
Article:
Scott, Hannah, Davies, Gideon J. orcid.org/0000-0002-7343-776X and Armstrong, Zachary (2022) The structure of Phocaeicola vulgatus sialic acid acetylesterase. Acta Crystallographica Section D: Structural Biology. pp. 647-657. ISSN 2059-7983

https://doi.org/10.1107/S2059798322003357

The structure of *Phocaeicola vulgatus* sialic acid acetylesterase

Hannah Scott[a], Gideon J. Davies[a], Zachary Armstrong[a,b],

[a] Department of Chemistry, University of York, Heslington, York, YO10 5DD, U.K.

[b] Department of Bioorganic Synthesis, Leiden University, Einsteinweg 55, 2333 CC Leiden, The Netherlands

Synopsis The sialic acid esterase from *Phocaeicola vulgatus* was produced heterologously in *E. coli*, purified and crystallized in two different crystal forms at 1.44 Å and 2.06 Å resolution.

Abstract Sialic acids terminate many N- and O-glycans and are widely distributed on cell surfaces. There are a diverse range of enzymes which interact with these sugars throughout the tree of life. They can act as receptors for influenza and specific β-coronaviruses in viral binding and their cleavage is important in virion release. Sialic acids are also exploited by both commensal and pathogenic bacteria for nutrient acquisition. A common modification of sialic acid is 9-O-acetylation, which can limit the action of sialidases. Some bacteria, including human endosymbionts, employ esterases to overcome this modification. However, few bacterial 9-O-acetyl sialic acid esterases (9-O-SAEs) have been structurally characterised. Here, the crystal structure of a 9-O-SAE from *Phocaeicola vulgatus* (PvSAE) is reported. The structure for PvSAE was determined from two different crystallization conditions to resolutions of 1.44 Å and 2.06 Å. Structural characterization revealed PvSAE as a dimer with a SGNH-fold – so named for the conserved sequence motif of this family – and a Ser-His-Asp catalytic triad. These structures also reveal flexibility in the most N-terminal α-helix, which provides a barrier to active site accessibility. We also show through biochemical assays that PvSAE deacetylates both mucin and the acetylated chromophore *para*-nitrophenyl acetate. This structural and biochemical characterization of PvSAE furthers our understanding of 9-O-SAEs and may aid in the discovery of small molecules targeting this class of enzyme.

Keywords: Acetylesterase; sialic acid esterase; SGNH-fold, Serine hydrolase

1. Introduction

Sialic acids are abundant monosaccharides that terminate glycolipids and glycoproteins on cell surfaces (Barnard *et al.*, 2020). As a result, they play key roles in a range of functions within humans, bacteria and viruses (Tortorici *et al.*, 2019). Sialic acids are also highly diverse, and frequently 9-O acetylated. In humans, 9-O-acetylation is regulated by CasD1 and sialate O-acetylesterase (SIAE), where CASD1 acetylates and SIAE deacetylates sialic acid (Baumann *et al.*, 2015, Orizio *et al.*, 2015), see Figure 1A. A key target molecule for SIAE is the 9-O-aceylated GD3 (Neu5,9Ac_2-α2,8-Neu5Ac- α2,3-Gal-β1,4-Glc-β1-ceramide), this ganglioside has an impact on a variety of biological functions. For example, elevated levels of 9-O-acetyl-sialic acid has been linked to apoptotic resistance by tumour cells, such as in medulloblastoma (a form of brain cancer) (Mather *et al.*, 2019).

Viruses, such as influenza C and specific β-coronaviruses (human OC43 and HKU1 coronavirus), use 9-O-acetylated sialylglycans as receptors. To promote viral release after infection these viruses use 9-O-SAEs as receptor destroying enzymes (RDEs). The RDE for influenza C is the hemagglutinin

esterase fusion protein (HEF) and for β-coronaviruses, hemagglutinin esterase (HE). These enzymes also bind to 9-O-acetyl sialic acid during viral entry. To date, HEFs and HEs from four β-coronaviruses have been structurally characterized and their 9-O-acetyl esterase domains have SGNH hydrolase-like folds (Rosenthal *et al.*, 1998, Hurdiss *et al.*, 2020, Zeng *et al.*, 2008, Langereis *et al.*, 2012).

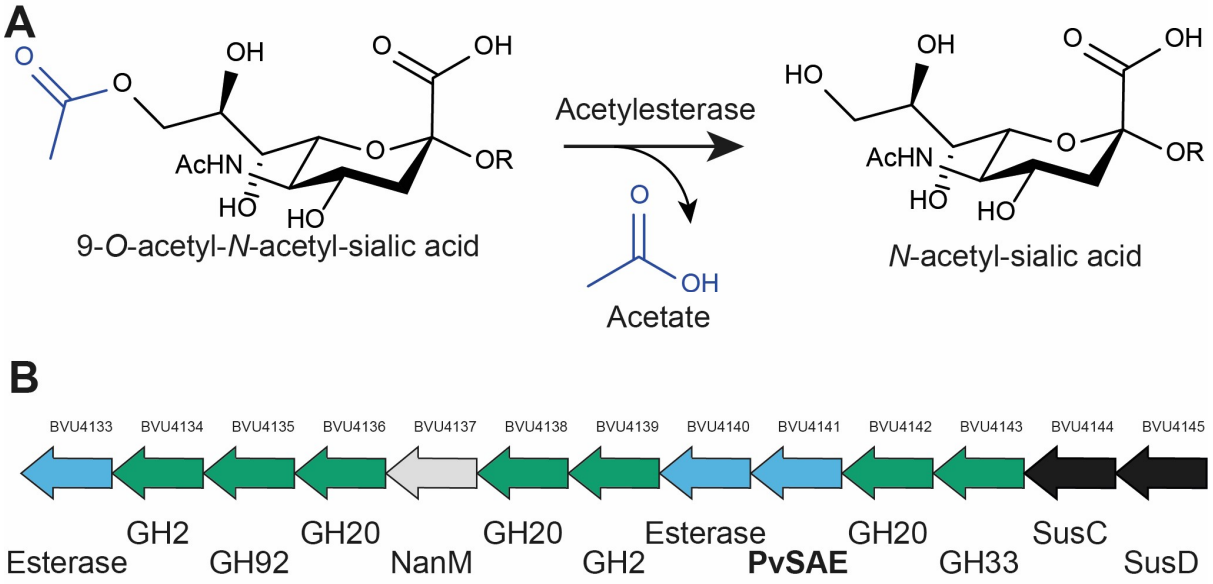

Figure 1 **A**) Sialic acid acetylesterases remove the acetyl groups from various sialylated glycoproteins and glycolipids. **B**) Genomic context of PvSAE from *P. vulgatus*. Locus tags are given above the arrows and predicted annotation is given below. Genes predicted to be esterases are highlighted in blue, while glycoside hydrolases (GHs) are highlighted in green, and SusC-like and SusD-like proteins are shown in black. Genes are not shown to scale.

In bacteria, esterases play an important role in glycan foraging and catabolism of sialic acid in the gut (Robinson *et al.*, 2017). There is a high level of sialic acid diacetylation in humans, such as in the intestine distal column and rectum (Robinson *et al.*, 2017). However, sialidases encoded by bacteria cannot access diacetylated sialic acid. Therefore, bacteria employ esterases to make acetylated sialic acid more accessible for sialidases to cleave. Carbohydrate active esterases have been classified into several different families, based on sequence homology (Drula *et al.*, 2022, Lenfant *et al.*, 2012) and can have domain architectures such as an α/β-hydrolase, $(\beta/\alpha)_7$ barrel, and SNGH folds. As with HE and HEF, the structurally characterized 9-O-acetyl sialic acid esterase NanS, encoded by enterohemorrhagic *Escherichia coli* O157:H7, also has a SGNH-hydrolase fold (Rangarajan *et al.*, 2011).

