YTNVLIGEVWLCSGQSNMEFMLRQTITGKKDIP---QAADEQLRLYDMKARWRTDAVQWD 351
BT_0457            YQNVLAGEVWLCSGQSNMEFMLKQASTARADIP---RAVDQQLRLYDMKARWRTNAVEWE 353
BACCOPRO_02673     FRHVAVGEVWLCSGQSNMEFLLSQATTGKEDIP---QANHPDLRLYNMKGNWATDNQAWQ 354
BACPLE_00585       FHHVAAGEIWLCSGQSNMQFMLQQTQDGSTAVS---QAYNSQIRLFHMKGKWETNASEWP 354
Bache_2780         YNNILAGEVWLCSGQSNMEFRLNNALTAKADIP---EAVNENIRLFDMKGRWATNAVEWD 351
BACSTE_01984       CKNVLVAGEVWLCSGQSNMEFMLKQAATAQTDIP---EANDDKLRLFDMKARWRTNAVAWN 211
BACINT_02833       IKDILIGDVWLCSGQSNMQWVVNNVTNAEVEKK--------NANYPQIRTLNIPRRME 143
BT_0447            LKNILIGEVWFCSGQSNMEMPVKGFRGQPVYGS----------QPYIVSANPKRPLRLYT 148
Ecoli              FNDIIPLTHCPHDVQDMQGYHHPLATNHQTQYG------------------------- 91
```

Block II

```
BACDOR_00666       RPLEQFPVKGFIWYQGESNAHN--RETHEKLFRLLVESWRKNWGDAELPFYFVQLSSIDR 529
BVU_4140           RPLEQFPVKGFIWYQGESNAHN--REAHEKLFGLLVESWRKNWGDAELPFYFVQLSSIDR 529
PRABACTJOHN_00552  RPLEKFPIRGVIWYQGESNAHN--MEAHERLFHLLTKNWRENW-AEELPFYYVQLSSIDR 529
PARMER_00328       QPLEKFPIRGIIWYQGESNAHN--MEAHEKLFHLLTKNWRENW-AEELPFYYVQLSSIDR 529
BDI_2948           QPLHRYPIKGIIWYQGESNAHN--MEVHERLFPLLVNSWRQNW-NADLPFYYVQLSSIDR 532
BACCAC_01087       RPLEQYPVRGVIWYQGESNAHN--CEAHEKLFKLLVGSWRKNWKNEDLPFYYVQLSSIAR 529
BT_0457            RPLEQYPIRGVIWYQGESNAHN--WEAHEKLFKLLVNSWRKNWNDACLPFYYVQLSSLNR 531
BACCOPRO_02673     RTLEQYPIKGVIWYQGESNAHN--MDAHKKLFRLLVESWRRNWQSPELPFYFVQLSSLNR 532
BACPLE_00585       LPLQQYPLKGIIWYQGESNAHN--MEAHYRLFKLLVNGWRDNWNNPDMPFYFVQLSSLNR 532
Bache_2780         RPLEQFPLKGVIWYQGESNAHN--KDAHGRLFRLLVDSWRRNWDNADMPFYFVQLSSIDR 529
BACSTE_01984       LPLSKYPLKGVINYQGESNTHN--KEAHEKLFKLLIDSWRSNWEQPNLPFYYVQLSSIDR 389
BACINT_02833       HPLLSFSIKGVLWYQGENNVGN--AEPYTDWLTCMIGDWRNRWN-SELPFYIIQLPNFDS 307
BT_0447            NPWEGFPVKGVIWYQGEANTPD--PTLYKKLFPAMVSQWRNFFHNAEMPFYYVQIAPWKS 306
Ecoli              AKNPQNKFLGACWMQGEFDLMTSDYASHPQHFNHMVEAFRR-----DLKQYHSQLNNITD 219
```

B

Block I

```
BACCAC_01087       PHRIL----KYTNVLIGEVWLCSGQSNMEFMLRQTITGKKD------IPQAADEQLRLYD 338
BT_0457            DETGF----QYQNVLAGEVWLCSGQSNMEFMLKQASTARAD------IPRAVDQQLRLYD 340
BACDOR_00666       EKRKL----VYENVLAGEVWLCSGQSNMEFYLNWSATAAQD------VPQAANSNIRFYD 338
BVU_4140           EKRKL----VYENVLAGEVWLCSGQSNMEFYLNWSATAAQD------VPQAANSNIRFYD 338
Bache_2780         VGRKL----VYNNILAGEVWLCSGQSNMEFRLNNALTAKAD------IPEAVNENIRLFD 338
BACCOPRO_02673     GKEKR----TFRHVAVGEVWLCSGQSNMEFLLSQATTGKED------IPQANHPDLRLYN 341
BACPLE_00585       PQQSH----TFHHVAAGEIWLCSGQSNMQFMLQQTQDGSTA------VSQAYNSQIRLFH 341
PRABACTJOHN_00552  ASGKL----DYTNVLIGEVWLCSGQSNMAFQVSSAVDSQRKAFL---EFAARKPEIRLFD 339
PARMER_00328       ESGKL----DYTNILIGEVWLCSGQSNMAFQVSSAVDSQRKAFL---EFAARKPQIRLFD 339
BDI_2948           PSRTL----NYRDVVAGEVWLCSGQSNMAFRVNESVKEEQQQQ----DYAKQHSQIRLFD 342
BACSTE_01984       GKQTL----SCKNVLVGEVWLCSGQSNMEFMLKQAATAQTD------IPEANDDKLRLFD 198
BACINT_02833       KENPI----SIKDILIGDVWLCSGQSNMQWVVNNVTNAEVE------KKNANYPQIRTLN 137
SIAE1              TLEKINFTLRVHDVLFGDVWLCSGQSNMQMTVLQIFNATRELS-----NTAAYQSVRILS 121
sIAE2              TLEKINFTLRVHDVLFGDVWLCSGQSNMQMTVLQIFNATRELS-----NTAAYQSVRILS 156
BT_0447            GEELT-----LKNILIGEVWFCSGQSNMEMPVKGFRGQPVYGSQPYIVSANPKRPLRLYT 148
```

Block II

```
BACCAC_01087       HPYEPCYLYESGIRPLEQYPVRGVIWYQGESNAHN-CEAHEKLFKLLVGSWRKNWKN--- 514
BT_0457            HPYEPCYLYESGIRPLEQYPIRGVIWYQGESNAHN-WEAHEKLFKLLVNSWRKNWND--- 516
BACDOR_00666       HPYEPCYLYESGIRPLEQFPVKGFIWYQGESNAHN-RETHEKLFRLLVESWRKNWGD--- 514
BVU_4140           HPYEPCYLYESGIRPLEQFPVKGFIWYQGESNAHN-REAHEKLFGLLVESWRKNWGD--- 514
Bache_2780         HPYEPCYLYEAGIRPLEQFPLKGVIWYQGESNAHN-KDAHGRLFRLLVDSWRRNWDN--- 514
BACCOPRO_02673     HPYEPCYLYEAGIRTLEQYPIKGVIWYQGESNAHN-MDAHKKLFRLLVESWRRNWQS--- 517
BACPLE_00585       HPYEPCYLYEAGILPLQQYPLKGIIWYQGESNAHN-MEAHYRLFKLLVNGWRDNWNN--- 517
PRABACTJOHN_00552  HPYEPCYLYESGIRPLEKFPIRGVIWYQGESNAHN-MEAHERLFHLLTKNWRENWA---- 514
PARMER_00328       HPYEPCYLYESGIQPLEKFPIRGIIWYQGESNAHN-MEAHEKLFHLLTKNWRENWA---- 514
BDI_2948           HPYEPCYLFEAGIQPLHRYPIKGIIWYQGESNAHN-MEVHERLFPLLVNSWRQNWN---- 517
BACSTE_01984       HPYEPCFLFESGILPLSKYPLKGVIWYQGESNTHN-KEAHEKLFKLLIDSWRSNWEQ--- 374
BACINT_02833       RNDYPSLVYNAMIHPLLSFSIKGVLWYQGENNVGN-AEPYTDWLTCMIGDWRNRWN---- 292
SIAE1              -PSKHSVLWNAMIHPLCNMTLKGVVWYQGESNINYNTDLYNCTFPALIEDWRETFHRGSQ 274
sIAE2              -PSKHSVLWNAMIHPLCNMTLKGVVWYQGESNINYNTDLYNCTFPALIEDWRETFHRGSQ 309
BT_0447            AAGTPTLLWNAMVNPWEGFPVKGVIWYQGEANTPD-PTLYKKLFPAMVSQWRNFFHN--- 291
```

Figure 6 - Sequence allignment comparing *E. coli* and human NanS with a select few examples of *Bacteroides* NanS proteins. A. Block I and Block II *Bacteroides* conserved sequences aligned with *E. coli*. **B.** Block I and Block II *Bacteroides* conserved sequences aligned with human sialate-*O*-acetylesterase (SIAE).

Block I

```
BACCOPRO_02673    GKWEVTLNPLQTGND-YELNISAGKEKRTFRHVAVGEVWLCSGQSNMEFLLSQATTGKED 328
BACPLE_00585      GKWEVILDPIKAGDN-YELSIQTPQQSHTFHHVAAGEIWLCSGQSNMQFMLQQTQDGSTA 328
BACDOR_00666      GKWAVDVRPLKAGGP-YTLTVSTEKRKLVYENVLAGEVWLCSGQSNMEFYLNWSATAAQD 325
BVU_4140          GKWAVDVRPLKAGGP-YTLTVSTEKRKLVYENVLAGEVWLCSGQSNMEFYLNWSATAAQD 325
BFAG_01068        GDWAITLQPLKAGGP-YTLTVSAGKQKQAFNNVLAGEVWLCSGQSNMEFYLGWSKTAKRD 325
BT_0457           GKWSVTLPPLKAGGP-YTLKISTDETGFQYQNVLAGEVWLCSGQSNMEFMLKQASTARAD 327
BDI_2948          GKWTVTLDPLRVSGKSYTLTVSTPSRTLNYRDVVAGEVWLCSGQSNMAFRVNESVKEEQQ 326
Tannerella        GRWTVTLDPMPAQTG-LVLTVATPSRKLTYRQVAVGEVWLCSGQSNMAFRTSEADAAERA 326
BACINT_02833      GYWEVNLKALETGGP-YQLQIIGKENPISIKDILIGDVWLCSGQSNMQWVVNNVTNAEVE 124
BT_0447           PFYYVQIAPWKSEGNDKLDWAWFRQCQLELMKEVPGVGMVTTGDAGSELFIHSPYKIKVG 354
                                 .                             .         :
```

Block II

```
BACCOPRO_02673    KGVIWYQGESNAH-------NMDAHKKLFRLLVESWRRNWQSPELPFYFVQLSSLN---- 531
BACPLE_00585      KGIIWYQGESNAH-------NMEAHYRLFKLLVNGWRDNWNNPDMPFYFVQLSSLN---- 531
BACDOR_00666      KGFIWYQGESNAH-------NRETHEKLFRLLVESWRKNWGDAELPFYFVQLSSID---- 528
BVU_4140          KGFIWYQGESNAH-------NREAHEKLFGLLVESWRKNWGDAELPFYFVQLSSID---- 528
BFAG_01068        KGIIWYQGESNAH-------NREAHEKLFKLLVESWRKNWENKDLPFYYVQLSSIN---- 528
BT_0457           RGVIWYQGESNAH-------NWEAHEKLFKLLVNSWRKNWNDACLPFYYVQLSSLN---- 530
BDI_2948          KGIIWYQGESNAH-------NMEVHERLFPLLVNSWRQNWN-ADLPFYYVQLSSID---- 531
Tannerella        KGVLWYQGESNAH-------NLEAYSKLFPMLINSWRSYWR-ETLPFYFVQLSSLN---- 531
BACINT_02833      KGVLWYQGENNVG-------NAEPYTDWLTCMIGDWRNRWN-SELPFYIIQLPNFDSINK 310
BT_0447           KGIIPITEEKHKYGAMINPLNEEYRTHILNVLKEVVTKYPDLDGLMLDRVRYDGITADFS 654
                  :*.:      *.:        *  :     :  :           : :  ::  .:
```

Figure 7 - Sialate-*O*-acetylesterases contain consensus sequences expected of SGNH hydrolases. Sequence alignment of a select few *Bacteroides* species in comparison to *T. forsythia* shows the consensus sequence in both Block I and Block II required for enzymatic activity (boxed residues).

40
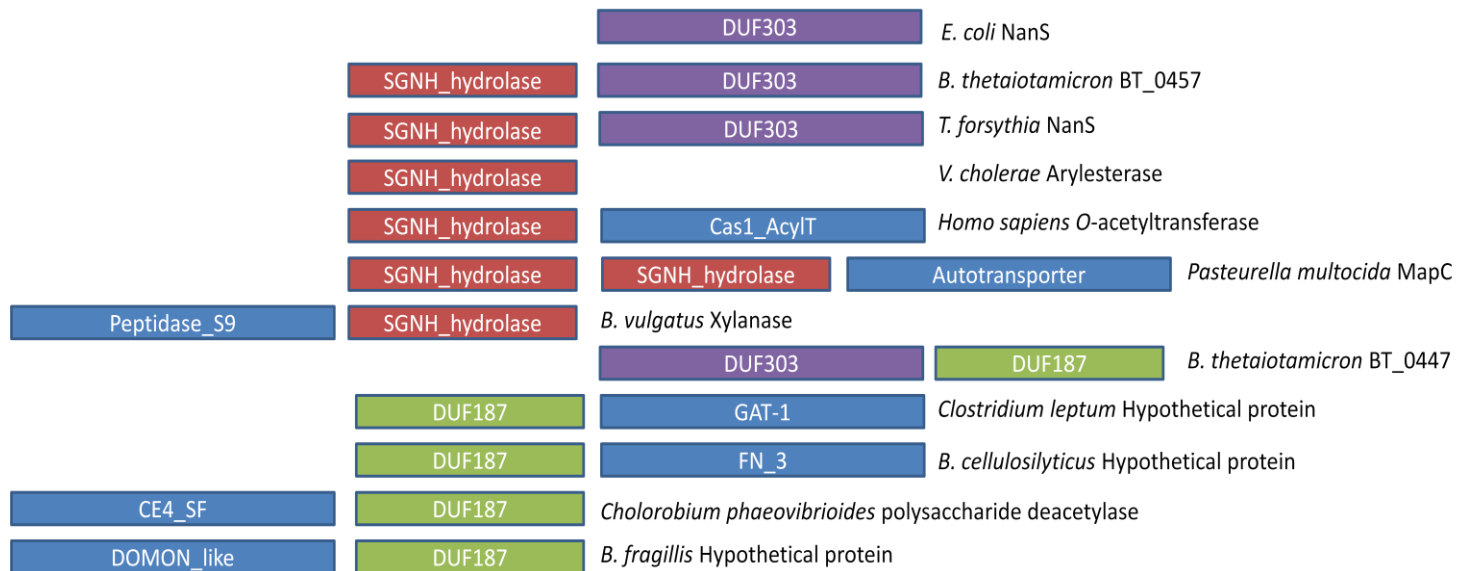
Figure 8 - Comparison of the *E. coli* NanS protein domain to those observed in *Bacteroides* and their presence in other proteins. The various additional domains found in the *Bacteroidies* NanS proteins which are absent in *E. coli* are found elsewhere in nature. In many cases these domains are associated with other domains involved in carbohydrate utilization. GAT-1 is a type 1 glutamineamidotransferase, CES4_SF belongs to the carbohydrate esterase family 4, DOMON_like domain is a ligand binding domain which interacts with sugars and hemes, FN3 (fibronectin type 3 domain) is found in many bacterial hydrolases. Not to scale.
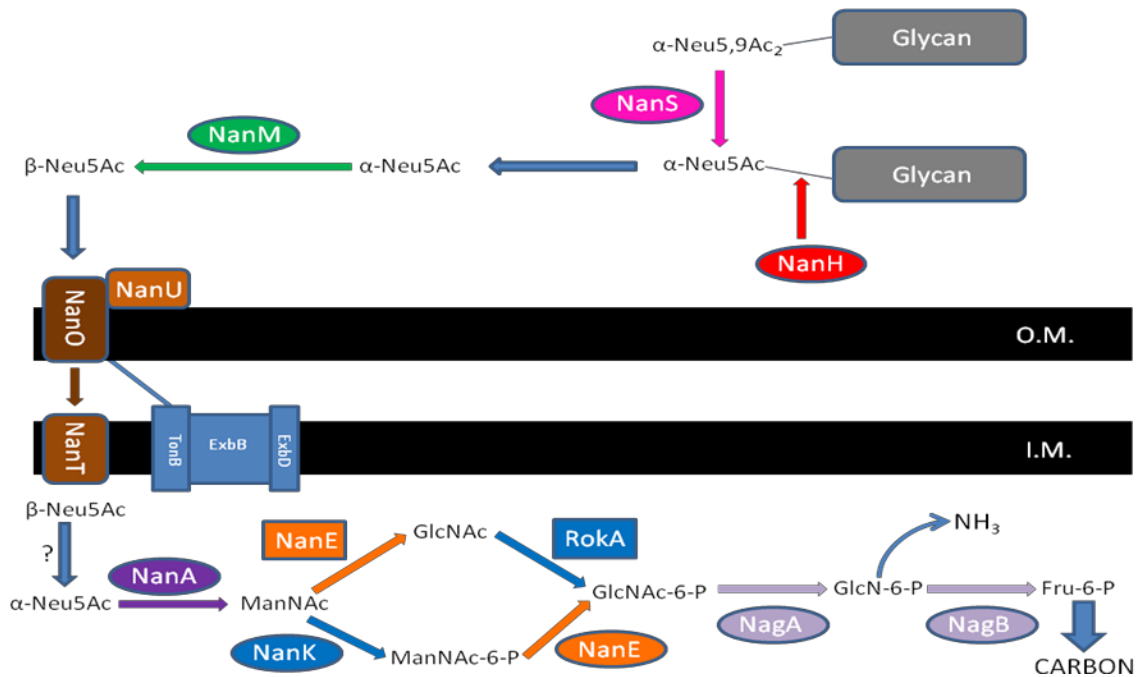
Figure 9 – A general sialic acid catabolism pathway in *Bacteroides*. In *Bacteroides* a novel NanE has been identified which does not require ManNAc-6-P as a substrate. Instead the NanE directly acts on the ManNAc to produce GlcNAc which is then phosphorylated to GlcNAc-6-P by RokA (Brigham *et al*, 2009). In addition to this *Bacteroides* is also predicted to contain a secreted sialidase and esterase involved in scavenging of sialic acids.

