JBC Papers in Press. Published on July 15, 2020 as Manuscript RA120.014454
The latest version is at https://www.jbc.org/cgi/doi/10.1074/jbc.RA120.014454

Uncovering a novel molecular mechanism for scavenging sialic acids in bacteria

Andrew Bell[1,#], Emmanuele Severi[2,#,//], Micah Lee[3], Serena Monaco[4], Dimitrios Latousakis[1], Jesus Angulo[4,5,6], Gavin H. Thomas[2], James Naismith[3] and Nathalie Juge[1,*]

[1]The Gut Microbes and Health Institute Strategic Programme, Quadram Institute Bioscience, Norwich Research Park, Norwich NR4 7UQ, UK

[2]Department of Biology, University of York, York, YO10 5DD, UK

[3]Division of Structural Biology, University of Oxford, Roosevelt Drive, Headington, Oxford OX3 7BN

[4]School of Pharmacy, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK

[5]Departamento de Química Orgánica, Universidad de Sevilla, C/ Prof. García González, 1, 41012 Sevilla, Spain.

[6]Instituto de Investigaciones Químicas (CSIC-US), Avda. Américo Vespucio, 49, 41092 Sevilla, Spain.

[#] contributed equally to this work

[//] current address: Biosciences Institute, Newcastle University, Faculty of Medical Sciences Baddiley-Clark Building, Richardson Road, Newcastle upon Tyne NE2 4AX

*Corresponding author: Nathalie Juge

Email: nathalie.juge@quadram.ac.uk

Running title: Structural basis of 2,7-anhydro-Neu5Ac catabolism

Keywords: sialic acid; oxidoreductase; 2,7-anhydro-Neu5Ac; STD NMR; gut microbiota; mucin glycosylation; *Ruminococcus gnavus*; *Escherichia coli*; sialic acid transporters; symbiosis

Abstract

The human gut symbiont *Ruminococcus gnavus* scavenges host-derived *N*-acetylneuraminic acid (Neu5Ac) from mucins, by converting it to 2,7-anhydro-Neu5Ac. We previously showed that 2,7-anhydro-Neu5Ac is transported into *R. gnavus* ATCC 29149 before being converted back to Neu5Ac for further metabolic processing. However, the molecular mechanism leading to the conversion of 2,7-anhydro-Neu5Ac to Neu5Ac remained elusive. Using 1D and 2D nuclear magnetic resonance (NMR), we elucidated the multistep enzymatic mechanism of the oxidoreductase (*Rg*NanOx) that leads to the reversible conversion of 2,7-anhydro-Neu5Ac to Neu5Ac through formation of a 4-keto-DANA intermediate and NAD^+ regeneration. The crystal structure of *Rg*NanOx in complex with the NAD^+ cofactor showed a protein dimer with a Rossman fold. Guided by the *Rg*NanOx structure, we identified catalytic residues by site-directed mutagenesis. Bioinformatics analyses revealed the presence of *Rg*NanOx homologues across Gram negative and Gram positive bacterial species and co-occurrence with sialic acid transporters. We showed by electrospray ionisation spray mass spectrometry (ESI-MS) that the *Escherichia coli* homologue YjhC displayed activity against 2,7-anhydro-Neu5Ac and that *E. coli* could catabolise 2,7-anhydro-Neu5Ac. Differential scanning fluorimetry (DSF) analyses confirmed the binding of YjhC to the substrates 2,7-anhydro-Neu5Ac and Neu5Ac, as well as to co-factors NAD and NADH. Finally, using *E. coli* mutants and complementation growth assays, we demonstrated that 2,7-anhydro-Neu5Ac catabolism in *E. coli* was dependent on YjhC and on the predicted sialic acid transporter YjhB. These results revealed the molecular mechanisms of 2,7-anhydro-Neu5Ac catabolism across bacterial species and a novel sialic acid transport and catabolism pathway in *E. coli*.

Introduction

The sialic acids comprise a family of 9-carbon sugar acids found predominantly on cell surface glycans of humans and other animals (1). Sialic acids are subject to a remarkable number of modifications, generating more than 50 structurally distinct molecules. Their terminal location makes them a preferential target for interaction with viruses and microorganisms at mucosa surfaces across the body (1). *N*-acetylneuraminic acid (Neu5Ac), the most common form of sialic acid in humans, is a major epitope of mucin glycans which can serve as a metabolic substrate to the gut bacteria which have adapted to the mucosal environment (2). Most of our knowledge on sialic acid metabolism in bacteria comes from the study of Neu5Ac using the model organism *E. coli* (3). We previously reported that sialic acid metabolism is vital to the ability of *Ruminococcus gnavus* strains to utilise mucin as a nutrient source through the production of 2,7-anhydro-Neu5Ac derivative (4,5).

R. gnavus is a human gut symbiont which plays a major role in human health and disease. *R. gnavus* is widely distributed amongst individuals being represented in the most common 57 species present in ≥90% of individuals (6). Colonisation by *R. gnavus* has been found in infants during the first days of life. *R. gnavus* is in the top 15 species showing abundance in both adult and infant gut-enriched genes, supporting *R. gnavus* adaptation to the intestinal habitat throughout life (6,7). Further, *R. gnavus* has been associated with an increasing number of intestinal or extra-intestinal diseases including inflammatory bowel disease (8-16).

The mucin-foraging strategy of *R. gnavus* is strain specific (5) and associated with the expression of an intramolecular *trans*-sialidase (IT-sialidase) that targets and cleaves off terminal α2-3-linked Neu5Ac from glycoproteins, releasing 2,7-anhydro-Neu5Ac instead of Neu5Ac (4,17,18). We unravelled the molecular pathway leading to the transport and metabolism of 2,7-anhydro-Neu5Ac in *R. gnavus* ATCC 29149 (19). The 2,7-anhydro-Neu5Ac compound binds specifically to the substrate binding protein (*Rg*SBP), which forms part of an ATP-binding cassette (ABC) sialic acid transporter in *R. gnavus*. Once inside the cell, 2,7-anhydro-Neu5Ac is converted into Neu5Ac *via* a novel enzymatic reaction catalysed by an oxidoreductase, *Rg*NanOx. Following this conversion, Neu5Ac is then catabolised into *N*-acetylmannosamine (ManNAc) and pyruvate *via* the action of a Neu5Ac-specific aldolase (19). We confirmed the importance of this metabolic pathway *in vivo* by generating a *R. gnavus nan* cluster

deletion mutant that lost the ability to grow on sialylated substrates. We showed that in gnotobiotic mice colonised with *R. gnavus* wild-type and mutant strains, the fitness of the *nan* mutant was significantly impaired as compared to the wild-type strain with a reduced ability to colonise the mucus layer (19).

The novel oxidoreductase identified in our work showed to catalyse the conversion 2,7-anhydro-Neu5Ac to Neu5Ac in the presence of NAD^+ in a reversible manner (19). Recently, the crystal structure of the *Rg*NanOx homologue in *Escherichia coli*, YjhC, was solved in complex with NAD^+ (20). However, the mechanism of action of these newly identified enzymes activity remained undefined. Here, using a combination of *in silico*, molecular, biochemical and structural approaches, we elucidated the molecular mechanism of *Rg*NanOx and showed that homologous enzymes are present across both Gram positive and Gram negative bacteria, and associated with different classes of predicted transporters. We validated these data *in vitro* and further unravelled the 2,7-anhydro-Neu5Ac catabolism in *E. coli*.

Results

Characterisation of *R. gnavus* oxidoreductase (*Rg*NanOx) reaction revealed a novel mechanism involving a 4-keto-DANA intermediate

*Rg*NanOx (RUMGNA_02695) from *R. gnavus* ATCC 29149 catalyses the equilibration of 2,7-anhydro-Neu5Ac and Neu5Ac (19). To gain insights into the mechanism of action of the enzyme, the conversion of 2,7-anhydro-Neu5Ac to Neu5Ac was monitored by ^{1}H NMR for 24 hours. Comparison of samples differing only in the solvent (light water (H_2O) vs deuterated water (D_2O)) at the same reaction time points showed the loss of specific ^{1}H signals from protons attached at C3 and C5 in D_2O indicating solvent exchange. The loss of these signals led to simplification in the splitting of the neighbouring protons signals (**Supp. Fig. 1**).

Analysis of the reaction curve, obtained by monitoring the signals from the methyl protons of the acetamide group at C5, revealed the presence of a new (third) molecule (XY). This molecule is formed with a very fast kinetics, reaching its highest concentration at very early time points. It is consumed more slowly rate, suggesting that it is a reaction intermediate (**Fig. 1A**). The reaction mixture was analysed by 2D NMR during the interval at which the intermediate was at its highest concentration, using a 600 MHz spectrometer equipped with cryoprobe. The comparison of the 2D ^{1}H,^{13}C HSCQ spectra of the substrate (2,7-anhydro-Neu5Ac), the product (Neu5Ac) and the reaction mixture at 30 min, allowed the identification of an additional set of cross peaks that did not belong to the substrate nor to the product, and were therefore assigned to the intermediate (**Fig. 1B** and **C**). Specifically, the presence of a heteronuclear cross peak at 2.73 ppm/43.1 ppm (^{1}H/^{13}C) (Signal 2 in **Fig. 1B**), characteristic of a proton in α to a keto group, strongly suggested the intermediate to be a keto-sugar. Homonuclear 2D ^{1}H,^{1}H COSY and ^{1}H,^{1}H TOCSY experiments showed connectivity of this proton to one and two other protons, respectively (**Supp. Fig. 2**). Signal 2, at 2.73 ppm, corresponds then to a proton neighbouring a carbonyl on one side and, at least, two other protons on the other side with chemical shifts in the 3-4 ppm region. Analysis of the 3-4 ppm spectral region, where standard sugar-ring C-H signals typically show, suggests that the intermediate chemical shifts are closer to Neu5Ac than to 2,7-anhydro-Neu5Ac (**Supp. Fig. 3**). Based on this information, we proposed that in this species the glycerol moiety is in the open form, meaning that the 2,7-anhydro bond is broken in the intermediate. The formation of such an oxidised intermediate (i.e. the keto-sugar), as detected by NMR, is compatible with the regeneration of NAD+ previously observed (19). As the final result of the reaction is a reductive opening of the ring through the O7-C2 bond, a preliminary oxidation must occur earlier on in the mechanism.

