Published in final edited form as:
Nat Microbiol. 2019 December ; 4(12): 2393–2404. doi:10.1038/s41564-019-0590-7.

Elucidation of a unique sialic acid metabolism pathway in mucus-foraging *Ruminococcus gnavus* unravels mechanisms of bacterial adaptation to the gut

Andrew Bell[1], **Jason Brunt**[1,#], **Emmanuelle Crost**[1], **Laura Vaux**[1], **Ridvan Nepravishta**[2], **C. David Owen**[3], **Dimitrios Latousakis**[1], **An Xiao**[4], **Wanqing Li**[4], **Xi Chen**[4], **Martin A. Walsh**[3], **Jan Claesen**[5], **Jesus Angulo**[2], **Gavin H. Thomas**[6], **Nathalie Juge**[1,*]

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

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

[3]Diamond Light Source Ltd, Harwell Science and Innovation Campus, Didcot, OX11 0DE, UK & Research Complex at Harwell, Harwell Science and Innovation Campus, Didcot, OX11 0FA, UK

[4]Department of Chemistry, University of California-Davis, Davis, CA 95616, USA

[5]Department of Cardiovascular and Metabolic Sciences, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44195, USA

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

Abstract

Sialic acid (Neu5Ac) is commonly found in terminal location of colonic mucins glycans where it is a much-coveted nutrient for gut bacteria including *Ruminococcus gnavus. R. gnavus* is part of

[*]**Corresponding authors**: Correspondence to Nathalie Juge. nathalie.juge@quadram.ac.uk.
[#]present address: Department of Chemical Engineering and Biotechnology, University of Cambridge, Philippa Fawcett Drive, Cambridge, CB3 0AS

Data Availability

Genome and protein sequences are available from NCBI and referenced within the text or supplementary information. Accession numbers of all genomes used for multigene alignments, are available in Supplementary Table 5. Raw FASTQ files for the RNA-seq libraries were deposited to the NCBI Sequence Read Archive (SRA), and have been assigned BioProject accession PRJNA559470. The crystal structures described in this paper have been deposited in the protein data bank (ODB) with the following identifiers 6RAB (WT), 6RB7 (K167A), and 6RD1 (K167A Neu5Ac complex). All other data are available upon request from the corresponding author. Computer code for statistical analysis is available on request from the corresponding author.

Contributions

NJ conceived the study and wrote the manuscript with contribution from co-authors. AB carried out bioinformatics analyses, transcriptomics, heterologous expression, site-directed mutagenesis, enzymatic assays, analytical product characterisation (HPLC, MS), protein-ligand interaction experiments (ITC and fluorescence spectroscopy) and *R. gnavus* mutagenesis. JB supervised the Clostron mutagenesis. GHT supervised the fluorescence spectroscopy experiments. DL developed the HPLC and MS analysis protocols. CDO carried out the X-ray crystallography under MAW's supervision. LV carried out the immuno-histochemistry experiments, EC contributed to the mouse study and RNASeq analyses. AB, EC, LV, DL worked under NJ's supervision. RN carried out the NMR experiments under JA's supervision, AX and WL synthesized the 2,7-anydro-Neu5Ac used in this study under XC's supervision. JC carried out the cluster bioinformatics analyses. All authors reviewed and corrected the final manuscript.

Competing interests

The authors declare no competing interests.

Europe PMC Funders Author Manuscripts

the healthy gut microbiota in humans but shows a disproportionate representation in diseases. There is therefore a need in understanding the molecular mechanisms underpinning its adaptation to the gut. Previous *in vitro* work demonstrated that *R. gnavus* mucin glycan-foraging strategy is strain-dependent and associated with the expression of an intramolecular *trans*-sialidase releasing 2,7-anhydro-Neu5Ac instead of Neu5Ac from mucins. Here, we have unravelled the metabolism pathway of 2,7-anhydro-Neu5Ac in *R. gnavus* which is underpinned by the exquisite specificity of the sialic transporter for 2,7-anhydro-Neu5Ac, and by the action of an oxidoreductase converting 2,7-anhydro-Neu5Ac into Neu5Ac which then becomes substrate of a Neu5Ac-specific aldolase. Having generated 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 with a reduced ability to colonise the mucus layer. Overall, our study revealed a unique sialic acid pathway in bacteria, with significant implications for the spatial adaptation of mucin-foraging gut symbionts in health and disease.

Keywords

Sialic acid; mucus; mucin; *Ruminococcus gnavus*; intramolecular *trans*-sialidase, 2,7-anhydro-Neu5Ac, sialic acid transporter; carbohydrate binding; ABC transporter; sialic acid aldolase; colonisation; sialic acid metabolism

Introduction

The gastrointestinal (GI) tract is heavily colonized with bacteria that play a vital role in human health. The gut microbiota composition varies longitudinally along the GI tract but also transversally from the mucosa to the lumen[1]. In the colon, the epithelium is covered with a bi-layer of mucus, with the outer mucus layer providing a natural habitat for the commensal bacteria whereas the stratified inner mucus layer restricts bacterial access to the epithelium[2]. Mucin proteins that form the mucus layer are highly glycosylated with a diverse and complex array of *O*-glycan structures containing *N*-acetylgalactosamine, galactose and *N*-acetylglucosamine (GlcNAc), and usually terminated by fucose, sialic acid (Neu5Ac) residues, and sulfate[3,4]. The terminal mucin glycans have been proposed to serve as metabolic substrates, providing a nutritional advantage to bacteria that have adapted to the GI mucosal environment[5–7]. The proportion of these terminal glycan epitopes varies along the GI tract with a decreasing gradient of fucose and an increasing gradient of sialic acid from the ileum to the rectum in humans[8,9]. Therefore, sialic acid represents a much-coveted source of nutrients for the gut bacteria inhabiting the mucus niche in the large intestine.

In bacteria, the genes involved in sialic acid metabolism are usually found clustered together forming different *nan* gene clusters[10–12]. The canonical *nanATEK* cluster was first described in *Escherichia coli* encompassing genes encoding the enzymes *N*-acetylneuraminate lyase (NanA), epimerase (NanE), and kinase (NanK), necessary for the catabolism of sialic acid into *N*-acetylglucosamine-6-P (GlcNAc-6-P) following its transport through the major facilitator superfamily transporter NanT[13,14]. An alternative pathway for sialic acid metabolism has been discovered later in *Bacteroides fragilis,* relying on the action of an

MFS transporter (NanT), an aldolase (NanL), a novel ManNAc-6-P epimerase (also named NanE), encoded in the *nanLET* operon and a hexokinase (RokA), converting Neu5Ac into GlcNAc-6-P[15]. GlcNAc-6-P is then converted into fructose-6-P, which is a substrate in the glycolytic pathway by genes encoding NagA (GlcNAc-6-P deacetylase) and NagB (glucosamine-6-P deaminase)[16]. The majority of bacteria that harbour a *nan* cluster colonize mucus regions of the human body[10–12]. To gain access to this substrate, bacteria are dependent on sialic acid release and uptake. Several gut bacteria species, including strains of *Clostridia*, *Bacteroides*, *Bifidobacterium longum*, *Vibrio cholerae*, *Ruminococcus gnavus* or *Akkermansia muciniphila* express sialidases to release sialic acid from their terminal location in mucins[10].

Since sialic acid cleavage takes place outside of the cell, bacteria have evolved multiple mechanisms to capture this important nutrient from their environment[12,17]. Such transport mechanisms involve the aforementioned NanT MFS transporter used by *E. coli* and *B. fragilis*, which in *E. coli* has been demonstrated biochemically to be a H^+-coupled symporter[18] or secondary transporters from the sodium solute symport (SSS) family, present in *C. difficile* and *S.* typhimurium[19,20]. High-affinity transport of sialic acid is mediated by substrate-binding protein-dependent systems, including a tripartite ATP-independent periplasmic (TRAP) transporter, SiaPQM, and ATP-binding cassette (ABC) transporters[21–25]. The sialic acid ABC transporters are classified into 3 types, SAT, SAT2 and SAT3[12,17]. To date all these transporters have been shown to transport Neu5Ac, with some being able to also move the related sialic acid Neu5Gc and KDN[26,27].

R. gnavus is an early coloniser of the infant gut[28] but persists in adults where it belongs to the 57 species detected in more than 90% of human faecal samples[29]. *R. gnavus* belongs to the Firmicutes division, Clostridia class and XIVa cluster, Lachnospiraceae family[30] and is considered as a prevalent member of the ‘normal’ gut microbiota[29,31]. Further, *R. gnavus* shows a disproportionate representation in a number of diseases such as inflammatory bowel disease[32–40]. The ability of *R. gnavus* strains to utilise mucin glycans as a source nutrient is associated with the expression of an intramolecular *trans*-sialidase (IT-sialidase) that specifically cleaves off terminal α2–3-linked Neu5Ac from glycoproteins, releasing 2,7-anhydro-Neu5Ac instead of Neu5Ac[41–45]. In *R. gnavus* ATCC 29149 and ATCC 35913 strains, the IT-sialidase (*Rg*NanH) is part of a *nan* cluster, which is induced when the cells are grown in the presence of mucin and absent in non-mucin glycan-degrading strains such as *R. gnavus* E1[43,45]. We enzymatically synthesised 2,7-anhydro-sialic acid derivatives[46], that were used to confirm the ability of IT-sialidase expressing *R. gnavus* strains to grow on 2,7-anhydro-Neu5Ac as sole carbon source[43]. We proposed that the ability of *R. gnavus* strains to produce and metabolise 2,7-anhydro-Neu5Ac, provide them with a competitive nutritional advantage in mucus by scavenging sialic acid from mucins in a form that others do not have access to[43,44].

In order to test this hypothesis and gain insights into *R. gnavus* 2,7-anhydro-Neu5Ac metabolism pathway, we identified candidate genes of the *nan* cluster involved in 2,7-anhydro-Neu5Ac transport and metabolism and characterised the proteins. Using fluorescence spectroscopy, STD-NMR and ITC, we showed that the solute binding protein (SBP) from *R. gnavus* ABC transporter was specific for 2,7-anhydro-Neu5Ac. Further

biochemical analyses uncovered an oxidoreductase activity allowing the conversion of 2,7-anhydro-Neu5Ac into Neu5Ac and confirmed the specificity of the sialic acid aldolase for Neu5Ac. Finally, we showed that the *nan* cluster was essential to support anaerobic growth of the bacteria on sialoconjugates *in vitro* and for *in vivo* fitness using gnotobiotic mice colonised with *R. gnavus* wild-type or *nan* mutant. These data demonstrate a unique sialic acid metabolism pathway in bacteria, which provides *R. gnavus* with a competitive strategy to colonise the mucus niche.

Results

Identification of genes involved in 2,7-anhydro-Neu5Ac metabolism in *R. gnavus nan* cluster

We first analysed the transcriptional activity of the *nan* cluster by qRT-PCR in *R. gnavus* ATCC 29149 grown on 2,7-anhydro-Neu5Ac or α2–3-sialyllactose (3'SL) as the sole carbon source. Expression of all genes constituting the *nan* cluster was induced upon bacterial growth on 2,7-anhydro-Neu5Ac or 3'SL as compared to glucose whereas the expression of the two genes flanking the cluster (RUMGNA_02702, RUMGNA_02690) remained unchanged (Figure 1). The 3'SL and 2,7-anhydro-Neu5Ac induced the transcription of the *nan* genes between 10 and 80-fold. Both substrates induced similar changes, which is not unexpected as 2,7-anhydro-Neu5Ac is the sialic acid form produced by *R. gnavus* ATCC 29149 from 3'SL. These results indicate that the *R. gnavus nan* operon is dedicated to the metabolism of 2,7-anhydro-Neu5Ac from host sialoglycans.

A sequence similarity network (SSN) analysis was then conducted to identify the proteins encoded by the *nan* cluster, which are associated with the ability of the bacteria to metabolise 2,7-anhydro-Neu5Ac over Neu5Ac. As expected, the IT-sialidase from *R. gnavus* strains clustered together with proteins from *S. pneumoniae* strains whose genomes are known to encode IT-sialidases (in addition to other sialidases)[47,48] (Supplementary Figure 1a). Other co-occurring bacterial species include *Rumminococcus torques, Lactobacillus salivarius, Staphylococcus pseudintermedius, Streptococcus infantis* and *Streptococcus mitis*. Bacterial species clustering for *Rg*NanH, also shared clusters for proteins encoding RUMGNA_02698, the predicted solute binding protein (SBP) giving specificity to ABC transporters, RUMGNA_02692 (sialic acid aldolase), the first protein of the canonical Neu5Ac metabolism, and RUMGNA_02695, a putative oxidoreductase, suggesting that these proteins may be associated with 2,7-anhydro-Neu5Ac metabolism (Supplementary Figure 1 and Supplementary Table 1). In contrast, RUMGNA_02701 with homology to sialic acid esterase proteins and RUMGNA_02700 with homology to the YhcH protein family did not cluster with proteins from the same set of bacteria (Supplementary Figure 1 and Supplementary Table 1). The candidate genes were then heterologous expressed, and the recombinant proteins purified as described in Methods.