Discovery of novel sialic acid utilization genes in the *Bacteroides*

As mentioned previously, sialic acid can be modified with various other groups aside from acetylation (Shauer, 2004). In a similar manner to acetylation of Neu5Ac causing a decrease in NanH activity, it is likely that addition of various other groups can also hinder sialidase activity. If this is the case it would be logical for Neu5Ac utilizing bacteria to have a way to overcome this. The presence of genes encoding potential enzymes involved in priming Neu5Ac for NanH function was observed in certain *nan* gene clusters in *Bacteroides* (Fig. 4). These genes could be of significant importance given the diversity of sialic acid and the presence and positioning of these genes in the *Bacteroides nan* clusters.

As indicated in figure 4, *P. merdae* appears to have arylsulphatase situated upstream of the sialidase. Interestingly homologues of the arylsulphatase were observed mainly in *Sphingobacterium*, *Blastopirellula*, *Rhodopirellula* and *Pedobacter.* Although direct homologues were rarely seen in *Bacteroides*, the species with the maximum sequence identity belonged to the *Bacteroidetes* phylum (the phylum to which *Bacteroides* belongs). In addition to the presence of this sulphatase within the sialic acid catabolism cluster, only one other *Bacteroides* (of those investigated) was seen to have a sulphatase within the cluster. Given that sialic acid can be modified by the addition of a sulphate group, the presence of a sulphatase would increase the sialic acid utilization potential of these *Bacteroides.*

A Ser/Thr phosphatase was observed adjacent to one of the potential NanS enzymes of *B. thetaiotamicron*. This protein was indicated to be part of the Calcineurin-like phosphoesterase family. Upon further analysis, homologues of this protein was observed in many different *Bacteroides* species but not seen within any other sialic acid *nan* operon in the species investigated. Homologues of this protein were also seen in many *Prevotella* and *Tannerella* species, both of which utilise sialic acid as a nutrient source. The presence of a phosphatase with sialic acid specificity would give these bacteria a competitive edge within their environment.

Cloning of *nanS* candidates from *B. thetaiotamicron*

Genomic analysis has indicated that homologues of the *E. coli* NanS enzymes are ubiquitous in bacteria of the *Bacteroides* genus. The most well studied commensal in the *Bacteroides* genus is *B. thetaiotamicron*, it is also one of the most dominant *Bacteroides* species in the gut accounting for 6% of all bacteria and 12% of all *Bacteroides* (Eckburg *et al*, 2005). Using amino acid sequences for

gene products involved in sialic acid catabolism from *E. coli, Tannerella forsythia* and *B. fragilis* as Blast queries, a gene map of potential *B. thetaiotamicron* sialic acid operon was constructed (Fig. 5). This analysis indicated the presence of two potential *nanS* genes within the same cluster.

Analysing the amino acid sequences of these putative NanS proteins showed only one of the two to contain a strong potential signal peptide (BT_0457). Both BT_0447and BT_0457 are longer in length in comparison to the *E.coli* NanS. BT_0457 not only has a signal peptide but also contains an N-terminal extension while the other (BT_0447) appears to have a C-terminal extension of approximately 200 amino acids (Fig. 7).

These finding formed the basis for the initial experiments. The aim was to successfully clone both genes into two vectors; the pBADcLIC vector which tags the C-terminal of the protein with decahistidin and the pETYSBLIC-3C which tags the N-terminal with hexahistidine. Cloning into the pETYSBLIC-3C vector requires the removal of the signal peptide (if one is predicted to be present) for cytoplasmic expression.

First we designed primers for cloning of the entire gene length and LIC cassette for both genes into the two above described vectors. These were then used to amplify the genes from genomic DNA using a KOD polymerase PCR reaction at a 55°C annealing temperature. The PCR products were purified and genes of interest were introduced to the expression vectors, followed by transformation initially into *E. coli* supercompetent cells (as described in Material and Methods).

Colonies present following transformations were screened using whole cell PCR for the uptake of the correct plasmid. The presence of the recombinant plasmid was determined by the appearance of bands corresponding to the expected size of the BT_0447 and BT_0457 inserts. The plasmids from the colonies shown to potentially contain the recombinant plasmids were purified and correct orientation and position of insert was established using the appropriate restriction enzyme digests (Fig. 8 and Table 2). It is important to note that although all the appropriate band were seen faintly during early stages of migration during electropherisis analysis of the restriction enzyme digests, the bands are not visible during later stages of migration due to the poor staining of the cybersafe staining that was used during the initial stages of the study.

Using whole cell PCR and restriction enzyme digest it was concluded that the pBAD-BT0457 and pBAD-BT0447 constructs were as expected. Confirmation of recombinant plasmids was achieved through sequencing before the plasmids were transformed into the appropriate expression *E. coli* strains (as described in Material and Methods).

44
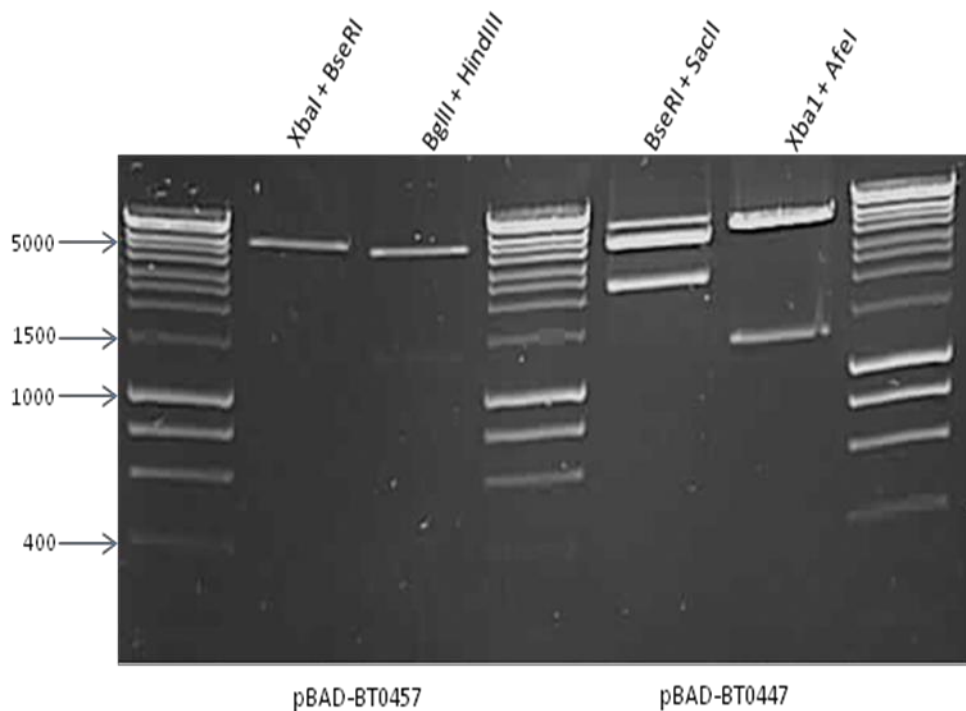
Figure 10 – Restriction enzyme digests of pBAD-BT0457 and pBAD-BT0447. Two double restriction enzyme digests were carried out on each of the recombinant plasmids. The XbaI and BseRI digest of pBAD-BT0457 was expected to produce bands of sizes 5514bp and 704bp. Although observed each time the digests were repeated, the smaller band (at approximately 700bp) was difficult to visualise due to poor staining. BglII and HindIII digest of pBAD-BT0457 was expected to produce bands of sizes 1297bp and 4921bp. BseRI and SacII digest of pBAD-BT0447 was expected to produce bands of sizes 2282bp and 4512bp. In this case a third band is also present which correlates to the size expected of an undigested plasmid. XbaI and AfeI digest of pBAD-BT0457 was expected to produce bands of sizes 1239bp and 5555bp.

B. *thetaiotamicron* candidate sialate-*O*-acetylesterases can be successfully expressed in E*. coli* but are insoluble

Once the recombinant plasmids were successfully transformed into the expression strains of *E. coli*, protein expression was attempted. Optimal expression was determined through bacterial growth at 37°C, 30°C and 21°C with growth measured hourly over a three hour period. To achieve this cells were grown in LB to mid log phase and expression induced with either arabinose (pBAD vector constructs) or IPTG (pET vector constructs). Cell samples taken hourly were then tested for protein expression through SDS-PAGE analysis (as described in Material and Methods).

Analysis of Coomassie-stained SDS-PAGE gels indicated that both BT_0457 (78.62 Kda) and BT_0447 (101.38 Kda) accumulated to higher levels after growth at 30°C compared to 21°C when expressed in either the pBAD or the pET vectors. There is a clear band in each case at the expected band size with increased expression at every hour for three hours following induction (Fig. 10 & 11). To maximise protein expression, solubility tests were carried out at both temperatures for all constructs (Fig. 12 & 13). Analysing these tests indicated only BT_0457 expressed from the pBAD vector appeared to be soluble when expressed at 21°C (Fig. 12).

The pBAD-BT0457 construct was then expressed (at 30°C) and purified using nickel affinity chromatography as described in Material and Methods (Fig. 13). Following gel analysis elution fraction 2, containing a band corresponding to the expected size of the BT_0457 protein, was gel extracted and the protein identified through digestion with trypsin by the proteomics lab at the University of York Technology Facility. Unfortunately, the excised protein in this case was not as expected and instead was identified as a KatE enzyme.

Due to the difficulties faced in obtaining soluble proteins of interest using the vectors described above an alternative approach was taken in an attempt to increase protein solubility.
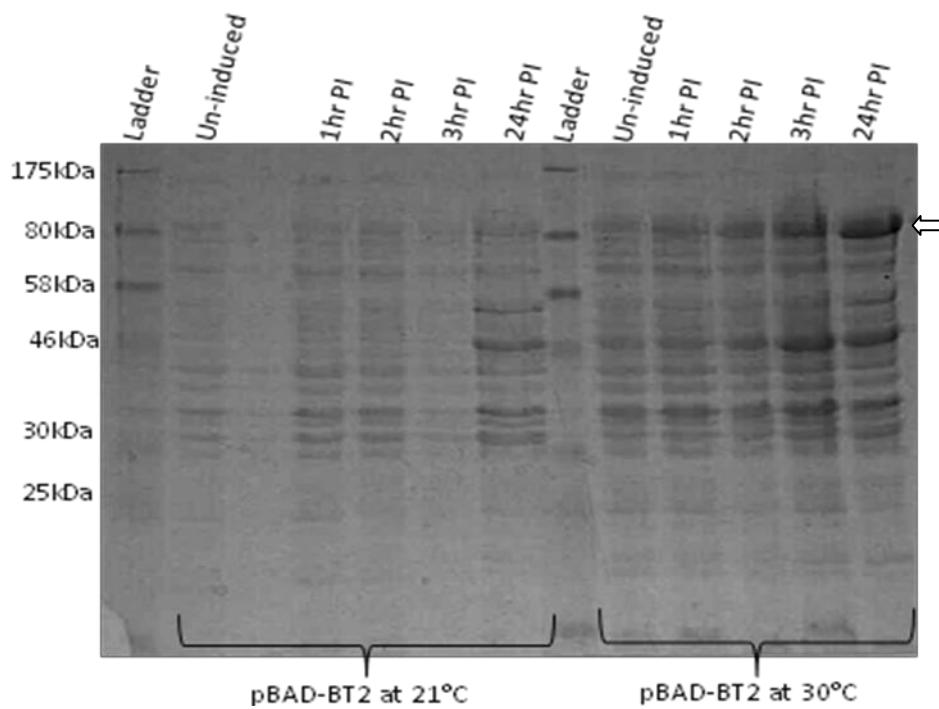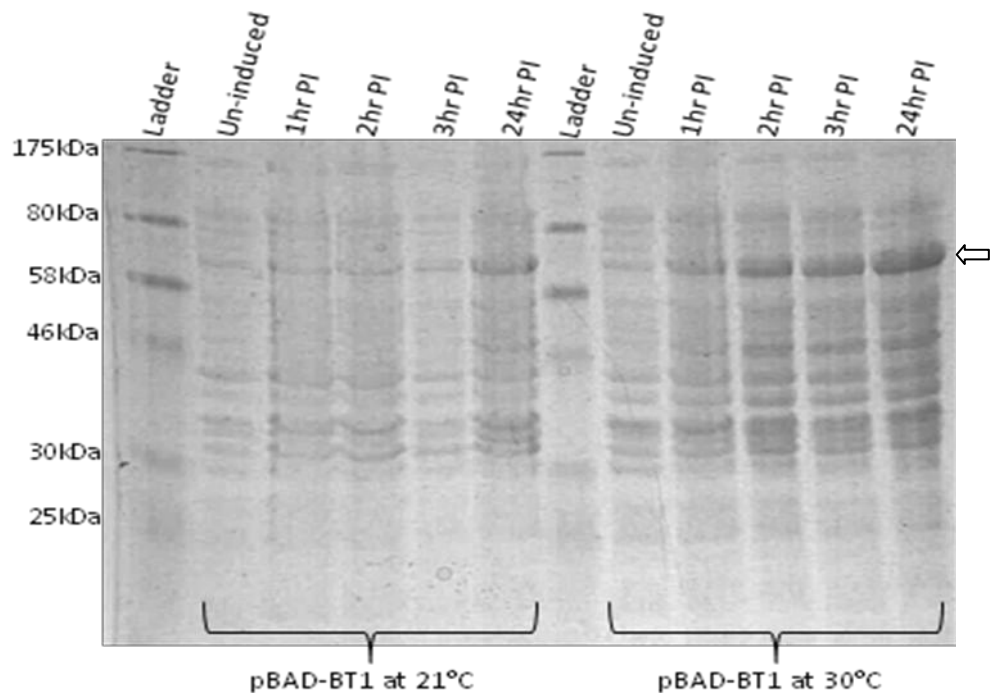
Figure 11 - Expression of pBAD-BT0447 and pBAD-BT0457 at varying temperatures. Protein expression was tested 21 °C and 30 °C over three hours post induction (PI) and after 24 hours. BT1 is protein BT_0457 and BT2 is protein BT_0447. Protein of interest is indicated by an arrow.