Based on this analysis, we propose a multi-step mechanism for the reversible conversion of 2,7-anhydro-Neu5Ac to Neu5Ac by *Rg*NanOx (**Fig. 2**). In the first step, 2,7-anhydro-Neu5Ac is oxidised at the C4 keto by NAD^+ cofactor. The proton at C3 is now α to the keto group, thus more acidic and can be abstracted in an anti-periplanar elimination reaction in which the 2,7-anhydro bond is also broken. The resulting compound, 4-keto-2-deoxy-2,3-dehydro-N-acetylneuraminic acid (4-keto-DANA), is consistent with 2D NMR, and the conjugation would be expected to stabilise the

intermediate (full NMR assignment of 4-keto-DANA is given in **Table S1**). 4-keto-DANA exists as an equilibrium between two ring flipped forms, however the form with the equatorial glycerol and *N*-acetyl substituents would be expected to predominate. In this ring flipped form, a Michael addition of water to C2, hydrates the C2-C3 double bond with of a proton from the solvent added to C3. The addition reaction would be expected to follow the expected anti-periplanar geometry. This is consistent with the NMR observation that axial not equatorial on proton undergoes exchange (**Supp. Fig. 1**). The product of the addition (2) is a 4-keto-Neu5Ac, in which the proton at C5 is now α to the keto and acidic. This acidic proton will exchange with solvent by the well-known keto enol tautomerisation reaction, as seen the NMR data (**Supp. Fig. 1**). The 4-keto-Neu5Ac (2) is finally reduced by NADH to yield the final Neu5Ac product and NAD^+. The regeneration of NAD^+ also explains why no net change in the ratio of NAD+ and NADH was observed in the enzymatic reaction (19).

The crystal structure of *Rg*NanOx in complex with NAD^+ confirms structural homology with short chain dehydrogenase/reductases

The crystal structure of the recombinant *Rg*NanOx in complex with the NAD^+ cofactor was solved at 2.58 Å and subsequently at 1.74 Å using molecular replacement with an oxidoreductase from *Agrobacterium radiobacter* as a model (PDB 5UI9). The protein shows a Rossman fold typical of NAD binding protein of the Gfo/Idh/MocA class (21) characterised by a central β-sheet with helices on either side. *Rg*NanOx forms a dimer in the asymmetric unit, and the NAD^+ cofactor can be seen bound to the protein (**Fig. 3A**). Structural similarity search (22) identifies multiple oxidoreductase enzymes, all of which share the same Rossman fold and location of the nucleotide bind site Oxidoreductase proteins are typically have a catalytic triad of K (found in EKP motif), D and H (often found as DxxxH motif, in some enzymes Y replaces the H) and a fourth residue which is positively charged (21). *Rg*NanOx has K93 and H178 which correspond to the K and H of the catalytic triad and K163 occupies the "fourth" position (23). However, *Rg*NanOx has H175 which occupies the position typical for the D in the catalytic triad (23). The closest structural match (0.8 Å over 341 residues) is the recently solved crystal structure of YjhC oxidoreductase from *Escherichia coli* (**Fig. 3B)** (20) which also has HXXH motif. In the *Rg*NanOx structure there is additional density adjacent to the nicotinamide ring that given the high resolution of the second structure we were able to unambiguously identify as citric acid from the crystallisation buffer.

Using the high-resolution structure, we use a simple modelling approach to place a molecule of 2,3-didehydro-2-deoxy-N-acetylneuraminic acid (DANA), a transition state analog inhibitor of sialidases, in *Rg*NanOx active site by overlapping the carboxylate acid of the DANA with each of the three carboxylate groups of citric acid. Next we manually positioned the sugar such that the H3 of atom ring pointed towards the C4' of the nicotinamide as would be required for hydride transfer. These models were then minimised and inspected as to whether the H4 atom of DANA was still able to transfer to nicotinamide. Only one position (**Fig. 3C**) where the DANA carboxylate was placed on the 2-carboxylic acid of citric remained positioned for hydride transfer. The goal of modelling was not to generate a precise model of the substrate protein interactions but rather along with sequence comparisons, generate hypothesis for site-directed active site mutagenesis. The model indicated that H178, predicted to be a catalytic residue, is positioned to remove the proton from O4 during the oxidation step. H176 is plausibly positioned to undertake proton transfer with the O7 of the substrates glycerol that the mechanism requires. H175 interacts the negatively charged carboxylic acid in the model but could play a role in proton transfer at the substrate C3 atom. K93 binds to NAD^+ and is part of the catalytic constellation is also within a plausible distance for proton transfer to C3. K163 binds to the NAD^+ co-factor as well as interacting with the substrate carboxylate. The following mutants of *Rg*NanOx were constructed K93A, K163A, H175A, H176A, H178A. The purified mutants lost enzymatic activity as demonstrated by electrospray ionisation spray mass spectrometry (ESI-MS) (**Fig. 4A**), supporting the hypothesis that they are catalytically important. K93A lost the ability to bind NAD^+ as to some extent so did H176A (bound NAD^+ in the presence of Neu5Ac and 2,7-anhydro-

Neu5Ac). The other mutants did appear able to retain substrate binding (**Fig. 4B**).

Catabolism of 2,7-anhydro-Neu5Ac by *Escherichia coli* requires a *Rg*NanOx homologue (YjhC) and a novel transporter

Sequence similarity network (SSN) analysis of the *R. gnavus* Nan cluster (responsible for 2,7-anhydro-Neu5Ac metabolism) identified the presence of *Rg*NanOx homologues in a number of organisms (19). One such example was the model Gram negative human commensal *Escherichia coli* K-12, the organism in which the genes for Neu5Ac catabolism were first discovered (24,25). In *E. coli*, the homologue of *Rg*NanOx is part of a two-gene operon, *yjhBC*, which is one of only three operons in *E. coli* regulated by the transcription factor NanR as previously reported (25) (**Fig. 5A**). Here, we demonstrated that *E. coli* could grow on 2,7-anhydro-Neu5Ac as a sole carbon source (**Fig. 5B**), reaching growth yields similar to that obtained when *E. coli* was grown on Neu5Ac. Deletion of *yjhC* resulted in loss of growth on 2,7-anhydro-Neu5Ac but not Neu5Ac (**Fig. 5C**), which could be complemented *in trans* with *yjhC* (**Fig. 5D**), suggesting that the gene encodes an equivalent protein to *Rg*NanOx. To test this hypothesis, the YjhC protein was recombinantly expressed, purified and its activity against 2,7-anhydro-Neu5Ac and Neu5Ac was analysed by ESI-MS. The purified enzyme was shown to be active against both substrates, reaching an equilibrium of 2:1 Neu5Ac:2,7-anhydro-Neu5Ac, in line with that of *Rg*NanOx (19) (**Fig. 6A**). DSF analyses confirmed the binding of YjhC to the substrates 2,7-anhydro-Neu5Ac and Neu5Ac, as well as to co-factors NAD or NADH (**Fig. 6B**), as previously showed for *Rg*NanOx (19). The knowledge of the biochemical and physiological function of YjhC is particularly insightful given that the 3D structure of this protein has also recently been solved (20).

The first gene in the *yjhBC* operon, *yjhB*, encodes a major facilitator superfamily (MFS) transporter protein that shows homology (35% identify, 55% similarity) to NanT, the known Neu5Ac transporter in *E. coli* (24,26,27). Deletion of *nanT* leads to a complete loss of growth on Neu5Ac, suggesting that YjhB cannot transport this particular sialic acid (28) (**Fig. 7A**). Similar to the phenotype observed with the *ΔyjhC* strain, the *ΔyjhB* mutant was also unable to grow on 2,7-anhydro-Neu5Ac but could grow on Neu5Ac (**Fig. 7B**). The co-expression of these two genes and the requirement of YjhB for growth on2,7-anhydro-Neu5Ac suggest that YjhB is a novel MFS transporter for 2,7-anhydro-Neu5Ac and that these two gene together form an 'accessory' operon to allow *E. coli* to scavenge a wider range of sialic acids that are available in the human gut. We propose to rename these genes *nanXY*, since the function of the final NanR-regulated operon has been elucidated through this work.

Catabolism of 2,7-anhydro-Neu5Ac is widespread across the bacterial kingdom

Having demonstrated that NanOx-like genes are functional in both Gram positive and Gram negative bacteria, functioning with different classes of transporters, we extended our analysis of the diversity of likely 2,7-anhydro-Neu5Ac catabolic genes. Genes encoding proteins with high similarity to *Rg*NanOx (ID% was ≥49%) were found in diverse micro-organisms across the Firmicutes, Proteobacteria, and Actinobacterial phyla and were most often co-localised with other genes for sialic acid catabolism (**Fig. 8**). Interestingly, we showed co-occurrence of NanOx genes with known sialic acid transporters belonging to the major facilitator superfamily (MFS) transporters, sodium-solute symporter (SSS) transporter or ABC SAT transporters. To test the hypothesis that other bacteria can act as 'scavengers' of 2,7-anhydro-Neu5Ac, we heterologously expressed and purified the NanOx protein from *Haemophilus haemoglobinophilus* and showed that the recombinant protein was active against 2,7-anhydro-Neu5Ac (**Fig. 6**). The analysis also revealed two additional couplings of NanOx-like genes to likely 2,7-anhydro-Neu5Ac transporters, namely to transporters of the SSS family, for example in *Streptococcus pneumoniae* TIGR4 and a transporter of the GPH family in *Lactobacillus salivarius* (**Fig. 8**), which together with the phylogenetically broad occurrence of the NanOx-like genes, suggests that 2,7-anhydro-Neu5Ac use is not a new trait in bacteria but the result of a symbiotic evolution of bacteria in the mammalian gastrointestinal tract.

Discussion

Given the accessibility of sialic acids in mucus rich environments, their utilisation offers

pathogenic and commensal bacteria a competitive advantage to colonise and persist within the gut (2,19). The ability of *R. gnavus* strains to produce and metabolise 2,7-anhydro-Neu5Ac, provide them with a nutritional advantage, by scavenging sialic acid from host mucus in a form that they have preferentially access to (4,17-19). We previously showed that the oxidoreductase *Rg*NanOx plays a key role in the catabolism of 2,7-anhydro-Neu5Ac inside the cells by converting it into Neu5Ac, before being catabolised into GlcNAc-6-P following the canonical pathway by the successive action of NanA (Neu5Ac aldolase), NanK (ManNAc kinase) and NanE (ManNAc-6-P epimerase). However, the molecular mechanism of *Rg*NanOx enzymatic reaction remained unknown (19). Here we showed that *Rg*NanOx acts through a multi-step mechanism involving a keto intermediate and cycling of NADH / NAD^+. The creation of a keto intermediate in sugars is widespread biology, perhaps it is most well-known for the SDR enzyme UDP-glucose / galactose epimerase (29,30). In this enzyme, the oxidation and reduction of the sugar occur so as to invert the chirality at C4. In other enzymes, including RmlB of dTDP-L-rhamnose biosynthetic pathway (31) and the multi-step enzyme GDP-mannose 3,5 epimerase (32), the creation of a keto group and the consequent acidification of the α proton(s) allows a range of chemical reactions. In *Rg*NanOx by creating a keto group, the enzyme has acidified the C3 proton, this facilitates an elimination reaction and formation of a conjugated intermediate 4-keto-DANA which we detected by NMR. The NAD^+ co-factor is tightly bound, since in the overall reaction it is unchanged (acting as co-catalyst) there is no need to recycle it with an additional enzyme. Instead the NAD^+ is reduced to NADH when it creates a keto group at C4 of the substrate in the first step and in the last step NAD^+ is regenerated by reduction of the keto group.