Specificity of *R. gnavus* sialic acid transporter for 2,7-anhydro-Neu5Ac

We first investigated the ligand specificity of the recombinant SBP (RUMGNA_02698), *Rg*SBP, by measuring changes in the intrinsic protein fluorescence upon addition of potential ligands. Addition of 10 μM or 20 μM 2,7-anhydro-Neu5Ac resulted in a significant

shift at 350 nm, causing an ~16% quench in the fluorescence (Figure 2a). In marked contrast, addition of Neu5Ac at 10 μM, 20 μM or 70 μM caused no change in the spectrum intensity, suggesting an absence of binding (Figure 2b). Titration of 0.5 μM *Rg*SBP with 2,7-anhydro-Neu5Ac resulted in a hyperbolic curve with a *K*d of 1.349 μM (+/- 0.046) (Figure 2c). To confirm the specificity of 2,7-anhydro-Neu5Ac over Neu5Ac we monitored sequential changes in fluorescence following additions of 10 μM of the two ligands. When Neu5Ac was added first, no change in fluorescence was observed and a quench was observed following addition of 2,7-anhydro-Neu5Ac (Figure 2d). Conversely, when 2,7-anhydro-Neu5Ac was added first the quench was observed and additions of 10 μM Neu5Ac caused no further change in the intensity (Figure 2d), indicating that Neu5Ac is unable to displace 2,7-anhydro-Neu5Ac, and further supporting the specificity of the interaction between *Rg*SBP and 2,7-anhydro-Neu5Ac.

The affinity of the interaction between *Rg*SBP and sialic acid ligands was further assessed by isothermal titration calorimetry (ITC). *Rg*SBP bound to 2,7-anhydro-Neu5Ac with a K_d of 2.42 ± 0.27 μM (Figure 3a) and no binding was observed when Neu5Ac was used as the ligand (Figure 3b), in agreement with the findings from fluorescence spectroscopy. The binding of 2,7-anhydro-Neu5Ac revealed a thermodynamic signature with both entropic (-TΔS -7.05 ± 0.08 kcal mol^{-1}) and enthalpic (ΔH -0.93 ± 0.03 kcal mol^{-1}) components contributing favourably to the binding process (ΔG -7.99 ± 0.05 kcal mol^{-1} Figure 3a).

To gain structural insights into the unique ligand specificity of *Rg*SBP, saturation transfer difference nuclear magnetic resonance spectroscopy (STD NMR) studies were conducted with *Rg*SBP in the presence of 2,7-anhydro-Neu5Ac or Neu5Ac. The transfer of magnetization as saturation from the protein to the ligand was clearly observed for 2,7-anhydro-Neu5Ac but not for Neu5Ac, confirming that *Rg*SBP preferentially selects 2,7-anhydro-Neu5Ac (Supplementary Figure 2). STD NMR epitope binding revealed that protons H3, H4 and H6 showed the highest STD (%) factors, indicating the close contacts present at the interface of binding (Figure 3c). On the other hand, protons H7, H8, H9 and protons belonging to the CH_3 group showed lower STD (%) and are expected to be more exposed to solvent. For the DEEP-STD NMR experiment, TEMPOL was used to gain insights into *Rg*SBP binding pocket (Supplementary Figure 3). We found that protons H4, H6, H7, H8, H9' were preferentially oriented toward aromatic residues while H3 and protons belonging to the CH_3 group were oriented toward aliphatic residues (Figure 3d).

Together these data demonstrate that *Rg*SBP specifically binds to 2,7-anhydro-Neu5Ac but not to Neu5Ac, in line with the growth profile of *R. gnavus* ATCC 29149 on these substrates[43].

Specificity of *R. gnavus* sialic acid aldolase for Neu5Ac

The substrate specificity of recombinant sialic acid aldolase (RUMGNA_02692; *Rg*NanA), was determined using a coupled activity assay where pyruvate released during the conversion of Neu5Ac to ManNAc is converted to lactate by a lactate dehydrogenase and the subsequent decrease in absorbance at 340 nm measured as NADH is converted to NAD^+. *Rg*NanA and *Ec*NanA (*E. coli* Neu5Ac lyase/aldolase used as a control) showed activity against Neu5Ac whilst neither enzyme showed activity against 2,7-anhydro-Neu5Ac (Figure

4a). The product of the reaction with Neu5Ac was confirmed to be ManNAc by HPLC (Supplementary Figure 4). *Rg*NanA showed a k_{cat} of 2.757 ± 0.033 s^{-1} and a K_M of 1.473 ± 0.098 mM (Figure 4b). These kinetic parameters are consistent with values from other bacterial sialic acid aldolases characterised to date (Supplementary Table 2).

The *Rg*NanA crystal structure presents as a (β/α.8) TIM barrel with an adjacent three-helix bundle (for data collection and refinement statistics see Supplementary Table 3), a fold shared with characterised Neu5Ac aldolases from *Staphylococcus aureus, E. coli, Fusobacterium nucleatem, Pasteurella multocida, and Haemophilus influenzae*[49–54] (Supplementary Figure 5a). The active site residues in *Ec*NanA, Ser47, Tyr110, and Tyr137, identified to be catalytically important are conserved in *Rg*NanA[55] (Figure 4c and Supplementary Figure 5b), supporting Neu5Ac specificity. The crystal structure of the complex between an inactive mutant, *Rg*NanA K167A, and Neu5Ac showed Neu5Ac in the open-chain ketone form, with the *N*-acetyl group oriented out of the active site (Figure 4d). Neu5Ac forms extensive interactions with the enzyme active site (Supplementary Table 4). The Tyr139 α-carbon was shifted 1.8 Å in the mutant compared to wild-type. This movement is also present in the apo crystal structure, therefore presumably due to the absence of Lys167 rather than the presence of Neu5Ac (Supplementary Figure 5c).

Conversion of 2,7-anhydro-Neu5Ac to Neu5Ac by RUMGNA_02695

RUMGNA_02695 is a putative oxidoreductase with a predicted Rossman fold. Therefore, the activity of the recombinant protein was determined in the presence or absence of NAD^+, NADH or FAD as potential cofactors. Reaction products were analysed by HPLC following DMB labelling of the sialic acid[46]. When 2,7-anhydro-Neu5Ac (which cannot be labelled by DMB) was used as a substrate, Neu5Ac was produced in the presence of NAD^+ or NADH, but not in the presence of FAD or in the absence of a cofactor (Figure 5a). Mass spectrometry analyses showed a ratio of 1:2 for 2,7-anhydro-Neu5Ac:Neu5Ac (Supplementary Figure 6a), indicating that the reaction may be reversible. This was confirmed enzymatically by assaying RUMGNA_02695 against Neu5Ac in the presence of NAD^+ or NADH, producing a 1:2 for 2,7-anhydro-Neu5Ac:Neu5Ac. Additionally, these data indicate that Neu5Ac is the favourable product (Supplementary Figure 6b). No net change in NADH concentration was observed during the conversion reaction using 2,7-anhydro-Neu5Ac or Neu5Ac as substrate, suggesting that the enzyme mechanism may involve oxidation and reduction of NADH cofactor (Supplementary Figure 6c). The kinetic parameters of the enzymatic reaction were therefore determined using the coupled reaction described above in the presence of an excess of sialic acid aldolase and increasing concentrations of 2,7-anhydro-Neu5Ac substrate (Figure 5b). Using these conditions, the k_{cat} was calculated to be 0.0824 ± 0.0043 s^{-1} and the K_M 0.074 ± 0.014 mM. Taken together these data indicate that RUMGNA_02695 is an oxidoreductase required for the conversion of 2,7-anydro-Neu5Ac into Neu5Ac, which will then become a substrate for *Rg*NanA. We will refer to RUMGNA_02695 as *Rg*NanOx in the rest of the study.

Impact of *R. gnavus nan* cluster on *in vitro* growth and *in vivo* colonisation of mice

The ClosTron transformation method[56] was successfully applied to *R. gnavus* ATCC 29149, enabling the generation of *nan* deletion mutants with an erythromycin resistance gene

present in either the sense or antisense direction (relative to *Rg*NanH). The recombination event was confirmed by PCR (Supplementary Figure 7a) and the expression of the full cluster tested by qPCR (Supplementary Figure 7b). The expression of the genes flanking the cluster, RUMGNA_02690 and RUMGNA_02702, showed levels comparable to the wild-type strain, as also observed for the first three genes of the *nan* cluster, RUMGNA_02701-02699, however, the *nan* cluster genes RUMGNA_02698-02691 showed significantly reduced expression compared to the wild-type strain. *R. gnavus* ATCC 29149 wild-type strain was able to utilise both 3'SL and 2,7-anhydro-Neu5Ac as a sole carbon source, while no growth was detected using the *nan* deletion mutants on these substrates (Supplementary Figure 8).

To assess the impact of the *nan* cluster on the fitness of *R. gnavus in vivo,* germ-free C57BL/6J mice were gavaged with $1x10^8$ CFU *R. gnavus* ATCC 29149 or *R. gnavus* antisense *nan* deletion mutant or a mixture of wild-type and *nan* mutant strains at $1x10^8$ CFU each (Figure 6). During mono-colonisation experiments, both strains were detectable in the faecal content at day 3, 7 and 14 post-gavage at mean levels of between $1x10^6$ and $1x10^7$ bacteria per mg of material (Figure 6a). Both strains were also detected in the caecal content of mono-colonised mice sacrificed at day 14. The absence of the *nan* cluster did not affect the mouse expression response, as shown by RNA seq (Supplementary Figure 9). In competition experiments, the wild-type strain reached mean colonisation levels comparable to the levels obtained during mono-colonisation, whereas the mutant strain was severely outcompeted, reaching only $2x10^4$ copies per mg at day 3, before decreasing further at day 7 and day 14 below the level of detection in the faecal and caecal contents (Figure 6b). The impact of the *nan* deletion on the location of *R. gnavus* within the mucus layer was determined in mono-colonised mice by measuring the distance of the *nan* mutant or wild-type *R. gnavus* strains to the epithelial layer throughout the colon by fluorescent *in situ* hybridization (FISH) staining using confocal microscopy. The data showed that the *nan* mutant resided 19.70 μm from the epithelial layer, 5.06 μm further away than the wild-type strain, 14.64 μm (Figure 6c&d).

Discussion

Sialic acid comprises a family of 9-carbon acidic sugar found predominantly on the cell-surface glycans of humans and other animals. 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[6,10]. *In vivo*, sialic acids may be modified by *O*-acetylation, *O*-methylation [9,10,57–59]. Sialic acid metabolism is vital to the ability of *R. gnavus* strains to utilise mucin as a nutrient source[43–45]. *R. gnavus* ATCC 29149 strain encodes an extended *nan* operon dedicated to the metabolism of 2,7-anhydro-Neu5Ac from host sialoglycans.

Before being metabolised, a functional sialic acid transporter is essential for the uptake of sialic acid derivatives into the bacterial cell. The *R. gnavus* ATCC 29149 *nan* cluster contains a single ABC transporter, orthologous to the uncharacterised *Steptococcus pneumoniae* SAT2 system (Sp_1690-2), including two permeases (RUMGNA_02696 and 02697) and *Rg*SBP (RUMNGA_02698). *R. gnavus* SAT2 transporter is expected to be

coupled with an MsiK-like ATPase encoded elsewhere in the genome, with RUMGNA_03040 sharing 59% identity with the *S. pneumoniae* MsiK. Interestingly, in contrast to *S. pneumoniae, R. gnavus* does not encode SAT or SAT3 transporters which are known to recognise Neu5Ac with SAT3 being required for growth on Neu5Ac [60–62].

By studying *Rg*SBP subunit, we have discovered that SAT2 is a specific transporter for 2,7-anhydro-Neu5Ac with a K_d of 2.42 ± 0.27 μM, which does not bind Neu5Ac. Using STD NMR and DEEP-STD NMR, we characterized the orientation of the ligand in the binding site and the contribution of aromatic and aliphatic residues in *Rg*SBP 2,7-anhydro-Neu5Ac binding pocket. The lower affinity as compared to bacterial SAT (SatA) transporters specific for Neu5Ac characterised to date, which bind in the nM range[63], might be consistent with the 'exclusive' access of the bacteria to the 2,7-anhydro-Neu5Ac substrate. Taken together these findings indicate that the ability of *R. gnavus* strains to grow on 2,7-anhydro-Neu5Ac (and not on Neu5Ac) can be explained by the exquisite specificity of *Rg*SBP (RUMGNA_02698) *Rg*SBP is also orthologous (72% identity/86% similarity) with the SBP from the putative sialic transporters in *Streptococcus sanguinis* SK36 (SSA_0076) and *Streptococcus gordonii* str. Challis substr. CH1 (SGO_0122)[12]. It would therefore be of interest to determine the specificity of *Streptococcus* SBPs towards 2,7-anhydro-Neu5Ac.

Once inside the cell, 2,7-anhydro-Neu5Ac needs to be converted back into Neu5Ac to become a substrate for the sialic acid aldolase. *Rg*NanOx (RUMGNA_02695) was identified as the oxidoreductase catalyzing the conversion of 2,7-anhydro-Neu5Ac into Neu5Ac, following a mechanism of action which remains to be determined. Bioinformatic analysis identified close homologous of this protein in a range of bacterial species, including YjhC from *E. coli* (Supplementary Figure 1, Supplementary Table 1). Neu5Ac is then converted into ManNAc and pyruvate *via* the action of *Rg*NanA (RUMGNA_02692), a Neu5Ac-specific aldolase with conserved structural features with NanA proteins from the *nan* canonical pathway.

MultiGeneBlast analysis revealed that predicted homologs of the *R. gnavus nan* cluster are shared by a limited number of species, including 37 homologous clusters in *S. pneumoniae* isolates, *S. suis* A7, *Blautia hansenii* DSM 20583, *Blautia* sp. YL58 and *Intestinimonas butyriciproducens* AF211 (Supplementary Figure 10 & 11 and Supplementary Table 5). The presence of this cluster in *S. pneumoniae* suggests that it can also transport this unusual sialic acid into the cell. A major difference between NanB/NanH IT-sialidase and NanC sialidase cluster types is the associated transporter class, a carbohydrate ABC transporter for NanB/NanH as opposed to a sodium:solute symporter in NanC clusters[47], which may indicate a difference in the form of sialic acid being transported. These bioinformatics analyses support the specialization of the *R. gnavus nan* cluster.