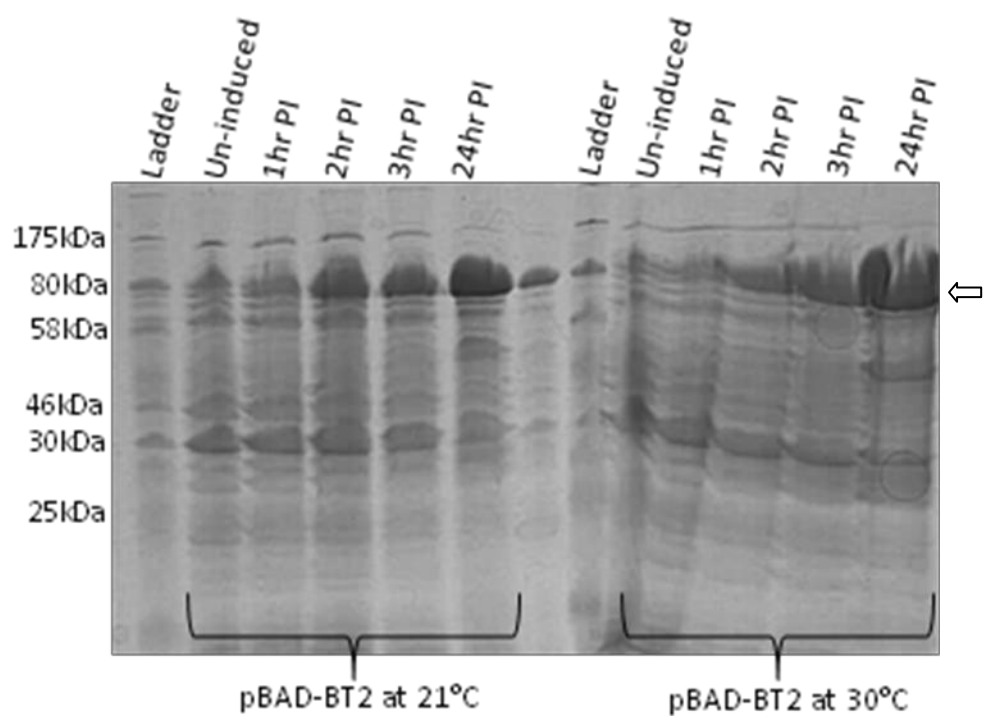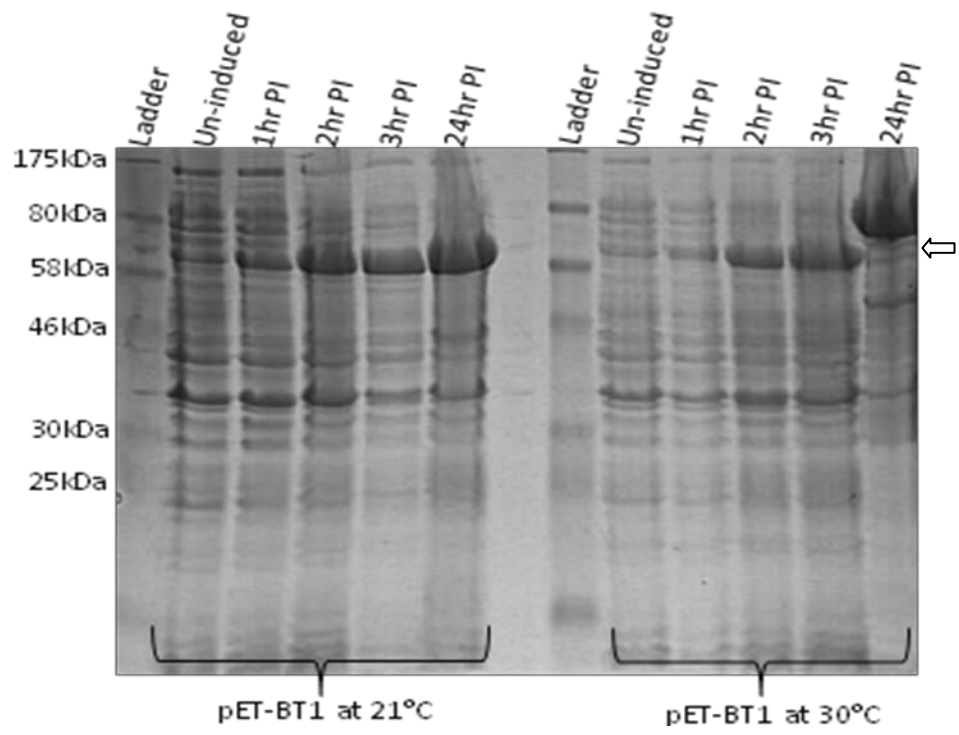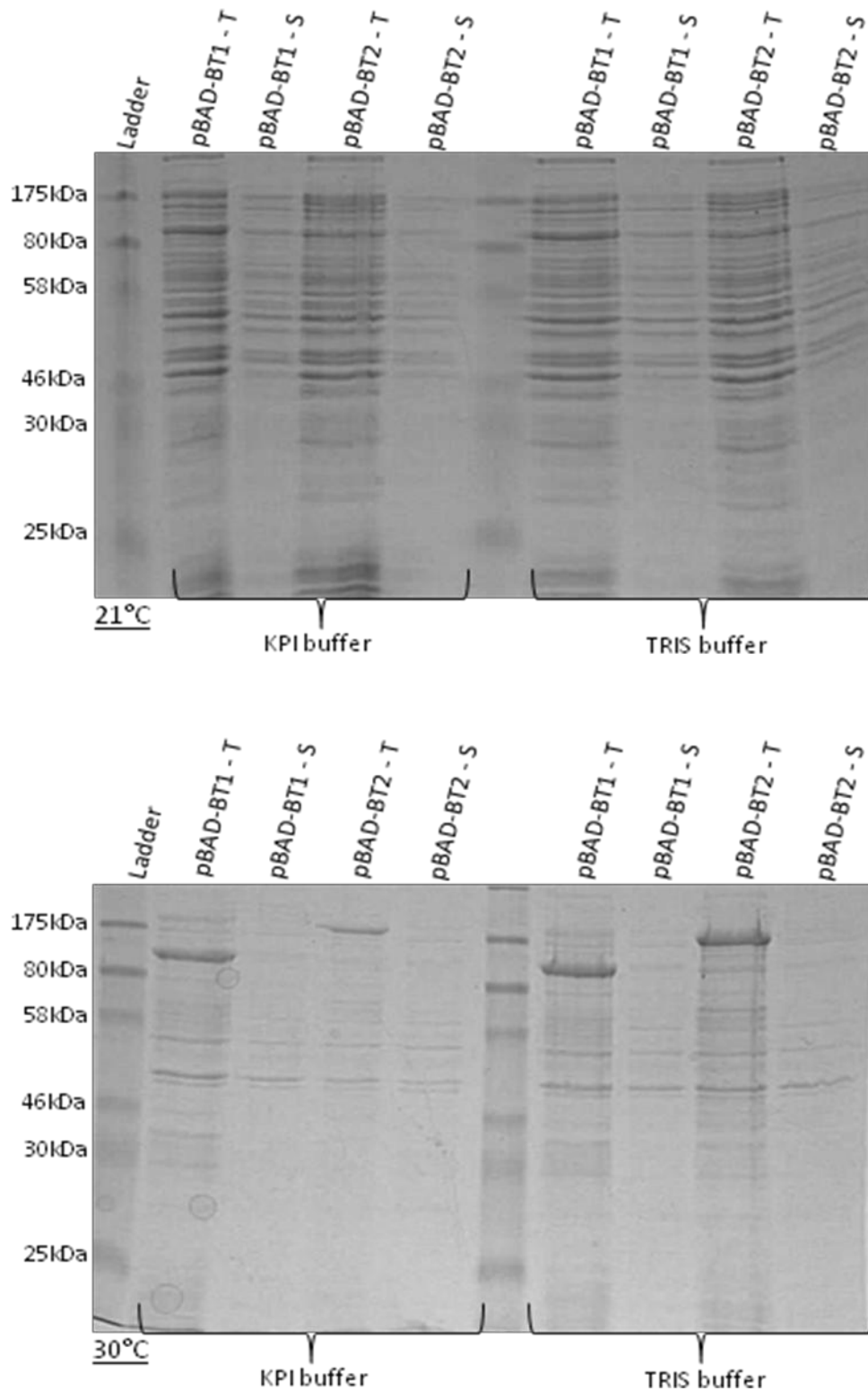
Figure 12 - Expression of pET-BT0447 and pET-BT0457 at varying temperatures. Protein expression was tested 21 °C and 30 °C over three hours post induction (PI) and after 24 hours. BT1 is protein BT_0457 and BT2 is protein BT_0447. Protein of interest is indicated by an arrow.

Figure 13 - Protein solubility of proteins expressed in pBAD vectors under various conditions.

Protein solubility was tested with Tris and a phosphate buffer at 21 °C and 30 °C. BT1 is protein BT_0457 and BT2 is protein BT_0447.
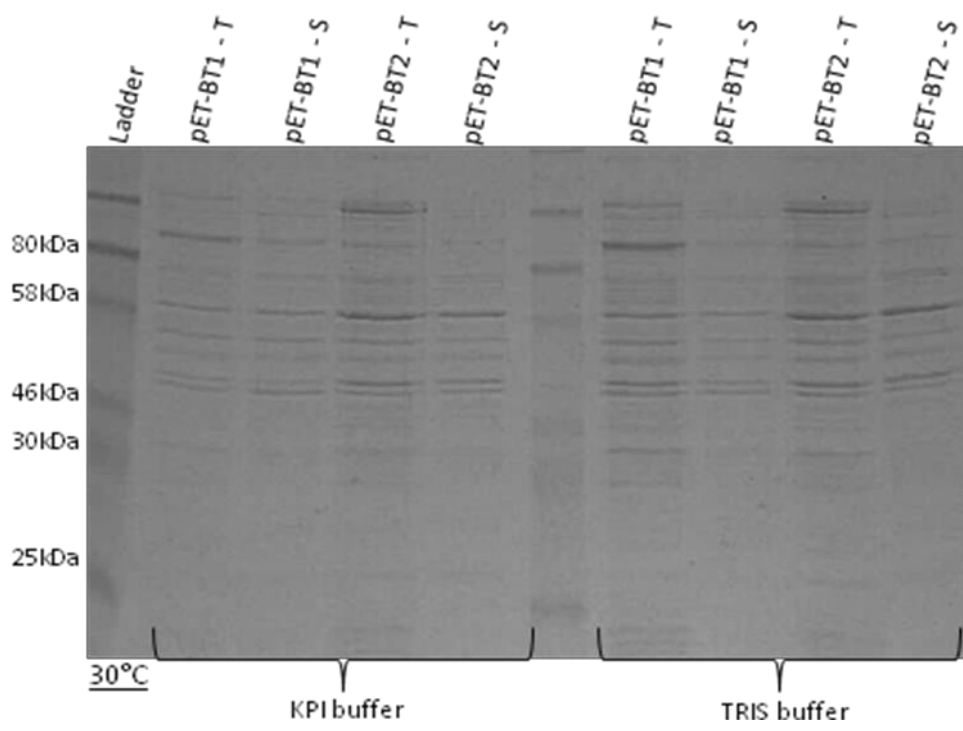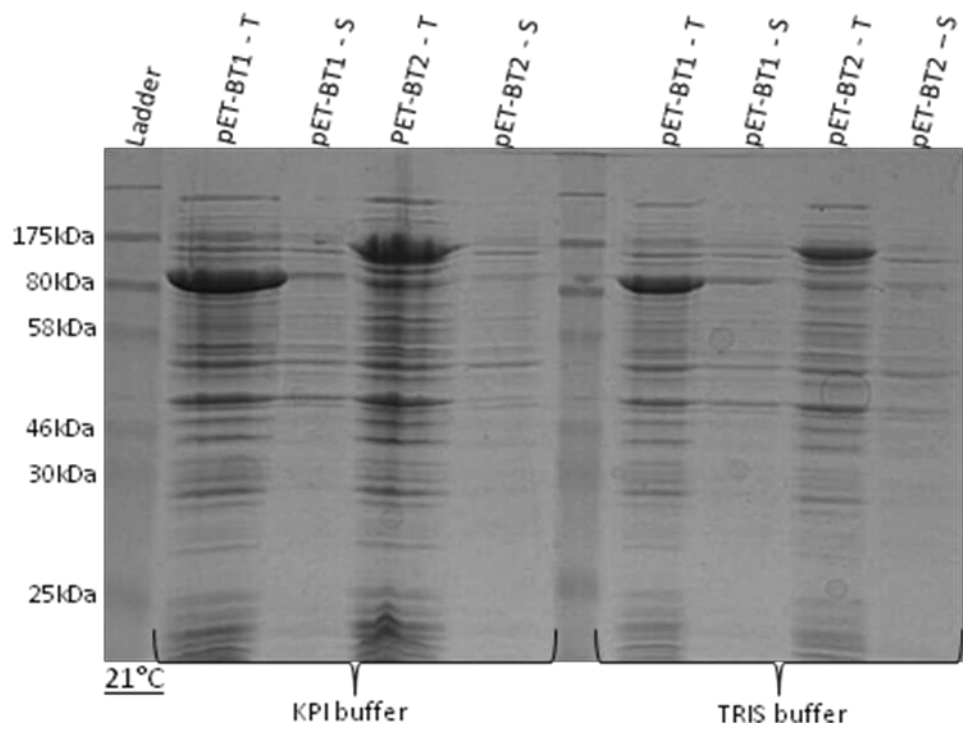
Figure 14 - Protein solubility of proteins expressed in pET vectors under various conditions. Protein solubility was tested with Tris and a phosphate buffer at 21 °C and 30 °C. BT1 is protein BT_0457 and BT2 is protein BT_0447.
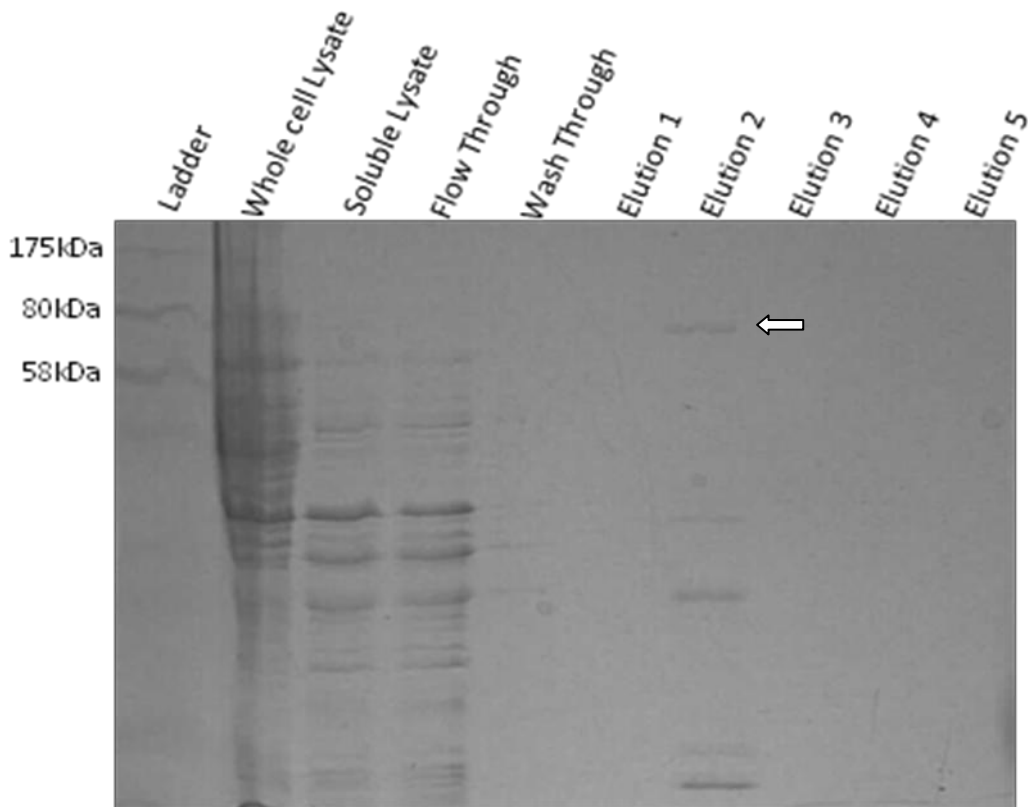
Figure 15 - Purification of pBAD-BT0457. Following large scale expression, cells were lysed in phosphate buffer. The resultant lysate was purified using nickel affinity chromatography and each fraction analysed on a 12% acrylamide gel. A band corresponding to the expected protein size was seen in the second elution (as indicated by the arrow).

Identification of additional putative acetyl esterases in *B. thetaiotamicron*

During the initial analysis of the *Bacteroides* NanS proteins, it was established that the well characterized bacterial NanS (from *E. coli*) has limited sequence similarity with those seen in *Bacteroides* (Fig. 6A). Aligning the *Bacteroides* NanS proteins with various mammalian sialate-O-acetylesterases, however, showed greater similarity at 44% maximum sequence identity when comparing the *B. thetaiotamicron* NanS with the SIAE protein in humans (Fig. 6B). Although still low this maximum sequence identity is higher than that observed when *Bacteroides* NanS is compared to the *E. coli* NanS where a maximum sequence identity of 34%, for the highest BLAST search result, was observed.

Mammalian sialate-O-acetylesterases are found in two forms – cytosolic and lysosomal – both of which are coded by the same gene and expressed differently in various tissues (Takematsu *et al*, 1999). Although conserved residues and enzymatic residues have been identified, very little is known about the structure of the protein (Zhu *et* al, 2003). Upon further analysis using Pfam, it was found that the only domain annotated to be present in the human sialate 9-O-acetylesterase (SIAE) was a domain of unknown function (DUF303) which almost spanned the entire length of the protein. This domain was also observed in various *Bacteroides* candidate NanS enzymes.

A deeper analysis of the DUF303 domain within *Bacteroides* identified its presence in 204 protein sequences from 45 species of *Bacteroides* (Fig. 17). These were designated by Pfam as either sialic acid 9-O-acetylesterase, acetyl xylan esterase or as uncharacterized proteins. Looking specifically at the *B. thetaiotamicron* genome, there are likely 10 proteins that contain the DUF303 domain. These not only aligned well with each other but also showed significant sequence similarity with other NanS proteins in various other *Bacteroides* species.

It was decided a single representative would be taken from each of the major clades observed when a tree of the proteins containing the DUF303 domain was constructed (Fig. 16). In addition to cloning the full length protein for all the chosen proteins, in the case of BT_0447 and BT_0457, each different domain was also decided to be cloned and expressed individually.

Use of a range of protein fusion partners to increase solubility of recombinant sialate *O*-acetylesterases

Due to poor solubility of recombinant esterases using simple expression systems, we chose to investigate the solubility using In-Fusion cloning. This utilised a modified set of pET-based plasmids

that have been engineered in the Protein production lab of the York Technology Facilty (TF) by Dr Jared Cartwright. This allowed the same amplified PCR product to be ligated into one of 4 similar vectors that create genetic fusions to Maltose Binding Protein (MBP), Glutathion S-transferase (SGT), Immunity protein 9 (Im9) and Green Fluorescent Protein (GFP).

Achieving In-Fusion cloning involved designing primers which not only amplified the gene of interest during the PCR process, but also incorporated vector complimentary regions flanking the gene in the final PCR product. The vectors were treated with BseR1 to produce these overhangs for which the adaptors on the PCR products are complementary. During the In-Fusion ligation independent cloning process, the In-Fusion enzyme then anneals the complementary sequences and so inserts the gene of interest at a specific position within the vector (as described in Material and Methods).

Each target, amplified from genomic DNA using high fidelity PCR, was treated with Dpn1 and PCR cleaned up using a mini-prep kit. PCR products were then cloned into the four vectors and transformed into XLI-blue supercompetant *E. coli* cells. Insert presence was determined through restriction enzyme digests of purified plasmids from colonies present following the transformation.
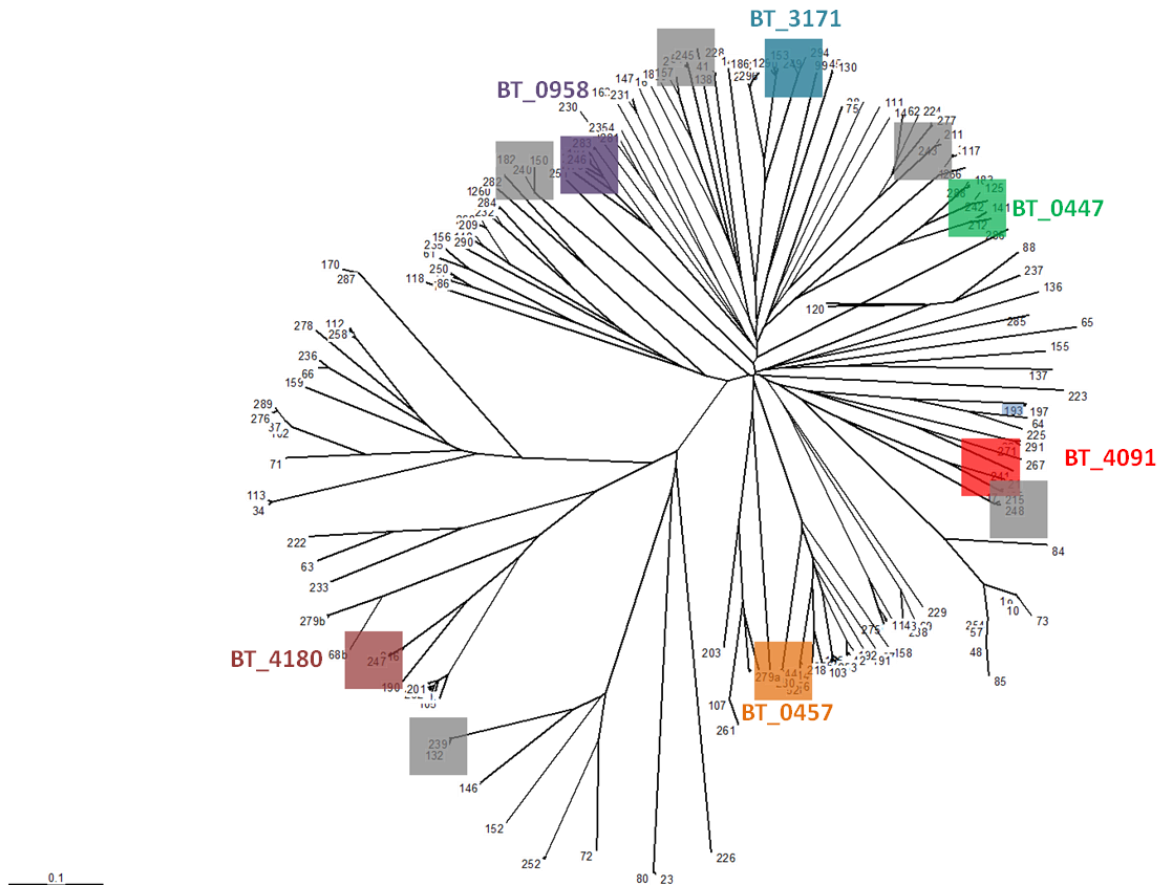
Figure 16 - The DUF303 domain is found across many species of the Bacteroides. Out of the 11 proteins identified in *B. thetaiotamicron* (grey and colored blocks) 5 targets were chosen for characterization (colored blocks). Each of the targets was chosen as a representative of the various clades observed.