Previous work using spectrophotometric assays reported that *E. coli* YjhC showed a very small rate for Neu5Ac with an apparent K_M of 68.8 mM (20). Here we used ESI-MS to assess the activity of YjhC against 2,7-anhydro-Neu5Ac or Neu5Ac. This analysis supported the earlier findings that YjhC could act on Neu5Ac (20) but also revealed that enzyme was able to utilise 2,7-anhydro-Neu5Ac as a substrate in the same manner as *Rg*NanOx. Given the structural resemblance of *Rg*NanOx to YjhC, it is likely that the *E. coli* oxidoreductase also uses the same mechanism of action for the reversible conversion of 2,7-anhydro-Neu5Ac to Neu5Ac.

Further findings from our work support that the catabolism of 2,7-anhydro-Neu5Ac is not restricted to *R. gnavus* strains. *E. coli* can transport and catabolise the common sialic acid, Neu5Ac, as a sole source of carbon and nitrogen but also related sialic acids, N-glycolylneuraminic acid (Neu5Gc) and 3-keto-3-deoxy-D-glycero-D-galactononononic acid (KDN), which are transported via the sialic acid transporter NanT and catabolized using the sialic acid aldolase NanA (33). Here we showed that *E. coli* BW25113 strain was able to grow on 2,7-anhydro-Neu5Ac as a sole carbon source and that the two-gene NanR-regulated operon *nanXY (yjhBC*) encodes both the transporter and oxidoreductase enzyme required for *E. coli* to uptake and catabolise 2,7-anhydro-Neu5Ac. This also now completes the functional characterisation of all NanR-regulated genes (25) in *E. coli*, giving us a broader picture of the sialic acid molecules it likely encounters in its natural environment. This strategy contrasts with *R. gnavus* strains which can only grow on 2,7-anhydro-Neu5Ac but not Neu5Ac (19) and is consistent with *E. coli* being able to integrate diverse sialic acids into its core catabolic pathway (33). Beyond *E. coli*, our bioinformatics analyses revealed *Rg*NanOx homologues across many bacterial species that also co-occurred with predicted sialic acid transporters.

Bacteria have evolved multiple mechanisms to capture sialic acid from their environment (34,35). To date, 6 different classes of sialic acid transporters have been described (35). These include the NanT MFS transporters from *E. coli* and *B. fragilis*, which in *E. coli* has been demonstrated biochemically to be a H+-coupled symporter (36). Secondary transporters implicated in sialic acid transport are from the SSS family present in pathogens such as *C. difficile*, *S. typhimurium* or *P. mirabilis* (28,37,38). High-affinity transport of sialic acid is mediated by substrate-binding protein-dependent systems, including a tripartite ATP-independent periplasmic (TRAP) transporter, SiaPQM, and a number of different ABC transporters (39-44). The sialic acid ABC transporters have been further classified into 3

types, SAT, SAT2 and SAT3 (34,35). To date all these transporters have been shown to transport Neu5Ac, with some being able to also transport the related sialic acids Neu5Gc and KDN (33,37). The exception is *R. gnavus* ABC SAT2 transporter, which we demonstrated was specific for 2,7-anhydro-Neu5Ac (19). Here we showed that in *E. coli* YjhC (NanY) is associated with a predicted sialic acid transporter, YjhB (NanX). Our genetic data suggest that YjhB could transport 2,7-anhydro-Neu5Ac but not Neu5Ac, whether the previously characterised Neu5Ac transporter NanT was not able to transport 2,7-anhydro-Neu5Ac. Based on its sequence, NanX (YjhB) is classified in the MFS class of sugar transporters and is a close homologue of NanT. Our bioinformatics provide striking evidence for two additional families of secondary transporters having evolved to recognise 2,7-anhydro-Neu5Ac, namely those of the SSS and GPH families, bringing the total number of transporter families for 2,7-anhydro-Neu5Ac to four.

It is of note that *E. coli* does not encode an IT-sialidase releasing 2,7-anhydro-Neu5Ac, therefore the ability of this strain to use 2,7-anhydro-Neu5Ac as a metabolic substrate *in vivo* would likely rely on cross-feeding in the mucosal environment. To date, only *R. gnavus* strains have been reported to produce 2,7-anhydro-Neu5Ac from Neu5Ac terminally bound glycoconjugates in the gut (4,5,18). Resource sharing is an important ecological feature of microbial communities living in the gut (45). Some bacteria present in the mucus might not be primary degrader but might cross-feed on mucin glycan degradation products released by other bacteria. This concept involves the ability of bacteria to benefit from substrate degradation products but also from fermentation products and plays a crucial role in microbial community shaping in the gut (46). Such cross-feeding activities have been reported in the gut mucosal environment for the utilisation of Neu5Ac. For example, *Bacteroides thetaiotaomicron* VPI-5482 encodes a sialidase and can release free Neu5Ac but lacks the nan operon required to metabolise the liberated monosaccharide (47). On the other hand, most *Clostridium difficile* and *Salmonella typhimurium* subsp. enterica strains encode the *nan* operon but lack the sialidase (48) and benefit from sialidase-producing organisms such as *B. thetaiotaomicron* to acquire this nutrient from the mucosal environment (49). Our bioinformatics analyses suggest a similar diversity for 2,7-anhydro-Neu5Ac metabolism across bacterial species. For example, while *R. gnavus* possesses the full complement of genes to produce and utilise 2,7-anhydro-Neu5Ac, including the IT-sialidase (*Rg*NanH), the 2,7-anhydro-Neu5Ac SAT2 transporter and the oxido-reductase (*Rg*NanOx) within the otherwise canonical Nan cluster, *E. coli* harbours a transporter with specificity for 2,7-anhydro-Neu5Ac (NanX), the homologous NanOx (NanY) , but does not express an IT-sialidase. *Streptococcus pneumoniae* strains on the other hand, may express up to three sialidases (neuraminidases), NanA, NanB and NanC, of which the first two are part of a universally conserved *nan* gene cluster (42), while the third one is part of an additional locus present in some strains but not others (50). The conserved *nan* cluster is well studied in strain D39 (42,51), and is divided in three operons that include operon I (*nanA* monocistronic), operon II (the *nanB* locus), and operon III (the *nanE* locus carrying the catabolic genes) (51). Transcriptomic response of *S. pneumoniae* D39 to Neu5Ac clearly demonstrated that NanR acts as a transcriptional activator of the *nan* operons I and III the presence of Neu5Ac, but not of operon II, for which regulation mechanisms remained unknown (51). Since NanB has been functionally characterised as an IT-sialidase in *S. pneumoniae* (52) and that the *nan* operon II also contains a gene encoding an oxidoreductase and a SAT2 ABC transporter (as in the case of *R. gnavus*), our results strongly suggest that the *nan* operon II is dedicated to 2,7-anhydro-Neu5Ac utilisation. This is also in agreement with the reported growth assays of *S. pneumoniae* transporter mutants, showing that SAT3 was required for Neu5Ac transport but that growth on Neu5Ac was unaffected in SAT2 mutant (42), suggesting that SAT2 may be involved in 2,7-anhydro-Neu5Ac, although this remains to be tested experimentally. The existence of multiple transporters with different specificities for sialic acid derivatives within the same species (e.g. *E. coli* NanT/YjhB) or restricted to 2,7-anhydro-Neu5Ac (e.g. *R. gnavus* SAT2) points towards divergent evolution of a common ancestor. Together, these data demonstrate that 2,7-anhydro-Neu5Ac

catabolism is not exclusive to *R. gnavus* and may help shape microbial communities in the gut. From an ecological point of view, since *R. gnavus* is the only strain reported to produce 2,7-anhydro-Neu5Ac in the gut, the strict specificity of its sialic acid transporter may give it a nutritional advantage while maintaining its keystone status in the mucus niche by providing an important nutrient to the microbial community.

Experimental Procedures

Materials and strains

All chemicals were obtained from Sigma (St Louis, MO, USA) unless otherwise stated. The Auto-induction media (AIM) 'Terrific Broth Base with Trace Elements' autoinduction media was purchased from ForMedium (Dundee, UK). 2,7-anhydro-Neu5Ac was produced as reported in Bell *et al.* (19). The *E. coli* Δ*nanT*, Δ*yjhB* (JW5768), and Δ*yjhC* (JW5769) mutants all come from the KEIO collection (53) and are derivatives of BW25113, which we used as WT strain. The Δ*nanT* strain, which we have previously characterised (28,36) carries an un-marked deletion, while the ΔyjhB and ΔyjhC mutants are un-modified and retain the original Kan marker.

Bacterial growth assays

Growth curves of *E. coli* BW25113 and sialometabolism mutants (Δ*nanT*, Δ*yjhC* and Δ*yjhB*) before or after complementation were carried out in M9 medium (without glucose) supplemented with 11.1 mM Neu5Ac, 2,7-anhydro-Neu5Ac or glucose using 200 µl cultures in 96-well microtiter plates. The OD_{595nm} was measured every 30 or 60 min for 16 h in a FLUOstar OPTIMA (BMG LABTECH).

Complementation of *E. coli* Δ*yjhC*

Plasmid pES156 is a derivative of the low-copy plasmid pWKS30 (Wang and Kushner, 1991) carrying *E. coli yjhBC* under the control of the lac promoter. To make pES156 a PCR product for the *yjhBC* genes was amplified with primers E549 and E550 (**Table S2**) cut with Eco31I (the primers carried sites for this type IIS enzyme that were designed to produce Acc65I- and BamHI-compatible ends) and ligated in to pWKS30. The constructs were verified by sequencing. The resulting plasmid was transformed into chemically competent Δ*yjhC* cells using heat shock. For bacterial growth curves 1mM IPTG was added to the growth medium for induction of the *yjhBC* proteins.

Cloning, site-directed mutagenesis, heterologous expression and protein purification

*Rg*NanOx was heterologously expressed and purified as previously described (19). Mutants of *Rg*NanOx were generated using the NZYMutagenesis kit (NZYTech) using the primers listed in **Table S2** following the manufacturer's instructions, expressed and purified as previously described for the wild-type enzyme (19). Fractions collected following gel filtration were analysed using NuPAGE Novex 4-12 % Bis-Tris gels (Life Technologies, UK). Fractions were pooled and concentrated using a 10-kDa MWCO Vivaspin column (Vivaspin, Germany). Protein concentration was determined by NanoDrop (Thermo Scientific, USA) using extinction coefficients calculated by Protparam (ExPASy-Artimo, 2012) from the peptide sequence.