We confirmed the importance of this metabolic pathway (Supplementary Figure 12) by generating a *R. gnavus nan* deletion mutant that was tested *in vitro* and *in vivo* using germ-free mice. In *in vivo* competition experiments, the fitness of the mutant was impaired as compared to the wild-type strain with a reduced ability to colonise the mucus layer. The *nan* cluster is therefore important to maintain the spatial distribution of *R. gnavus* strains in the gut. The ability for *R. gnavus* strains harbouring a *nan* cluster to penetrate further down into

the mucus layer may contribute to protect the bacteria from the constant mucus turner-over. This mechanism may serve as a determinant underlying *R. gnavus* success as one of the most largely shared species among individuals[29,31].

Together these findings provide robust biochemical and *in vivo* evidence for the role of *R. gnavus nan* cluster in the adaptation of this important gut symbiont to the mucosal environment in the gut, providing defined molecular targets for biomarkers and therapeutic strategies.

Methods

Materials

All chemicals were obtained from Sigma (St Louis, USA) unless otherwise stated. D-glucose (Glc), N-acetylneuraminic acid (Neu5Ac), were purchased from Sigma-Aldrich (St Louis, MO). 3′-sialyllactose (3′SL) was purchased from Carbosynth Limited (Campton, UK). 2,7-anhydro-Neu5Ac was prepared as previously described[46,64].

Bacterial strains and media

R. gnavus ATCC 29149 was routinely grown in an anaerobic cabinet (Don Whitley, Shipley, UK) in BHI-YH as previously described[45]. Growth on single carbon sources utilized anaerobic basal YCFA medium[65] supplemented with 11.1 mM of specific mono- or oligosaccharides (2,7-anhydro-Neu5Ac, 3'Sialyllactose (3'SL) or glucose). The bacteria were grown to late exponential phase for RNA extraction, the culture was performed in 14 ml tubes. Growth was determined spectrophotometrically by monitoring changes in optical density at 600 nm compared to the same medium without bacterium (Δ $OD_{595\ nm}$) hourly for 10 hours.

Quantitative real-time PCR (qRT-PCR)

Total RNA was extracted from 3 ml of mid- to late exponential phase cultures of *R. gnavus* ATCC 29149 in YCFA supplemented with one carbon source (Glc, 3'SL or 2,7-anhydro-Neu5Ac). Three biological replicates were performed for each carbon source. The RNA was stabilized prior to extraction by using RNAprotect Bacteria Reagent (Qiagen, Crawley, UK) according to the manufacturer's instructions. The RNA was then extracted after an enzymatic lysis followed by a mechanical disruption of the cells, using the RNeasy Mini Kit (Qiagen) according to manufacturer's instructions with an on-column DNAse treatment. The purity and quantity of the extracted RNA was assessed with NanoDrop 1000 UV-Vis Spectrophotometer (Thermo Fischer Scientific, Wilmington, DE) and with Qubit 2.0 (Invitrogen).

qPCR was carried out in an Applied Biosystems 7500 Real-Time PCR system (Life Technologies Ltd). One pair of primers was designed for each target gene using ProbeFinder version 2.45 (Roche Applied Science, Penzberg, Germany) to obtain an amplicon of around 60–80 bp long. The primers were between 18 and 23 nt-long, with a T_m of 59–60°C (Supplementary Table S6). Calibration curves were prepared in triplicate for each pair of primers using 2.5-fold serial dilutions of *R. gnavus* ATCC 29149 genomic DNA. The

standard curves showed a linear relationship of log input DNA vs. the threshold cycle (C_T), with acceptable values for the slopes and the regression coefficients (R^2). The dissociation curves were also performed to check the specificity of the amplicons. Each DNAse-treated RNA (1 μg) was converted into cDNA using QuantiTect® Reverse Transcription kit (Qiagen) according to the manufacturer's instructions. DNAse-treated RNA was also treated the same way but without addition of the reverse-transcriptase (RT−). Each qPCR reaction (10 μl) was then carried out in triplicate with 1 μl of 1 ng/μl (cDNA or RT−) and 0.2 μM of each primer, using the QuantiFast SYBR Green PCR kit (Qiagen) according to the manufacturer's instructions (except for the combined annealing/extension step which was extended to 35 s). Data obtained with cDNA were analyzed only when C_T values above 36 were obtained for the corresponding RT−. For each cDNA sample, the 3 C_T values obtained for each gene were analyzed using the $2^{-\Delta\Delta CT}$ method using housekeeping *gyrB* (RUMGNA_00867) gene as a reference gene and glucose as a reference condition. For each gene in each condition, the final value of the relative level of transcription (expressed as a fold change in gene transcription compared to glucose) is an average of 3 biological replicates, 1-way Annova was used for statistical analysis, using Graph Pad Prism (V 5.03).

Cloning, expression, mutagenesis and purification of recombinant proteins

R. gnavus ATCC 29149 genomic DNA (gDNA) was purified from the cell pellet of a bacterial overnight culture (1 ml) following centrifugation (5,000 g, 5 min) using the GeneJET Genomic DNA Purification Kit (ThermoFisher, UK), according to the manufacturer's instructions.

The full-length *Rg*SBP excluding the signal sequence (residues 1–29), the full length *Rg*NanA and full length RUMGNA_02695 were amplified from *R. gnavus* ATCC 29149 gDNA, and cloned into the pEHISTEV[66] expression system, introducing a His-tag at the N terminus using primers listed in Supplementary Table S6. DNA manipulation was carried out in *E. coli* DH5α cells. Sequences were verified by DNA sequencing by Eurofins MWG (Ebersberg, Germany) following plasmid preparation using the Monarch Plasmid Miniprep kit (New England Biolabs). The *Rg*NanA active site mutant, K167A, was generated using the QuikChange Lightning mutagenesis kit (Agilent) and primers listed in Supplementary Table S6. *E. coli* BL21 (New England BioLabs) cells were transformed with the recombinant plasmid harbouring the gene of interest according to manufacturer's instructions. Expression was carried out in 800 ml 'Terrific Broth Base with Trace Elements' autoinduction media (ForMedium, Dundee, UK) growing cells for 3 h at 37 °C and then at 16 °C for 48 h, with shaking at 250 rpm. The cells were harvested by centrifugation at 10,000 *g* for 20 min. The His-tagged proteins were purified by immobilized metal affinity chromatography (IMAC) and further purified by gel filtration (Superdex 75 column) on an Akta system (GE Health Care Life Sciences, Little Chalfont, UK). Protein purification was assessed by standard SDS–polyacrylamide gel electrophoresis using NuPAGE Novex 4–12% Bis-Tris gels (Life Technologies, Paisley, UK). Protein concentration was measured with NanoDrop 1000 UV-Vis Spectrophotometer (Thermo Fischer Scientific, Wilmington, DE) and using the extinction coefficient calculated by Protparam (ExPASy-Artimo, 2012) from the peptide sequence.

Fluorescence spectroscopy

All protein fluorescence experiments used a FluoroMax 3 fluorescence spectrometer with connecting water bath at 37°C. Because of the presence of 15 tyrosine residues, the protein was excited at 297 nm with slit widths of 5 nm. Under these conditions, the protein has a maximal emission at 331 nm. *Rg*SBP was used at a concentration of 0.2 μM in 50 mM Tris pH 7.5 for all fluorescence experiments. Cumulative fluorescence changes from titration of the protein with ligand were plotted in GraphPad and fitted to a single rectangular hyperbola. The K_d values reported were averaged from three separate ligand titration experiments.

Isothermal titration calorimetry (ITC)

Isothermal titration calorimetry (ITC) experiments were performed using the PEAQ-ITC system (Malvern, Malvern, UK) with a cell volume of 200 μl. Prior to titration, protein samples were exhaustively dialysed into 50 mM Tris-HCl pH 7.5. The ligand was dissolved in the dialysis buffer. The cell protein concentration was 100 μM and the syringe ligand concentration was 2 mM. Controls with titrant (sugar) injected into the buffer only were subtracted from the data. The analysis was performed using the Malvern software, using a single-binding site model. Experiments were carried out in triplicate.

Sialic acid aldolase activity assays

Aldolase activity was measured by monitoring the decrease in absorbance at 340 nm (A_{340nm}) as NADH is converted to NAD by lactate dehydrogenase in a coupled reaction where pyruvate is released from sialic acid by the aldolase. Reactions were performed in a 100 μl volume with final concentrations of 150 μM NADH (Sigma, St Louis, USA), 0.5 U LDH (Sigma, St Louis, USA), 10 mM sialic acid (Neu5Ac or 2,7-anhydro-Neu5Ac) and 1.5 μg purified *Rg*NanA or *Ec*NanA (*E. coli* aldolase CAS: 9027-60-5, Carbosynth, UK) in 50 mM Na-phosphate buffer (pH 7.0). The reactions were performed at 37 °C and monitored using FLUOstar OPTIMA (BMG LABTECH). For kinetics experiments, the sialic acid concentration was varied at 20, 10, 5, 4, 2, 1, 0.4, 0.2, 0.1 mM and the initial rate of reaction determined for each concentration in triplicate before analysis was performed by fitting the data to a Michaelis-Menten using Graph Pad Prism (V 5.03).

To monitor the production of ManNAc during the aldolase-catalyzed reactions, 2-AB labelling was carried out on the products from the above reactions. Briefly, 50 ng GlcNAc was added to 10 μl of each sample as an internal reference, before drying using a Concentrator Plus (Eppendorf). 5 μl of labelling reagent was added and incubated at 65 °C for 3 h. The labelling reagent was prepared by dissolving 50 mg 2-aminobenzamide in a solution containing 300 μl acetic acid and 700 μl DMSO, before 60 mg sodium cyanoborohydride is added. Following addition of H_2O to reach 100 μl total volume, the sample was transferred to a HPLC vial and 10 μl loaded onto a HyperClone 3u ODS (C18) 120A 150x4.6 mm 3 μ column. Mobile phases of 0.25% n-butylamine, 0.5% phosphoric acid, 0.1% Tetrahydrofurane; 50% methanol; Acetonitrile and H_2O were used at a 0.7 ml/min flow rate.

Bioinformatics analyses

Sequence Similarity Networks (SSN)—The InterPro families for *Rg*NanH (Glycoside Hydrolase, family 34; IPR001860) and *Rg*NanA (N-acetylneuraminate lyase; IPR005264) were identified using the UniProt database, this family identifier was used to extract protein sequences using Enzyme Function Initiative (EFI) Enzyme Similarity tool[67]. For the other proteins, the families found in the InterPro database were too large to be analysed, so the sequence BLAST tool was used with a maximum of 2500 protein sequences extracted. From this sequence similarity networks were generated and viewed in Cytoscape version 3.6[68].

Cluster analysis—Homologous gene clusters were identified for the *R. gnavus* ATCC 29149 *nan* cluster[45] using MultiGeneBlast[69]. The BCT (Bacteria) GenBank subdivision was queried with the sequence spanning locus tags RUMGMA_RS11835 – RUMGNA_RS11885 (from scaffold AAYG02000020_1). The data was manually curated, excluding all clusters that do not contain a predicted sialidase or are homologous to the functionally characterized *S. pneumoniae* NanC cluster[47,70] and the clusters are summarized by organism and predicted gene content in Supplementary Table S5.

RUMGNA_02695 enzymatic activity assay

To assay RUMGNA_02695 activity against 2,7-anhydro-Neu5Ac, the purified recombinant protein was incubated in 100 µl reactions at 37 °C overnight with 1 mM 2,7-anhydro-Neu5Ac, 50 mM sodium phosphate buffer pH 7.0 and 500 µM NADH, NAD, FAD or no cofactor. The reactions were dried using a Concentrator Plus (Eppendorf) for 1 h. Samples were then resuspended in 50 µl of water and 50 µl of reaction buffer (1.74 mg of 1,2-Diamino-4,5-methylenedioxybenzene dihydrochloride (DMB, Carbosynth, UK), 324.6 µl MilliQ water, 88.6 µl glacial acetic acid, 58.2 µl of β-Mercaptoethanol and 79.3 µl of sodium hydrosulphite) and incubated for 2 h at 55 °C in the dark. The samples were then centrifuged for 1 min and filtered using a 0.45 µm filter into a glass HPLC vial and directly analysed by HPLC.

DMB-labelled samples were analysed by injecting 10 µl onto a Luna 5 µm C-18(2) LC column 250x4.6 mm (Phenomenex) at 1 ml/min. Mobile phases methanol/acetonitrile/water were used for separation of fluorescently labelled sialic acids[46]. The settings of the fluorescence detector were 373 nm excitation and 448 nm emission. Samples were run alongside a Neu5Ac standard.

To determine the kinetic parameters of RUMGNA_02695 enzymatic reaction, a coupled reaction with lactate dehydrogenase and sialic acid aldolase was carried out as described above but with 15 µg of *Rg*NanA and 10 µg RUMGNA_02695 in each reaction. For the kinetics assays, 1, 0.4, 0.2, 0.1, 0.04, 0.02 and 0.01 mM 2,7-anhydro-Neu5Ac was used and the initial rate of reaction determined for each concentration in triplicate before analysis was performed by fitting the data to a Michaelis-Menten using Graph Pad Prism (V 5.03).

Electrospray ionisation spray mass spectrometry (ESI-MS) analysis was performed using the Applied Biosystems 4000 Q-TRAP. The full 100 µl reaction was diluted with 500 ul of 50%

Acetonitrile and 0.1 % formic acid and samples analysed in negative ion mode using direct injection.