Putative acetylesterases were successfully expressed using the Infusion cloning

Following PCR amplification of the selected targets electrophoresis was used to determine correct product size (Fig. 17A). The PCR products were extracted and purified to ensure maximum efficiency during the plasmid construction process. Following successful plasmid construction as determined by plasmid restriction enzyme digests (Fig. 17B); plasmids were purified and transformed into XL1-blue supercompetent cells. Successful transformations were used to purify recombinant plasmids, which were introduced to BL21 (DE3) *E. coli* cells for overexpression of proteins. Following growth and protein expression induction, cells were harvested and the lysate analysed using SDS PAGE (as described in Material and Methods). Figure 18 shows an example of the results obtained during this process. Using this method a series of 40 plasmid constructs were created to express the 10 targets with their various fusion partners (Table 3).

Once we had observed evidence for recombinant protein synthesis (Fig. 18), we tested protein solubility (Fig. 19). Initially the cells were lysed using Bugbuster with the addition of a phosphate buffer (Fig. 20A). Using this method resulted in partial solubility or insolubility of the proteins. This lead to the possibility of salts within the phosphate buffer interfering with the solubility and so the samples were re-tested following lysis with Bugbuster alone (Fig. 20B). As shown the removal of the phosphate buffer appeared to make no difference in the solubility.

Due to limited success of solubility using Bugbuster as a lysis method, various other methods were investigated to determine if solubility could be increased. A basic lysis buffer was made to include lysozyme, DNAse and a detergent (Fig. 21 and Fig. 22) which solubilised proteins that were previously insoluble. Sonication and a different version of Bugbuster (HT protein extraction reagent) was also tested (data not shown) and it was determined for large culture lysis the detergent based lysis buffer would be used.

Using this method a total of 5 proteins were expressed in large cell cultures before the cells were lysed and the clear supernatant containing the soluble proteins was purified. Purification was carried out using nickel column affinity and analysed using SDS PAGE (Fig. 23 and Fig. 24). Elutions containing the protein of interests were dialysed into Tris buffer and stored for enzymatic analysis.
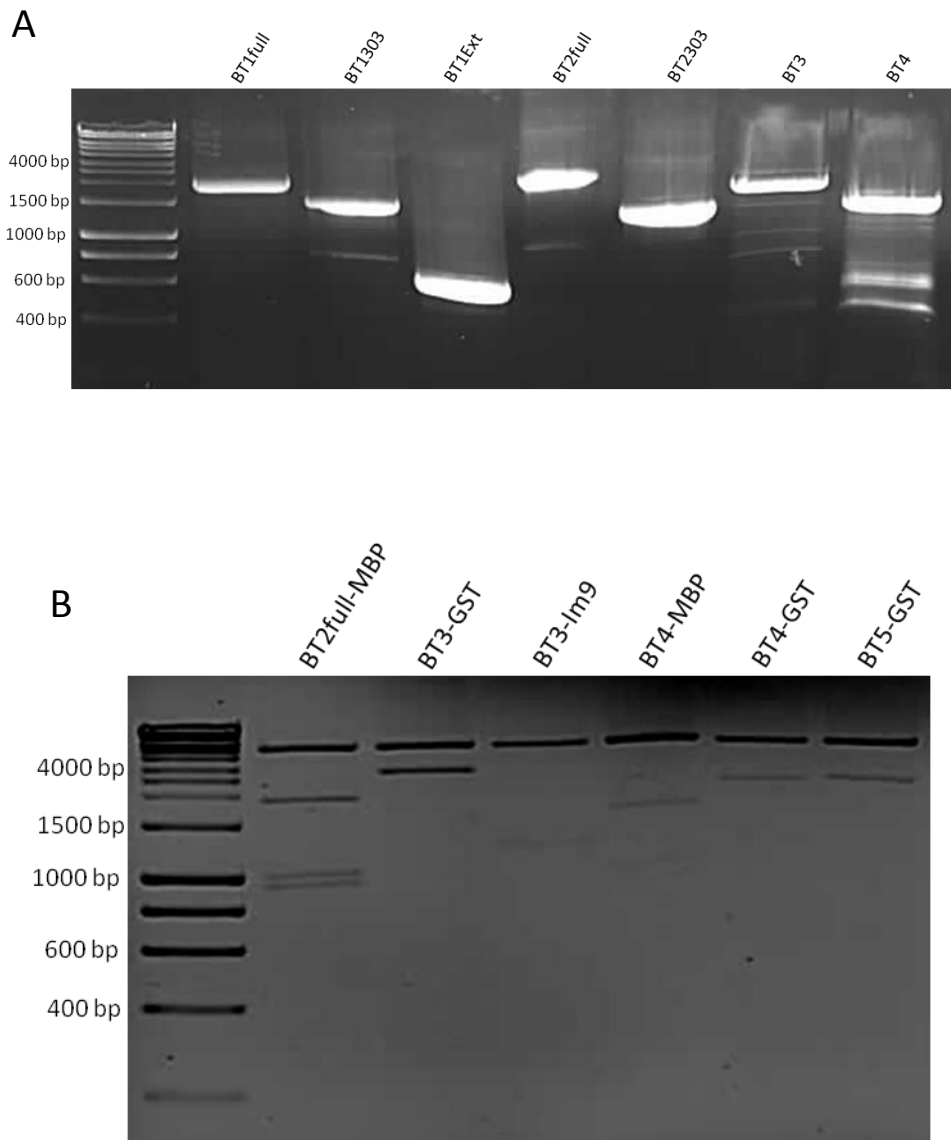
Figure 17 - PCR amplification and restriction enzyme digests. PCR (A) and restriction enzyme digests (B) ensured successful plasmid construction. Refer to Table 3 for abbreviations of gene names.
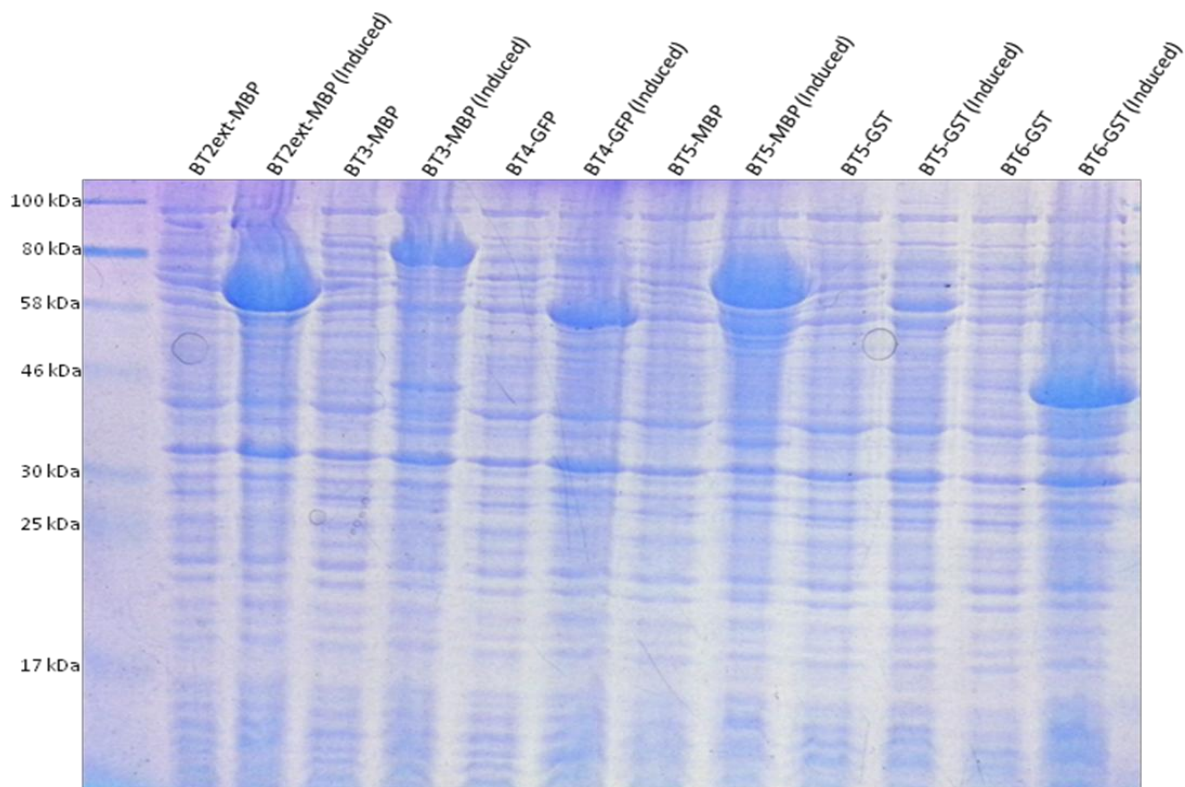
Figure 18 - Protein expression was determined using SDS-PAGE. Samples taken previous to induction with IPTG and post induction with IPTG were analyzed for correct expression. As predicted samples post induction had increased levels of protein expression in comparison to pre induction. Refer to Table 3 for name abbreviations and expected protein sizes.
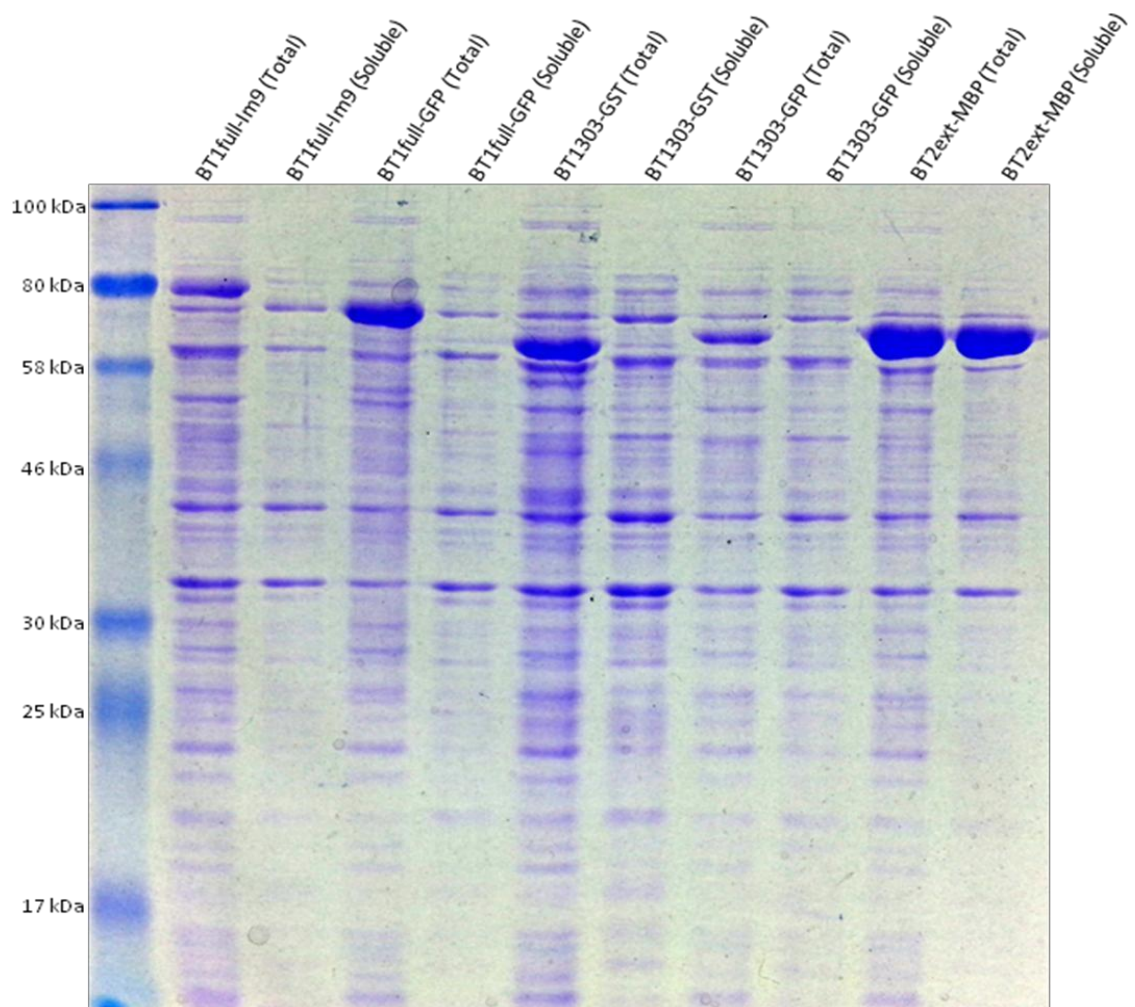
Figure 19 - Protein solubility was determined using SDS-PAGE analysis. Samples taken after 3 hours post induction were lysed using Bugbuster in the presence of a phosphate buffer. Using this method of lysis only a small sample of proteins tested was found to be soluble. Refer to Table 3 for name abbreviations and expected protein sizes.
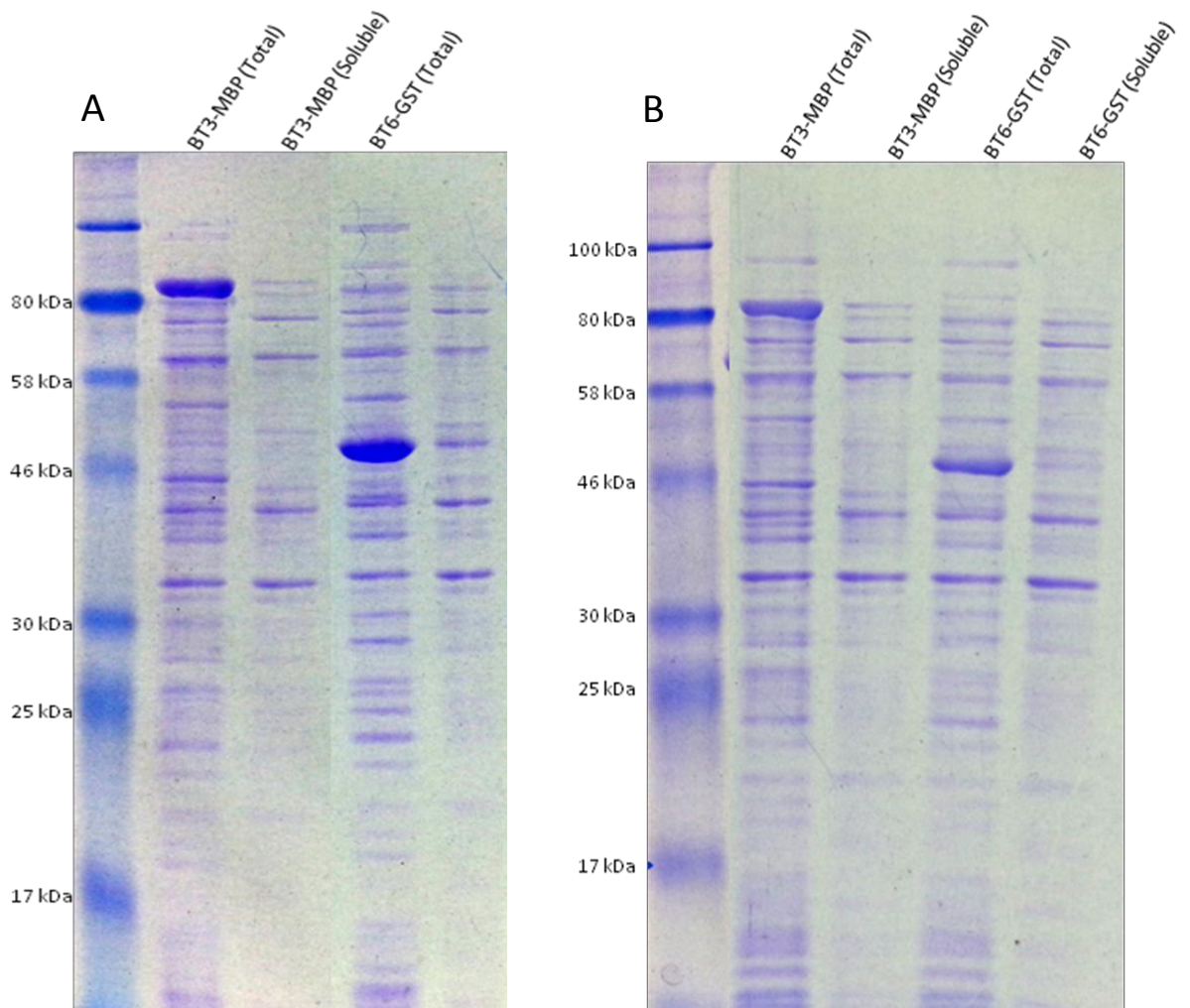
Figure 20 - Protein solubility was unaffected by the addition of a phosphate buffer. Samples 3 hour after induction were solubilised using Bugbuster in the presence of a phosphate buffer (A) and absence of a phosphate buffer (B). The removal of the phosphate buffer made no difference in the solubility of the proteins. Refer to Table 3 for name abbreviations and expected protein sizes.