The oxidoreductase genes encoding YjhC from *E. coli* BW25113 and the homologue in *H. haemoglobinophilus* CCUG 3714 (B0186_5960; strain purchased from DSMZ, ref DSM21241) were amplified from genomic DNA with HerculaseIIFusion (Agilent) using primer pairs E521 and E522, and ES525 and ES526, respectively (**Table S2**).The PCR products, each flanked by sites for Type IIS enzyme Eco31I (BsaI) designed to produce NcoI and XhoI compatible ends, were digested with Eco31I and ligated in pET28a digested with NcoI and XhoI. The resulting constructs were confirmed by sequencing.

For purification of these recombinant proteins, the corresponding expression plasmids were transformed into BL21(DE3) pLysS, and single colonies were grown overnight in 10 ml LB with Cm15 Kan25. These starters were used to inoculate 1 L of Terrific Broth (TB 47.6 g/L, glycerol 0.5% v/v) with antibiotics. Cultures were grown in baffled flasks at 37 °C and 200 rpm until 0.7 OD_{650} before being quickly chilled on ice-water for about 20 min and finally

induced overnight with 0.2 mM IPTG at 18 °C and 220 rpm. After harvest, pellets were resuspended in equilibration buffer (50 mM KPi, 200 mM NaCl, 20% glycerol, 40 mM imidazole, pH 7.8) and disrupted with sonication. The clarified lysate was run through an IMAC column to elute the C-terminally His6-tagged proteins, which eluted in sharp peaks with a single 500 mM imidazole step. These were further purified by SEC with elution in 50 mM KPi, 200 mM NaCl, pH 7.8, dialysed against assay buffer (20 mM Tris, 150 mM NaCl, 2 mM TCEP, pH 7.5), and finally concentrated to 0.5-1 mM for storage at 4 °C. Purity was assessed on SDS PAGE as above.

ESI-MS analysis of the oxidoreductase reaction

To assay for oxidoreductase activity, the purified recombinant proteins were incubated in 100 µl reactions at 37 °C overnight with 1 mg/ml 2,7-anhydro-Neu5Ac or Neu5Ac in 20 mM sodium phosphate buffer pH 7.5, in the presence 500 µM NADH. The conversion of 2,7-anhydro-Neu5Ac to Neu5Ac or Neu5Ac to 2,7-anhydro-Neu5Ac was monitored by ESI-MS. Briefly, 100 µl acetonitrile was added to each reaction, vortexed and centrifuged to remove particles. The resulting supernatants were loaded onto an AmaZon Speed ETD (Bruker) mass spectrometer and analysed by direct injection in negative mode.

NMR analyses of the oxidoreductase reaction

All 1D NMR experiments were performed using a Bruker Advance I 500MHz spectrometer with a 5 mm PATXI 1H/D-13C/15N Z-GRD probe at 293 K. To follow the kinetics of the reaction and assess the position of deuteration, two samples containing 2 mM 2,7-anhydro-Neu5Ac, 100 µM NADH and 60 µM *Rg*NanOx were used, one in deuterated PBS buffer (PBS/D_2O) and one in standard PBS buffer (PBS/H_2O, containing 10% D_2O for locking purposes). The reaction was followed by acquiring 1D NMR experiments at 15 min intervals, over 24 h. The standard *zg* pulse sequence was used for the D_2O sample, while excitation sculpting was used to remove the strong solvent signals for the H_2O sample (pulse sequence: *zgesgp*). 1 mM DSS-d_6 was added to each sample as an internal reference and the signal calibrated to 0 ppm. All 2D-NMR experiments were performed using an Avance NEO 600 MHz NMR spectrometer equipped with He-cooled TCI cryoprobe. Typically, a spectral width of 12 ppm was used for ^1H and 165 ppm for ^{13}C, acquiring 8 scans per experiment with a TD of 2K data points in the direct dimension and 128 experiments in the indirect dimension. To characterise the reaction intermediate, a sample containing 3 mM 2,7-anhydro-Neu5Ac, 100 µM NADH, 15 µM *Rg*NanOx and 1 mM DSS-d_6 in PBS/D_2O was prepared and analysed at 293 K. A full set of 2D-NMR experiments, including HSQC (*hsqcetgpsi*), COSY (*cosygpqf*), TOCSY (*mlevph*) and TOCSY-HSQC (*hsqcdietgpsi*), was acquired in order to fully assign ^1H and ^{13}C signals of the intermediate.

Differential scanning fluorimetry (DSF)

The Applied Biosystems StepOnePlus Real-Time PCR system (Life Technologies Ltd) was used to record the thermal stability of the purified proteins with and without substrates or cofactors. Reactions were performed in 20 mM sodium phosphate pH7.5 and consisted of 5 µM protein, 5x SYPRO Orange (prepared as a 40x stock), 10 mM substrate (2,7-anhydro-Neu5Ac or Neu5Ac), 1 mM cofactor (NAD or NADH) in a 20 µL final reaction volume. Samples were heated from 10 °C to a 99 °C in 0.5 °C increments, taking fluorescent readings at each time point. Triplicate measurements were performed for each sample. Negative controls were included for each component of the experiment individually as well as a dye only control well. The melting point (Tm) of each sample in °C was obtained from the lowest point of the first derivative plot.

Crystallization, data collection and structure determination of *Rg*NanOX

Sitting drop vapour diffusion crystallisation experiments of *Rg*NanOx were set up at a concentration of 20 mg/ml. The structure was acquired from a crystal grown in the JCSG Plus screen (100 mM sodium citrate pH 5.5, 20% PEG 3000). The diffraction experiment was performed on the I04 beamline at Diamond Light Source Ltd at 100K using a wavelength of 0.9795 Å. The data were processed with Xia2

pipeline. The structure was phased using online MrBump Sculptor pipeline (54) with PDB 5UI9. Refinement was carried out using Phenix AutoBuild (55), Refmac (56), and PDB redo (57). Coot (58) was used for manual model building and Molprobity (59) for structure validation. It has not yet been possible to obtain well diffracting crystals of any substrate analogue complexed the protein. For modelling, after manual positioning in COOT the structure was idealised using REFMAC5 (56).

Bioinformatics

To search and compare protein sequences for *Rg*NanOx, the basic local alignment search tool (BLAST) program and BLASTp (60) were used. The DALI server (61) was used for protein structure comparison. Amino acid sequences and atomic structures of homologues were sourced from the NCBI/UniProt and Protein Data Bank (PDB) databases, respectively. BLAST searches were initiated with *Rg*NanOx/YjhC as queries followed by manual annotation of the ORFs going outwards. To map complete sialometabolic pathways within individual micro-organisms, BLAST searches were performed against all known Neu5Ac transporters (35) as well as for the Neu5Ac aldolase NanA and N-acetylmannosamine-6-phosphate epimerase NanE (using queries of different organismal origin). This was carried out to include annotation of genes for *Rg*NanOx homologues residing outside the cluster, or where multiple sialometabolic clusters were present.

Data availability

The atomic coordinates have been deposited in the Protein Data Bank under accession codes 6Z3B (2.58 Å resolution structure of *R*gNanOx), and 6Z3C (1.74 Å resolution structure of *R*gNanOx). The authors declare that all other data supporting the findings of this study are available within the article and its supplementary information files.

Acknowledgments

The authors gratefully acknowledge the support of the Biotechnology and Biological Sciences Research Council (BBSRC); this research was funded by the BBSRC Institute Strategic Programme Gut Microbes and Health BB/R012490/1 and its constituent project BBS/E/F/000PR10353 (Theme 1, Determinants of microbe-host responses in the gut across life). AB was supported by the BBSRC Norwich Research Park Biosciences Doctoral Training Partnership grant number BB/M011216/1. SM and JA acknowledge financial support from BBSRC (BB/P010660/1). JA also acknowledges financial support from the Universidad de Sevilla (Acciones Especiales del VI Plan Propio de Investigación y Transferencia). We would like to acknowledge Reyme Herman for assistance with the purification of YjhC and B0186_05690. We would like to Diamond Light Source beamlines I04 and I24 for beamtime and assistance.

Conflicts of interest

The authors declare that they have no conflicts of interest with the contents of this article.