ClosTron mutagenesis

R. gnavus mutants were generated using the ClosTron methodology[56], which inserts an erythromycin resistance cassette into the gene of interest. The target site (270a) was identified using the Perutka method[71]. The re-targeted introns were synthesised and ligated into the pMTL007C-E2 vector by ATUM (MenloPark, USA). The plasmids were then transformed into *E. coli* CA434 using the heat-shock 42°C for 45 seconds followed by 2 min on ice before the recombinant clones were selected for chloramphenicol resistance (25 μg/ml). Recombinant *E. coli* cells were grown overnight aerobically in 10 ml LB, 1 ml of the overnight culture was pelleted and washed with PBS. Continuing under anaerobic conditions the *E. coli* cell pellet was resuspended in 200 μl of an *R. gnavus* overnight culture and the cell suspension spotted onto a non-selective BHI-YH plate. Following incubation for 8 h at 37 °C the bacteria were washed from the plate using PBS and plated onto BHI-YH supplemented with cycloserine (250 μg/ml) and thiamphenicol (15 μg/ml) and grown for 72 h to select against *E. coli* and for transfer of the plasmid to *R. gnavus.* Individual colonies were grown in non-selective BHI-YH broth overnight to allow expression of the plasmid and genomic recombination. The culture was then plated onto a BHI-YH medium containing cycloserine (250 μg/ml) and erythromycin (10 μg/ml) to select clones with successful genomic recombination. PCR and sequencing were used to confirm recombination in the gene of interest.

Expression of the *nan* cluster genes in the generated mutants was assessed as described above using RNA samples from growth on YCFA supplemented with glucose.

The ability of the mutants to utilise sialic acids and sialoconjugates was assessed by supplementing YCFA with 11.1 mM of 2,7-anhydro-Neu5Ac, 3'SL, glucose or Neu5Ac in triplicate 200 μl cultures in 96-well microtiter plates. The $OD_{595\ nm}$ was measured hourly for 10 h in an infinite F50 plate reader (Tecan, UK) housed within an anaerobic cabinet connected to Magellan V7.0 software.

Saturation Transfer Difference (STD) NMR Spectroscopy

An amicon centrifuge filter unit with a 10 kDa MW cut-off was used to exchange the protein in 25 mM d_{19}-2,2-bis(hydroxymethyl)-2,2′,2″-nitrilotriethanol pH* 7.4 (uncorrected for the deuterium isotope effect on the pH glass electrode) D_2O buffer and 50 mM NaCl. 2,7-anhydro-Neu5Ac and Neu5Ac were dissolved in 25 mM d19-2,2-bis(hydroxymethyl)-2,2′,2″-nitrilotriethanol pH 7.4, 50 mM NaCl. Characterization of ligand binding by Saturation Transfer Difference NMR Spectroscopy[72] was performed on a Bruker Avance 800.23 MHz at 298 K. The on- and off-resonance spectra were acquired using a train of 50 ms Gaussian selective saturation pulses using a variable saturation time from 0.5 s to 4 s, for binding epitope mapping determination while only 0.5 s of saturation time for each selected frequency was used to perform the DEEP-STD NMR experiments[42]. The water signal was suppressed by using the excitation sculpting technique[73], while the remaining protein resonances were filtered using a T_2 filter of 40 ms. All the spectra were performed with a

spectral width of 10 KHz and 32768 data points using 256 or 512 scans. This time due to the absence of a 3D structure it was impossible to derive the resonances for saturation of aliphatic and aromatic residues found in the binding site as required by the DEEP-STD NMR technique. Moreover, *Rg*SBP being a high molecular weight protein the NMR spectra assignment is precluded. For this we adopted a search for druggable sites strategy using 4-hydroxy-1-oxyl-2,2,6,6-tetramethylpiperidine (TEMPOL) as previously described[74].

1H-1H TOCSY spectra of the protein (500 μM) were acquired in the presence and in the absence of TEMPOL (2.5 mM and 12.5 mM). The spectra were performed with a spectral width of 10 kHz using a time domain of 2056 data points in the direct dimension and 32 scans. The indirect dimension was acquired using the non-uniform sampling (NUS) technique acquiring a NUS amount of 50% of the original 256 increments resulting in 64 hypercomplex points. The spectra were processed with the Topspin 3.1 compressed sensing (cs) routine. The final selected resonances were those identified by the TEMPOL PRE effect, and not overlapping with ligand signals. The DEEP-STD NMR data obtained were used to derive the average orientation of the ligand bound to *Rg*SBP by averaging the DEEP-STD factors obtained from each saturated region. The DEEP-STD NMR and binding epitope mapping analysis were performed using previously published procedures[42,74,75].

Crystal structure determination

Sitting drop vapour diffusion crystallisation experiments of *Rg*NanA wt were set up at a concentration of 20 mg/ml and monitored using the VMXi beamline at Diamond Light Source[76]. The described *Rg*NanA wild-type crystal structure was acquired from a crystal grown in the Morpheus screen (Molecular Dimensions), 0.2 M 1,6-hexandiol, 0.2 M 1-butanol, 0.2 M 1,2-propanediol, 0.2 M 2-propanol, 0.2 M 1,4-butanediol, 0.2 M 1,3-propanediol, 0.1 M Hepes/MOPS pH 6.5, 20% ethylene glycol, 10% PEG 8000. The diffraction experiment was performed on beamline I24 beamline at Diamond Light Source Ltd at 100K using a wavelength of 0.9686 Å. The data were processed with Xia2 making use of aimless, dials, and pointless (for data collection and refinement statistics see Supplementary Table S3). The structure was phased using MrBump through CCP4 online and Molrep[77–79], by CdNaI from *C. difficile* (PDB 4woq) prepared using Chainsaw. Refinement was carried out using Refmac, Buster, and PDB redo[80–84]. Coot and ArpWarp were used for model building and Molprobity for structure validation[85]. It was not possible to crystallise *Rg*NanA wt in the presence of Neu5Ac as it caused protein precipitation and *Rg*NanA crystals dissolved in Neu5Ac soaking experiments, as also observed previously with *P. multocida* Neu5Ac aldolase[53], possibly due to conformational changes during substrate binding or catalysis. Experiments with *R*gNanA K167A mutant were set up at 25 mg/ml. Diffracting crystals grew in 0.1 M Tris/BICINE pH 8.5, 20% ethylene glycol, 100 mM $MgCl_2$, 10% PEG 8000 and diffraction experiments performed on beamline I04 at Diamond Light Source using a wavelength of 0.9795 Å. The crystal structure was phased with PHASER using the *Rg*NanA wild-type crystal structure[86]. A 60 second 5 mM Neu5Ac soak prior to freezing generated the *Rg*NanA K167A Neu5Ac complex. Due to data anisotropy, we processed the data in autoPROC[86,87] with the STARANISO option[88] and used these data for refinement with Buster using the previously obtained models of *Rg*NanA wild-type and K167A Neu5Ac complex.

In vivo colonisation and analyses

The impact of the *nan* deletion mutation on *R. gnavus* fitness was assessed by its ability to colonise germ-free C57BL/6J mice. Groups containing four 7-9 week old germ-free mice (two male, two female) were gavaged with $1x10^8$ CFU of *R. gnavus* ATCC 29149 wild-type or antisense *nan* mutant in 100 μl PBS, individually or in combination. Sample size was selected following the 3 R's principles of reduction, replacement, refinement, whilst ensuring data collected allowed for statistical analysis, randomization was not possible due to the constraints of germ-free isolators, scientists were blinded for the FISH analysis. Care and treatment of animals was in accordance guidelines from and approval by the University of East Anglia Disease Modelling Unit and all animal experiments were conducted in strict accordance with the Home Office Animals (Scientific Procedures) Act 1986. Faecal samples were collected from each mouse at 3,7 and 14 days post gavage, and caecal content taken at day 14. DNA was extracted from these samples using the MP Biomedicals Fast DNA™ SPIN kit for Soil DNA extraction with the following modifications. The samples were resuspended in 978 μl of sodium phosphate buffer before being incubated at 4 °C for one hour following addition of 122 μl MT Buffer. The samples were then transferred to the lysing tubes and homogenised in a FastPrep® Instrument (MP Biomedicals) 3 times for 40 s at a speed setting of 6.0 with 5 min on ice between each bead-beating step. The protocol was then followed as recommended by the supplier.

Colonisation was quantified using qPCR carried out in an Applied Biosystems 7500 Real-Time PCR system (Life Technologies Ltd). In competition experiments, primers based on the insertion in the *Rg*NanH gene were used to distinguish between wild-type and *nan* mutant. One pair of primers was designed to specifically target *R. gnavus* wild-type strain by spanning the area of insertion into the *nan* cluster and one pair of primers was designed to specifically amplify the inserted DNA, therefore targeting the *nan* mutant (Supplementary Table S6). The primers were between 18 and 23 nt-long, with a T_m of 59–60°C. Standard curves were prepared in triplicate for both primer pairs using a 10-fold serial dilution of DNA corresponding to $1x10^7$ copies of *Rg*NanH/2ul to $1x10^2$ copies/2ul diluted in 5 μg/ml Herring sperm DNA. The standard curves showed a linear relationship of log input DNA vs. the threshold cycle (C_T), with acceptable values for the slopes and the regression coefficients (R^2). The dissociation curves were also performed to check the specificity of the amplicons. Each qPCR reaction (10 μl) was then carried out in triplicate with 2 μl of 1 ng/μl DNA (diluted in 5 μg/ml Herring sperm DNA) and 0.2 μM of each primer, using the QuantiFast SYBR Green PCR kit (Qiagen) according to the manufacturer's instructions (except that the combined annealing/extension step was extended to 35 s instead of 30 s). Data obtained were analysed using the prepared standard curves.

RNAseq analysis

For RNAseq analysis, the colonic tissues from mono-colonised mice were gently washed and stored in RNAlater at -80ºC until extraction. RNA extraction was performed using the RNeasy mini kit (QIAGEN) following the manufacturer's instructions for purification of total RNA from animal tissues, including the on-column DNase digestion. Homogenisation was achieved with acid washed glass beads using the FastPrep®-24 (MP Biomedicals, Solon, USA) by 3 intermittent runs of 30 s at 6 m/s speed every 5 min, at room temperature.

Elution was performed as recommended with 50 µl RNAse-free water. The quality and concentration of the RNA samples was assessed using NanoDrop 2000 Spectrophotometer Nanodrop, the Qubit RNA HS assay on Qubit® 2.0 fluorometer (Life Technologies) and Agilent RNA 600 Nano kit on Agilent 2100 Bioanalyzer (Agilent Technologies, Stockport, UK).

RNAseq was carried out by Novogene (HK) (Hong Kong). Briefly, mRNA was enriched using oligo(dT) beads, fragmented randomly in fragmentation buffer, followed by cDNA synthesis using random hexamers and reverse transcriptase. After first-strand synthesis, a custom second-strand synthesis buffer (Illumina) was added with dNTPs, RNase H and *Escherichia coli* polymerase I to generate the second strand by nick-translation. The final cDNA library was obtained after a round of purification, terminal repair, A-tailing, ligation of sequencing adapters, size selection and PCR enrichment. Library concentration was first quantified using a Qubit® 2.0 fluorometer (Life Technologies), and then diluted to 1 ng/µl before checking insert size on an Agilent 2100 and quantifying to greater accuracy by qPCR (library activity >2 nM). Sequencing of the library was carried out on Illumina Hiseq platform and 125/150 bp paired-end reads were generated.

FASTQ files containing base calls and quality information for all reads that passed quality filtering were generated. Reads were mapped to the mouse reference genome using TopHat2[89]. The mismatch parameter was set to two, and other parameters were set to default. Appropriate parameters were also set, such as the longest intron length. Filtered reads were used to analyze the mapping status of RNA-seq data to the reference genome. The HTSeq software was used to analyze the gene expression levels, using the union mode[90]. In order for the gene expression levels estimated from different genes and experiments to be comparable, the FPKM (Fragments Per Kilobase of transcript sequence per Millions base pairs sequenced) was used to take into account the effects of both sequencing depth and gene length. The differential gene expression analysis was carried out using the DESeq package[90] and the readcounts from gene expression level analysis as input data. An adjusted p value (padj) cut-off of 0.05 was used to determine differential expressed transcripts.