|	MBP – 42.5Kda	GST – 26Kda	IM9 10Kda	GFP- 26.9Kda
BT_0457 (BT1full) 76.5Kda — SGNH_Hydrolase — DUF303	119kDa	102.5kDa	86.5kDa	103.4kDa
BT_0457 (BT1duf303) 56.96 kDa — DUF303	99.46kDa	82.96kDa	66.96kDa	83.86kDa
BT_0457 (BT1ext) 24.21kDa — SGNH_Hydrolase	66.71Kda	50.21kDa	34.21kDa	51.11kDa
BT_0447 (BT2full) 101.38 kDa — DUF303 — DUF187	143.88kDa	127.38kDa	111.38kDa	128.28kDa
BT_0447 (BT2duf303) 54.4 kDa — DUF303	96.9kDa	80.4kDa	64.4kDa	81.3kDa
BT_0447 (BT2ext) 47.79 kDa — DUF187	90.29kDa	73.79kDa	57.79kDa	74.69kDa
BT_4091 (BT3) 71.83 kDa — DUF303 — Glyco_hydro_2_N — DUF303	114.33kDa	97.83kDa	81.83kDa	98.73kDa
BT_0958 (BT4) 50.5 kDa — Big_3_4 — DUF303	93kDa	76.5kDa	60.5kDa	77.4kDa
BT_3171 (BT5) 53.44 kDa — DUF303	95.94kDa	79.44kDa	63.44kDa	80.34kDa
BT_4180 (BT6) 27.15 kDa — DUF303	69.65kDa	53.15kDa	37.15kDa	54.05kDa

Table 3 - Table of final protein constructs. Each of the individual proteins was cloned into four of expression vectors to include fusion partners. The left of the table shows the various protein domains that were expressed. (Colours match those in figure 14). The right side of the table shows the expected protein sizes with the addition of the fusion partners.
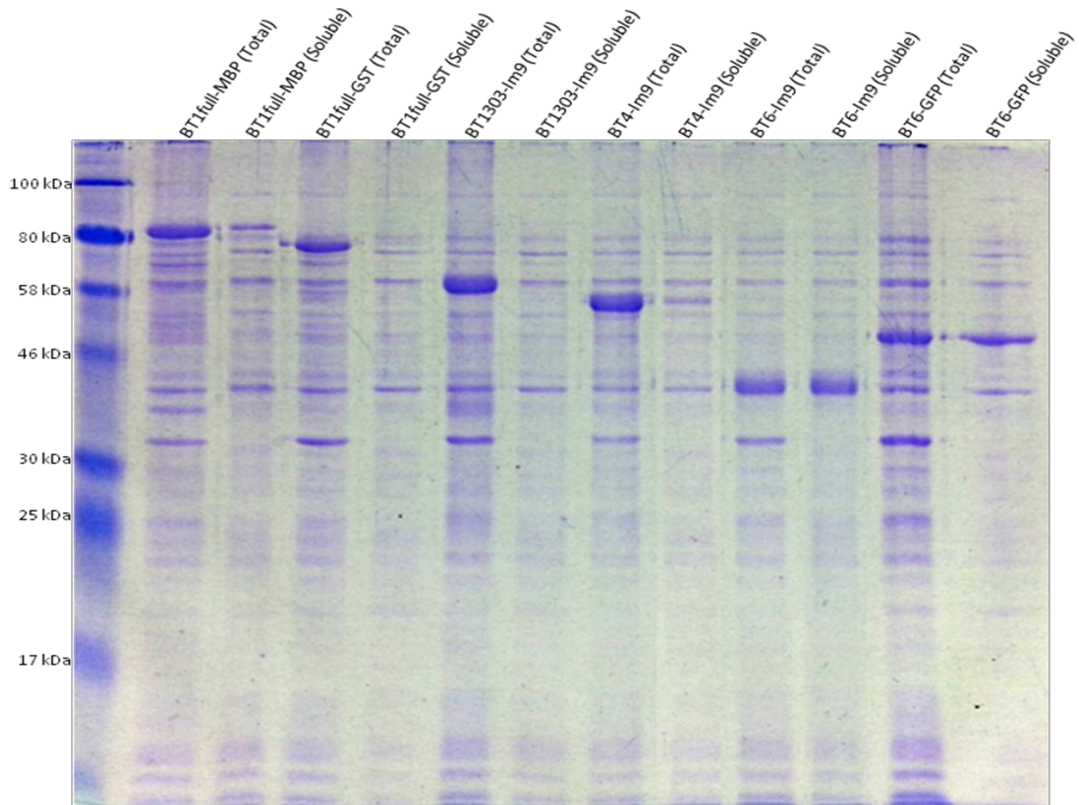
Figure 21 - Protein solubility improved when cells were lysed using a detergent based lysis buffer. Samples 3 hour after induction were solubilised using detergent based lysis buffer which produced partial (BT1full-MBP) or complete (BT6-Im9) protein solubility. Refer to Table 3 for name abbreviations and expected protein sizes.
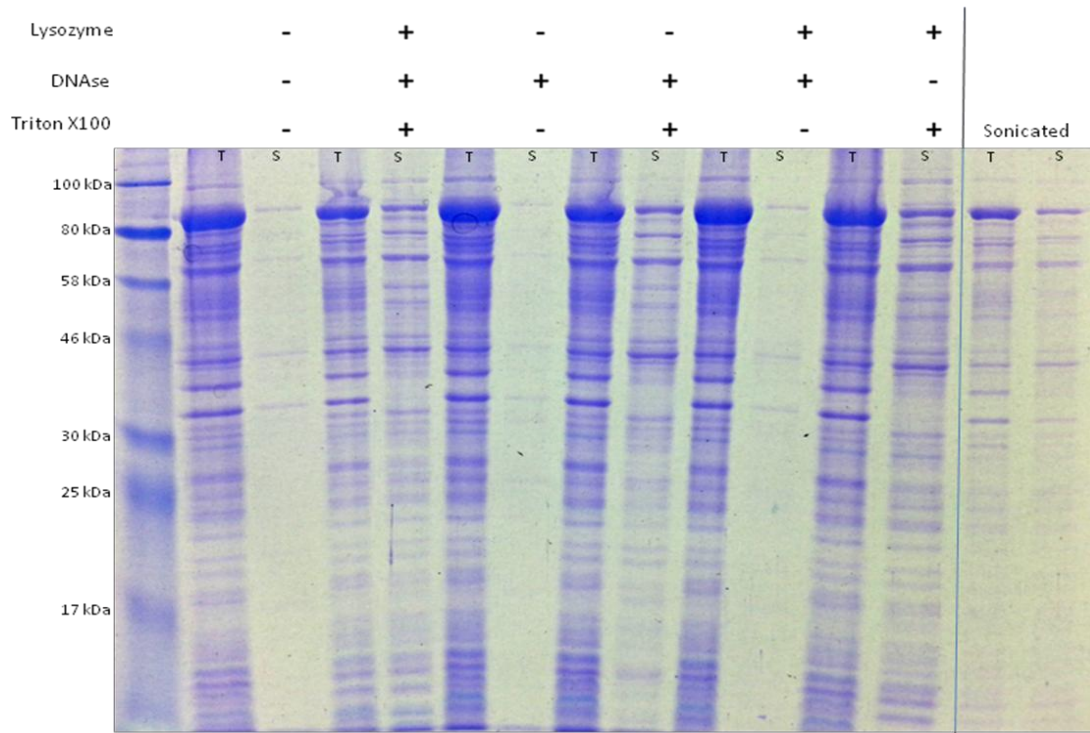
Figure 22 - Improved protein solubility was dependent on the presence of detergent in the lysis buffer. Following analysis of the lysis buffer it is evident that solubility of the proteins is unaffected by the presence of DNAse or Lysozyme and is reliant on the detergent Triton X100. This is exemplified by diminished protein solubility in the absence of Triton X100. Cell lysis using sonication also appears to make no significant difference in the amount of soluble protein when compared to lysis buffer usage. T is total lysate and S is soluble lysate.
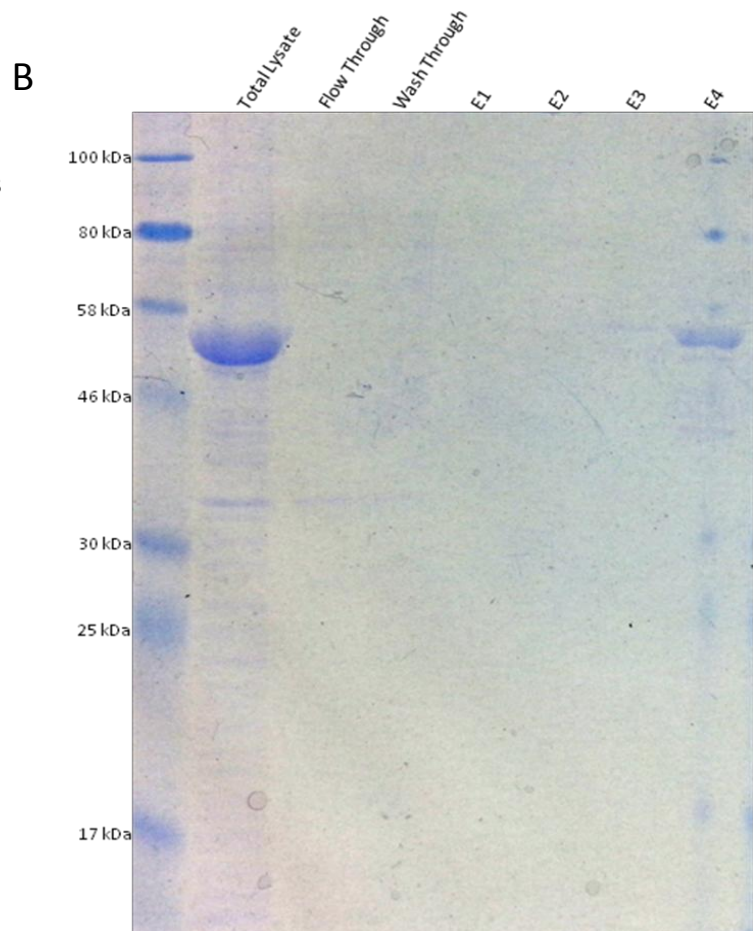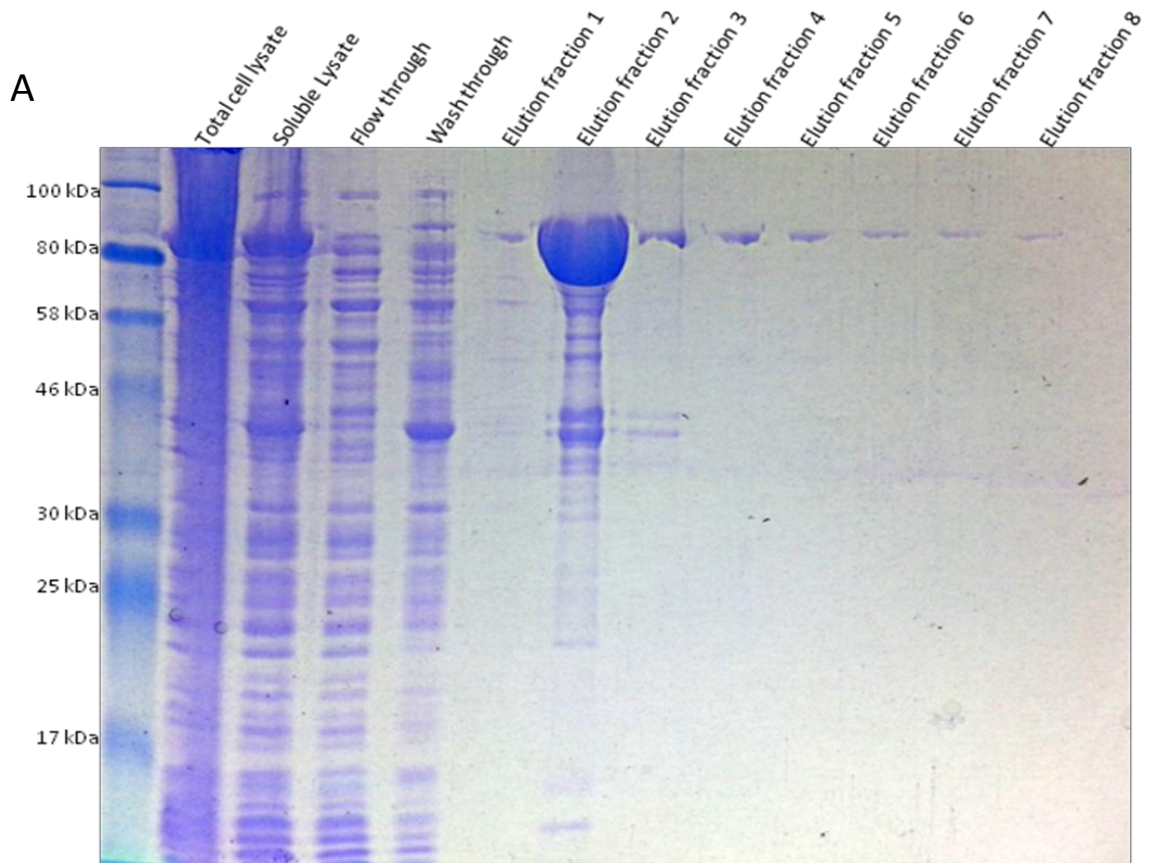
Figure 23 - Purification of proteins using Ni-affinity chromatography. Example of SDS-PAGE gels following purification. Presence of protein is clearly visible in various elations for BT1-MBP (A) and BT1ext-MBP (B). Refer to Table 3 for name abbreviations and expected protein sizes.
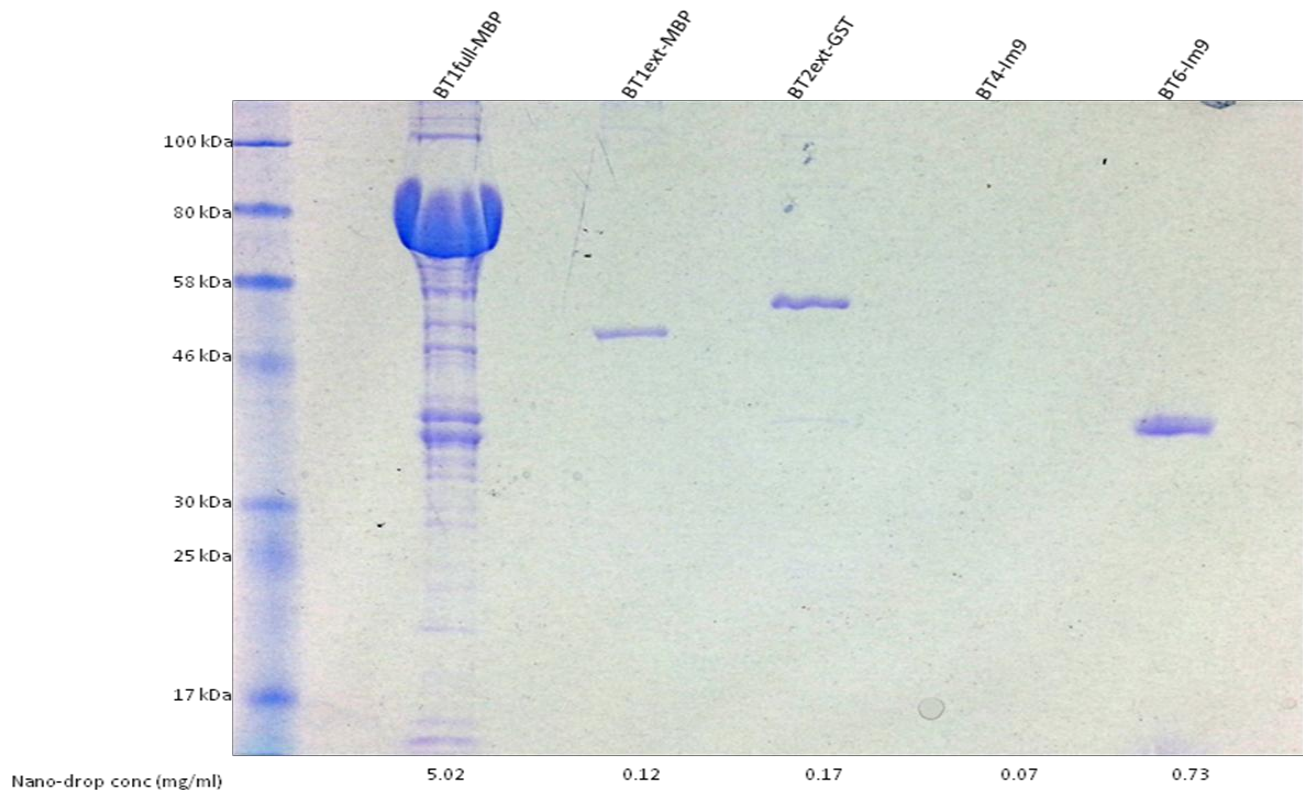
Figure 24 - Purification of five proteins. Each of the proteins that was successfully expressed and found to be soluble was purified using Ni-column affinity. Elations from this process containing the most protein were then used to determine concentration using a nano-drop. Proteins were also analyzed on an SDS-PAGE gel prior to enzyme assays. Refer to Table 3 for name abbreviations and expected protein sizes.

An enzyme based assay using an acetylated substrate showed each of the purified proteins to have enzymatic activity

Determining enzymatic activity of the purified enzyme was achieved using an enzyme assay based on p-Nitrophenyl Acetate (pNPA). An increase in OD of the samples incubated with the PNPA correlates to de-acetylation of the PNPA by the enzyme. This is also visibly evident by the change in colour from clear to yellow of the test sample. Following incubation at 40°C of the purified enzyme, OD_{405} readings were taken every five minutes for a minimum of 30 minutes. In addition to testing enzyme activity of purified enzymes of interest, a xylan esterase at various dilutions was tested on the PNPA substrate.

In the control experiment it is evident that at higher concentrations of enzymes the acetyl group from the PNPA is removed faster by the esterase enzyme in comparison to enzymes at lower concentrations (Fig. 25). It is also evident that pre-incubation of the enzyme at 40°C for 10 minutes before addition of the pNPA substrate seemed to cause no significant difference in the rate at which the conversion occurred. In both cases maximum acetyl release was achieved at 5 minutes indicated by a consistent OD_{405} reading of approximately 0.3 between 5 and 15 minutes. When the concentration of the substrate is doubled the OD_{405} readings of the reaction also doubles, despite the enzyme concentration remaining the same. Interestingly when the pNPA concentration was doubled the rate of the conversion appeared to remain constant, with the OD_{405} reaching approximately 0.6 at 5 minutes and changing very little over the next 10 minutes.

Each of the purified proteins were also tested against the pNPA substrate to determine their enzymatic activity (Fig. 26). Controls samples using 1 unit of the xylan esterase and 10 units of the xylan esterase were also run in parallel to the assays carried out using the purified proteins. Unexpectedly there appears to be no increase in OD_{405} readings when the enzyme is present 10 fold. There does appear to be a small decrease in the OD_{405} reading when 10 units of the control enzyme was present in the sample in comparison to the 1 unit sample.

It is evident from the results obtained that each protein was able to convert the pNPA to pNP, but at different rates. In each case the maximum enzymatic activity was reached within 5 minutes with little change in OD405 after this time. In the case of BT_0457 full length protein this maximum was reached within 30 seconds of pNPA addition, following a similar trend observed in the control xylan esterases.