References

1. Varki, A., Schnaar, R. L., and Schauer, R. (2017) Sialic Acids and Other Nonulosonic Acids. in *Essentials of Glycobiology* (Varki, A., Cummings, R. D., and Esko, J. D. eds.), 3rd edition Ed., Cold Spring Harbour (NY). pp
2. Juge, N., Tailford, L., and Owen, C. D. (2016) Sialidases from Gut Bacteria: A Mini-Review. *Biochemical Society Transactions* 44, 166-75
3. Vimr, E. R., Kalivoda, K. A., Deszo, E. L., and Steenbergen, S. M. (2004) Diversity of Microbial Sialic Acid Metabolism. *Microbiology and Molecular Biology Reviews* 68, 132-53
4. Crost, E. H., Tailford, L. E., Monestier, M., Swarbreck, D., Henrissat, B., Crossman, L. C., and Juge, N. (2016) The Mucin-Degradation Strategy of Ruminococcus Gnavus: The Importance of Intramolecular Trans-Sialidases. *Gut Microbes* 7, 302-312
5. Crost, E. H., Tailford, L. E., Le Gall, G., Fons, M., Henrissat, B., and Juge, N. (2013) Utilisation of Mucin Glycans by the Human Gut Symbiont Ruminococcus Gnavus Is Strain-Dependent. *PLoS One* 8, e76341
6. Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D. R., Li, J., Xu, J., Li, S., Li, D.*, et al.* (2010) A Human Gut Microbial Gene Catalogue Established by Metagenomic Sequencing. *Nature* 464, 59-65
7. Kraal, L., Abubucker, S., Kota, K., Fischbach, M. A., and Mitreva, M. (2014) The Prevalence of Species and Strains in the Human Microbiome: A Resource for Experimental Efforts. *PLoS One* 9, e97279
8. Olbjorn, C., Cvancarova Smastuen, M., Thiis-Evensen, E., Nakstad, B., Vatn, M. H., Jahnsen, J., Ricanek, P., Vatn, S., Moen, A. E. F., Tannaes, T. M., Lindstrom, J. C., Soderholm, J. D., Halfvarson, J., Gomollon, F., Casen, C.*, et al.* (2019) Fecal Microbiota Profiles in Treatment-Naive Pediatric Inflammatory Bowel Disease - Associations with Disease Phenotype, Treatment, and Outcome. *Clin Exp Gastroenterol* 12, 37-49
9. Sokol, H., Jegou, S., McQuitty, C., Straub, M., Leducq, V., Landman, C., Kirchgesner, J., Le Gall, G., Bourrier, A., Nion-Larmurier, I., Cosnes, J., Seksik, P., Richard, M. L., and Beaugerie, L. (2018) Specificities of the Intestinal Microbiota in Patients with Inflammatory Bowel Disease and Clostridium Difficile Infection. *Gut Microbes* 9, 55-60
10. Nishino, K., Nishida, A., Inoue, R., Kawada, Y., Ohno, M., Sakai, S., Inatomi, O., Bamba, S., Sugimoto, M., Kawahara, M., Naito, Y., and Andoh, A. (2018) Analysis of Endoscopic Brush Samples Identified Mucosa-Associated Dysbiosis in Inflammatory Bowel Disease. *J Gastroenterol* 53, 95-106
11. Machiels, K., Sabino, J., Vandermosten, L., Joossens, M., Arijs, I., de Bruyn, M., Eeckhaut, V., Van Assche, G., Ferrante, M., Verhaegen, J., Van Steen, K., Van Immerseel, F., Huys, G., Verbeke, K., Wolthuis, A.*, et al.* (2017) Specific Members of the Predominant Gut Microbiota Predict Pouchitis Following Colectomy and Ipaa in Uc. *Gut* 66, 79-88
12. Hall, A. B., Yassour, M., Sauk, J., Garner, A., Jiang, X., Arthur, T., Lagoudas, G. K., Vatanen, T., Fornelos, N., Wilson, R., Bertha, M., Cohen, M., Garber, J., Khalili, H., Gevers, D.*, et al.* (2017) A Novel Ruminococcus Gnavus Clade Enriched in Inflammatory Bowel Disease Patients. *Genome Medicine* 9, 103
13. Fuentes, S., Rossen, N. G., van der Spek, M. J., Hartman, J. H., Huuskonen, L., Korpela, K., Salojarvi, J., Aalvink, S., de Vos, W. M., D'Haens, G. R., Zoetendal, E. G., and Ponsioen, C. Y. (2017) Microbial Shifts and Signatures of Long-Term Remission in Ulcerative Colitis after Faecal Microbiota Transplantation. *Isme j* 11, 1877-1889
14. Joossens, M., Huys, G., Cnockaert, M., De Preter, V., Verbeke, K., Rutgeerts, P., Vandamme, P., and Vermeire, S. (2011) Dysbiosis of the Faecal Microbiota in Patients with Crohn's Disease and Their Unaffected Relatives. *Gut* 60, 631-7
15. Willing, B. P., Dicksved, J., Halfvarson, J., Andersson, A. F., Lucio, M., Zheng, Z., Jarnerot, G., Tysk, C., Jansson, J. K., and Engstrand, L. (2010) A Pyrosequencing Study in Twins Shows That

Gastrointestinal Microbial Profiles Vary with Inflammatory Bowel Disease Phenotypes. *Gastroenterology* 139, 1844-1854.e1
16. Png, C. W., Linden, S. K., Gilshenan, K. S., Zoetendal, E. G., McSweeney, C. S., Sly, L. I., McGuckin, M. A., and Florin, T. H. (2010) Mucolytic Bacteria with Increased Prevalence in Ibd Mucosa Augment in Vitro Utilization of Mucin by Other Bacteria. *Am J Gastroenterol* 105, 2420-8
17. Owen, C. D., Tailford, L. E., Monaco, S., Šuligoj, T., Vaux, L., Lallement, R., Khedri, Z., Yu, H., Lecointe, K., Walshaw, J., Tribolo, S., Horrex, M., Bell, A., Chen, X., Taylor, G. L.*, et al.* (2017) Unravelling the Specificity and Mechanism of Sialic Acid Recognition by the Gut Symbiont Ruminococcus Gnavus. *Nature Communications* 8, 2196
18. Tailford, L. E., Owen, C. D., Walshaw, J., Crost, E. H., Hardy-Goddard, J., Le Gall, G., de Vos, W. M., Taylor, G. L., and Juge, N. (2015) Discovery of Intramolecular Trans-Sialidases in Human Gut Microbiota Suggests Novel Mechanisms of Mucosal Adaptation. *Nat Commun* 6, 7624
19. Bell, A., Brunt, J., Crost, E., Vaux, L., Nepravishta, R., Owen, C. D., Latousakis, D., Xiao, A., Li, W., Chen, X., Walsh, M. A., Claesen, J., Angulo, J., Thomas, G. H., and Juge, N. (2019) Elucidation of a Sialic Acid Metabolism Pathway in Mucus-Foraging Ruminococcus Gnavus Unravels Mechanisms of Bacterial Adaptation to the Gut. *Nat Microbiol*
20. Horne, C. R., Kind, L., Davies, J. S., and Dobson, R. C. J. (2020) On the Structure and Function of Escherichia Coli Yjhc: An Oxidoreductase Involved in Bacterial Sialic Acid Metabolism. *Proteins: Structure, Function, and Bioinformatics* 88, 654-668
21. Taberman, H., Parkkinen, T., and Rouvinen, J. (2016) Structural and Functional Features of the Nad(P) Dependent Gfo/Idh/Moca Protein Family Oxidoreductases. *Protein Science* 25, 778-86
22. Krissinel, E., and Henrick, K. (2004) Secondary-Structure Matching (Ssm), a New Tool for Fast Protein Structure Alignment in Three Dimensions. *Acta Crystallogr D Biol Crystallogr* 60, 2256-68
23. Persson, B., Kallberg, Y., Bray, J. E., Bruford, E., Dellaporta, S. L., Favia, A. D., Duarte, R. G., Jörnvall, H., Kavanagh, K. L., Kedishvili, N., Kisiela, M., Maser, E., Mindnich, R., Orchard, S., Penning, T. M.*, et al.* (2009) The Sdr (Short-Chain Dehydrogenase/Reductase and Related Enzymes) Nomenclature Initiative. *Chem Biol Interact* 178, 94-8
24. Vimr, E. R., and Troy, F. A. (1985) Regulation of Sialic Acid Metabolism in Escherichia Coli: Role of N-Acylneuraminate Pyruvate-Lyase. *Journal of Bacteriology* 164, 854-60
25. Kalivoda, K. A., Steenbergen, S. M., and Vimr, E. R. (2013) Control of the Escherichia Coli Sialoregulon by Transcriptional Repressor Nanr. *Journal of Bacteriology* 195, 4689-701
26. Martinez, J., Steenbergen, S., and Vimr, E. (1995) Derived Structure of the Putative Sialic Acid Transporter from Escherichia Coli Predicts a Novel Sugar Permease Domain. *Journal of Bacteriology* 177, 6005-10
27. Vimr, E. R. (2013) Unified Theory of Bacterial Sialometabolism: How and Why Bacteria Metabolize Host Sialic Acids. *ISRN Microbiol* 2013, 816713
28. Severi, E., Hosie, A. H., Hawkhead, J. A., and Thomas, G. H. (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 Microbiology Letters* 304, 47-54
29. Kang, U. G., Nolan, L. D., and Frey, P. A. (1975) Uridine Diphosphate Galactose-4-Epimerase. Uridine Monophosphate-Dependent Reduction by Alpha- and Beta-D-Glucose. *The Journal of biological chemistry* 250, 7099-7105
30. Thoden, J. B., Frey, P. A., and Holden, H. M. (1996) Molecular Structure of the Nadh/Udp-Glucose Abortive Complex of Udp-Galactose 4-Epimerase from Escherichia Coli: Implications for the Catalytic Mechanism. *Biochemistry* 35, 5137-5144
31. Allard, S. T. M., Beis, K., Giraud, M.-F., Hegeman, A. D., Gross, J. W., Wilmouth, R. C., Whitfield, C., Graninger, M., Messner, P., Allen, A. G., Maskell, D. J., and Naismith, J. H.

(2002) Toward a Structural Understanding of the Dehydratase Mechanism. *Structure* 10, 81-92
32. Major, L. L., Wolucka, B. A., and Naismith, J. H. (2005) Structure and Function of Gdp-Mannose-3',5'-Epimerase: An Enzyme Which Performs Three Chemical Reactions at the Same Active Site. *Journal of the American Chemical Society* 127, 18309-20
33. Hopkins, A. P., Hawkhead, J. A., and Thomas, G. H. (2013) Transport and Catabolism of the Sialic Acids N-Glycolylneuraminic Acid and 3-Keto-3-Deoxy-D-Glycero-D-Galactonononic Acid by Escherichia Coli K-12. *FEMS Microbiology Letters* 347, 14-22
34. Almagro-Moreno, S., and Boyd, E. F. (2009) Insights into the Evolution of Sialic Acid Catabolism among Bacteria. *BMC Evolutionary Biology* 9, 118
35. Thomas, G. H. (2016) Sialic Acid Acquisition in Bacteria - One Substrate, Many Transporters. *Biochemical Society Transactions* 44, 760-765
36. Mulligan, C., Geertsma, E. R., Severi, E., Kelly, D. J., Poolman, B., and Thomas, G. H. (2009) The Substrate-Binding Protein Imposes Directionality on an Electrochemical Sodium Gradient-Driven Trap Transporter. *Proceedings of the National Academy of Sciences* 106, 1778
37. North, R. A., Wahlgren, W. Y., Remus, D. M., Scalise, M., Kessans, S. A., Dunevall, E., Claesson, E., Soares da Costa, T. P., Perugini, M. A., Ramaswamy, S., Allison, J. R., Indiveri, C., Friemann, R., and Dobson, R. C. J. (2018) The Sodium Sialic Acid Symporter from Staphylococcus Aureus Has Altered Substrate Specificity. *Frontiers in Chemistry* 6, 233
38. Wahlgren, W. Y., Dunevall, E., North, R. A., Paz, A., Scalise, M., Bisignano, P., Bengtsson-Palme, J., Goyal, P., Claesson, E., Caing-Carlsson, R., Andersson, R., Beis, K., Nilsson, U. J., Farewell, A., Pochini, L.*, et al.* (2018) Substrate-Bound Outward-Open Structure of a Na(+)-Coupled Sialic Acid Symporter Reveals a New Na(+) Site. *Nat Commun* 9, 1753
39. Post, D. M., Mungur, R., Gibson, B. W., and Munson, R. S., Jr. (2005) Identification of a Novel Sialic Acid Transporter in Haemophilus Ducreyi. *Infection and Immunity* 73, 6727-35
40. Severi, E., Randle, G., Kivlin, P., Whitfield, K., Young, R., Moxon, R., Kelly, D., Hood, D., and Thomas, G. H. (2005) Sialic Acid Transport in Haemophilus Influenzae Is Essential for Lipopolysaccharide Sialylation and Serum Resistance and Is Dependent on a Novel Tripartite Atp-Independent Periplasmic Transporter. *Molecular Microbiology* 58, 1173-85
41. Muller, A., Severi, E., Mulligan, C., Watts, A. G., Kelly, D. J., Wilson, K. S., Wilkinson, A. J., and Thomas, G. H. (2006) Conservation of Structure and Mechanism in Primary and Secondary Transporters Exemplified by Siap, a Sialic Acid Binding Virulence Factor from Haemophilus Influenzae. *Journal of Biological Chemistry* 281, 22212-22
42. Marion, C., Burnaugh, A. M., Woodiga, S. A., and King, S. J. (2011) Sialic Acid Transport Contributes to Pneumococcal Colonization. *Infection and Immunity* 79, 1262-1269
43. Mulligan, C., Leech, A. P., Kelly, D. J., and Thomas, G. H. (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. *Journal of Biological Chemistry* 287, 3598-608
44. Gangi Setty, T., Cho, C., Govindappa, S., Apicella, M. A., and Ramaswamy, S. (2014) Bacterial Periplasmic Sialic Acid-Binding Proteins Exhibit a Conserved Binding Site. *Acta Crystallogr D Biol Crystallogr* 70, 1801-11
45. Pereira, F. C., and Berry, D. (2017) Microbial Nutrient Niches in the Gut. *Environmental Microbiology* 19, 1366-1378
46. Hoek, M. J. A. v., and Merks, R. M. H. (2017) Emergence of Microbial Diversity Due to Cross-Feeding Interactions in a Spatial Model of Gut Microbial Metabolism. *BMC Systems Biology* 11, 56
47. Marcobal, A., Barboza, M., Sonnenburg, E. D., Pudlo, N., Martens, E. C., Desai, P., Lebrilla, C. B., Weimer, B. C., Mills, D. A., German, J. B., and Sonnenburg, J. L. (2011) Bacteroides in the