Fluorescent in situ hybridization (FISH) staining

For FISH analysis, the colonic tissue was fixed in methacarn (60% dry methanol, 30% chloroform and 10% acetic acid), processed and embedded in paraffin as previously described[2]. Tissue sections were prepared at 8-10 µm. Paraffin sections were dewaxed and washed in 95% ethanol. The tissue sections were incubated with 100 µl of Alexa Fluor 555-conjugated Erec482 probe (5' – GCTTCTTAGTCARGTACCG -3') at a concentration of 10 ng/µl, in hybridisation buffer (20 mM Tris-HCl, pH 7.4, 0.9M NaCl, 0.1% SDS) at 50°C overnight. The sections were then incubated in a 50°C prewarmed wash buffer (20m M Tris-HCl, pH 7.4, 0.9 M NaCl) for 20 min. All subsequent steps were performed at 4°C. The sections were washed with PBS, the blocked with TNB buffer (0.5% w/v blocking reagent in 100 mM Tris-HCl, pH 7.5, 150 mM NaCl) supplemented with 5% goat serum. To detect mucin, the sections were then counterstained with a Muc2 antibody (sc-15334) at 1:100 dilution in TNB buffer overnight. The sections were washed in PBS, then goat anti-rabbit

antibodies (diluted 1:500) were used for immunodetection. The sections were counterstained with Sytox blue (S11348, ThermoFisher) diluted 1:1000 in PBS and mounted in Prolong gold anti-fade mounting medium. The slides were imaged using a Leica TCS SP2 confocal microscope with a x63 objective. The distance between the leading front of bacteria and the base of the mucus layer was measured with FIJI[91]. A total of 70 images from 8 mice were analysed, and scientists were blinded for the analysis due the subjectivity of determining the leading front of bacteria and base of the mucus layer. The association between genotype and distances was estimated by a linear mixed model, including fixed effects of genotype and area and random effects of mouse and each individual image. There was substantial spatial correlation between adjacent observations and so an AR(1) correlation structure was added. The resulting model had no residual autocorrelation as judged by visual inspection of autocorrelation function. The nmle package version 3.1-137 using R version 3.5.3 was used to estimate the model.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

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) and by the BBSRC responsive mode grant BB/P008895/1. AB was supported by the BBSRC Norwich Research Park Biosciences Doctoral Training Partnership grant number BB/M011216/1. We would like to acknowledge Martin Rejzek (JIC) for help with purification of 2,7-anhydro-Neu5Ac, Nigel Minton (Univ. Nottingham) for access to ClosTron technology and Haiyang Wu, Mark Philo, and George Savva at QIB for their help with the SSN, MS analysis and statistical analyses, respectively and Arlaine Brion and Andrew Goldson for their technical help with gnotobiotic mouse experiments. We would like to Diamond Light Source beamlines VMXi, I03, I04 and I24 for beamtime and assistance. Furthermore, we would like to thank Ronan Keegan of CCP4 and STFC for assistance with phasing the *Rg*NanA crystal structure.

References

1. Donaldson GP, Lee SM, Mazmanian SK. Gut biogeography of the bacterial microbiota. Nature Reviews Microbiology. 2016; 14:20–32. DOI: 10.1038/nrmicro3552 [PubMed: 26499895]
2. Johansson MEV, Larsson JMH, Hansson GC. The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108:4659–4665. DOI: 10.1073/pnas.1006451107 [PubMed: 20615996]
3. Etienne-Mesmin L, et al. Experimental models to study intestinal microbes-mucus interactions in health and disease. FEMS Microbiol Rev. 2019; doi: 10.1093/femsre/fuz013
4. Jensen PH, Kolarich D, Packer NH. Mucin-type O-glycosylation--putting the pieces together. FEBS J. 2010; 277:81–94. DOI: 10.1111/j.1742-4658.2009.07429.x [PubMed: 19919547]
5. Ndeh D, Gilbert HJ. Biochemistry of complex glycan depolymerisation by the human gut microbiota. FEMS Microbiol Rev. 2018; 42:146–164. DOI: 10.1093/femsre/fuy002 [PubMed: 29325042]
6. Tailford LE, Crost EH, Kavanaugh D, Juge N. Mucin glycan foraging in the human gut microbiome. Frontiers in genetics. 2015; 6:81–81. DOI: 10.3389/fgene.2015.00081 [PubMed: 25852737]
7. Martens EC, Chiang HC, Gordon JI. Mucosal Glycan Foraging Enhances Fitness and Transmission of a Saccharolytic Human Gut Bacterial Symbiont. Cell Host & Microbe. 2008; 4:447–457. DOI: 10.1016/j.chom.2008.09.007 [PubMed: 18996345]

8. Robbe C, Capon C, Coddeville B, Michalski JC. Structural diversity and specific distribution of 0-glycans in normal human mucins along the intestinal tract. Biochem J. 2004; 384:307–316. DOI: 10.1042/bj20040605 [PubMed: 15361072]
9. Robbe C, et al. Evidence of regio-specific glycosylation in human intestinal mucins - Presence of an acidic gradient along the intestinal tract. J Biol Chem. 2003; 278:46337–46348. DOI: 10.1074/jbc.M302529200 [PubMed: 12952970]
10. Juge N, Tailford L, Owen CD. Sialidases from gut bacteria: a mini-review. Biochem Soc Trans. 2016; 44:166–175. DOI: 10.1042/bst20150226 [PubMed: 26862202]
11. Lewis AL, Lewis WG. Host sialoglycans and bacterial sialidases: a mucosal perspective. Cell Microbiol. 2012; 14:1174–1182. DOI: 10.1111/j.1462-5822.2012.01807.x [PubMed: 22519819]
12. Almagro-Moreno S, Boyd EF. Insights into the evolution of sialic acid catabolism among bacteria. BMC Evol Biol. 2009; 9:118.doi: 10.1186/1471-2148-9-118 [PubMed: 19470179]
13. Plumbridge J, Vimr E. Convergent pathways for utilization of the amino sugars N-acetylglucosamine, N-acetylmannosamine, and N-acetylneuraminic acid by Escherichia coli. J Bacteriol. 1999; 181:47–54. [PubMed: 9864311]
14. Martinez J, Steenbergen S, Vimr E. Derived structure of the putative sialic acid transporter from Escherichia coli predicts a novel sugar permease domain. J Bacteriol. 1995; 177:6005–6010. [PubMed: 7592358]
15. Brigham C, et al. Sialic Acid (N-Acetyl Neuraminic Acid) Utilization by Bacteroides fragilis Requires a Novel N-Acetyl Mannosamine Epimerase. J Bacteriol. 2009; 191:3629–3638. DOI: 10.1128/jb.00811-08 [PubMed: 19304853]
16. Vimr ER, Kalivoda KA, Deszo EL, Steenbergen SM. Diversity of microbial sialic acid metabolism. Microbiol Mol Biol Rev. 2004; 68:132–153. DOI: 10.1128/mmbr.68.1.132-153.2004 [PubMed: 15007099]
17. Thomas GH. Sialic acid acquisition in bacteria - one substrate, many transporters. Biochem Soc Trans. 2016; 44:760–765. DOI: 10.1042/bst20160056 [PubMed: 27284039]
18. Mulligan C, et al. The substrate-binding protein imposes directionality on an electrochemical sodium gradient-driven TRAP transporter. Proceedings of the National Academy of Sciences. 2009; 106:1778.doi: 10.1073/pnas.0809979106
19. Wahlgren WY, et al. Substrate-bound outward-open structure of a Na(+)-coupled sialic acid symporter reveals a new Na(+) site. Nat Commun. 2018; 9:1753.doi: 10.1038/s41467-018-04045-7 [PubMed: 29717135]
20. Severi E, Hosie AH, Hawkhead JA, Thomas GH. Characterization of a novel sialic acid transporter of the sodium solute symporter (SSS) family and in vivo comparison with known bacterial sialic acid transporters. FEMS Microbiol Lett. 2010; 304:47–54. DOI: 10.1111/j.1574-6968.2009.01881.x [PubMed: 20100283]
21. Gangi Setty T, Cho C, Govindappa S, Apicella MA, Ramaswamy S. Bacterial periplasmic sialic acid-binding proteins exhibit a conserved binding site. Acta crystallographica Section D, Biological crystallography. 2014; 70:1801–1811. DOI: 10.1107/s139900471400830x [PubMed: 25004958]
22. Mulligan C, Leech AP, Kelly DJ, Thomas GH. 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. 2012; 287:3598–3608. DOI: 10.1074/jbc.M111.281030 [PubMed: 22167185]
23. Muller A, et al. Conservation of structure and mechanism in primary and secondary transporters exemplified by SiaP, a sialic acid binding virulence factor from Haemophilus influenzae. J Biol Chem. 2006; 281:22212–22222. DOI: 10.1074/jbc.M603463200 [PubMed: 16702222]
24. Severi E, et al. Sialic acid transport in Haemophilus influenzae is essential for lipopolysaccharide sialylation and serum resistance and is dependent on a novel tripartite ATP-independent periplasmic transporter. Mol Microbiol. 2005; 58:1173–1185. DOI: 10.1111/j.1365-2958.2005.04901.x [PubMed: 16262798]
25. Post DM, Mungur R, Gibson BW, Munson RS Jr. Identification of a novel sialic acid transporter in Haemophilus ducreyi. Infect Immun. 2005; 73:6727–6735. DOI: 10.1128/IAI.73.10.6727-6735.2005 [PubMed: 16177350]

26. North RA, et al. The Sodium Sialic Acid Symporter From Staphylococcus aureus Has Altered Substrate Specificity. Frontiers in Chemistry. 2018; 6:233. [PubMed: 30023356]
27. Hopkins AP, Hawkhead JA, Thomas GH. 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 Microbiol Lett. 2013; 347:14–22. DOI: 10.1111/1574-6968.12213 [PubMed: 23848303]
28. Sagheddu V, Patrone V, Miragoli F, Puglisi E, Morelli L. Infant Early Gut Colonization by Lachnospiraceae: High Frequency of Ruminococcus gnavus. Frontiers in pediatrics. 2016; 4:57.doi: 10.3389/fped.2016.00057 [PubMed: 27313996]
29. Qin J, et al. A human gut microbial gene catalogue established by metagenomic sequencing. Nature. 2010; 464:59–65. DOI: 10.1038/nature08821 [PubMed: 20203603]
30. Ludwig, W, Schleifer, K-H, Whitman, WB. Bergey's Manual® of Systematic Bacteriology: Volume Three The Firmicutes. De Vos, Paul; , et al., editors. Springer; New York: 2009. 1–13.
31. Kraal L, Abubucker S, Kota K, Fischbach MA, Mitreva M. The Prevalence of Species and Strains in the Human Microbiome: A Resource for Experimental Efforts. PLOS ONE. 2014; 9:e97279.doi: 10.1371/journal.pone.0097279 [PubMed: 24827833]
32. Olbjorn C, et al. Fecal microbiota profiles in treatment-naive pediatric inflammatory bowel disease - associations with disease phenotype, treatment, and outcome. Clinical and experimental gastroenterology. 2019; 12:37–49. DOI: 10.2147/ceg.S186235 [PubMed: 30774408]
33. Sokol H, et al. Specificities of the intestinal microbiota in patients with inflammatory bowel disease and Clostridium difficile infection. Gut Microbes. 2018; 9:55–60. DOI: 10.1080/19490976.2017.1361092 [PubMed: 28786749]
34. Nishino K, et al. Analysis of endoscopic brush samples identified mucosa-associated dysbiosis in inflammatory bowel disease. Journal of gastroenterology. 2018; 53:95–106. DOI: 10.1007/s00535-017-1384-4 [PubMed: 28852861]
35. Machiels K, et al. Specific members of the predominant gut microbiota predict pouchitis following colectomy and IPAA in UC. Gut. 2017; 66:79–88. DOI: 10.1136/gutjnl-2015-309398 [PubMed: 26423113]
36. Hall AB, et al. A novel Ruminococcus gnavus clade enriched in inflammatory bowel disease patients. Genome Medicine. 2017; 9:103.doi: 10.1186/s13073-017-0490-5 [PubMed: 29183332]
37. Fuentes S, et al. Microbial shifts and signatures of long-term remission in ulcerative colitis after faecal microbiota transplantation. The ISME journal. 2017; 11:1877–1889. DOI: 10.1038/ismej.2017.44 [PubMed: 28398347]
38. Joossens M, et al. Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. Gut. 2011; 60:631–637. DOI: 10.1136/gut.2010.223263 [PubMed: 21209126]
39. Willing BP, et al. A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. Gastroenterology. 2010; 139:1844–1854.e1841. DOI: 10.1053/j.gastro.2010.08.049 [PubMed: 20816835]
40. Png CW, et al. Mucolytic bacteria with increased prevalence in IBD mucosa augment in vitro utilization of mucin by other bacteria. The American journal of gastroenterology. 2010; 105:2420–2428. DOI: 10.1038/ajg.2010.281 [PubMed: 20648002]
41. Owen CD, et al. Unravelling the specificity and mechanism of sialic acid recognition by the gut symbiont Ruminococcus gnavus. Nature Communications. 2017; 8:2196.doi: 10.1038/s41467-017-02109-8
42. Monaco S, Tailford LE, Juge N, Angulo J. Differential Epitope Mapping by STD NMR Spectroscopy To Reveal the Nature of Protein-Ligand Contacts. Angewandte Chemie (International ed in English). 2017; 56:15289–15293. DOI: 10.1002/anie.201707682 [PubMed: 28977722]
43. Crost EH, et al. The mucin-degradation strategy of Ruminococcus gnavus: The importance of intramolecular trans-sialidases. Gut Microbes. 2016; 7:302–312. DOI: 10.1080/19490976.2016.1186334 [PubMed: 27223845]
44. Tailford LE, et al. Discovery of intramolecular trans-sialidases in human gut microbiota suggests novel mechanisms of mucosal adaptation. Nat Commun. 2015; 6:7624.doi: 10.1038/ncomms8624 [PubMed: 26154892]