The SGNH-hydrolase superfamily was identified by Upton and Buckley through sequence comparisons of various motifs and classified from members of the GDSL family (Upton & Buckley, 1995). GDSL esterases are so named for the distinct GDSL motif near the N-terminus of the sequence that contains

the catalytic nucleophile (Upton & Buckley, 1995). SGNH-hydrolases have been defined by the structure of rhamnogalacturonan acetyl esterase and classified in regards to the presence of residues Ser-Gly-Asn-His (SGNH), within four blocks of conserved sequences (I, II, III, V) (Mølgaard *et al.*, 2000). The residues play important roles in the catalytic activity of the enzymes. The nucleophilic serine is found in block I of the sequences. The glycine and asparagine residues, within blocks II and III, respectively, stabilize the oxyanion in the oxyanion hole. The catalytic histidine within block V acts as a base by deprotonating the nucleophilic serine (Mølgaard *et al.*, 2000). Block V also contains an aspartic acid residue, forming the third member of the Ser-His-Asp catalytic triad.

Phocaeicola vulgatus (previously known as *Bacteroides vulgatus*) is a commensal gut bacterium that has a cooperative role in the gut microbiome (Huang *et al.*, 2015). Sialidases encoded by several *Bacteriodetes* species cleave sialic acids from the intestinal tissue, which in turn, has been shown to increase *E. coli* growth resulting in 'enterobacterial blooms' which are apparent during inflammation (Huang *et al.*, 2015). A close homolog of PvSAE from *Bacteroides fragilis* (estA; 81 % identity), has been shown to have a role in the complex network of interactions involving polysaccharides in the gut. Robinson et al. developed further understanding of the role of *B. fragilis* in this network and highlighted the community scavenging of sialic acid in the gut (Robinson *et al.*, 2017). *E. coli* growth significantly increased in the presence of *B. fragilis* esterase with either *B. fragilis* or *B. thetaiotaomicron* sialidase, demonstrating how the microorganisms cooperate. Furthermore, PvSAE is found in a genomic locus that contains multiple glycoside hydrolases, including a sialidase (see Figure 1B), that has synteny with loci from several Bacteroidetes including *P. massiliensis, B. fragilis,* and *B. thetaiotaomicron* (Briliūtė *et al.*, 2019). The syntenic loci from *B. thetaiotaomicron* has recently been characterized as part of the sialylated N-glycan degrading machinery of this host-associated bacteria (Briliūtė *et al.*, 2019).

The structure of 9-O-SAEs, including PvSAE, should provide insight into the function of sialic acid esterases and add to our understanding of host-glycan degradation. Furthermore, as this class of enzyme is also preserved in pathogens such as *Tanerella forsythia* (Albers *et al.*, 2021) investigating the structure of these enzymes should aid in the discovery of therapeutics for both bacterial and viral enzymes. Herein we show through SEC-MALLS that the 9-O-SAE from *P. vulgatus* (PvSAE) exists as a homo-dimer in solution. We also show through biochemical assays that PvSAE is a genuine SAE with activity on both chromogenic and natural mucins. The structure of PvSAE has previously been solved and deposited in the PDB (ID: 6NJC). Here we have solved the structure of the PvSAE in two new crystal forms to 1.44 Å and 2.06 Å resolution, which reveal previously unidentified conformational flexibility of this enzyme. We also show through docking studies this enzyme may bind sialic acids.

2. Materials and methods

2.1. Enzyme production and purification

The gene encoding PvSAE (GenBank accession ABR41743.1) was predicted to contain a signal peptide with a cleavage site between amino acids 19 and 20 by signal_P5.0 (Almagro Armenteros *et al.*, 2019). A codon optimized version of this gene with a His_6-tag (MGSSHHHHHHGTAENLYFQG) in place of the signal peptide was synthesized and cloned into a pET-28 vector with restriction sites NcoI and BamHI by GenScript (Leiden, Netherlands). Plasmid was transformed into chemically competent BL21(DE3) gold cells that were then plated on LB agar containing 50 µg mL^{-1} of kanamycin. A single colony was used to inoculate 1 L of LBE-5052 autoinduction media (Studier, 2005) containing 50 µg mL^{-1} of kanamycin that was incubated overnight with shaking at 300 rpm, 37 °C. Expression cultures were harvested by centrifugation (5,000 x *g*, 30 min, 4 °C). Pellets were resuspended in 25 mL of resuspension buffer (20 mM HEPES, 200 mM NaCl, pH adjusted to 7.4 with NaOH) and stored at -70 °C until use.

Frozen pellets were thawed and 1.8 mL of buffer B (20 mM HEPES, 500 mM imidazole, 200 mM NaCl, pH adjusted to 7.4 with HCl) was added with lysozyme (4 mg) and Dnase I (1 mg). Volume was increased to 50 mL with buffer A (20 mM HEPES, 20 mM imidazole, 200 mM NaCl, 1 mM DTT, pH adjusted to 7.4 with HCl). After warming to room temperature, the cell suspension was vortexed and sonicated on ice for 1 min 40 s (1 s sonication, 2 s rest), in a sonicator (Qsonica) set with 40 % amplitude. Lysed cells were centrifuged (18,000 x *g*, 30 min, 4 °C) and the supernatant was directly loaded onto a 5 mL His-trap Excel column (GE Healthcare). Bound protein was washed with 10 columns of buffer A then subsequently eluted using a 0 - 100 % linear gradient of buffer B over 20 column volumes. Fractions containing PvSAE were determined by SDS-PAGE and peak fractions pooled. Some precipitate was present, so pooled fractions were centrifuged and filtered (0.22 µm filter). Filtrate was concentrated using a VivaSpin 10 kDa concentrator (GE Healthcare) to a 2 mL final volume and further purified by gel filtration (HiLoad 16/600 Superdex 75 pg) in buffer C (20 mM HEPES, 200 mM NaCl, 1 mM DTT, pH adjusted to 7.4 with NaOH), at a flow rate of 1 mL min^{-1}. Fractions containing PvSAE were determined using SDS PAGE, peak fractions were pooled and concentrated using a VivaSpin 10 kDa concentrator (GE Healthcare). Concentrated protein was washed with buffer C and diluted to 50 mg mL^{-1}. Aliquots (40 µL) were flash frozen in liquid nitrogen and stored at -70 °C until use. Protein concentrations were determined spectrophotometrically using a calculated A_{280} extinction coefficient of 32,890 M^{-1} cm^{-1}. This purification yielded 45.6 mg of protein per litre of culture.