Infant Gut Consume Milk Oligosaccharides Via Mucus-Utilization Pathways. *Cell Host & Microbe* 10, 507-514
48. Sebaihia, M., Wren, B. W., Mullany, P., Fairweather, N. F., Minton, N., Stabler, R., Thomson, N. R., Roberts, A. P., Cerdeno-Tarraga, A. M., Wang, H., Holden, M. T., Wright, A., Churcher, C., Quail, M. A., Baker, S.*, et al.* (2006) The Multidrug-Resistant Human Pathogen Clostridium Difficile Has a Highly Mobile, Mosaic Genome. *Nature Genetics* 38, 779-86
49. Ng, K. M., Ferreyra, J. A., Higginbottom, S. K., Lynch, J. B., Kashyap, P. C., Gopinath, S., Naidu, N., Choudhury, B., Weimer, B. C., Monack, D. M., and Sonnenburg, J. L. (2013) Microbiota-Liberated Host Sugars Facilitate Post-Antibiotic Expansion of Enteric Pathogens. *Nature* 502, 96-9
50. Pettigrew, M. M., Fennie, K. P., York, M. P., Daniels, J., and Ghaffar, F. (2006) Variation in the Presence of Neuraminidase Genes among Streptococcus Pneumoniae Isolates with Identical Sequence Types. *Infection and Immunity* 74, 3360-5
51. Afzal, M., Shafeeq, S., Ahmed, H., and Kuipers, O. P. (2015) Sialic Acid-Mediated Gene Expression in Streptococcus Pneumoniae and Role of Nanr as a Transcriptional Activator of the Nan Gene Cluster. *Applied and Environmental Microbiology* 81, 3121-31
52. Xu, G., Potter, J. A., Russell, R. J., Oggioni, M. R., Andrew, P. W., and Taylor, G. L. (2008) Crystal Structure of the Nanb Sialidase from Streptococcus Pneumoniae. *Journal of Molecular Biology* 384, 436-49
53. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., and Mori, H. (2006) Construction of Escherichia Coli K-12 in-Frame, Single-Gene Knockout Mutants: The Keio Collection. *Molecular Systems Biology* 2, 2006.0008
54. Keegan, R. M., McNicholas, S. J., Thomas, J. M. H., Simpkin, A. J., Simkovic, F., Uski, V., Ballard, C. C., Winn, M. D., Wilson, K. S., and Rigden, D. J. (2018) Recent Developments in Mrbump: Better Search-Model Preparation, Graphical Interaction with Search Models, and Solution Improvement and Assessment. *Acta Crystallogr D Struct Biol* 74, 167-182
55. Liebschner, D., Afonine, P. V., Baker, M. L., Bunkoczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L. W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J.*, et al.* (2019) Macromolecular Structure Determination Using X-Rays, Neutrons and Electrons: Recent Developments in Phenix. *Acta Crystallogr D Struct Biol* 75, 861-877
56. Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) Refmac5 for the Refinement of Macromolecular Crystal Structures. *Acta Crystallogr D Biol Crystallogr* 67, 355-67
57. Joosten, R. P., Long, F., Murshudov, G. N., and Perrakis, A. (2014) The Pdb_Redo Server for Macromolecular Structure Model Optimization. *IUCrJ* 1, 213-20
58. Emsley, P., and Cowtan, K. (2004) Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallogr D Biol Crystallogr* 60, 2126-32
59. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) Molprobity: All-Atom Contacts and Structure Validation for Proteins and Nucleic Acids. *Nucleic Acids Research* 35, W375-83
60. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic Local Alignment Search Tool. *Journal of Molecular Biology* 215, 403-10
61. Holm, L. (2019) Benchmarking Fold Detection by Dalilite V.5. *Bioinformatics* 35, 5326-5327
62. Plumbridge, J., and Vimr, E. (1999) Convergent Pathways for Utilization of the Amino Sugars N-Acetylglucosamine, N-Acetylmannosamine, and N-Acetylneuraminic Acid by Escherichia Coli. *Journal of Bacteriology* 181, 47-54
63. Condemine, G., Berrier, C., Plumbridge, J., and Ghazi, A. (2005) Function and Expression of an N-Acetylneuraminic Acid-Inducible Outer Membrane Channel in Escherichia Coli. *Journal of Bacteriology* 187, 1959-65

64. Severi, E., Hood, D. W., and Thomas, G. H. (2007) Sialic Acid Utilization by Bacterial Pathogens. *Microbiology* 153, 2817-22
65. Steenbergen, S. M., Jirik, J. L., and Vimr, E. R. (2009) Yjhs (Nans) Is Required for Escherichia Coli to Grow on 9-O-Acetylated N-Acetylneuraminic Acid. *Journal of Bacteriology* 191, 7134-7139
66. Tettelin, H., Nelson, K. E., Paulsen, I. T., Eisen, J. A., Read, T. D., Peterson, S., Heidelberg, J., DeBoy, R. T., Haft, D. H., Dodson, R. J., Durkin, A. S., Gwinn, M., Kolonay, J. F., Nelson, W. C., Peterson, J. D*., et al.* (2001) Complete Genome Sequence of a Virulent Isolate of Streptococcus Pneumoniae. *Science* 293, 498-506

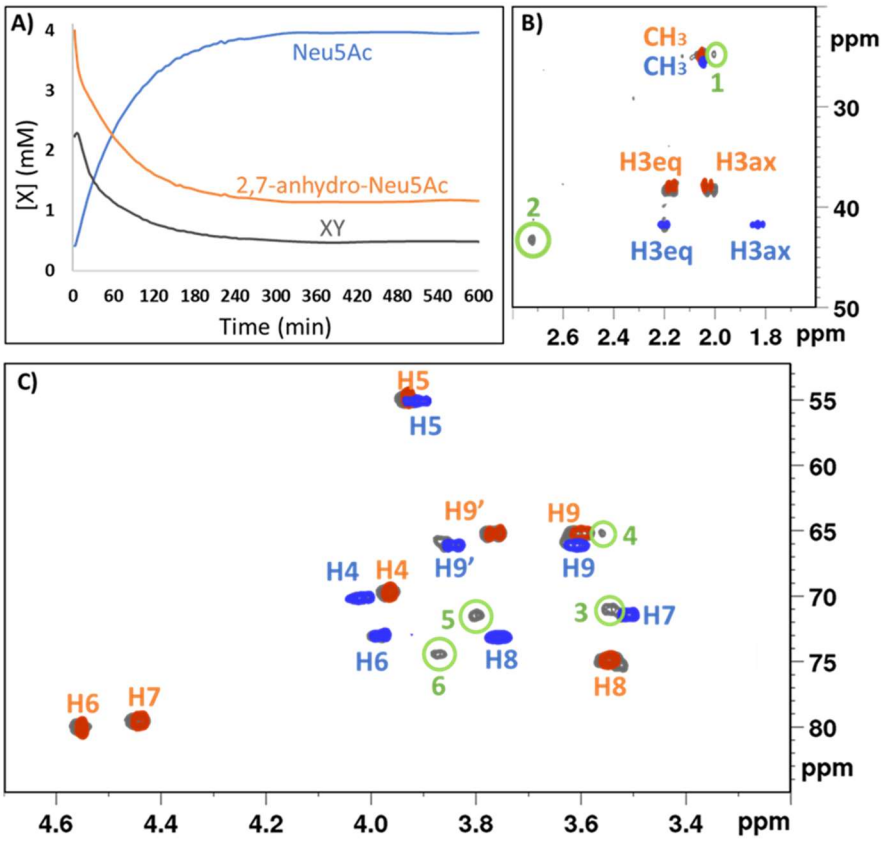

Figure 1. NMR characterisation of the *Rg*NanOx-catalysed enzymatic reaction. **A)** ^{1}H 1D-NMR analysis of the conversion of 2,7-anhydro-Neu5Ac to Neu5Ac via the XY intermediate; evolution of the concentrations of substrate, product, and intermediate of the reaction, [X](mM), by integration of the acetyl methyl groups for Neu5Ac (blue), 2,7-anhydro-Neu5Ac (orange) or XY intermediate (grey). **B,C)** Superposition of the ^{1}H,^{13}C HSQC reference spectra of Neu5Ac (blue) and 2,7-anhydro-Neu5Ac (orange), and ^{1}H,^{13}C HSQC reaction mixture at 30 min (where the XY peak is observed to peak, grey). Proton-carbon peaks not corresponding to Neu5Ac nor to 2,7-anhydro-Neu5Ac are numbered from 1 to 6 and circled in green.