45. Crost EH, et al. Utilisation of mucin glycans by the human gut symbiont Ruminococcus gnavus is strain-dependent. PLoS One. 2013; 8:e76341.doi: 10.1371/journal.pone.0076341 [PubMed: 24204617]
46. Monestier M, et al. Membrane-enclosed multienzyme (MEME) synthesis of 2,7-anhydro-sialic acid derivatives. Carbohydr Res. 2017; 451:110–117. DOI: 10.1016/j.carres.2017.08.008 [PubMed: 28851488]
47. Xu G, et al. Three Streptococcus pneumoniae sialidases: three different products. J Am Chem Soc. 2011; 133:1718–1721. DOI: 10.1021/ja110733q [PubMed: 21244006]
48. Xu G, et al. Crystal structure of the NanB sialidase from Streptococcus pneumoniae. J Mol Biol. 2008; 384:436–449. DOI: 10.1016/j.jmb.2008.09.032 [PubMed: 18835278]
49. Kumar JP, Rao H, Nayak V, Ramaswamy S. Crystal structures and kinetics of N-acetylneuraminate lyase from Fusobacterium nucleatum. Acta Crystallographica Section F. 2018; 74:725–732. DOI: 10.1107/S2053230X18012992
50. Campeotto I, et al. Pathological macromolecular crystallographic data affected by twinning, partial-disorder and exhibiting multiple lattices for testing of data processing and refinement tools. Scientific Reports. 2018; 8doi: 10.1038/s41598-018-32962-6
51. North RA, et al. Structure and inhibition of N-acetylneuraminate lyase from methicillin-resistant Staphylococcus aureus. FEBS Lett. 2016; 590:4414–4428. DOI: 10.1002/1873-3468.12462 [PubMed: 27943302]
52. Timms N, et al. Structural insights into the recovery of aldolase activity in N-acetylneuraminic acid lyase by replacement of the catalytically active lysine with gamma-thialysine by using a chemical mutagenesis strategy. ChemBioChem. 2013; 14:474–481. DOI: 10.1002/cbic.201200714 [PubMed: 23418011]
53. Huynh N, et al. Structural Basis for Substrate Specificity and Mechanism of N-Acetyl-D-neuraminic Acid Lyase from Pasteurella multocida. Biochemistry. 2013; 52:8570–8579. DOI: 10.1021/bi4011754 [PubMed: 24152047]
54. Barbosa JA, et al. Active site modulation in the N-acetylneuraminate lyase sub-family as revealed by the structure of the inhibitor-complexed Haemophilus influenzae enzyme. J Mol Biol. 2000; 303:405–421. DOI: 10.1006/jmbi.2000.4138 [PubMed: 11031117]
55. Daniels AD, et al. Reaction mechanism of N-acetylneuraminic acid lyase revealed by a combination of crystallography, QM/MM simulation, and mutagenesis. ACS Chem Biol. 2014; 9:1025–1032. DOI: 10.1021/cb500067z [PubMed: 24521460]
56. Heap JT, et al. The ClosTron: Mutagenesis in Clostridium refined and streamlined. J Microbiol Methods. 2010; 80:49–55. DOI: 10.1016/j.mimet.2009.10.018 [PubMed: 19891996]
57. Mandal, C, Schwartz-Albiez, R, Vlasak, R. Sialoglyco Chemistry and Biology I: Biosynthesis, Structural Diversity and SialoglycopathologiesTopics in Current Chemistry. GerardySchahn, R, Delannoy, P, VonItzstein, M, editors. Vol. 366. Springer Int Publishing Ag; 2015. 1–30.
58. Vimr ER. Unified theory of bacterial sialometabolism: how and why bacteria metabolize host sialic acids. ISRN Microbiol. 2013; 2013:816713.doi: 10.1155/2013/816713 [PubMed: 23724337]
59. Robbe-Masselot C, Maes E, Rousset M, Michalski JC, Capon C. Glycosylation of human fetal mucins: a similar repertoire of O-glycans along the intestinal tract. Glycoconjugate journal. 2009; 26:397–413. DOI: 10.1007/s10719-008-9186-9 [PubMed: 18807179]
60. Pezzicoli A, Ruggiero P, Amerighi F, Telford JL, Soriani M. Exogenous Sialic Acid Transport Contributes to Group B Streptococcus Infection of Mucosal Surfaces. J Infect Dis. 2012; 206:924–931. DOI: 10.1093/infdis/jis451 [PubMed: 22829646]
61. Bidossi A, et al. A Functional Genomics Approach to Establish the Complement of Carbohydrate Transporters in Streptococcus pneumoniae. Plos One. 2012; 7doi: 10.1371/journal.pone.0033320
62. Marion C, Burnaugh AM, Woodiga SA, King SJ. Sialic Acid Transport Contributes to Pneumococcal Colonization. Infect Immun. 2011; 79:1262–1269. DOI: 10.1128/iai.00832-10 [PubMed: 21189320]
63. Gangi Setty T, et al. Molecular characterization of the interaction of sialic acid with the periplasmic binding protein from Haemophilus ducreyi. J Biol Chem. 2018; 293:20073–20084. DOI: 10.1074/jbc.RA118.005151 [PubMed: 30315109]

64. Xiao A, et al. Streptococcus pneumoniae Sialidase SpNanB-Catalyzed One-Pot Multienzyme (OPME) Synthesis of 2,7-Anhydro-Sialic Acids as Selective Sialidase Inhibitors. J Org Chem. 2018; doi: 10.1021/acs.joc.8b01519

65. Duncan SH, Hold GL, Harmsen HJ, Stewart CS, Flint HJ. Growth requirements and fermentation products of Fusobacterium prausnitzii, and a proposal to reclassify it as Faecalibacterium prausnitzii gen. nov., comb. nov. Int J Syst Evol Microbiol. 2002; 52:2141–2146. DOI: 10.1099/00207713-52-6-2141 [PubMed: 12508881]

66. Liu H, Naismith JH. A simple and efficient expression and purification system using two newly constructed vectors. Protein Expr Purif. 2009; 63:102–111. DOI: 10.1016/j.pep.2008.09.008 [PubMed: 18845260]

67. Gerlt JA, et al. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. Biochim Biophys Acta. 2015; 1854:1019–1037. DOI: 10.1016/j.bbapap.2015.04.015 [PubMed: 25900361]

68. Shannon P, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13:2498–2504. DOI: 10.1101/gr.1239303 [PubMed: 14597658]

69. Medema MH, Takano E, Breitling R. Detecting Sequence Homology at the Gene Cluster Level with MultiGeneBlast. Mol Biol Evol. 2013; 30:1218–1223. DOI: 10.1093/molbev/mst025 [PubMed: 23412913]

70. Owen CD, et al. Streptococcus pneumoniae NanC: STRUCTURAL INSIGHTS INTO THE SPECIFICITY AND MECHANISM OF A SIALIDASE THAT PRODUCES A SIALIDASE INHIBITOR. J Biol Chem. 2015; 290:27736–27748. DOI: 10.1074/jbc.M115.673632 [PubMed: 26370075]

71. Perutka J, Wang W, Goerlitz D, Lambowitz AM. Use of computer-designed group II introns to disrupt Escherichia coli DExH/D-box protein and DNA helicase genes. J Mol Biol. 2004; 336:421–439. [PubMed: 14757055]

72. Mayer M, Meyer B. Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. Angewandte Chemie International Edition. 1999; 38:1784–1788. DOI: 10.1002/(SICI)1521-3773(19990614)38:12<1784::AID-ANIE1784>3.0.CO;2-Q [PubMed: 29711196]

73. Hwang TL, Shaka AJ. Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. Journal of Magnetic Resonance, Series A. 1995; 112:275–279. DOI: 10.1006/jmra.1995.1047

74. Nepravishta R, Walpole S, Tailford L, Juge N, Angulo J. Deriving Ligand Orientation in Weak Protein-Ligand Complexes by DEEP-STD NMR Spectroscopy in the Absence of Protein Chemical-Shift Assignment. ChemBioChem. 2019; 20:340–344. DOI: 10.1002/cbic.201800568 [PubMed: 30379391]

75. Mayer M, James TL. NMR-based characterization of phenothiazines as a RNA binding scaffold. J Am Chem Soc. 2004; 126:4453–4460. DOI: 10.1021/ja0398870 [PubMed: 15053636]

76. Sanchez-Weatherby J, et al. VMXi: a fully automated, fully remote, high-flux in situ macromolecular crystallography beamline. Journal of synchrotron radiation. 2019; 26:291–301. DOI: 10.1107/s1600577518015114 [PubMed: 30655497]

77. Krissinel E, Uski V, Lebedev A, Winn M, Ballard C. Distributed computing for macromolecular crystallography. Acta crystallographica Section D, Structural biology. 2018; 74:143–151. DOI: 10.1107/s2059798317014565 [PubMed: 29533240]

78. Vagin A, Teplyakov A. Molecular replacement with MOLREP. Acta crystallographica Section D, Biological crystallography. 2010; 66:22–25. DOI: 10.1107/s0907444909042589 [PubMed: 20057045]

79. Keegan RM, Winn MD. MrBUMP: an automated pipeline for molecular replacement. Acta crystallographica Section D, Biological crystallography. 2008; 64:119–124. DOI: 10.1107/s0907444907037195 [PubMed: 18094475]

80. van Beusekom B, Joosten K, Hekkelman ML, Joosten RP, Perrakis A. Homology-based loop modelling yields more complete crystallographic protein structures. bioRxiv. 2018; doi: 10.1101/329219

81. Emsley P. Tools for ligand validation in Coot. Acta crystallographica Section D, Structural biology. 2017; 73:203–210. DOI: 10.1107/s2059798317003382 [PubMed: 28291755]
82. Smart OS, et al. Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. Acta crystallographica Section D, Biological crystallography. 2012; 68:368–380. DOI: 10.1107/s0907444911056058 [PubMed: 22505257]
83. Langer G, Cohen SX, Lamzin VS, Perrakis A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nature protocols. 2008; 3:1171–1179. DOI: 10.1038/nprot.2008.91 [PubMed: 18600222]
84. Winn, MD, Murshudov, GN, Papiz, MZ. Methods Enzymol. Vol. 374. Academic Press; 2003. 300–321.
85. Williams CJ, et al. MolProbity: More and better reference data for improved all-atom structure validation. Protein Sci. 2018; 27:293–315. DOI: 10.1002/pro.3330 [PubMed: 29067766]
86. McCoy AJ, et al. Gyre and gimble: a maximum-likelihood replacement for Patterson correlation refinement. Acta crystallographica Section D, Structural biology. 2018; 74:279–289. DOI: 10.1107/s2059798318001353 [PubMed: 29652255]
87. Vonrhein C, et al. Data processing and analysis with the autoPROC toolbox. Acta crystallographica Section D, Biological crystallography. 2011; 67:293–302. DOI: 10.1107/s0907444911007773 [PubMed: 21460447]
88. Tickle, IJ, Flensburg, C, Keller, P, Paciorek, W, Sharff, A, Vonrhein, C, Bricogne, G. STARANISO. Cambridge, United kingdom: Global Phasing Ltd; 2018.
89. Kim D, et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology. 2013; 14:R36.doi: 10.1186/gb-2013-14-4-r36 [PubMed: 23618408]
90. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biology. 2010; 11:R106.doi: 10.1186/gb-2010-11-10-r106 [PubMed: 20979621]
91. Schindelin J, et al. Fiji: an open-source platform for biological-image analysis. Nat Methods. 2012; 9:676.doi: 10.1038/nmeth.2019 [PubMed: 22743772]

Europe PMC Funders Author Manuscripts

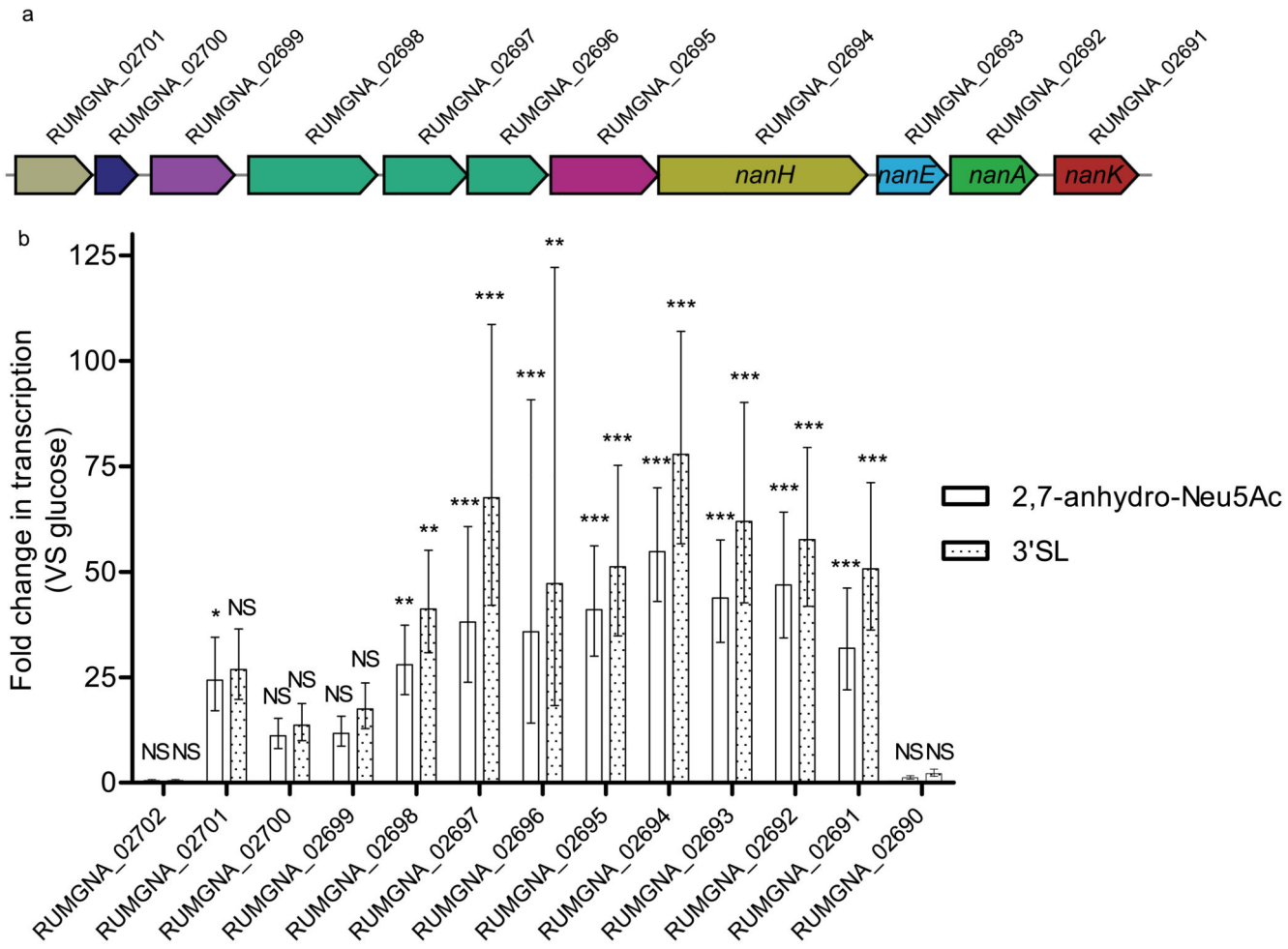

Figure 1. *R. gnavus* ATCC 29149 *nan* operon

a) Diagram depicting the genomic organisation of the *nan* operon RUMGNA_02701 (putative sialic acid esterase; tan) RUMGNA_02700 (putative YhcH family protein; dark blue), RUMGNA_02699 (predicted transcriptional regulator; purple), RUMGNA_02698 – 02696 (putative sialic acid ABC transporter of the SAT2 family; green), RUMGNA_02695 (putative oxidoreductase, pink), RUMGNA_02694 (RgNanH (Intramolecular *trans* sialidase), gold), RUMGNA_02693 (NanE (epimerase), blue), RUMGNA_02692 (NanA (aldolase), dark green), RUMGNA_02691 (kinase) (NanK, red). **b)** qPCR analysis showing fold changes in expression of *nan* genes when *R. gnavus* was grown with 3'SL or 2,7-anhydro-Neu5Ac compared to glucose using ΔΔCt calculation. Error represent standard deviation and are based on three biological replicates analysed in triplicate. Statistical significance was determined using a 1-way ANOVA with a Dunnett's multiple comparison test. NS – no significant change in expression ($p > 0.05$), * - p 0.05 – 0.01, ** - p 0.01 – 0.001, *** - $p < 0.001$.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