2.2. SEC-MALLS

Experiments were conducted on a system comprising a Wyatt HELEOS-II multi-angle light scattering detector and a Wyatt rEX refractive index detector linked to a Shimadzu HPLC system (SPD-20A UV

detector, LC20-AD isocratic pump system, DGU-20A3 degasser and SIL-20A autosampler). Work was conducted at room temperature (20 ± 2°C). Sample injection volume was at 100 µL at a protein concentration of 2 mg mL^{-1}. The sample was separated on a Superdex S200 Increase 10/300 GL column (G.E. Healthcare) using 20 mM HEPES, 200 mM NaCl, pH adjusted to 7.4 with NaOH (0.2 µm filtered) as buffer, a further 0.1 µm filter was present in the flow path. Shimadzu LabSolution software was used to control the HPLC and Astra 7 software for the HELEOS-II and rEX detectors. Data were analysed using the Astra 7 software. Molecular masses were estimated using the Zimm fit method with degree 1 (Zimm, 1945). A value of 0.182 was used for protein refractive index increment (dn/dc).

2.3. Crystallization

Initial crystallization conditions were identified using a range of commercial crystallization screens in 96-well format plates, including PEG/ION (Hampton Research), JCSG+ (Molecular Dimensions), AMSO4 (Qiagen), PACT (Molecular Dimensions), INDEX (Hampton Research) and Morpheus (Molecular dimensions). The condition of 0.2 M sodium malonate/malonic acid pH 5.0, 20 % w/v polyethylene glycol 3,350 with His_6-PvSAE at 30 mg mL^{-1} with 150 nL; 150 nL ratio, yielded thin rod-like crystals, this condition was further optimized. The optimized crystals were grown in maxi 48-well-plates using the sitting drop vapor-diffusion method at 20 °C with 0.2 M sodium malonate/malonic acid pH 4.5, 18 % w/v polyethylene glycol 3,350 with a protein:well solution ratio of 400:400 nL and His_6-PvSAE at 30 mg mL^{-1} in buffer C, this condition will be referred to as the Apo I condition. Crystals were soaked in well solution containing 0.2 M sodium malonate/malonic acid pH 4.5, 18 % w/v polyethylene glycol 3350, 15 % w/v ethylene glycol before flash cooling in liquid nitrogen.

Crystals were also observed in the Morpheus screen and were collected without cryoprotectant from the condition with 50 % precipitant mix 4 (25 % w/v PEG 1000, 25 % w/v PEG 3350, 25 % v/v MPD), 0.03 M sodium nitrate, 0.03 M sodium phosphate, 0.03 M ammonium sulphate, 0.1 M MOPS/HEPES-Na pH 7.5 (prepared according to Gorrec (Gorrec, 2009)), with His_6-PvSAE at 30 mg mL^{-1} in buffer C. This condition will be referred to as the Apo II condition. The crystals were harvested directly from the crystallization drop before flash cooling in liquid nitrogen.

Table 1 Macromolecule production information

Source organism	*Phocaeicola vulgatus*
DNA source	Synthetic
Cloning vector	pET28 with restriction sites: NcoI and BamHI
Expression vector	pET28
Expression host	*E. coli* BL21(DE3)

Complete amino acid sequence of the construct produced	MGSSHHHHHHGTAENLYFQGERKYSTFYEQRATLFEELPVTSKDI IFLGNSITNGCEWAELFQNKNVKNRGISGDICMGVYDRLDPIVKG KPAKIFLLIGINDVSRGTSADKIISEISMIVRKIKQESPKTKLYLQSV LPVNDCYGMFNGHTSRWQVVKQINDLLEPLAVKEGVAYIDLYSH FVEKETGKMNPVYTNDGLHLLGKGYLLWRDIVKPYVDQK
Molecular mass	25191.9 Daltons
Calculated pI	8.30
Calculated extinction coefficient	32890 M^{-1} cm^{-1}

Table 2 Crystallization

	Apo I	Apo II
Method	Sitting drop	Sitting drop
Plate type	48-well MRC Maxi	48-well MRC Maxi
Temperature (K)	293	293
Protein concentration (mg ml^{-1})	30	30
Buffer composition of protein solution	20 mM HEPES, 200 mM NaCl, 1 mM DTT adjusted to pH 7.5 with NaOH	20 mM HEPES pH 7.5, 200 mM NaCl, 1 mM DTT adjusted to pH 7.5 with NaOH
Composition of reservoir solution	18 % (w/v) PEG 3350, 200 mM sodium malonate/malonic acid pH 4.5	12.5 % w/v PEG 1000, 12.5 % w/v PEG 3350, 12.5 % v/v MPD, sodium nitrate, 0.03 M sodium phosphate, 0.03 M ammonium sulphate, 0.1 M MOPS/HEPES-Na pH 7.5
Volume and ratio of drop	0.8 µl (1:1)	1 µl (1:1)
Volume of reservoir (µl)	100	100

2.4. Data collection, processing, structure solution and refinement

All datasets were collected on the I03 beamline at the Diamond Light Source (DLS) and were integrated using the *DIALS* pipeline in *xia2*. All other calculations were carried out using the CCP4i2 suite (Winn *et al.*, 2011). Data reduction was performed using *AIMLESS (Evans & Murshudov, 2013)* and the data for Apo I and Apo I crystals were processed to resolutions 1.44 Å and 2.06 Å respectively. The structures were solved by expert mode molecular replacement with *Phaser* (McCoy *et al.*, 2007) using coordinates of BVU4141 from *Phocaeicola vulgatus* as a search model (PDB: 6njc). The models were corrected and completed manually with multiple rounds of model building using *Coot* (Emsley *et al.*, 2010) and refined using *REFMAC5* (Murshudov *et al.*, 2011). There was suspected twinning in the structure for

the Apo II condition, this was confirmed by an *L*-test score of 0.42. To take twinning into account, twin refinement was used for the Apo II structure in *REFMAC5* once R-free was below 40 %.

Water molecules for both models were added using *REFMAC5* (Murshudov *et al.*, 2011) and manually inspected after refinement using validation points in *Coot* (Emsley *et al.*, 2010). Final refinement statistics can be found in Table 3.

Table 3 Data collection and processing

Values for the outer shell are given in parentheses.

	Apo I (7PZG)	Apo II (7PZH)
Diffraction source	Beamline I03, DLS	Beamline I03, DLS
Wavelength (Å)	0.97625	0.97624
Temperature (K)	100	100
Detector	Eiger2 XE 16M	Eiger2 XE 16M
Crystal-detector distance (mm)	175	175
Rotation range per image (°)	0.1	0.1
Total rotation range (°)	220	220
Exposure time per image (s)	0.01	0.01
Space group	P 1 2_1 1	P 6_1
a, *b*, *c* (Å)	71.3, 90.2, 76.6	93.2, 93.2, 353.6
α, β, γ (°)	90.0, 99.3, 90.0	90.0, 90.0, 120.0
Resolution range (Å)	70.33-1.44 (1.46-1.44)	80.74-2.06 (2.10-2.06)
Total No. of reflections	1144420 (47744)	2186743 (108269)
No. of unique reflections	168248 (7900)	106903 (5267)
Completeness (%)	97.7 (92.8)	100 (100.0)
Redundancy	6.8 (6.0)	20.5 (20.6)
$CC_{1/2}$	1.0 (0.71)	0.99 (0.54)
$\langle I/\sigma(I)\rangle$	18.3 (1.6)	11.2 (1.8)
$R_{\text{r.i.m.}}$	0.018 (0.416)	0.039 (0.372)
Overall *B* factor from Wilson plot (Å^2)	17.91	31.11
L-test	0.49	0.42

Note: Apo I – Severe deviation from Wilson plot, 10.9 % of bins deviate. Possible ice rings. Apo II – Some deviation from the Wilson plot, 1.7 % of binds deviate. Data resolution cut-off criteria were: $\langle I/\sigma(I)\rangle > 1.5$ and $CC(1/2) > 0.5$.