Although each of the proteins appeared to have enzymatic activity, each enzyme has a different end point despite each sample containing the same PNPA and enzyme concentration. Of particular interest is the enzyme activity observed during the assay using only the purified extension domains of BT_0457 and BT_0447. This result was unexpected as the various domains were not the DUF303 domain expected to have acetylesterase activity. The BT_0457 extension has double the enzymatic activity when compared to the full length BT_0457 protein. The BT_0447 extension has even greater enzymatic activity reaching a maximum OD_{405} of approximately 0.7. However the initial rate of conversion appears to be almost identical to that observed in the presence of the BT_0457 extension domain, where both proteins reached an approximate OD_{405} of 0.55 within 30 seconds of pNPA addition.

The full length BT_0457 protein has the most similar enzyme activity to the 1U of the control xylan esterase. In contrast, BT_4180 has the slowest rate of enzyme activity and appears not to reach above OD_{405} 0.1 even after an hour of incubation at 40 °C. This is particularly surprising given that this protein has been described as a xylan esterase. Out of the purified enzymes, BT_0985 appears to have the greatest enzymatic activity, reaching a maximum OD_{405} of approximately 0.85. This is interesting given that this putative sialate-*O*-acetylesterase is the closest homologue found in *B. thetaiotamicron* to the SIAE found in humans. Although reaching the same OD_{405} as the BT_0447 extension domain within the first 5 minutes, the OD_{405} of BT_0985 continues to rise whereas that of BT_0447 extension plateaus.
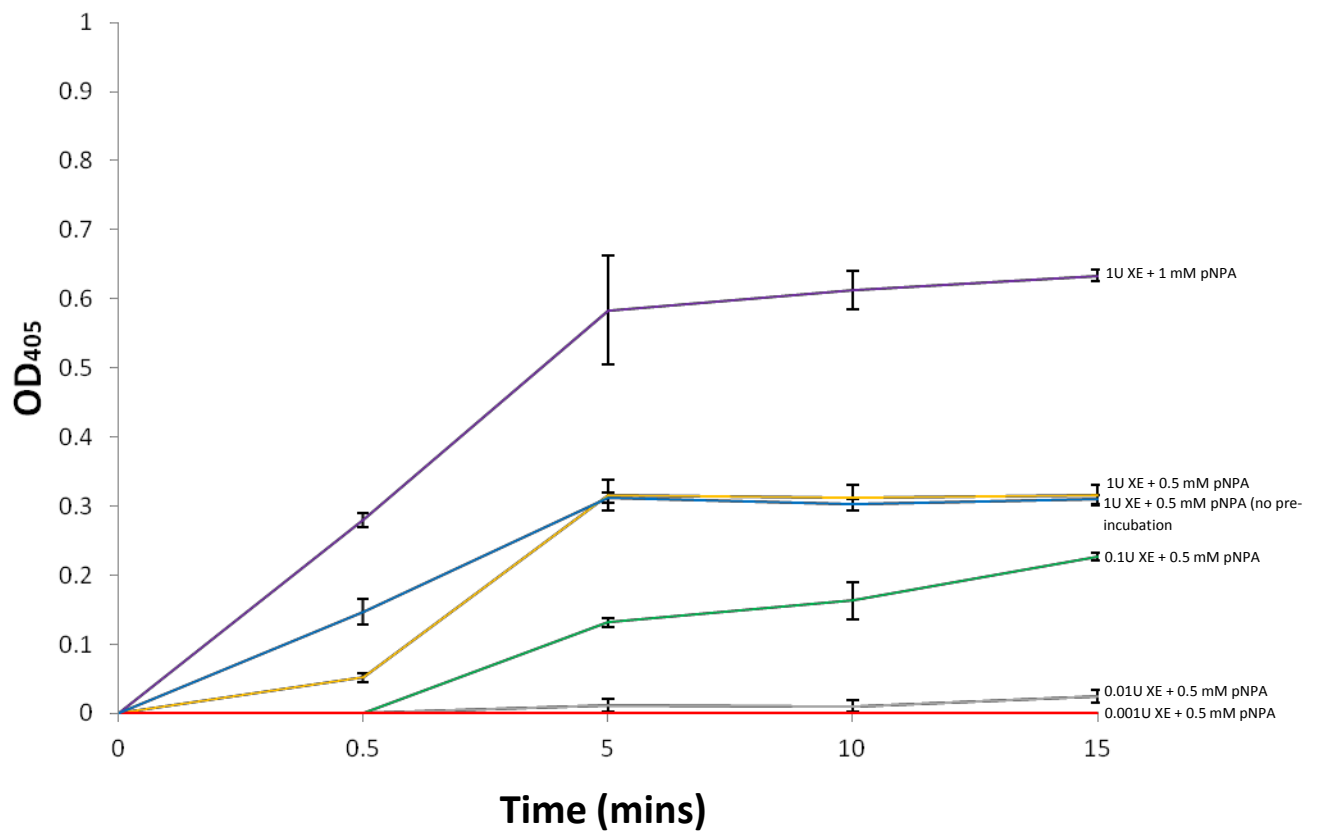
Figure 25 - Enzymatic activity of Xylan esterase. The activity of Xylan esterase (XE) against pNPA was tested at various XE concentrations. Decreasing concentration of XE showed a decrease in enzymatic activity (as determined by a change in $OD0_{405}$) against the pNPA substrate. Error bars represent standard deviation of triplicate measurements.
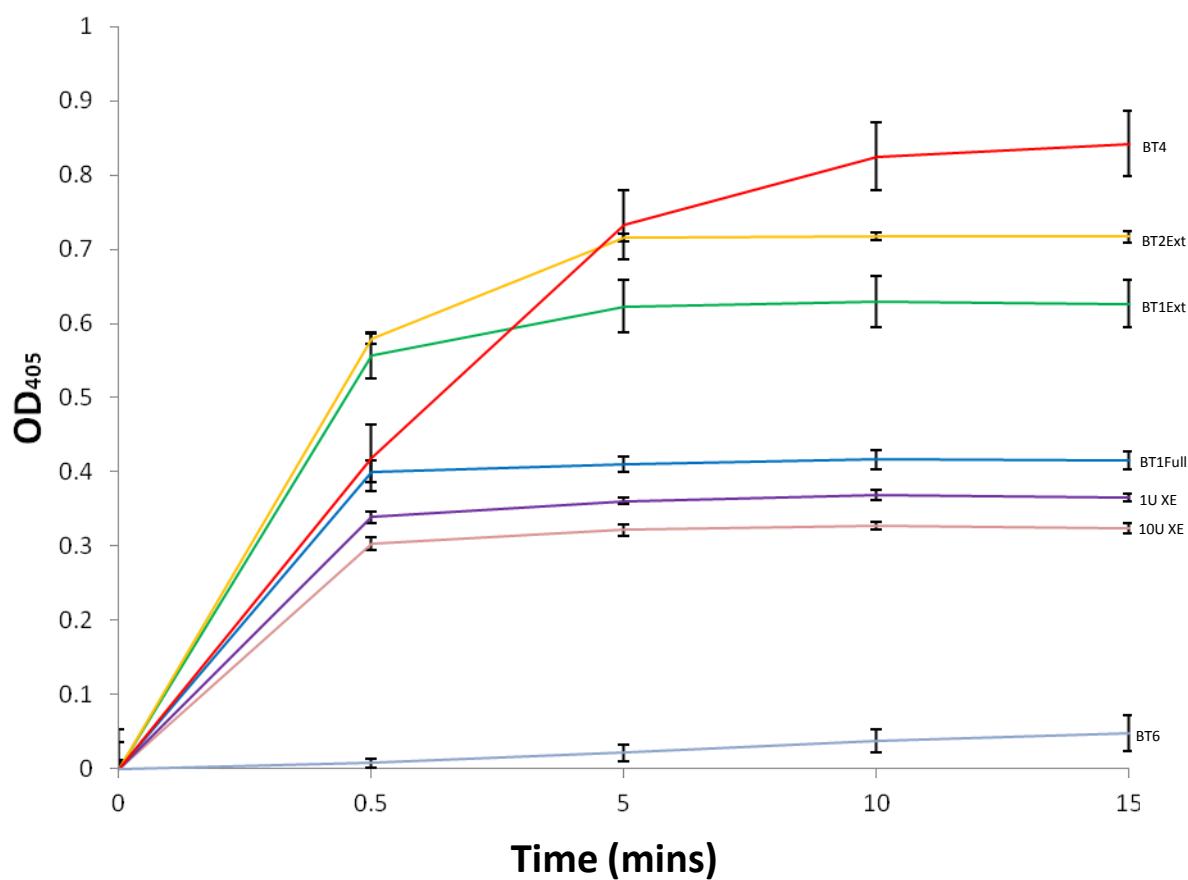
Figure 26 - Enzyme activity of purified proteins. Each of the purified proteins was tested for enzymatic activity against 0.5 mM of pNPA substrate. A change in OD_{405} was used to determine the enzymatic activity of 0.5 µM of enzyme. There is clear evidence of enzymatic activity against the substrate in all cases. Refer to Table 3 for name abbreviations and expected protein sizes. Error bars represent standard deviation of triplicate measurements.

Discussion

The human intestine is host to approximately 100 trillion microbial cells. This is tenfold higher than the number of human cells (Ley *et al*, 2006; Karlsson *et al*, 2010). These microbial cells make up what is known as the microbiota and not only do they impact human health, but have also been linked with inflammatory bowel disease and obesity (Ley *et al*, 2005; Manichanh *et al*, 2006; Qin *et al*, 2010; Tumbaugh *et al*, 2009). The microbiota in the human colon is composed of two main phyla of bacteria; *Bacteroidetes* and *firmicutes.* Together they make up a total of 95 – 99% of the gut bacteria with the relative abundance ranging from between 17 – 60% for *Bacteroidetes* and 35 – 80% for *Firmicutes* (Costello *et al*, 2009; Eckburg *et al*, 2005; Tap *et al*, 2009). It is an increase in the relative abundance of firmicutes to *Bacteroides* that has been linked with the obesity in both mice and humans (Tumbaugh *et al*, 2009; Tumbaugh *et al*, 2003).

The *Bacteroides* are the most abundant of the *Bacteroidites* phylum and make up the core component of the gastrointestinal tract (GI tract) microbiota (Costello *et al*, 2009; Eckburg *et al*, 2005; Tap *et al*, 2009). This microbiota complements the host enzymes within the gut with the activity of their own enzymes. The main function of these is the degradation of polysaccharides and the production of vitamins (Karlsson *et al*, 2010). Members of the *Bacteroides* have adapted to the surviving in the mammalian gut where there is a large abundance of undigested polysaccharides that the human enzymes are unable to digest.

The perfect example of such an adaptation is observed in *B. thetaiotamicron. B. thetaiotamicron* was the first species of *Bacteroides* to have its complete genome sequenced. This sequencing revealed the presence of 172 glycoside hydrolyses and 162 homolgues of the SusD and SusC outer membrane polysaccharide binding proteins for the utilization of polysaccharides (Xu *et al*, 2003). This repertoire of enzymes involved in acquisition and metabolism of polysaccharides makes *B. thetaiotamicron* an exemplary glycophile which is able to break down a large range of dietary polysaccharides (Sonnenburg *et al*, 2005; Xu and Gordon, 2003). Over half of the carbohydrate degrading enzymes are predicted to be secreted into the periplasmic or extracellular space leading to the release of oligo- and monosaccharides from the host mucus and dietary polysaccharides for the consumption by *B. thetaiotamicron* or other members of the microbiota (Sonnenburg *et al*, 2005).

The analysis indicated that although *B. thetaiotamicron* contained many of the genes required for sialic acid catabolism, it was missing two genes essential for the transport of sialic acid through the inner membrane let alone its catabolism (Fig. 5, 27). During this study it was also established that *B. thetaiotamicron* has a large range of potential sialate-*O*-acetylesterases.

Lateral gene transfer is a strong force of bacterial evolution. It has been suggested that the lateral gene transfer has played a major role in the evolution of bacteria in the human intestine (Doolittle, 1999; Xu *et al*, 2007). This could potentially explain the sequence similarity observed between the human sialate acetylesterase and those observed in *B. thetaiotamicron.* A good example of horizontal gene transfer is in fact observed in genes involved in sialic acid catabolism. The *nanA* genes from *Bacteroides* have been shown to branch closely to members of the Eukarya kingdom, providing evidence for the horizontal gene transfer between *Bacteroides* and eukaryotes (Almegro-Moreono and Boyd, 2009; De Koning *et al*, 2000).

Given that this bacterium is unable to catabolise sialic acid it is surprising that there is such a large repertoire of genes involved in the scavenging of sialic acid. Recently Vimr described *B. thetaiotamicron* as a "spitter" in the gut microbiota in regards to sialic acid. This fitting name describes the activity of *B. thetaiotamicron*, which cleaves the terminal sialic acid of glycoconjugates but allows the released sugar to enter the extracellular milieu (Vimr, 2013). While the sialic acid is not used by *B. thetaiotamicron*, it is the subterminal sugars that are hydrolysed by specific glycosidases, transported into the cell and used for growth. The sialic acid would then be utilized by other bacteria within the gut which lack the enzymes required for the release of sialic acid from glycoconjugates, such as *E. coli* (Vimr, 2013).

In addition to this explanation, it has been postulated that the ability of *B. thetaiotamicron* to cleave sialic acid from various structures within the human gut could confer to protection of the human host from viral infection. This was demonstrated by Varyukhina and colegues, who showed that the extent of rotavirus infection inversely correlated to the ability of bacteria to produce soluble factors (such as sialidases) that are able to specifically modify cell surface glycan structure. Their results demonstrated that the manipulation of cell surface glycan composition by *B. thetaiotamicron* was able to alter the infection of human intestinal cells by the rotavirus. Although in this study they specifically investigated the effect of an increase in galactose density caused by the action of *B. thetaiotamicron* enzymes, it was indicated that in addition to this increase, a decrease in sialic acid, mannose and fucose density correlated to a reduction in viral infection (Varyukhina *et al*, 2012). There is the possibility that this decreased is observed because a release of sialic acid hinders the attachment of the virus but this hypothesis is of yet to be explored.
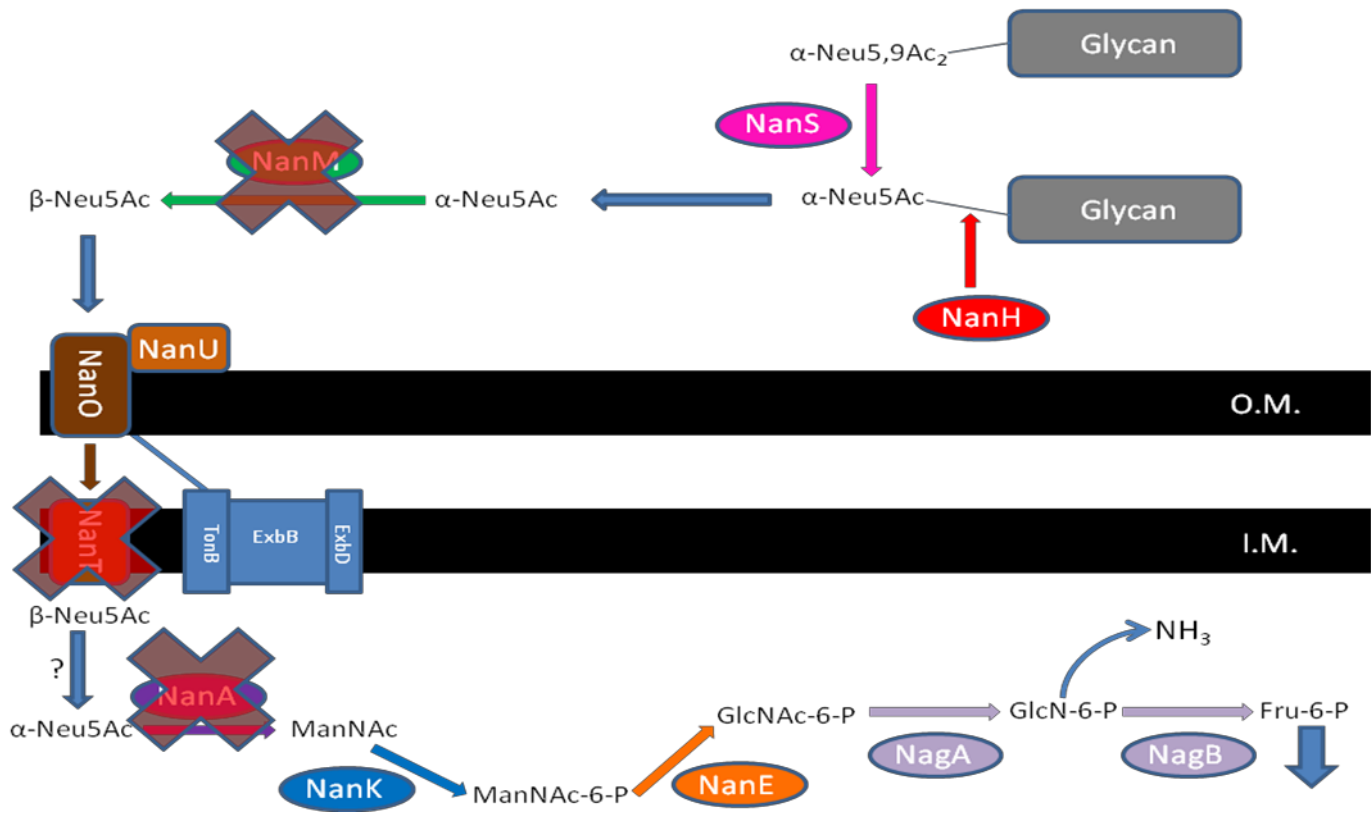
Figure 27 – Predicted sialic acid catabolism pathway in *B. thetaiotamicron*. Various aspects of the predicted *Bacteroides* sialic acid catabolism pathway appear to be missing in *B. thetaiotamicron* as indicated by red crosses.