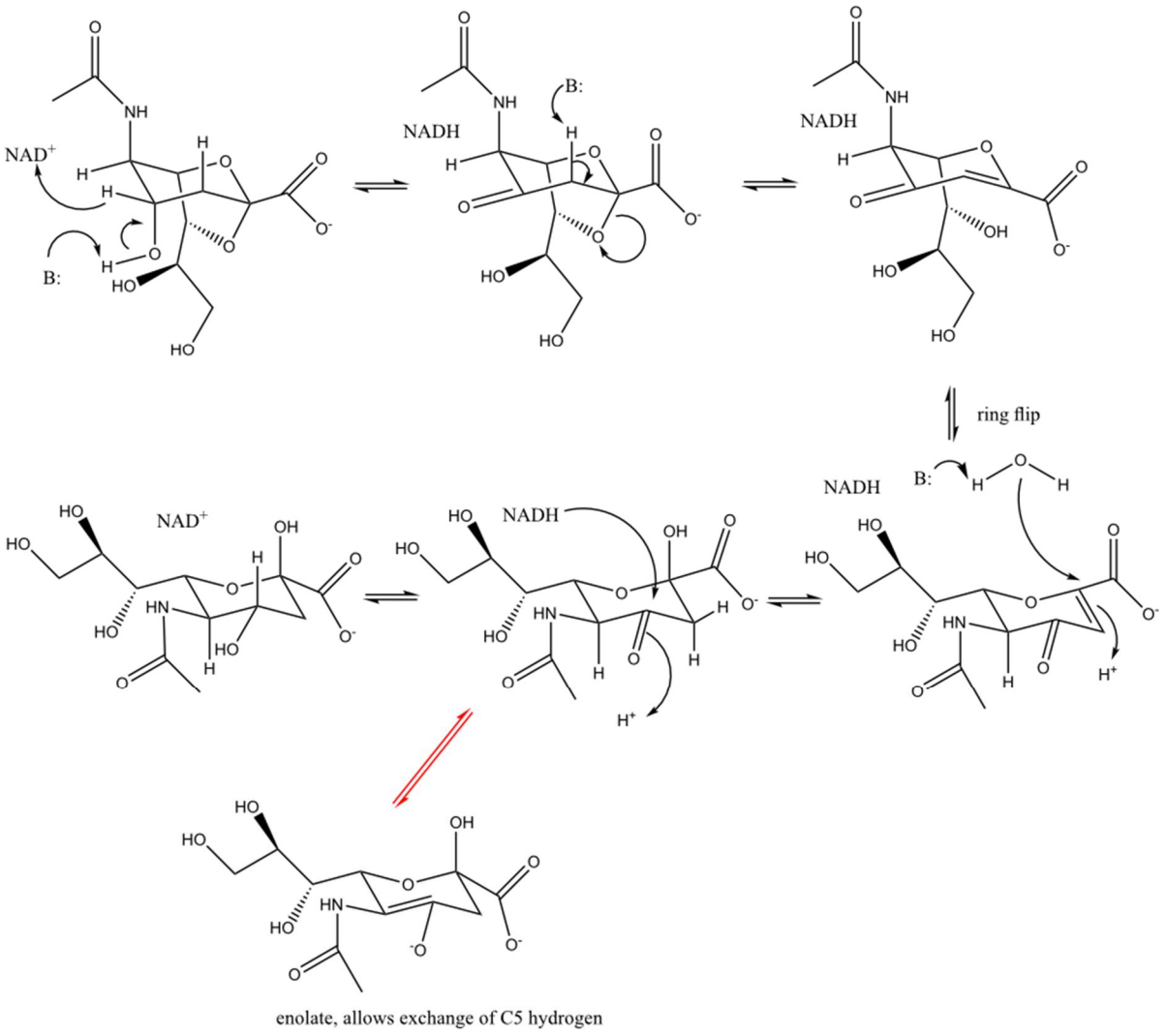

Figure 2. Proposed mechanism for the reversible conversion of Neu5Ac to 2,7-anhydro-Neu5Ac by *Rg*NanOx. The reaction is shown in the favourable direction converting 2,7-anhydro-Neu5Ac (1) to Neu5Ac (6). The order of events taking compound 2 to compound 4 including the opening of the 2,7 secondary ring and the primary ring flip has yet to be determined. The red arrows indicate the keto enol tautomerization of compound 5 that allows for the C5 hydrogen exchange.

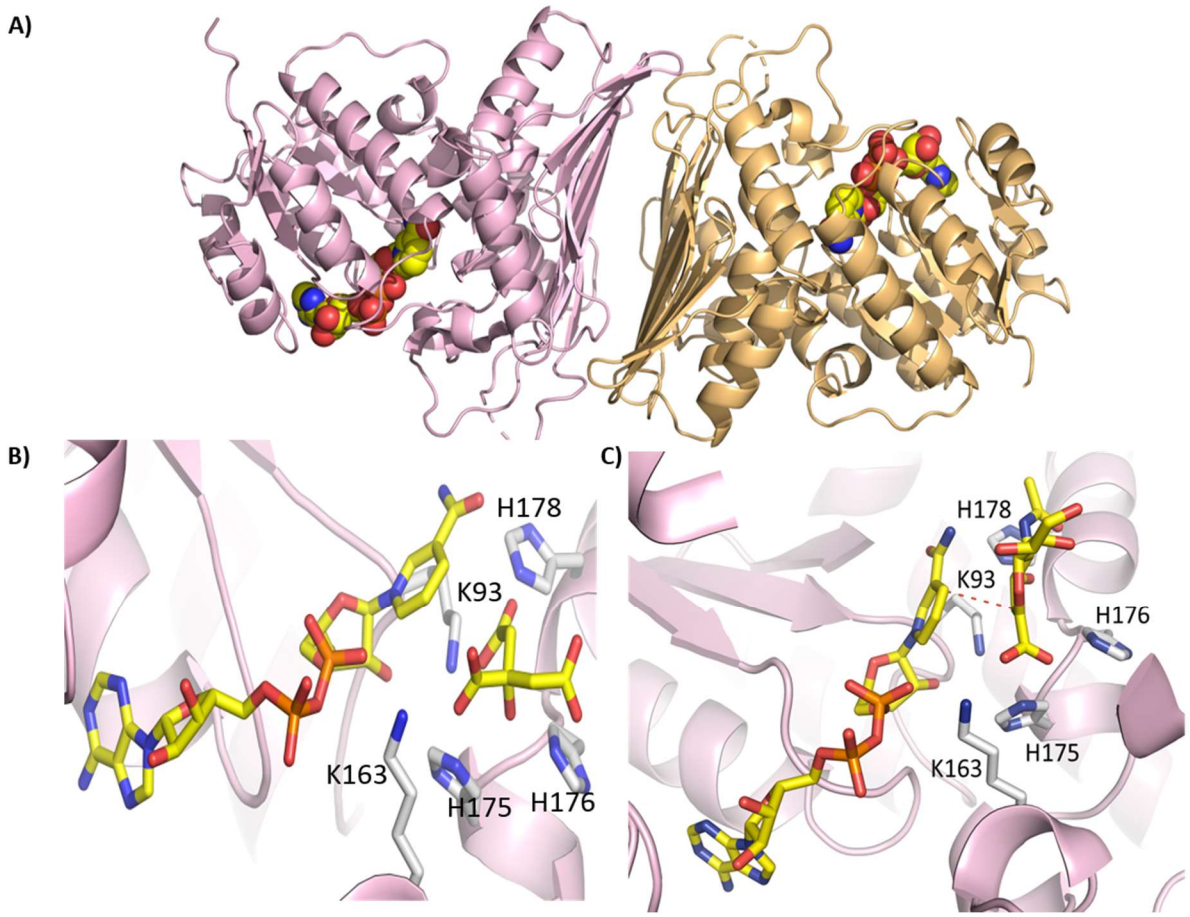

Figure 3. Crystal structure of *Rg*NanOx. **A)** Dimeric structure of *Rg*NanOx shown in cartoon format with the NAD cofactor bound (spheres). **B)** Structure of putative active site of *Rg*NanOx, protein backbone is shown in cartoon with residues NAD and citric acid shown in sticks. **C)** Crystal structure of *Rg*NanOx with key residues marked and DANA modelled into the active site, The red line marks the trajectory of hydride transfer.

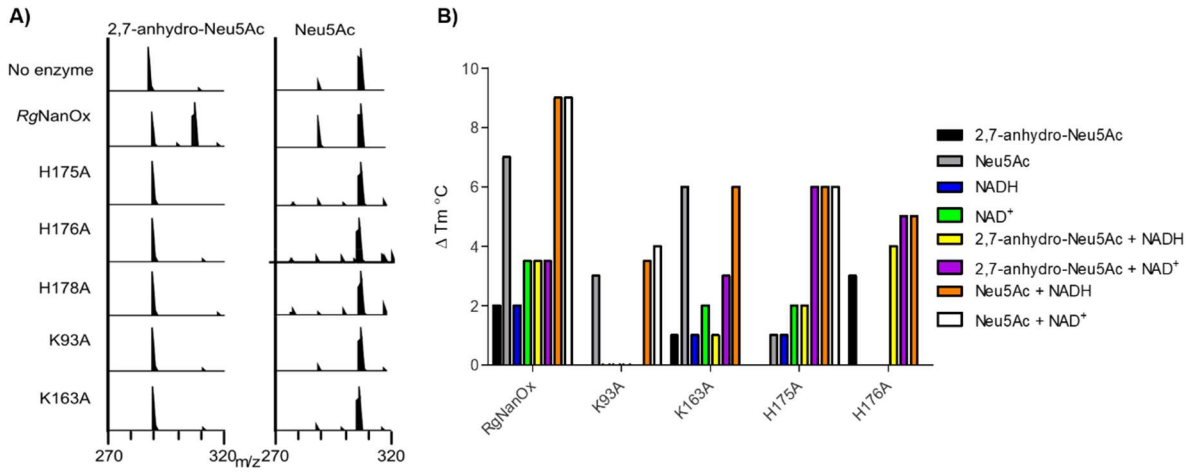

Fig. 4: Analysis of *Rg*NanOx mutants. **A)** ESI-Ms analysis of the enzymatic reaction between *Rg*NanOx mutants and 2,7-anhydro-Neu5Ac (290; left) or Neu5Ac (308; left). **B)** DSF analysis of *Rg*NanOx mutants binding to NAD/H cofactor and sialic acid substrates. The ΔTm is shown compared to the Tm of the protein alone.