Table 4 Structure solution and refinement

Values for the outer shell are given in parentheses.

	Apo I (7PZG)	Apo II (7PZH)
Resolution range (Å)	70.43-1.44 (1.46-1.44)	80.74-2.06 (2.10-2.06)
Completeness (%)	97.7 (92.8)	100 (100)
No. of reflections, working set	159915 (11313)	101414 (7559)
No. of reflections, test set	8262 (568)	5355 (369)
Final R_{cryst}	0.18 (0.27)	0.19 (0.26)

Final R_{free}	0.20 (0.27)	0.23 (0.30)
No. of non-H atoms		
Protein	6390	12810
Ion	4	8
Ligand	44	0
Water	698	260
Total	7136	13078
R.m.s. deviations		
Bonds (Å)	0.010	0.009
Angles (°)	1.8	1.6
Average *B* factors (Å^2)		
Protein	23	41
Ion	32	39
Ligand	37	-
Water	33	35
Ramachandran plot		
Most favoured (%)	97.5	95.8
Allowed (%)	2.3	3.9
Outliers (%)	0	0.1

2.5. Michaelis-Menten kinetics

Hydrolysis of chromogenic para-nitrophenyl acetate (pNP-Ac) was carried out in a 96 well plate at 25 ˚C. Reactions were carried out in a total volume of 100 µL. Each reaction contained PvSAE (10 nM), phosphate/citrate buffer (100 mM each) pH 7.0 and pNP-Ac in 1 % DMSO (1 mM – 50 µM). Reactions were initiated by the addition of 10 µL of 100 nM enzyme to a reaction containing 90 µL of water, buffer, and pNP-Ac. All kinetic assays were completed in triplicate.

Reactions were measured using a Clariostar microplate reader (BMG Labtech) (λ_{Ex} = 405 nm). Absorbance of released pNP was measured continuously over 15 minutes. The concentration of pNP released was calculated using a calibration curve of pNP standards (100 µM – 1 µM) present on the same plate. Rates of substrate hydrolysis were determined by a linear fit of initial data points. Plots of initial rate (v_o) *vs.* substrate concentration were fit with nonlinear regression using OriginPro 2021 (OriginLab) to the hyperbolic equation $v_o = k_{cat}[E][S]/(K_M+[S])$ to determine K_M and k_{cat}. Error values given for K_M and k_{cat} are fitting errors, determined in OriginPro when fitting all data. Error values given for the specificity constant (k_{cat}/K_M) were propagated from the error determined for K_M and k_{cat} using the formula: $\sigma_{\frac{k_{cat}}{K_M}} = \frac{k_{cat}}{K_M}\sqrt{(\frac{\sigma_{k_{cat}}}{k_{cat}})^2 + (\frac{\sigma_{K_M}}{K_M})^2}$.

2.6. Mucin activity assays

The hydrolysis of acetylated sialic acids present in mucin was assessed using bovine submaxillary mucin (Millipore, Lot:3716979). Purified PvSAE was added to a final concentration of 1 µM or 100 nM to a 100 µL reaction containing 6.25 mg/mL bovine submaxillary mucin, 20 mM HEPES, 100 mM

NaCl, pH 7.0 (adjusted with NaOH). The samples were incubated at 37 ˚C for 1 hour with shaking at 800 rpm. The concentration of acetate in each sample was quantified using a manual acetic acid assay kit (Megazyme, K-ACET) according to the manufacturer's instructions. HEPES buffer was used in place of phosphate buffer for this assay as phosphate interferes with coupled enzyme activity in the acetic acid assay kit. An acetate standard curve (0-800 μM Acetate) in the presence of 6.25 mg/mL bovine submaxillary mucin was used to calibrate the acetate concentration. All measurements were performed in at least quadruplicate.

2.7. Molecular Docking

Molecular docking of 9-*O*-acetyl sialic acid in the crystal structure of PvSAE was performed with AutoDock Vina (Trott & Olson, 2009). The ligand coordinates were made using AceDRG (Long *et al.*, 2017) and molecular charges were computed with AutoDock (Morris *et al.*, 2009). The ligand was docked into a grid box (30 x 20 x 20 Å) centred on the catalytic nucleophile. A search was performed on each of the active sites for the Apo I and Apo II structures. An exhaustiveness of 30 was used to compute 20 docking positions for each active site. The docked position with the best calculated energy, and the acetyl group positioned in the active site was used for further analysis.

3. Results and discussion

3.1. Purification and biochemical characterization

Phocaeicola vulgatus 9-*O*-acetyl sialic acid esterase (PvSAE) was produced in *E. coli* BL21(DE3) gold cells. The protein contained an N-terminal His_6-tag and was purified using immobilized nickel-affinity chromatography, followed by size-exclusion chromatography. The protein eluted earlier than expected for a monomer from a calibrated size-exclusion column. To determine if the protein was a multimer in solution, it was subjected to SEC-MALLS (size exclusion chromatography – multi-angle laser light scattering). A single peak was observed in the chromatogram, with a calculated molar mass of 51.6 kDa, indicating PvSAE forms a dimer in solution.

To gain further insight into the action of PvSAE, we investigated the activity of PvSAE on both a synthetic substrate and a natural mucin. We first tested the activity of PvSAE on the activated, chromogenic substrate para-nitrophenyl acetate (pNP-Ac). This substrate has previously been used to detect 9-O-SAE activity for *E. coli* NanS, *B. fragilis* EstA and NanS from *Tannerella forsythia* (Rangarajan *et al.*, 2011, Nakayama-Imaohji *et al.*, 2012, Albers *et al.*, 2021). pNP-Ac hydrolysis was monitored at pH 7.0 to limit unspecific acid- or base-catalyzed hydrolysis. The catalytic constants determined from a Michaelis-Menten plot of initial rates, see Figure 2A, revealed a high specificity rate constant (k_{cat}/K_M: 930 ± 70 mM^{-1} s^{-1}) for this non-natural substrate, with rates slightly higher than that determined for the bi-functional NanS from *T. forsythia* (k_{cat}/K_M: 434 mM^{-1} s^{-1})(Albers *et al.*, 2021).

To examine the activity of PvSAE on a more natural substrate we assessed its activity on bovine submaxillary mucin (BSM). BSM is purified directly from the submaxillary gland, allowing access to the natural substrate before it has been acted on by commensal microbes. As such it contains several percent of sialic acid by weight and is highly acetylated. We incubated either 100 nM or 1 µM of PvSAE with BSM for 1 hour at 37 ˚C. At both enzyme concentrations PvSAE showed significant release of acetate when compared to a buffer only control, Figure 2B. This indicates that PvSAE is indeed a genuine sialic acid esterase, capable of action on natural substrates.