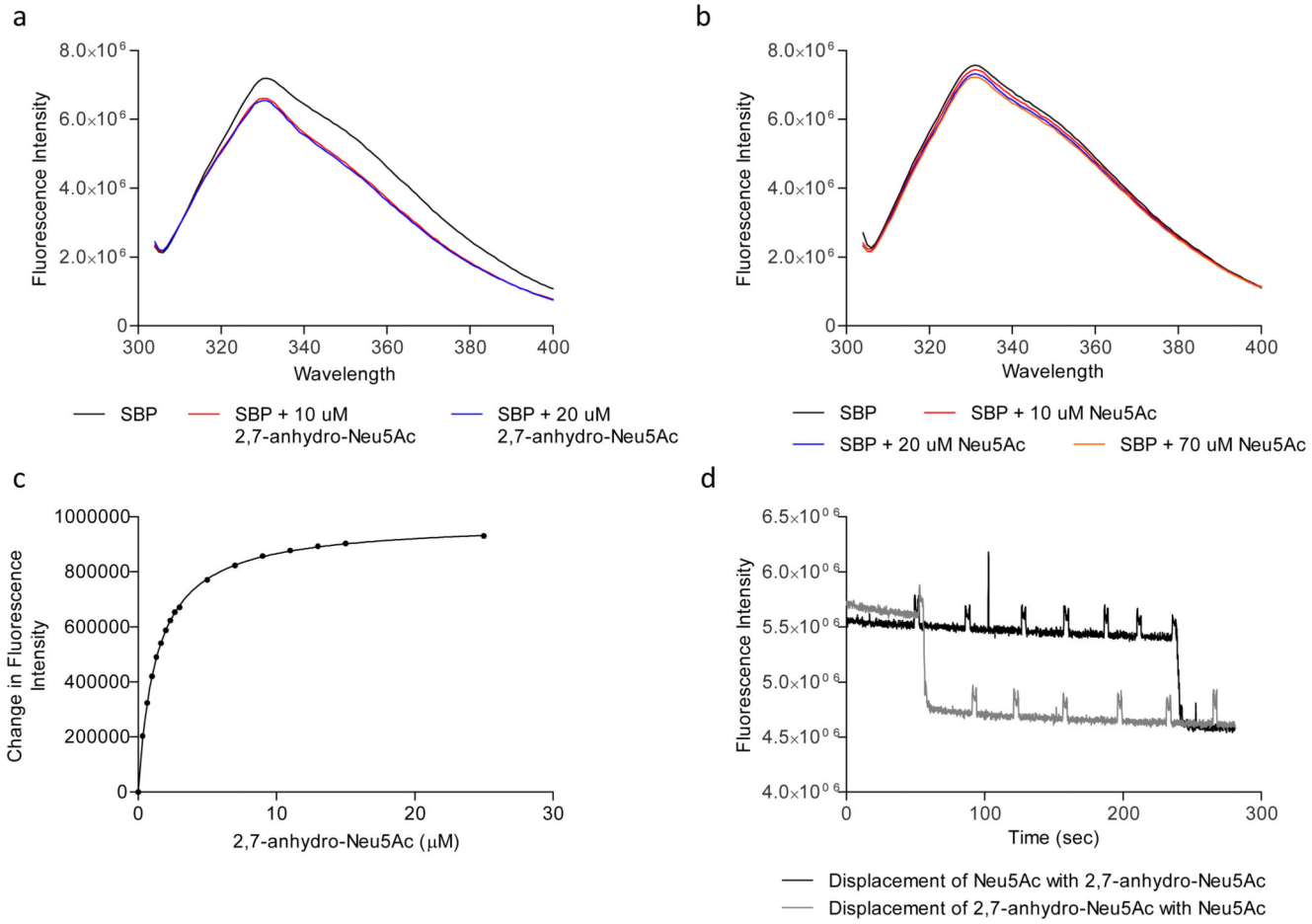

Figure 2. Steady-state fluorescence analysis of ligand binding to *Rg*SBP.

Fluorescence emission spectrum of 0.5 μM *Rg*SBP excited at 297 nm in the presence or absence of **a)** 2,7-anhydro-Neu5Ac or **b)** Neu5Ac. The data shown are representative of triplicate readings. **c)** Titration of 0.5 μM *Rg*SBP with 2,7-anhydro-Neu5Ac. The data represents the mean of triplicate readings. **d)** Displacement of Neu5Ac with 2,7-anhydro-Neu5Ac, six sequential additions of 10 μM Neu5Ac to 0.5 μM *Rg*SBP followed by one addition of 10 μM 2,7-anhydro-Neu5Ac, and displacement of 2,7-anhydro-Neu5Ac with Neu5Ac, one addition of 10 μM 2,7-anhydro-Neu5Ac followed by 6 subsequent additions of 10 μM Neu5Ac. The data shown are representative of triplicate experiments, the signal peaks are artefacts attributed to external light during sample addition.

Europe PMC Funders Author Manuscripts

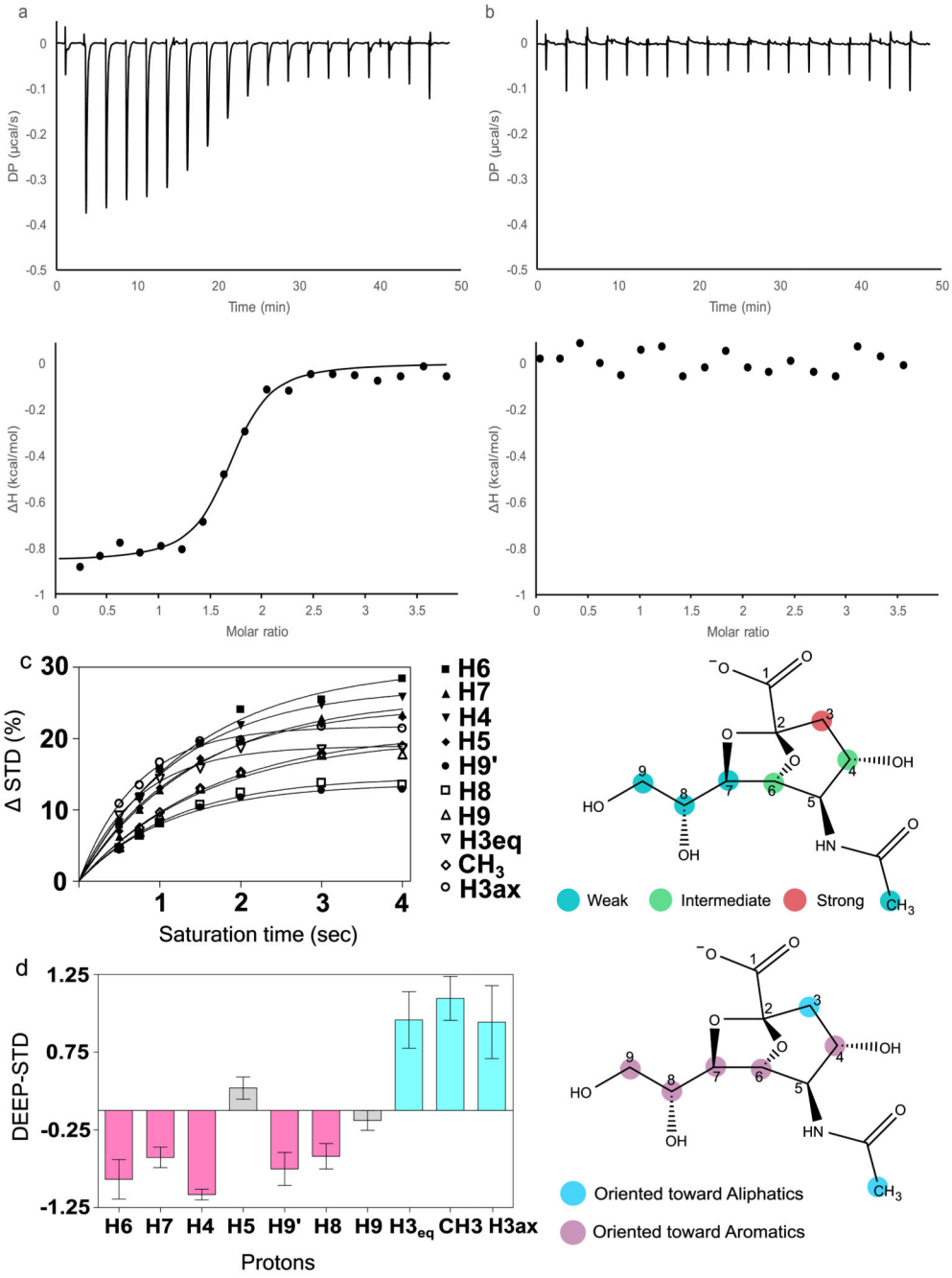

Figure 3. Biophysical analysis of ligand binding to *Rg*SBP.

ITC Isotherms of *Rg*SBP binding to **a**) 2,7-anhydro-Neu5Ac or **b**) Neu5Ac, showing both DP – differential power and ΔH – enthalpy change. The data shown are representative of triplicate experiments. **c)** Saturation Transfer Difference (STD) NMR binding epitope mapping of 2,7-anhydro-Neu5Ac interacting with *Rg*SBP. The initial slopes STD_0(%) were normalized against the highest STD_0, assigned as 100%. The obtained factors were then classified as weak (0-60 %), intermediate (60-80 %), and strong (80-100%) and used to identify the close contacts found at the interface of binding, data is representative of

triplicate readings **d)** Average Differential Epitope Mapping (DEEP) STD factors for 2,7-anhydro-Neu5Ac obtained saturating *Rg*SBP in spectral regions 0.6, 0.78, 1.44 ppm for aliphatic and 7.5, 7.23, 7.27 ppm for aromatic residues. Each differential mapping epitope obtained using different saturation frequencies are combined and the average DEEP STD is calculated resulting in five points for each frequency and a total of fifteen points for each proton receiving saturation. The data reported are the mean ± SEM of a sample of data of fifteen points for each proton receiving saturation.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