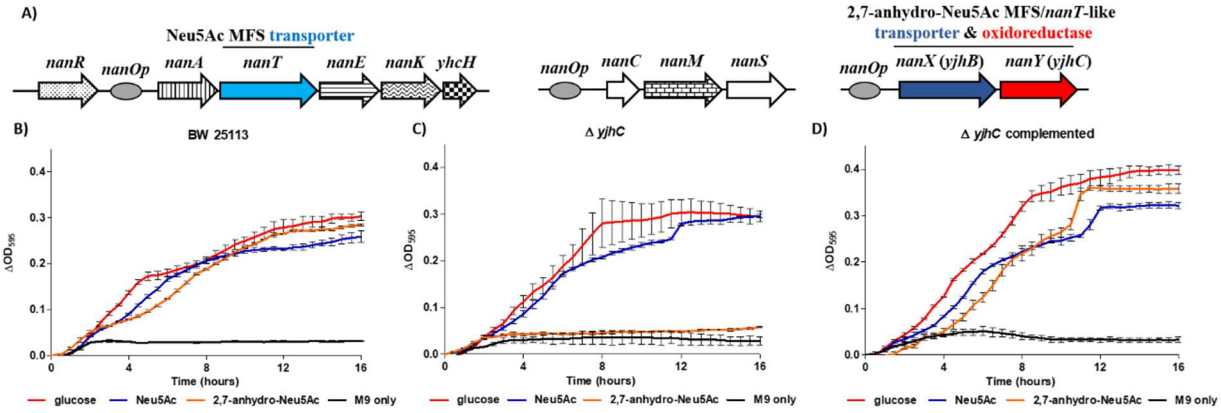

Figure 5. Structure of the sialometabolic *nan* regulon of *E. coli* K12 strains and the role of YjhC in sialometabolism by E. coli BW25113. **A)** Structure of the complete sialometabolic regulon of *E. coli* K12 strains. The repressor NanR controls expression of only these three loci in *E. coli* (25), with *nanOp* indicating the position of NanR-regulated promoters. The first locus is the core *nanATEKyhcH* operon for Neu5Ac uptake and dissimilation into the cytoplasm (62). The two "accessory" loci contain the *nanCMS* operon, for Neu5Ac uptake through the outer membrane, sialic acid mutarotation and processing of *O*-acetyled sialic acids in periplasm (63-65), and the *nanXY* (yjhBC) operon here characterised as being required for 2,7-anhydro-Neu5Ac uptake and utilisation. **B to D)** Growth of *E. coli* on different carbon sources. All strains were grown on 2,7-anhydro-Neu5Ac (red), Neu5Ac (blue), glucose (orange) or M9 media alone (black) in 200 µl microtitre plates. ΔOD_{595} for triplicate experiments is shown **B)** BW25113 **C)** Δ *yjhC*, **D)** complemented *yjhC*. Error bars display standard error of 3 biological repeats

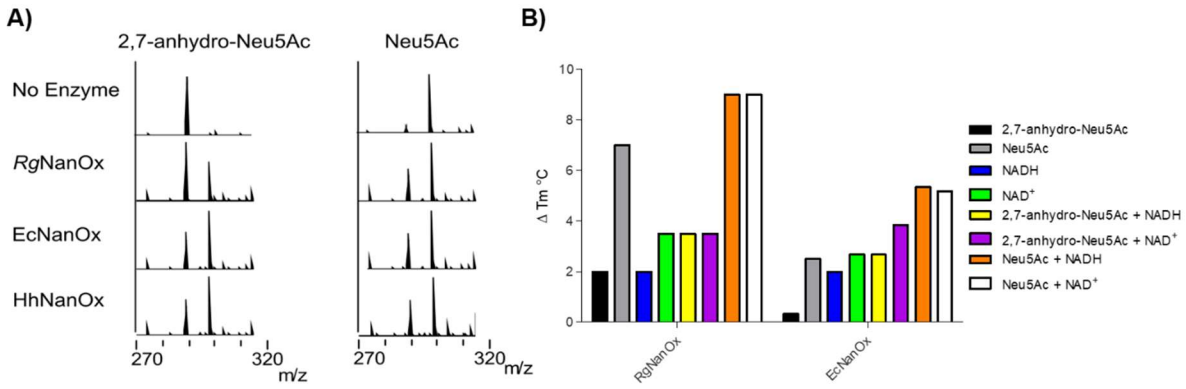

Fig. 6: Characterisation of *Ec*NanOx and *Hh*NanOx. **A)** ESI-MS analysis of the enzymatic reaction of *Rg*NanOx, *Ec*NanOx and *Hh*NanOx with 2,7-anhydro-Neu5Ac (left) or Neu5Ac (right). **B)** DSF analysis of *Rg*NanOx and *Ec*NanOx, the ΔTm is shown compared to the Tm of the protein alone.

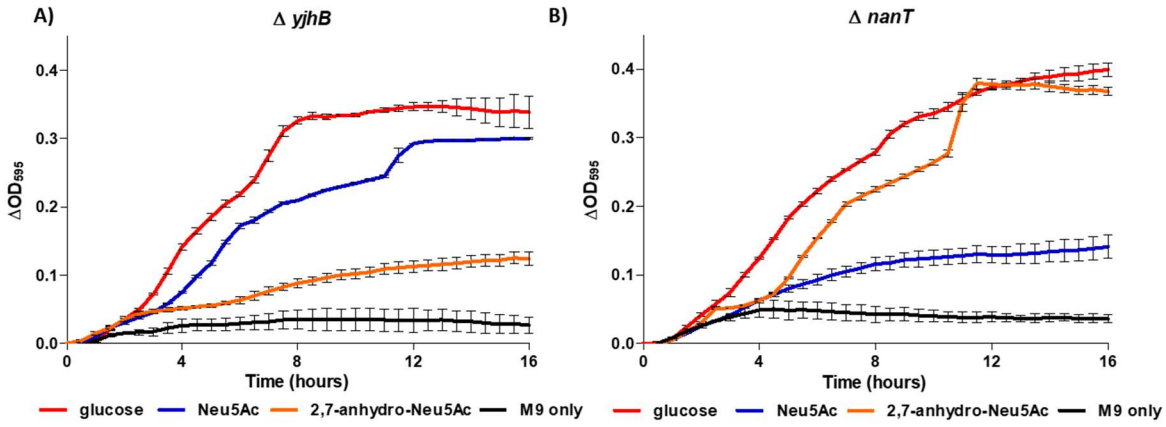

Figure 7. Growth of sialometabolism *E. coli* BW25113 transporter mutants. Both mutants were grown on 2,7-anhydro-Neu5Ac (red), Neu5Ac (blue), glucose (orange) or M9 media alone (black) in 200 µl microtitre plates. ΔOD_{595} for triplicate experiments is shown A) Δ *nanT* B) Δ *yjhB.* Error bars display standard error of 3 biological repeats

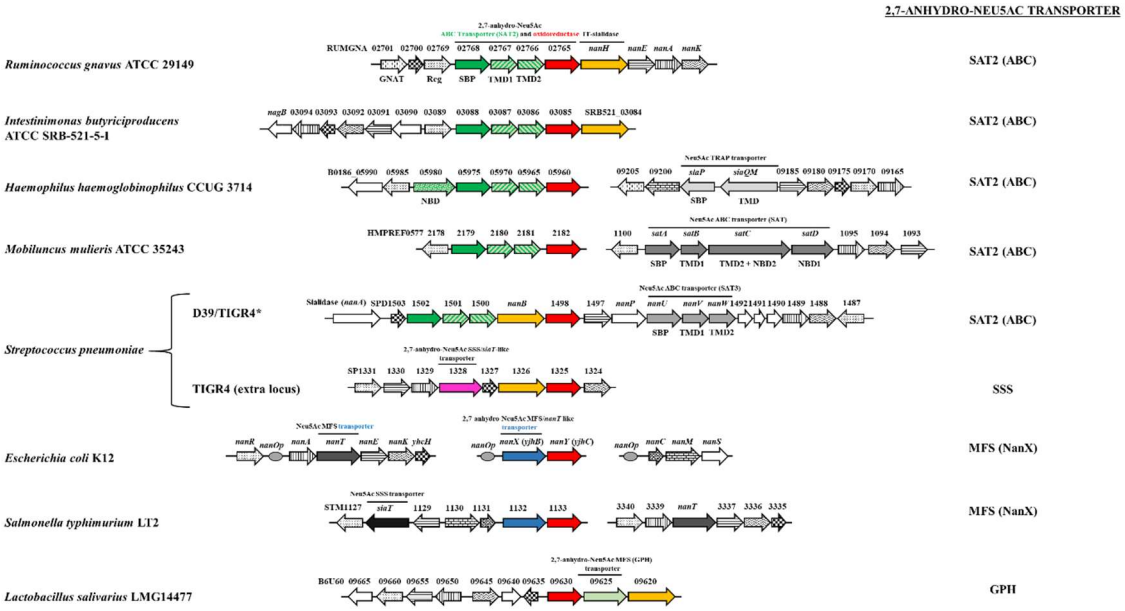

Fig. 8: Diversity of 2,7-anhydro-Neu5Ac catabolic clusters among sialic acid-utilising bacteria. Gene functions were inferred from BLAST searches followed by gene linkage and cluster analysis. Orthologous genes are identified with individual patterns. The genes encoding 2,7-anhydro-Neu5Ac transporters, 2,7-anhydro-Neu5Ac oxidoreductases, and IT-sialidases are distinguished by colour for emphasis. Nomenclature: *nanA*: Neu5Ac lyase; *nanK*: *N*-acetylmannosamine kinase; *nanE*: *N*-acetylmannosamine-6-phosphate epimerase; *nanC*: Neu5Ac outer membrane channel; *nanM*: Neu5Ac mutarotase; *nanS*: N-acetyl-9-O-acetylneuraminate esterase; *nagB*: glucosamine-6-phosphate deaminase; "GNAT": GCN5-related *N*-acetyltransferase; "Reg": regulator (please note that GNAT family proteins and regulator proteins, while recurrent within clusters, may belong to different clades thus function differently in each organism); SAT2: 2,7-anhydro-Neu5Ac transporter of the ABC superfamily; *siaPQM*: Neu5Ac transporter of the TRAP family; *satABCD*: Neu5Ac transporter of the ABC superfamily (SAT); *nanUVW* (SAT3): Neu5Ac transporter of the ABC superfamily (also named *satABC*); *nanT*: Neu5Ac transporter of the MFS superfamily; *siaT*: Neu5Ac transporter of the SSS family; *nanX* (*yjhB*): 2,7-anhydro-Neu5Ac transporter (*nanT*-like) of the MFS superfamily ABC: ATP-binding cassette; MFS: major facilitator superfamily; SSS: sodium-solute symporter family; GPH: Glycoside-pentoside-hexuronide:cation symporter family. SBP: solute binding protein; TMD: transmembrane domain; NBD: nucleotide binding domain. Note that the classification of ABC sialic acid transporters follows Almagro-Moreno & Boyd (34).

*: TIGR4 possesses both the conserved sialic acid "supercluster" as in strain D39, and an additional, candidate 2,7-anhydro-Neu5Ac cluster bearing the siaT-like transporter gene. In TIGR4, the conserved supercluster bears minor difference in the form of pseudogenes and insertions (66).

Uncovering a novel molecular mechanism for scavenging sialic acids in bacteria

Andrew Bell, Emmanuele Severi, Micah O Lee, Serena Monaco, Dimitrios Latousakis, Jesus Angulo, Gavin H. Thomas, James Naismith and Nathalie Juge

J. Biol. Chem. published online July 15, 2020