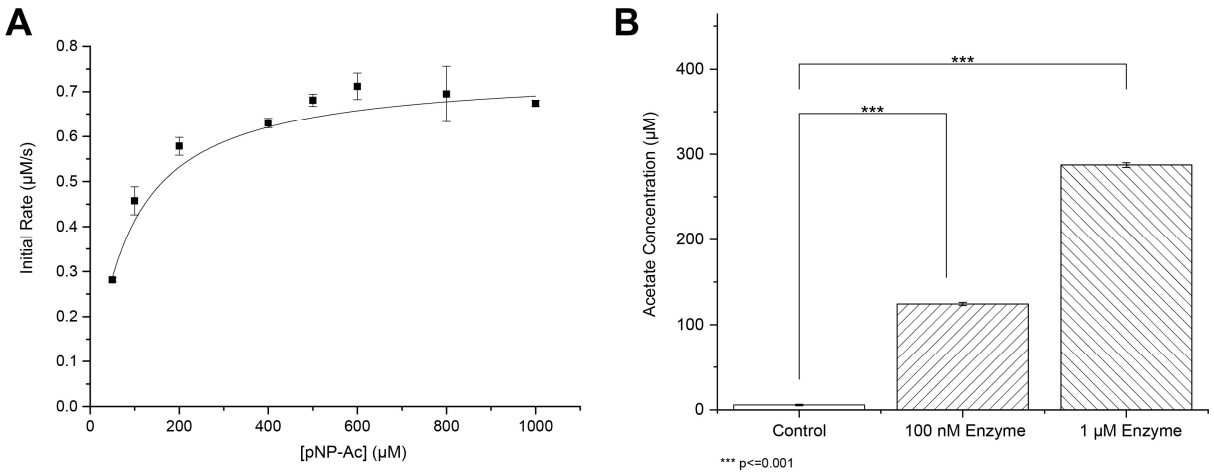

Figure 2 **A)** Michaelis-Menten kinetic plot for hydrolysis of pNP-Ac by PvSAE. Rates were established using a pNP standard curve. Experiments were performed in triplicate. **B)** Activity of PvSAE on bovine submaxillary mucin. Acetate release from 6.25 mg/ml bovine submaxillary mucin was quantified after 1 hour of incubation at 37 °C with either a buffer control, 100 nM or 1 µM of PvSAE.

Table 5 Kinetic parameters for hydrolysis of pNP-Ac catalysed by PvSAE at pH 7.0.

Parameter	Value
K_M	80 ± 6 µM
k_{cat}	75 ± 1 s^{-1}
k_{cat}/K_M	930 ± 70 mM^{-1} s^{-1}

3.2. Protein crystallization and structure solution

Purified PvSAE crystalized in two different forms (Apo I and Apo II), see Table 2 for crystallization conditions. A crystal from the Apo I condition diffracted to a resolution of 1.44 Å and had *P* 1 2_1 1 symmetry, while the crystal from the Apo II condition diffracted to 2.06 Å and had *P* 6_1 symmetry. Both structures were solved by molecular replacement using a structure deposited in the PDB (6NJC) of PvSAE at 1.9 Å resolution, that currently lacks a formal manuscript. The unit cells of these crystals

contained four and eight protomers, for the Apo I and Apo II forms respectively. These protomers were arranged as dimers in both crystal forms, recapitulating the dimeric form observed in solution.

The PvSAE Apo I and II structures displayed a near complete overlap when aligned (root-mean-square-deviation (RMSD) of 0.18 Å), with the secondary structures in the same conformation. The only major difference between these structures was a small region of the most N-terminal α-helix, which is discussed below. The individual protomers have a fold with a central five-stranded parallel β-sheet surrounded by α-helices, see Figure 3A, which is typical of an SNGH type I domain. This domain architecture is more similar to that of the β-coronavirus HEs (Bakkers *et al.*, 2017) than the domain organization observed for *E. coli* NanS (Rangarajan *et al.*, 2011), see Figure 3C. β-coronavirus HEs also contain a central five-stranded parallel β-sheet surrounded by α-helices. However, unlike PvSAE these viral enzymes contain a hemagglutinin domain inserted between the 3rd and 4th β-strand of the β-sheet. The structure of *E. coli* NanS, in contrast, contains a seven-stranded β-sheet with six parallel β-strands and one anti-parallel β-strand (Rangarajan *et al.*, 2011), typical of SNGH-type II domains. The suspected serine nucleophile of PvSAE, S52, is located at the beginning of the second α-helix, following the first β-strand from the N-terminus, a position that is preserved in both *E. coli* NanS and β-coronavirus HEs (see Figure 3C). The suspected catalytic histidine and aspartic acid, H202 and D199, are located on an Ω loop which is followed by the most C-terminal α-helix.