Although the release of sialic acid for the use by other gut bacteria rationalizes the inability of *B. thetaiotamicron* to grow on sialic acid, it does not explain the presence of the multiple putative acetylesterases scattered across the genome of the bacterium. Although those investigated in this study were shown to have esterase activity their substrate specificity has as of yet not been identified. It is entirely possible that each putative enzyme removed the acetyl group from different positions on the sialic acid to allow maximum activity of the NanH sialidase. Given that *B. thetaiotamicron* forges the mucosal surfaces in search of energy sources it is able to utilize it would make sense to maximize its potential by ensuring that a series of enzymes are present with varying substrate specificity to release the multitude of sialic acid modification seen in the gut. This would then allow the bacterium to reach a wider range of potential energy sources. During previous genetic analysis, *B. thetaiotamicron* has been described as a "genetic pack rat" due to the presence of multiple genes appearing to be involved in the same processes (Wexler *et al*, 2007).

Before the candidate sialate-*O*-acetylesterases of *B. thetaiotamicron* could be analysed for the enzymatic activity the insolubility of the proteins needed to be resolved. This was achieved partly through the construction of protein fusions and partially by adjusting the method of cell lysis from using Bugbuster to using a detergent based lysis buffer. When proteins were expression in standard vectors, each protein was insoluble. Initially we thought this was due to the large sizes of the expressed proteins and so during the alternative cloning process individual domains were also expressed. When, however, the majority of the expressed fusion proteins still remained insoluble until the lysis method was altered it indicated the possibility that using a detergent based lysis buffer might have increased the solubility of the originally expressed proteins.

When using Bugbuster to lyse the cells we had very limited success in protein solubility. Only 3 of the 40 expressed proteins were partially soluble while the rest remained insoluble. Interestingly those that were soluble were the individually expressed extensions of BT0447 and BT0457. This could possibly be due to Bugbuster interacting with the DUF303 domain found in the other proteins and preventing their solubility. In addition to this each of the soluble proteins had different fusion partners indicating that the presence of a specific fusion partner does not necessarily correlate to increased protein solubility.

The enzyme assay carried out during this study indicated that each of the purified proteins had enzymatic activity. Although the basic assay was able to determine this, it was unable to determine the substrate specificity of each of the putative acetylesterases. This is of particular interest given

that *B. thetaiotamicron* appears to have many such enzymes. It has previously been shown that sialate-*O*-acetylesterases from different species have various substrate specificities. It would be interesting to see if the two NanS proteins within the sialic acid utilization cluster have a stronger substrate specificity to sialate-9-*O*-acetylesterase given that Neu5,9Ac$_2$ is the most common di-acetylated for of sialic acid found in the human GI tract (Schauer, 2004).

The most surprising result obtained during the enzyme assay was the enzymatic activity of the purified extension domains. The possibility of errors in the process of the enzyme activity was eliminated during the control experiment using a brought xylan esterase against the substrate at various concentrations. This control ensured that the conversion of the pNPA to pNP was indeed enzyme dependent. It was also established that there was no detectable spontaneous release of the acetyl group in the absence of an enzyme during the time period of the assay. However, further experimental analysis is required to confirm these unusual findings.

Although it is predicted that the DUF303 domain observed within each of these proteins is the catalytic portion of the protein (due to its presence in known characterized sialate-*O*-acetylesterase) the N-terminal extension found in BT_0457 contains an SGNH hydrolase domain. In a similar manner the DUF187 extension domain of the BT_0447 protein is predicted to be related to putative glycosyl hydrolases. In both cases these domains have the potential to work as acetylesterases. This would be a surprising result indicating that these two proteins have two domains which are both involved in deacetylation, although it is possible each domain has different substrate specificity. Until the DUF303 domain of these proteins is expressed in the absence of the extensions and the enzyme activity tested, the possibility remains that the DUF303 domains of BT_0457 and BT_0447 are inactive.

It was interesting to see that during the enzyme assay each of the proteins investigated appeared to have different end point despite the concentration of added substrate remaining the same. Although maximum conversion was reached within 5 minutes in each case, this maximum was different for each enzyme. A potential explanation for this is the presence of an inhibition loop. This would rely on the released acetate binding the enzyme directly, causing a change in the enzyme and so preventing its enzymatic activity.

Another interesting observation was the higher enzymatic activity seen when the substrate was incubated in the presence of the extension in BT_0457 in comparison to that seen during incubation with the full length protein. A potential explanation for this would be that the extension domain has a higher affinity and so a higher enzymatic activity in comparison to the full length protein. Given

that each of the samples were normalised to contain the same concentration of total enzyme, the presence of the extension in the full length protein would be diluted in comparison to the concentration when the extension was expressed alone. Although unlikely, it is however, impossible to rule out the possibility that the extension domain somehow hinders the activity of the DUF303 domain without performing the enzyme assay using the DUF303 domain alone.

Another explanation for this increase in activity would rely on the presence of the potential inhibition loop described above. The binding site for the release acetate could potentially found in a part of the protein that is not present in the specific domain expressed from as the BT_0457 extension. The binding site would however be present in the full length protein potentially explaining the different maximum activities observed between the two. Given that the BT_0447 extension also appears to have a higher enzymatic activity in comparison to the BT_0457 full length protein, it would be interesting to see if the full length BT_0447 protein also has a lower enzymatic activity compared to its stand alone extension domain.

There is the possibility of incorrect cloning, despite PCR and restriction enzyme digest analysis indicating the correct genes were present in the recombinant plasmids. Confirming without a doubt that the purified proteins were in fact as expected would require plasmid sequencing as only partial sequencing evidence is currently available. In addition to this, to eliminate any potential errors during the construct preparations, the proteins require sequencing through peptide mass spectrometry.

During the genetic analysis of various *Bacteroides* species, the presence of putative sulphatases and phosphatases within the *nan* operons was unexpected but it is entirely logical for these species to contain these enzymes. Previous mass spectrometry of human mucin oligosaccharides isolated from the ileum, cecum, transverse colon and rectum showed that the human mucin oligosaccharides contains 46 neutral carbohydrate chains and 50 acidic carbohydrate chains (Robbe *et al*, 2004; Vimr, 2013). The neutral residues did not contain Nue5Ac or sulphate residues. The acid chains, on the other hand, included Neu5Ac, sulphate or sulphated Neu5Ac (Vimr, 2013). It was initially postulated in 1993 that some mucosal bacteria have glycosulphatases in addition to the acetylesterases and sialidases needed for growth on Neu5Ac (Corfield *et al*, 1993). This supports the idea that bacteria containing a putative sulphatase would be able to prime the Neu5Ac for release via the sialidase in a similar manner to that expected of NanS for utilization as a growth substrate or release to reach subterminal sugars on within mucin oligosaccharides.

In the *Bacteroides* species investigated during this study a single putative *phosphatase* was identified to be present within the sialic acid catabolism clusters. Modification of sialic acid with a phosphate group is found at residue O-9 (Schauer, 2004). The putative *phosphatase* was seen in *B. thetaiotamicron* and for the reasons discussed above about the role of this bacterium as a forger, the presence of an additional enzyme aiding its ability to release sialic acid would give provide a competitive advantage. However, Neu5Ac9P as of yet has only been found as a free cytosolic form which is present as an intermediate of sialic acid biosynthesis (Schauer, 2000). If this is the only version of a phosphated sialic acid, the presence of the sulphatase within the sialic acid catabolism cluster of *B. thetaiotamicron* is a mystery given its ability to utilize free sialic acid as a growth factor.

According to public databases, this *phosphatase* contain a metalophosphoesterase domain which is found in the range of phosphoesterases including phosphoserine *phosphatase*s and sphingomyelin phosphoesterases. This domain is found 846 times in *Bacteroides* and 15 times within *B. thetaiotamicron.*

In addition to a *phosphatase*, two predicted sulphatases were also observed in two individual *Bacteroides* species within the sialic acid catabolism clusters. One was seen in *B. fluxus* and the other in *P. merdae*. The public databases indicated that both observed sulphatases belonged to the same family. Interestingly this family also includes sulphatases which hydrolyse the release of sulphate groups from Neu5Gc (Arylsulphatase B) but none have been described which are specific for Neu5Ac. Within *Bacteroides* proteins of this family are seen 1458 times with 31 within *B. thetaiotamicron* alone. In nature the addition of a sulphate group to sialic acid is on residue O-8. Although a form of Neu5Gc8S has been identified, a sulphated form of Neu5Ac has as of yet not been isolated.

There is a close relationship between the *B. thetaiotamicron* and the gut health. It has been suggested that the relationship is one of mutualism rather than *B. thetaiotamicron* acting as a commensal as both the host and the bacteria benefit from the relationship. The high predominance of sialic acid within the gut makes it a prime growth substrate for the bacteria in the gut. The importance of sialic acid metabolism stems from the important roles played by sialic acid within the human including immune regulation and cell-cell interactions. A high proportion of the sialic acids within the gut are found as di-acetylated forms, the deacetylation of which has been associated with the progression of cancer. This study highlights the prevalence of sialate-*O*-acetylesterase across the *Bacteroides* species. In particular a series of putative acetylesterases in the most prevalent gut bacteria *B. thetaiotamicron* have been established to have acetylesterase activity.

References

Achyuthan, K. E. And A. M. Achyuthan (2001). Comparative enzymology, biochemistry and pathophysiology of human exo-alpha-sialidases (neuraminidases). *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology* 129(1) 29-64.

Almagro-Moreno, S. And E. F. Boyd (2009). Sialic Acid Catabolism Confers a Competitive Advantage to Pathogenic Vibrio cholerae in the Mouse Intestine. *Infection and Immunity* 77(9) 3807-3816.
Almagro-Moreno, S. And E. F. Boyd (2010). Bacterial catabolism of nonulosonic (sialic) acid and fitness in the gut. *Gut Microbes* 1(1) 45-50.

Altheide, T. K., et al. (2006). System-wide genomic and biochemical comparisons of sialic acid biology among primates and rodents - Evidence for two modes of rapid evolution. *Journal of Biological Chemistry* 281(35) 25689-25702.

Angata, T. And A. Varki (2002). Chemical diversity in the sialic acids and related alpha-keto acids: An evolutionary perspective. *Chemical Reviews* 102(2) 439-469.

Arming, S., et al. (2011). The human Cas1 protein: A sialic acid-specific O-acetyltransferase? *Glycobiology* 21(5) 553-564.

Avril, T., et al. (2006). Sialic acid-binding immunoglobulin-like lectin 7 mediates selective recognition of sialylated glycans expressed on Campylobacter jejuni lipooligosaccharides. *Infection and Immunity* 74(7) 4133-4141.

Becher, N., et al. (2009). The cervical mucus plug: Structured review of the literature. *Acta Obstetricia Et Gynecologica Scandinavica* 88(5) 502-513.

Blair, D. E., et al. (2006). Structure and mechanism of chitin deacetylase from the fungal pathogen Colletotrichum lindemuthianum. *Biochemistry* 45(31) 9416-9426.

Bouchet, V., et al. (2003). Host-derived sialic acid is incorporated into Haemophilus influenzae lipopolysaccharide and is a major virulence factor in experimental otitis media. *Proc Natl Acad Sci U S A* 100(15) 8898-8903.

Brigham, C., et al. (2009a). Sialic acid (N-acetyl neuraminic acid) utilization by Bacteroides fragilis requires a novel N-acetyl mannosamine epimerase. *J Bacteriol* 191(11) 3629-3638.

Brigham, C., et al. (2009b). Sialic Acid (N-Acetyl Neuraminic Acid) Utilization by Bacteroides fragilis Requires a Novel N-Acetyl Mannosamine Epimerase. *Journal of Bacteriology* 191(11) 3629-3638.

Bulai, T., et al. (2003). Diversity of the human erythrocyte membrane, sialic acids in relation with blood groups. *Febs Letters* 534(1-3) 185-189.

Butor, C., S. Diaz and A. Varki (1993). High level O-acetylation of sialic acids on N-linked oligosaccharides of rat liver membranes. Differential subcellular distribution of 7- and 9-O-acetyl groups and of enzymes involved in their regulation. *J Biol Chem* 268(14) 10197-10206.

Carlin, A. F., et al. (2007). Group B streptococcal capsular sialic acids interact with siglecs (immunoglobulin-like lectins) on human leukocytes. *Journal of Bacteriology* 189(4) 1231-1237.

Carr, P. D. And D. L. Ollis (2009). Alpha/beta Hydrolase Fold: An Update. *Protein and Peptide Letters* 16(10) 1137-1148.

Caufrier, F., et al. (2003). Carbohydrate esterase family 4 enzymes: substrate specificity. *Carbohydrate Research* 338(7) 687-692.

Chang, D. E., et al. (2004). Carbon nutrition of Escherichia coli in the mouse intestine. *Proceedings of the National Academy of Sciences of the United States of America* 101(19) 7427-7432.

Chowdhury, N., et al. (2012). The VC1777-VC1779 proteins are members of a sialic acid-specific subfamily of TRAP transporters (siapqm) and constitute the sole route of sialic acid uptake in the human pathogen Vibrio cholerae. *Microbiology* 158(Pt 8) 2158-2167.

Clarke, P. A., N. Mistry and G. H. Thomas (2012a). Synthesis of the complete series of mono acetates of N-acetyl-D-neuraminic acid. *Org Biomol Chem* 10(3) 529-535.

Claus, H., et al. (2004). Genetics of capsule O-acetylation in serogroup C, W-135 and Y meningococci. *Mol Microbiol* 51(1) 227-239.

Comb, D. G. And S. Roseman (1960). Sialic acids .1. Structure and enzymatic synthesis of n-acetylneuraminic acid. *Journal of Biological Chemistry* 235(9) 2529-2537.

Corfield, A. P., et al. (1992). Mucin degradation in the human colon - production of sialidase, sialate o-acetylesterase, n-acetylneuraminate lyase, arylesterase, and glycosulfatase activities by strains of fecal bacteria. *Infection and Immunity* 60(10) 3971-3978.

Corfield, A. P., et al. (1993). The roles of enteric bacterial sialidase, sialate o-acetyl esterase and glycosulfatase in the degradation of human colonic mucin. *Glycoconjugate Journal* 10(1).

Corfield, A. P., et al. (1999). Reduction of sialic acid O-acetylation in human colonic mucins in the adenoma-carcinoma sequence. *Glycoconjugate Journal* 16(6).

Corfield, T. (1992). Bacterial sialidases - roles in pathogenicity and nutrition. *Glycobiology* 2(6) 509-521.

Costello, E. K., et al. (2009). Bacterial community variation in human body habitats across space and time. *Science* 326(5960) 1694-1697.

Crennell, S., et al. (1994). Crystal structure of Vibrio cholerae neuraminidase reveals dual lectin-like domains in addition to the catalytic domain. *Structure* 2(6) 535-544.

De Koning, A. P., et al. (2000). Lateral gene transfer and metabolic adaptation in the human parasite Trichomonas vaginalis. *Molecular Biology and Evolution* 17(11) 1769-1773.

Degrassi, G., et al. (2000). The acetyl xylan esterase of Bacillus pumilus belongs to a family of esterases with broad substrate specificity. *Microbiology-Uk* 146 1585-1591.

Deli, A., et al. (2010). Lmbe proteins from Bacillus cereus are de-N-acetylases with broad substrate specificity and are highly similar to proteins in Bacillus anthracis. *Febs Journal* 277(13) 2740-2753.

Diaz, S., et al. (1989). O-acetylation and de-O-acetylation of sialic acids. 7- and 9-o-acetylation of alpha 2,6-linked sialic acids on endogenous N-linked glycans in rat liver Golgi vesicles. *J Biol Chem* 264(32) 19416-19426.

Domingues, R. M., et al. (1992). Sialic acid as receptor of Bacteroides fragilis lectin-like adhesin. *Zentralbl Bakteriol* 277(3) 340-344.

Donelli, G., et al. (2012). Biofilm-growing intestinal anaerobic bacteria. *Fems Immunology and Medical Microbiology* 65(2) 318-325.

Eckburg, P. B., et al. (2005). Diversity of the human intestinal microbial flora. *Science* 308(5728) 1635-1638.

Fabich, A. J., et al. (2008). Comparison of carbon nutrition for pathogenic and commensal Escherichia coli strains in the mouse intestine. *Infection and Immunity* 76(3) 1143-1152.

Faillard, H. (1989). The early history of sialic acids. *Trends in Biochemical Sciences* 14(6) 237-241.

Figueira, M. A., et al. (2007). Role of complement in defense of the middle ear revealed by restoring the virulence of nontypeable Haemophilus influenzae siab mutants. *Infection and Immunity* 75(1) 325-333.

Fukuda, M. N., et al. (1985). Embryonal lactosaminoglycan. The structure of branched lactosaminoglycans with novel disialosyl (sialyl alpha 2----9 sialyl) terminals isolated from PA1 human embryonal carcinoma cells. *J Biol Chem* 260(11) 6623-6631.

Godoy, V. G., et al. (1993). A role for Bacteroides fragilis neuraminidase in bacterial growth in two model systems. *Infect Immun* 61(10) 4415-4426.

Green, E. D., et al. (1988). The asparagine-linked oligosaccharides on bovine fetuin. Structural analysis of N-glycanase-released oligosaccharides by 500-megahertz 1H NMR spectroscopy. *J Biol Chem* 263(34) 18253-18268.

Gruteser, N., et al. (2012). Sialic acid utilization by the soil bacterium Corynebacterium glutamicum. *FEMS Microbiol Lett* 336(2) 131-138.

Harvey, H. A., W. E. Swords and R. A. Apicella (2001). The mimicry of human glycolipids and glycosphingolipids by the lipooligosaccharides of pathogenic Neisseria and Haemophilus. *Journal of Autoimmunity* 16(3) 257-262.

Higa, H. H., A. Manzi and A. Varki (1989). O-acetylation and de-O-acetylation of sialic acids. Purification, characterization, and properties of a glycosylated rat liver esterase specific for 9-O-acetylated sialic acids. *J Biol Chem* 264(32) 19435-19442.

Higa, H. H., et al. (1989). O-acetylation and de-O-acetylation of sialic acids. O-acetylation of sialic acids in the rat liver Golgi apparatus involves an acetyl intermediate and essential histidine and lysine residues--a transmembrane reaction? *J Biol Chem* 264(32) 19427-19434.

Holmgren, J., et al. (1975). Interaction of cholera toxin and membrane gm1 ganglioside of small-intestine. *Proceedings of the National Academy of Sciences of the United States of America* 72(7) 2520-2524.

Houliston, R. S., et al. (2006). Identification of a sialate O-acetyltransferase from Campylobacter jejuni: demonstration of direct transfer to the C-9 position of terminalalpha-2, 8-linked sialic acid. *J Biol Chem* 281(17) 11480-11486.