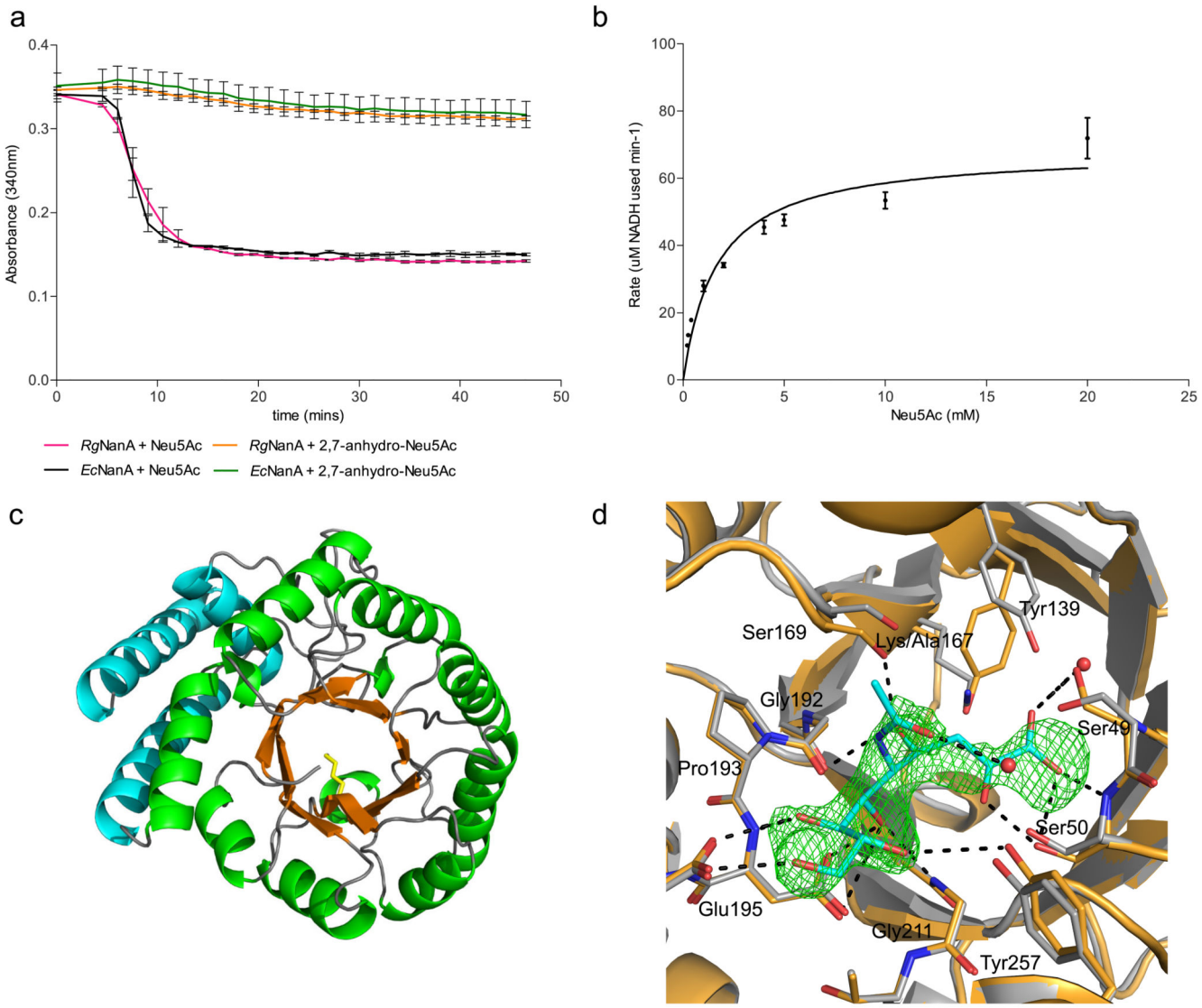

Figure 4. *R. gnavus* sialic acid aldolase enzymatic reaction.

a) Change of A_{340nm} over time using *R. gnavus* sialic acid aldolase (*Rg*NanA) with Neu5Ac (pink) or 2,7-anhydro-Neu5Ac (orange), or *E. coli* sialic acid aldolase (*Ec*NanA) with Neu5Ac (black) or 2,7-anhydro-Neu5Ac (green) reactions coupled to lactate dehydrogenase, error bars represent standard error from 3 independent experiments. **b)** Michaelis-Menten plot of *Rg*NanA rate of reaction with increasing concentration of Neu5Ac, error bars represent standard error. The rate of reaction at each concentration (μM NADH) was determined in triplicate by measuring A_{340nm} change using a standard curve. The mean value is plotted with standard error of the meaning shown with error bars **c)** Cartoon representation of the wild type *Rg*NanA crystal structure showing the (β/α.8) TIM barrel organisation and Lys167 as yellow sticks. **d)** The *Rg*NanA K167A active site is shown in orange with bound Neu5Ac in the open-chain ketone form shown in cyan. The green mesh represents the Neu5Ac F_o-F_c difference map at the 3σ level (for a stereo image of Neu5Ac and F_o-F_c difference map see Supplementary Figure 5d). Hydrogen bonding interactions are

depicted using black dashed lines. In addition, the unbound *Rg*NanA wt active site is shown in grey.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

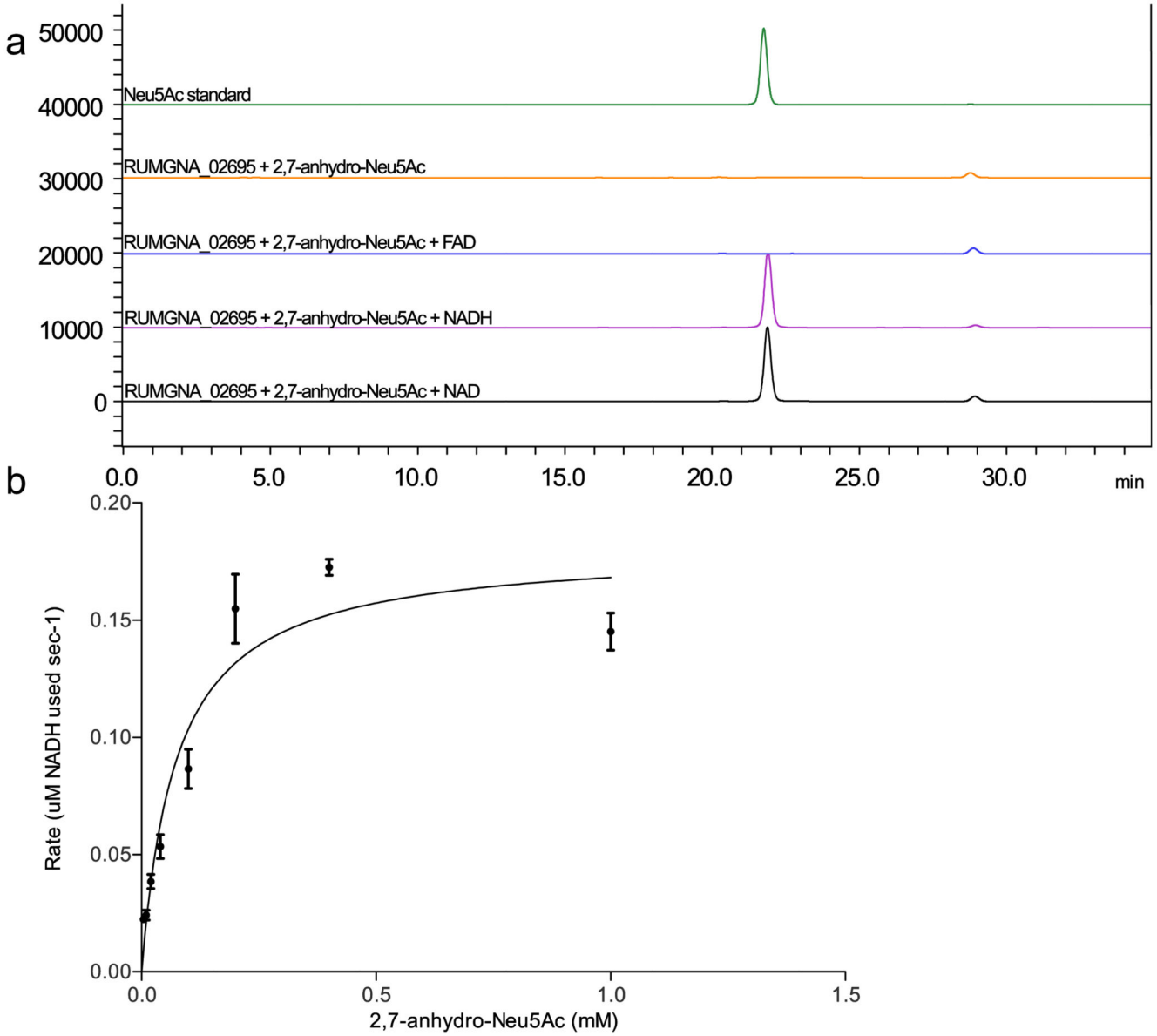

Figure 5. RUMGNA_02695 catalyses the conversion of 2,7-anhydro-Neu5Ac to Neu5Ac.

a) High Performance Liquid Chromatography (HPLC) expand acronyms number replicates with same outcome analysis of DMB labelled RUMGNA_02695 reactions with 2,7-anhydro-Neu5Ac using different co-factors. NAD (black), NADH (pink), FAD (blue), no co-factor (brown), and a Neu5Ac standard (green), data is representative of five independent experiments. **b**) Michaelis-Menten plot of the rate of reaction for RUMGNA_02695 with increasing concentration of 2,7-anhydro-Neu5Ac. The rate of reaction (μM NADH) at each concentration was determined in triplicate by measuring A_{340nm} change and using a standard curve, the mean value is plotted with standard error of the meaning shown with error bars

Europe PMC Funders Author Manuscripts

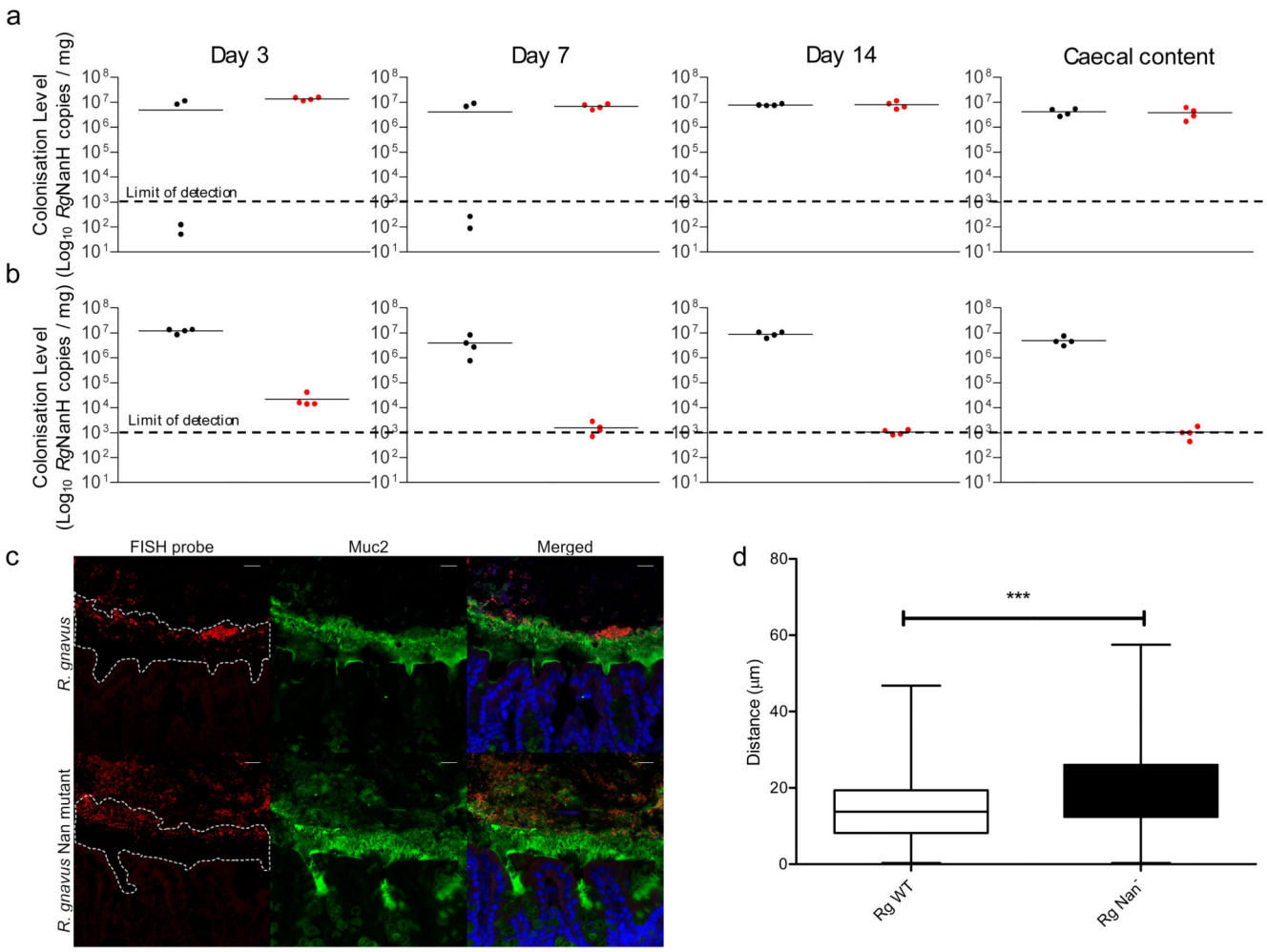

Figure 6. Colonisation of germ-free C57BL/6J mice with *R. gnavus* ATCC 29149 wild-type or *nan* mutant strains.

Mice were monocolonised with a and b sample size (n) and define centre measure mean (**a**) *R. gnavus* wild-type (black; n = 4) or *nan* mutant (red; n = 4) strains individually or (**b**) in competition (n = 4). Mice were orally gavaged with $1x10^8$ of each strain, faecal samples were analysed at 3,7 and 14 days after inoculation and caecal samples at 14 days after inoculation using qPCR, centre line denotes the mean. **(c)** Fluorescent in situ hybridisation (FISH) and immunostaining of the colon from *R. gnavus* monocolonised C57BL/6 mice. *R. gnavus* ATCC 29149 and *R. gnavus nan* mutant are shown in red. The mucus layer is shown in green and an outline of the mucus is shown in the first panels. Cell nuclei were counterstained with Sytox blue, shown in blue. Scale bar: 20 µm. Image is representative of 70 total images **(d)** Quantification of the distance between the leading front of bacteria and the base of the mucus layer. A total of 70 images of stained colon from 8 *R. gnavus* monocolonised mice were analysed. The asterisks (***) show the significance (P=0.0135, by linear mixed model analysis, including fixed effects of genotype and area and random effects of mouse and each individual image. There was substantial spatial correlation between adjacent observations and so an AR(1) correlation structure was added. The

resulting model had no residual autocorrelation as judged by visual inspection of autocorrelation function. The nmle package version 3.1-137 using R version 3.5.3 was used to estimate the model), centre point indicates the mean, box limits, upper and lower quartiles; whiskers, minimum and maximum.

Europe PMC Funders Author Manuscripts