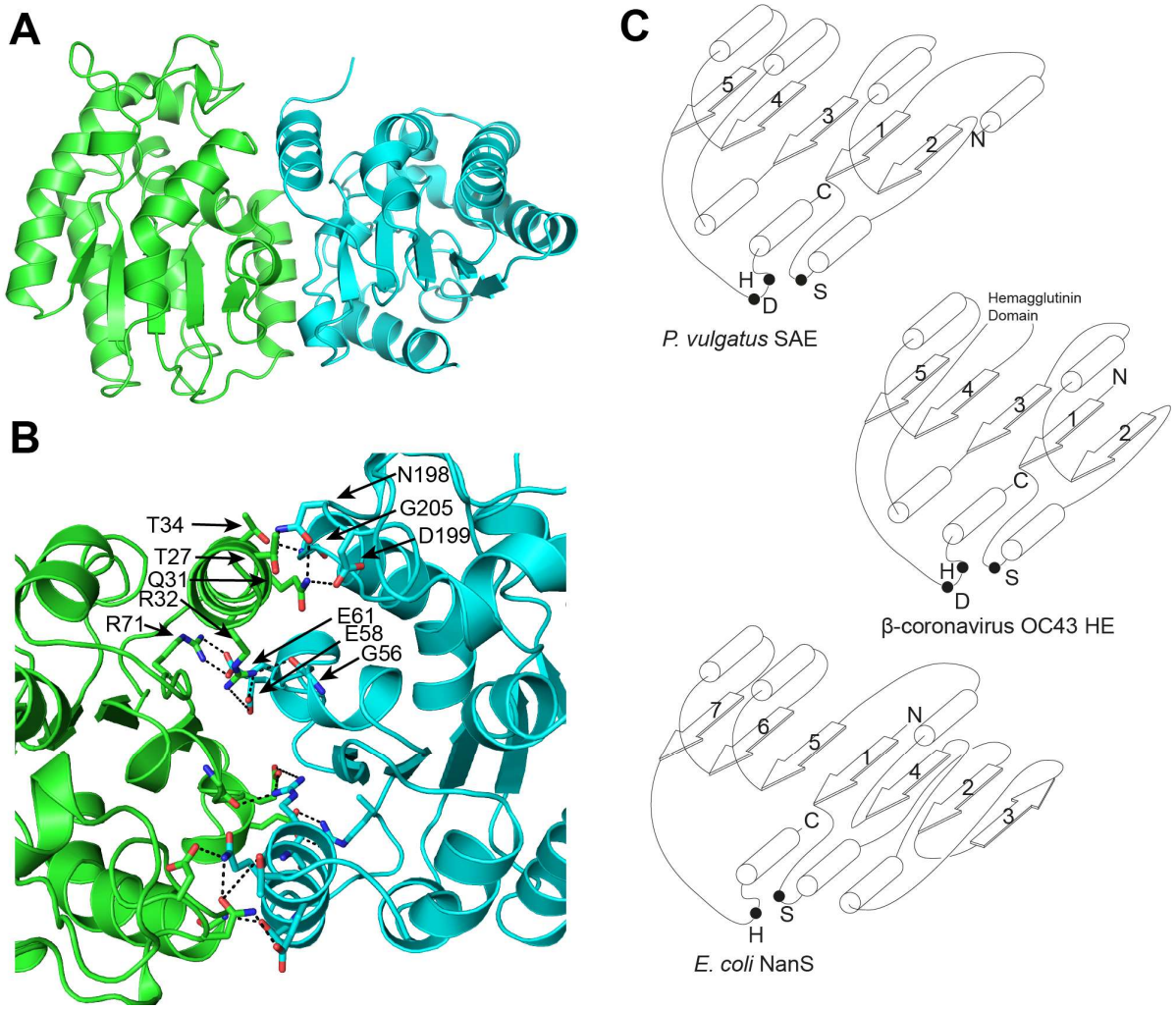

Figure 3 (**A**) Apo I PvSAE dimer structure shown in cartoon representation. (**B**) PvSAE dimer interface. Residues forming salt bridges and hydrogen bonds between two protomers are shown as sticks. Hydrogen bonding and salt bridges are shown as black dashed lines. The residues from one of the two symmetrical interaction interfaces are labelled. The two protomers are coloured green and cyan. (**C**) Domain organization of sialic acid esterases. β-strands are numbered according to their primary sequence positions. The β-strands present in the hemagglutinin domain of OC43 HE have been omitted for simplicity. The catalytic serine (S), histidine (H) and aspartate (D) are labelled. The N-terminus and C terminus are denoted with N and C respectively. Structural elements are not drawn to scale.

As stated above, each of the protomers is present in a dimer in both of the crystal structures determined here, and in the previously determined structure of PvSAE. These dimers have C2 rotational symmetry and interactions between the two interfaces were examined using PISA (Krissinel & Henrick, 2007). The dimer interface is composed mainly of hydrophilic residues that form salt bridges and hydrogen bonds (Figure 3B). A majority of these interactions are formed between residues on the most N-terminal α-helix (residues 27-34) and residues on the Ω loop containing the catalytic histidine and aspartic acid

(residues 198-205), see Figure 3B. This interaction also involves the aspartic acid D199 (a constituent of the Asp-His-Ser catalytic triad) which forms a hydrogen bond with Q31 in the neighbouring protomer. We expect that the dimerization of PvSAE observed here maintains the correct positioning of the catalytic residues. Dimerization results in a loss of 1300 $Å^2$ of solvent accessible surface for each monomer and leads to the formation of a 15 Å deep cleft. This cleft contains both active sites, which face each other and are separated by 15 Å, as determined by the distance between the proposed serine nucleophiles.

The most significant difference between the Apo I and Apo II structures was the positioning of the most N-terminal residues (residues 24-27) of the first α-helix. In the Apo II structure and a previously solved structure of PvSAE (PDB: 6njc) this alpha helix extends above the active site cleft restricting the entrance to the active site cleft to approximately 8 Å (see Figure S1). In the Apo I structure residues 24-27 of one of the protomers in the homodimer is no longer part of this α-helix structure, but instead forms a loop projecting out of the active site. This conformational change is supported by a hydrogen-bond between the hydroxyl of Tyrosine Y25 and the carboxylic acid of Aspartic acid D83 from the neighbouring protomer. The result of this change in conformation is that one active site is more accessible and has an 'open' conformation while the other is still 'blocked' by the N-terminal α-helix, see Figure S1. Although the observed conformational change may be an artifact of the crystallization conditions, this change indicates the residues present in the N-terminal have inherent plasticity. We suspect this conformational change may be important for the accommodation of large, complex glycans that contain acetylated sialic acid.

3.3. Structural comparisons and active site

To investigate proteins with structural homology to PvSAE, a search was performed through the DALI server (Holm, 2020). This further supported PvSAE as a member of the SGNH family, as the search revealed PvSAE shared low amino acid sequence identity, but high structural similarity, with members of the SGNH family, see Figure S2. Further comparison between sequences of the top hits showed conservation for the residues: S52, G75, N103, H202 and D199. Confirming that the catalytic triad corresponds to residues S52, H202 and D199, with residues G75 and N103 responsible for forming the oxyanion hole. The residues are also present within the conserved blocks of sequences common to the SGNH superfamily. However, the block I sequence varied, with a GNS motif (G50, N51, S52) compared to the classic GDS motif, see Figure S2. The presence of asparagine in the place of aspartate in PvSAE, however, does not change the positioning of this amino acid when compared to the structure of the top DALI result with a GDS motif (putative platelet activating factor from *Streptococcus pneumoniae*; PDB: 2hsj), see Figure S2. The remaining three sequence blocks (II, III and V) were conserved across the DALI hits. These blocks contain the conserved Glycine (block II) the conserved

Asparagine in a GXN motif (block III) and the catalytic serine and histidine in a DXXH motif (block V).

The PvSAE structure for both Apo I and Apo II had a Mg^{2+} metal ion in close proximity to the active site. This ion is also observed in the previously reported structure of PvSAE (PDB: 6njc). The electron density surrounding Mg^{2+}, bond lengths and octahedral coordination to six electron donors supports this annotation. The metal was likely incorporated into the structure during the production of the enzyme. Mg^{2+} interacts with the carbonyl groups of P194 and T197, and the side-chain hydroxyl group of Y146 and three waters. The T197 residue, in turn is positioned to form a hydrogen bond with the catalytic aspartate (D199), see Figure 4C.

One of the top DALI hits for the structure of PvSAE was PAF-AH (brain platelet-activating factor acetylhydrolase). A structure of PAF-AH (PDB: 3dt9) contains the covalent oganophosphorous inhibitor soman bound to the nucleophilic active-site serine (Epstein *et al.*, 2009). Overlay of PvSAE and pAF-AH displayed good overlap between the active site residues (RMSD of 0.37 Å for residues within 8 Å of the nucleophilic serine), with the residues in similar conformations, see Figure 4A. This suggests that organophosphorous inhibitors may also be active against PvSAE.

All of the top five DALI hits are homodimeric structures, as is seen for PvSAE. Like PvSAE the dimer interface for these five structures involves the N-terminal α-helix. However, the N-terminal α-helix does not interact with the Ω loop that contains the catalytic histidine, in all of these structures, see Figure S3. In three of these structures (PDB: 3p94, 4ppy, 4iyj) the N-terminal α-helix interacts with residues present in the 3rd and 4th α-helices which in turn interact with the 3rd and 4th α-helices on the neighbouring protomer. This results in more exposed active sites that are not part of the same active site cleft. The dimeric structure of *S. pneumonia* putative platelet activating factor (PDB: 2hsj) is more similar to PvSAE, with the dimer interface between the N-terminal α-helix and the C-terminal α-helix. However, in this structure the N-terminal α-helix does not involve the Ω loop. The structure of pAF-AH (PDB: 3dt9) also forms a dimer with an interface composed of salt bridges formed by hydrophilic residues between the loop containing the catalytic histidine and aspartate and the N-terminal α-helix (Epstein *et al.*, 2009). The angle of this interaction interface, however, differs to that seen for PvSAE, resulting in a deeper active site cleft than observed for PvSAE.