Hutchins, J. T., et al. (1988). Distribution of mono-, di, and tri-O-acetylated sialic acids in normal and neoplastic colon. *Cancer Res* 48(2) 483-489.

Iwasaki, M., et al. (1984). Novel oligosaccharide chains on polysialoglycoproteins isolated from rainbow trout eggs. A unique carbohydrate sequence with a sialidase-resistant sialyl group, dga1nac beta 1 leads to 4(neugc2 leads to 3)dga1nac. *Biochemistry* 23(2) 305-310.

Jeong, H. G., et al. (2009). The Capability of Catabolic Utilization of N-Acetylneuraminic Acid, a Sialic Acid, Is Essential for Vibrio vulnificus Pathogenesis. *Infection and Immunity* 77(8) 3209-3217.

Jones, C., M. Virji and P. R. Crocker (2003). Recognition of sialylated meningococcal lipopolysaccharide by siglecs expressed on myeloid cells leads to enhanced bacterial uptake. *Molecular Microbiology* 49(5) 1213-1225.

Karlsson, F. H., et al. (2011). A Closer Look at Bacteroides: Phylogenetic Relationship and Genomic Implications of a Life in the Human Gut. *Microbial Ecology* 61(3) 473-485.

Kean, E. L., A. K. Munster-Kuhnel and R. Gerardy-Schahn (2004). CMP-sialic acid synthetase of the nucleus. *Biochimica Et Biophysica Acta-General Subjects* 1673(1-2) 56-65.

Kitagawa, H., et al. (1991). Isolation and structural studies of human milk oligosaccharides that are reactive with a monoclonal antibody MSW 113. *J Biochem* 110(4) 598-604.

Kitazume, S., et al. (1994). Identification of polysialic acid-containing glycoprotein in the jelly coat of sea urchin eggs. Occurrence of a novel type of polysialic acid structure. *J Biol Chem* 269(36) 22712-22718.

Knirel, Y. A., et al. (2003). 5,7-diamino-3,5,7,9-tetradeoxynon-2-ulosonic acids in bacterial glycopolymers: chemistry and biochemistry. *Adv Carbohydr Chem Biochem* 58 371-417.

Kubo, H., et al. (1990). Gangliosides from the eggs of the sea urchin, Anthocidaris crassispina. *J Biochem* 108(2) 185-192.

Lewis, A. L., et al. (2007). Neua sialic acid O-acetylesterase activity modulates O-acetylation of capsular polysaccharide in Group B Streptococcus. *Journal of Biological Chemistry* 282(38) 27562-27571.

Lewis, A. L., et al. (2009a). Innovations in host and microbial sialic acid biosynthesis revealed by phylogenomic prediction of nonulosonic acid structure. *Proc Natl Acad Sci U S A* 106(32) 13552-13557.

Lewis, A. L., et al. (2009b). Innovations in host and microbial sialic acid biosynthesis revealed by phylogenomic prediction of nonulosonic acid structure. *Proceedings of the National Academy of Sciences of the United States of America* 106(32) 13552-13557.

Li, Y. And X. Chen (2012). Sialic acid metabolism and sialyltransferases: natural functions and applications. *Applied Microbiology and Biotechnology* 94(4) 887-905.

Lo, Y. C., et al. (2003). Crystal structure of Escherichia coli thioesterase I/protease I/lysophospholipase L-1: Consensus sequence blocks constitute the catalytic center of SGNH-hydrolases through a conserved hydrogen bond network. *Journal of Molecular Biology* 330(3) 539-551.

Logan, S. M., et al. (2009). Identification of novel carbohydrate modifications on Campylobacter jejuni 11168 flagellin using metabolomics-based approaches. *Febs j* 276(4) 1014-1023.

Malleron, A., et al. (2012). Chondroitin-4-O-sulfatase from Bacteroides thetaiotaomicron: exploration of the substrate specificity. *Carbohydrate Research* 353 96-99.

Mandal, C., R. Schwartz-Albiez and R. Vlasak (2012). Functions and Biosynthesis of O-Acetylated Sialic Acids. *Top Curr Chem*.

Mann, B., et al. (1997). Low O-acetylation of sialyl-Le(x) contributes to its overexpression in colon carcinoma metastases. *International Journal of Cancer* 72(2) 258-264.

Manzi, A. E., et al. (1990). Biosynthesis and turnover of O-acetyl and N-acetyl groups in the gangliosides of human melanoma cells. *J Biol Chem* 265(22) 13091-13103.

Marques, M. B., et al. (1992). Prevention of C3 deposition by capsular polysaccharide is a virulence mechanism of type III group B streptococci. *Infect Immun* 60(10) 3986-3993.

Mcnally, D. J., et al. (2006). Functional characterization of the flagellar glycosylation locus in Campylobacter jejuni 81-176 using a focused metabolomics approach. *J Biol Chem* 281(27) 18489-18498.

Mcnally, D. J., et al. (2007). Targeted metabolomics analysis of Campylobacter coli VC167 reveals legionaminic acid derivatives as novel flagellar glycans. *J Biol Chem* 282(19) 14463-14475.

Molgaard, A., S. Kauppinen and S. Larsen (2000). Rhamnogalacturonan acetylesterase elucidates the structure and function of a new family of hydrolases. *Structure with Folding & Design* 8(4) 373-383.

Moran, A. P., A. Gupta and L. Joshi (2011). Sweet-talk: role of host glycosylation in bacterial pathogenesis of the gastrointestinal tract. *Gut* 60(10) 1412-1425.

Mulligan, C., et al. (2012). The membrane proteins siaq and siam form an essential stoichiometric complex in the sialic acid tripartite ATP-independent periplasmic (TRAP) transporter siapqm (VC1777-1779) from Vibrio cholerae. *J Biol Chem* 287(5) 3598-3608.

Muralikrishna, G., et al. (1992). Identification of a new ganglioside from the starfish Asterias rubens. *Carbohydr Res* 236 321-326.

Nakayama-Imaohji, H., et al. (2012). Characterization of a gene cluster for sialoglycoconjugate utilization in Bacteroides fragilis. *The journal of medical investigation : JMI* 59(1-2) 79-94.

Nees, S., et al. (1976). PURIFICATION AND CHARACTERIZATION OF N-ACETYLNEURAMINATE LYASE FROM CLOSTRIDIUM-PERFRINGENS. *Hoppe-Seylers Zeitschrift Fur Physiologische Chemie* 357(6) 839-853.

Newstead, S. L., et al. (2008). The structure of Clostridium perfringens nani sialidase and its catalytic intermediates. *J Biol Chem* 283(14) 9080-9088.

Park, K. H., et al. (2013). Structural and biochemical characterization of the broad substrate specificity of Bacteroides thetaiotaomicron commensal sialidase. *Biochim Biophys Acta*.

Pillai, S., A. Cariappa and S. P. Pirnie (2009). Esterases and autoimmunity: the sialic acid acetylesterase pathway and the regulation of peripheral B cell tolerance. *Trends in Immunology* 30(10) 488-493.

Plumbridge, J. And E. Vimr (1999). Convergent pathways for utilization of the amino sugars N-acetylglucosamine, N-acetylmannosamine, and N-acetylneuraminic acid by Escherichia coli. *Journal of Bacteriology* 181(1) 47-54.

Ram, S., et al. (1998). A novel sialic acid binding site on factor H mediates serum resistance of sialylated Neisseria gonorrhoeae. *J Exp Med* 187(5) 743-752.

Rangarajan, E. S., et al. (2011). Structural and enzymatic characterization of nans (yjhs), a 9-O-Acetyl N-acetylneuraminic acid esterase from Escherichia coli O157:H7. *Protein Science* 20(7).

Ringenberg, M. A., S. M. Steenbergen and E. R. Vimr (2003). The first committed step in the biosynthesis of sialic acid by Escherichia coli K1 does not involve a phosphorylated N-acetylmannosamine intermediate. *Molecular Microbiology* 50(3) 961-975.

Rogemond, V. And R. M. Guinet (1986). Lectinlike adhesins in the Bacteroides fragilis group. *Infect Immun* 53(1) 99-102.

Roggentin, P., et al. (1993). The sialidase superfamily and its spread by horizontal gene-transfer. *Molecular Microbiology* 9(5) 915-921.

Salyers, A. A. (1984). Bacteroides of the human lower intestinal tract. *Annu Rev Microbiol* 38 293-313.

Salzberg, S. L., et al. (2001). Microbial genes in the human genome: Lateral transfer or gene loss? *Science* 292(5523) 1903-1906.

Schauer, R. (2000). Achievements and challenges of sialic acid research. *Glycoconjugate Journal* 17(7-9) 485-499.

Schauer, R. (2004). Victor Ginsburg's influence on my research of the role of sialic acids in biological recognition. *Archives of Biochemistry and Biophysics* 426(2) 132-141.

Schauer, R., et al. (1989). Partial-purification and characterization of sialate o-acetylesterase from bovine brain. *Journal of Biochemistry* 106(1) 143-150.

Schauer, R., et al. (2001). Metabolism and role of O-acetylated sialic acids. *In* A. M. Wu ed. *Molecular Immunology of Complex Carbohydrates-2*. 325-342.

Schirm, M., et al. (2003). Structural, genetic and functional characterization of the flagellin glycosylation process in Helicobacter pylori. *Mol Microbiol* 48(6) 1579-1592.

Severi, E., D. W. Hood and G. H. Thomas (2007). Sialic acid utilization by bacterial pathogens. *Microbiology-Sgm* 153 2817-2822.

Severi, E., et al. (2005). Sialic acid transport in Haemophilus influenzae is essential for lipopolysaccharide sialylation and serum resistance and is dependent on a novel tripartite ATP-independent periplasmic transporter. *Molecular Microbiology* 58(4) 1173-1185.

Severi, E., et al. (2008). Sialic acid mutarotation is catalyzed by the Escherichia coli beta-propeller protein yjht. *Journal of Biological Chemistry* 283(8) 4841-4849.

Severi, E., et al. (2008). Sialic acid mutarotation is catalyzed by the Escherichia coli beta-propeller protein yjht. *J Biol Chem* 283(8) 4841-4849.

Severi, E., et al. (2010). Characterization of a novel sialic acid transporter of the sodium solute symporter (SSS) family and in vivo comparison with known bacterial sialic acid transporters. *FEMS Microbiol Lett* 304(1) 47-54.

Shakhnovich, E. A., S. J. King and J. N. Weiser (2002). 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. *Infection and Immunity* 70(12) 7161-7164.

Sharma, S. K., et al. (2011). Functional characterization of VC1929 of Vibrio cholerae El Tor: role in mannose-sensitive haemagglutination, virulence and utilization of sialic acid. *Microbiology-Sgm* 157 3180-3186.

Shell, D. M., et al. (2002). The Neisseria lipooligosaccharide-specific alpha-2,3-sialyltransferase is a surface-exposed outer membrane protein. *Infection and Immunity* 70(7) 3744-3751.

Shen, Y. Q., et al. (2004). O-acetylation and de-O-acetylation of sialic acids in human colorectal carcinoma. *European Journal of Biochemistry* 271(2) 281-290.

Shi, W. X., R. Chammas and A. Varki (1996). Linkage-specific action of endogenous sialic acid O-acetyltransferase in Chinese hamster ovary cells. *J Biol Chem* 271(25) 15130-15138.

Sillanaukee, P., M. Ponnio and I. P. Jaaskelainen (1999). Occurrence of sialic acids in healthy humans and different disorders. *European Journal of Clinical Investigation* 29(5) 413-425.

Slomiany, A. And B. L. Slomiany (1978). Structures of the acidic oligosaccharides isolated from rat sublingual glycoprotein. *J Biol Chem* 253(20) 7301-7306.

Sohanpal, B. K., et al. (2004). Integrated regulatory responses of fimb to N-acetylneuraminic (sialic) acid and glcnac in Escherichia coli K-12. *Proc Natl Acad Sci U S A* 101(46) 16322-16327.

Sohanpal, B. K., et al. (2007). Multiple co-regulatory elements and IHF are necessary for the control of fimb expression in response to sialic acid and N-acetylglucosamine in Escherichia coli K-12. *Molecular Microbiology* 63(4) 1223-1236.

Sola-Carvajal, A., et al. (2012). Properties of boage2, a second N-acetyl-D-glucosamine 2-epimerase from Bacteroides ovatus ATCC 8483. *Biochimie* 94(1) 222-230.

Sonnenburg, J. L., et al. (2005). Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science* 307(5717) 1955-1959.

Soong, G., et al. (2006). Bacterial neuraminidase facilitates mucosal infection by participating in biofilm production. *Journal of Clinical Investigation* 116(10) 2828-2828.

Spinola, S. M., et al. (2012). Sialylation of Lipooligosaccharides Is Dispensable for the Virulence of Haemophilus ducreyi in Humans. *Infection and Immunity* 80(2) 679-687.

Stafford, G., et al. (2012). Sialic acid, periodontal pathogens and Tannerella forsythia: stick around and enjoy the feast! *Molecular Oral Microbiology* 27(1) 11-22.

Steenbergen, S. M., et al. (2005). Sialic acid metabolism and systemic pasteurellosis. *Infection and Immunity* 73(3) 1284-1294.

Steenbergen, S. M., et al. (2006). Separate pathways for O acetylation of polymeric and monomeric sialic acidsand identification of sialyl O-acetyl esterase in Escherichia coli K1. *Journal of Bacteriology* 188(17) 6195-6206.

Steenbergen, S. M., J. L. Jirik and E. R. Vimr (2009). Yjhs (nans) is required for Escherichia coli to grow on 9-O-acetylated N-acetylneuraminic acid. *J Bacteriol* 191(22) 7134-7139.

Stoddart, A., Y. Zhang and C. J. Paige (1996). Molecular cloning of the cdna encoding a murine sialic acid-specific 9-O-acetylesterase and RNA expression in cells of hematopoietic and non-hematopoietic origin. *Nucleic Acids Research* 24(20) 4003-4008.

Sundberg, M. And K. Poutanen (1991). Purification and properties of 2-acetylxylan esterases of trichoderma-reesei. *Biotechnology and Applied Biochemistry* 13(1) 1-11.

Surolia, I., et al. (2010). Functionally defective germline variants of sialic acid acetylesterase in autoimmunity. *Nature* 466(7303) 243-U118.

Takao, A., H. Nagamune and N. Maeda (2010). Sialidase of Streptococcus intermedius: a putative virulence factor modifying sugar chains. *Microbiology and Immunology* 54(10) 584-595.

Takasaki, S., et al. (1979). The sugar chains of cold-insoluble globulin. A protein related to fibronectin. *J Biol Chem* 254(17) 8548-8553.

Takematsu, H., et al. (1999). Lysosomal and cytosolic sialic acid 9-O-acetylesterase activities can be encoded by one gene via differential usage of a signal peptide-encoding exon at the N terminus. *Journal of Biological Chemistry* 274(36) 25623-25631.

Tap, J., et al. (2009). Towards the human intestinal microbiota phylogenetic core. *Environ Microbiol* 11(10) 2574-2584.

Taylor, G. (1996). Sialidases: Structures, biological significance and therapeutic potential. *Current Opinion in Structural Biology* 6(6) 830-837.

Thibault, P., et al. (2001). Identification of the carbohydrate moieties and glycosylation motifs in Campylobacter jejuni flagellin. *Journal of Biological Chemistry* 276(37) 34862-34870.

Thomas, F., et al. (2011). Environmental and gut bacteroidetes: the food connection. *Frontiers in microbiology* 2 93.

Thomas, G. H. And E. F. Boyd (2011). On sialic acid transport and utilization by Vibrio cholerae. *Microbiology-Sgm* 157 3253-3254.

Traving, C. And R. Schauer (1998). Structure, function and metabolism of sialic acids. *Cellular and Molecular Life Sciences* 54(12) 1330-1349.

Tsuji, S., A. K. Datta and J. C. Paulson (1996). Systematic nomenclature for sialyltransferases. *Glycobiology* 6(7) v-vii.

Twine, S. M., et al. (2008). Flagellar glycosylation in Clostridium botulinum. *Febs j* 275(17) 4428-4444.

Varghese, J. N., W. G. Laver and P. M. Colman (1983). Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 A resolution. *Nature* 303(5912) 35-40.

Varki, A. And S. Diaz (1985). The transport and utilization of acetyl coenzyme-a by rat-liver golgi vesicles - o-acetylated sialic acids are a major product. *Journal of Biological Chemistry* 260(11) 6600-6608.

Varyukhina, S., et al. (2012). Glycan-modifying bacteria-derived soluble factors from Bacteroides thetaiotaomicron and Lactobacillus casei inhibit rotavirus infection in human intestinal cells. *Microbes and Infection* 14(3) 273-278.

Vimr, E. And C. Lichtensteiger (2002). To sialylate, or not to sialylate: that is the question. *Trends in Microbiology* 10(6) 254-257.

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

Vimr, E. R. And F. A. Troy (1985a). Identification of an inducible catabolic system for sialic acids (nan) in Escherichia coli. *J Bacteriol* 164(2) 845-853.

Vimr, E. R. And F. A. Troy (1985b). Regulation of sialic-acid metabolism in escherichia-coli - role of n-acylneuraminate pyruvate-lyase. *Journal of Bacteriology* 164(2) 854-860.

Vimr, E. R. And S. M. Steenbergen (2006). Mobile contingency locus controlling Escherichia coli K1 polysialic acid capsule acetylation. *Molecular Microbiology* 60(4) 828-837.

Vimr, E. R., et al. (2004). Diversity of microbial sialic acid metabolism. *Microbiology and Molecular Biology Reviews* 68(1) 132-+.

Vimr, E., C. Lichtensteiger and S. Steenbergen (2000). Sialic acid metabolism's dual function in Haemophilus influenzae. *Molecular Microbiology* 36(5) 1113-1123.

Vogel, U. And M. Frosch (1999). Mechanisms of neisserial serum resistance. *Molecular Microbiology* 32(6) 1133-1139.

Weadge, J. T. And A. J. Clarke (2006). Identification and characterization of O-acetylpeptidoglycan esterase: A novel enzyme discovered in Neisseria gonorrhoeae. *Biochemistry* 45(3) 839-851.

Xu, J. And J. I. Gordon (2003). Honor thy symbionts. *Proc Natl Acad Sci U S A* 100(18) 10452-10459.

Xu, J., et al. (2003). A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. *Science* 299(5615) 2074-2076.