The 9-O-SAE NanS from *E. coli* has a SGNH-type II fold with catalytically active residues S19 and H301. This enzyme, however, lacks the catalytic aspartate residue from the catalytic triad. This interaction may be compensated for by a more remote carboxylate, whose action is mediated through an ordered water molecule (Rangarajan *et al.*, 2011). Overlay of the active site of NanS and PvSAE (RMSD = 0.32 Å for an 8 Å sphere surrounding the catalytic serine) reveals that the catalytic histidine and serine residues occupy a similar position, despite the absence of a catalytic acid in NanS, see Figure 4B. The oxyanion hole of NanS contains a glycine as is seen in PvSAE. However, in NanS the other

constituent of the oxyanion hole is Glutamine Q18, as opposed to Asparagine N103 in PvSAE, see Figure 4B. Unlike PvSAE, *E. coli* NanS forms a monomer in solution, indicating homodimerization is not conserved for 9-O-SAEs. The monomeric nature of *E. coli* NanS is likely due to the absence of the most N-terminal α-helix which forms a large part of the dimer interface in PvSAE, see Figure 3C.

EstA from *Streptococcus pneumoniae* has also been identified as a 9-O-SAE (Mitchell *et al.*, 2017). This enzyme displays a canonical α/β hydrolase topology with a Ser-His-Asp catalytic triad (Kim *et al.*, 2008). *S. pneumoniae* EstA, however, is not an SGNH-hydrolase as the catalytic triad (S120-H231-D202) residues do not align with the conserved blocks typical of a SGNH-hydrolase. The active site of EstA also differs from an SGNH-hydrolase, with the backbone nitrogen from two methionine residues (M45 an M121) forming the oxyanion hole (Kim *et al.*, 2008), whereas in PvSAE the backbone nitrogen of a glycine (G75) and the amide side chain of an asparagine residue (N103) form the oxyanion hole, see Figure S4.

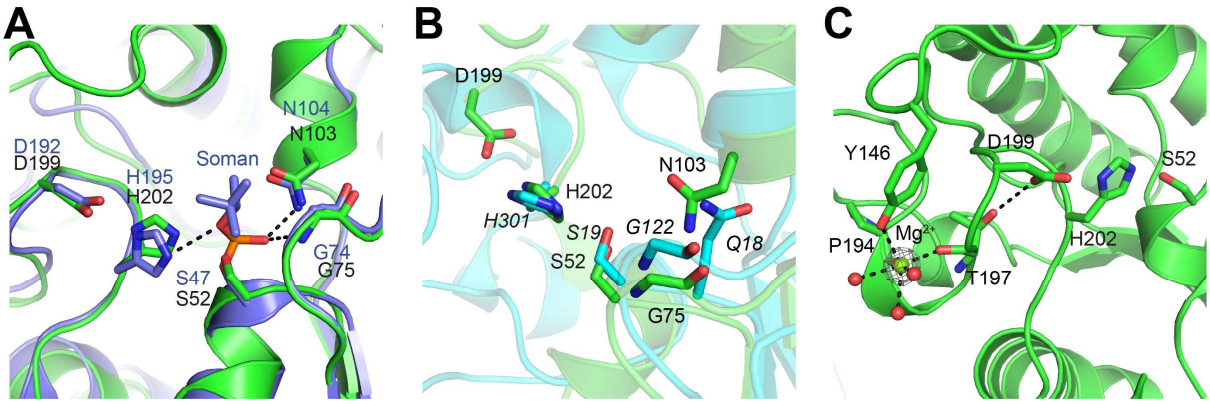

Figure 4 Active site structure of PvSAE (**A**) Active site overlay of PvSAE Apo I and the pAF-AH (PDB: 3dt9), with active site residues and soman inhibitor shown as sticks. PvSAE residues are labelled in black, pAF-AH residues are labelled in blue. Polar contacts between the soman inhibitor and active site residues are displayed with black dashed lines. pAF-AH and soman inhibitor carbons are coloured purple, while PvSAE carbons are green. (**B**) Active site overlay of PvSAE and NanS (3pt5) (Rangarajan *et al.*, 2011), with active site residues shown as sticks. NanS residues are labelled in italics and NanS is coloured in cyan. (**C**) Mg^{2+} coordination with surrounding electron density. Electron density mesh is σA-weighted $2F_o - F_c$ density contoured at 2 σ (0.65 e^- A^{-3}) and rendered using PyMol. Sphere scale is set as 0.25. PvSAE catalytic triad is also shown.

To further investigate phylogenetic conservation of amino acids, conservation scores were determined using CONSURF (Ashkenazy *et al.*, 2016) and mapped onto the PvSAE structure. This revealed absolute conservation of residues in the active site, see Figure 5. There was also absolute conservation between the interface amino acid residues involved in forming salt bridges between the two dimers,

namely R32, R71, E58 and E61. The central five-stranded β-sheet was relatively well conserved (CONSURF scores 6-9), with only K133 displaying low conservation (CONSURF score = 2). The level of conservation between the α-helices varied, helices distal to the active site and protein interface displayed increased variability, A111 – E128, N150 – E173. Loops at the protein surface were also poorly conserved, in contrast to those near the active site. The level of conservation for the Mg^{2+} binding site varied. The T197 residue, which may interact with the active site, was relatively well conserved (CONSURF score of 7). Conversely, Y146 and P194 residues displayed poor conservation (consurf scores of 1 and 2 respectively).

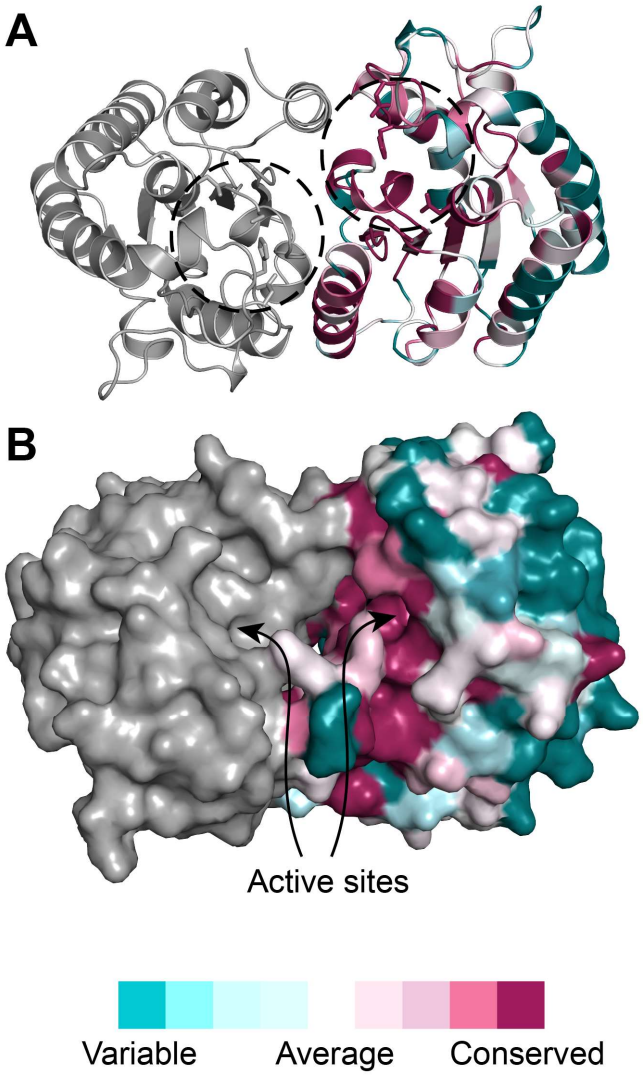

Figure 5 Conservation of PvSAE Apo I surface. In the ribbon representation of PvSAE Apo I active site residues are shown as sticks (S52, H202, D199, N103, G75). Chain A is coloured by sequence conservation, chain B is coloured grey. Figures were prepared using CONSURF (Ashkenazy *et al.*, 2016) using the top 500 BLAST hits(Altschul *et al.*, 1990) to the protein sequence of PvSAE. The figure was generated using PyMol (DeLano Scientific LLC).

The high resolution of the Apo I structure allowed identification of a malonate molecule present in the active sites of PvSAE. Malonate is positioned such that a carboxylate functional group is within hydrogen bonding distance of the oxyanion hole residues N103 and G75, see Figure 6A. This mirrors the positioning of acetate present in the previously determined structure of PvSAE, where the acetate interacts directly with the oxyanion hole, see Figure 6B. Automated docking of 9-O-acetyl sialic acid into each active site showed that this substrate could fit into both active sites of the Apo I structure and the Apo II structures, with the 9-acetyl group correctly positioned for catalysis (see Figures 6 & S6). As the most N-terminal a-helix of the structure of Apo I has different conformations, the two active sites accommodate the modelled 9-O-acetyl sialic acid in a different conformation. In the 'open' active site, the 9-O-acetyl sialic acid is positioned such that the 2-position projects towards the bulk solvent, indicating that this site could accommodate larger glycans. There also appears to be space for both acetylation at the 7-position, or an N-glycolyl modification at the 5-position, both modifications which occur on animal mucins (Park, 2019). The carboxylate on the modelled ligand is within hydrogen bonding distance of the backbone nitrogen of F28, which only becomes accessible once the N-terminal α-helix has taken the 'open' conformation. This interaction with the carboxylate contrasts with the interaction seen in viral sialic acid esterases, which employ an active site arginine as an interaction partner with the carboxylate (Rosenthal *et al.*, 1998). Docking into the 'blocked' site of the Apo I structure resulted in a ligand conformation with the 2-position projecting towards the N-terminal α-helix, suggesting that this active site could only accommodate monomeric substrates. A similar conformation to the blocked active site was observed in the Apo II structure, see Figure S6.

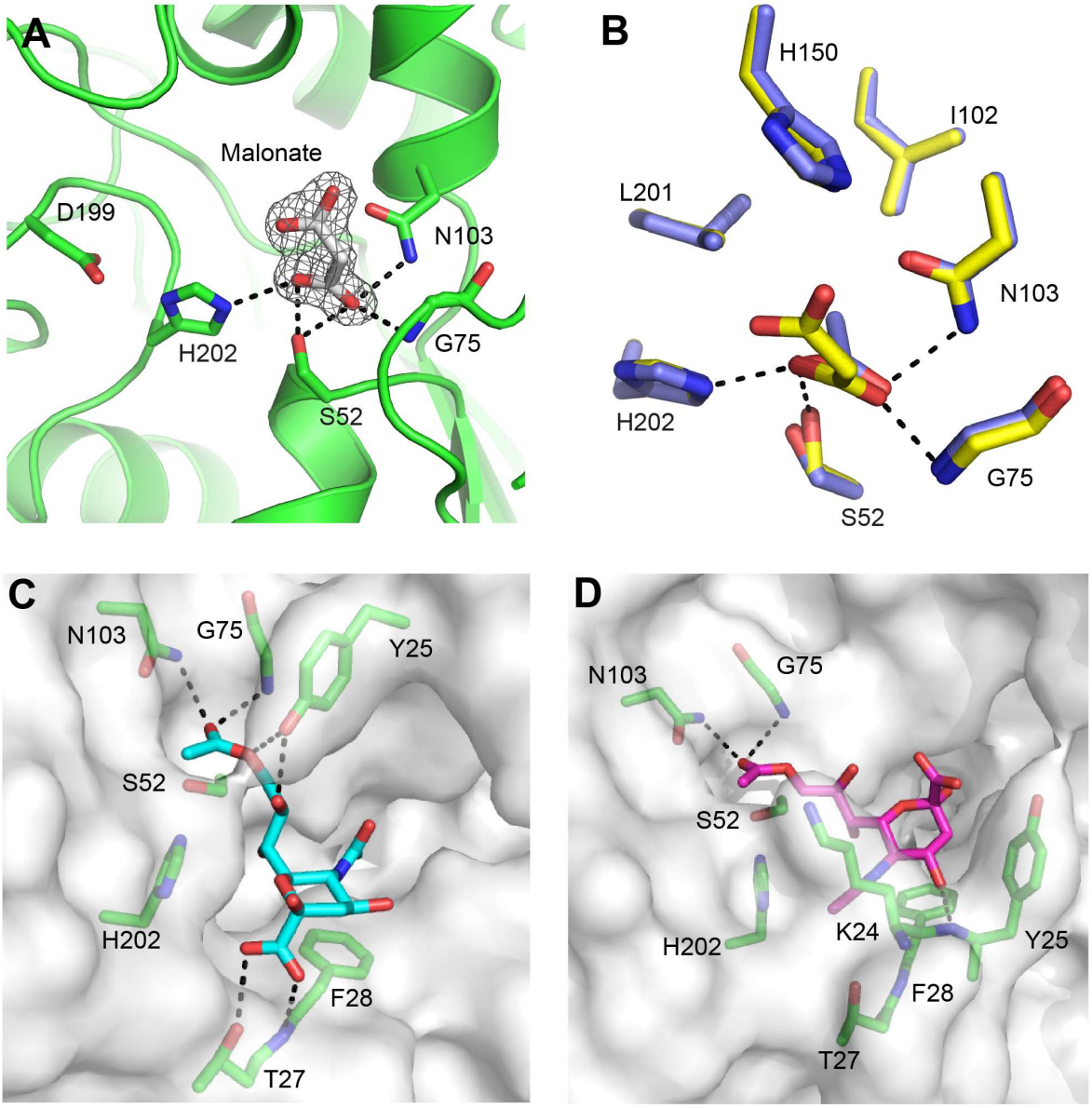

Figure 6 **A**) PvSAE Apo I complex with malonate. Active site residues and malonate are shown as sticks. Electron density mesh surrounding malonate is σA-weighted $2F_o - F_c$ density contoured at 1 σ, (0.33 e^- A^{-3}). **B**) Superposition of active sites containing malonate and acetic acid. The Apo I structure, containing malonate, is shown with carbons colored green. The PDB structure 6njc, containing acetate, is shown with carbons colored blue. **C**) 9-O-acetylsialic acid docked into the 'open' active site of the Apo I structure and **D**) the 'blocked' active site of Apo I. Polar interactions within 3.2 Å are shown as black dashed lines in all panels.

4. Conclusion

9-O-acetyl-N-acetyl sialic acid esterases are found throughout the tree of life, from viruses to bacteria and animals. Examination of the 9-O-SAE from the human commensal bacteria *P. vulgatus* (PvSAE) adds to our growing understanding of the structure and mechanisms underpinning this class of enzymes.

Three-dimensional structures of PvSAE determined by X-ray crystallography reveal the homodimerization of PvSAE, and the molecular architecture of the SGNH active site. This has revealed a canonical Ser-His-Asp catalytic triad encompassed in a binding cleft formed through dimerization. Additionally, the conformational flexibility of the N-terminal α-helix appears to mediate the accessibility of the active cleft to large oligosaccharide substrates. The structural analysis of PvSAE also contributes to our understanding of other predicted SGNH-hydrolases that are yet to be solved, such as the human SIAE. Looking forward, the structural analysis of 9-O-SAE enzymes such as PvSAE promises to enable the development of specific inhibitors targeting this class of enzymes. This could be extended to the human 9-O-SAE, bacterial pathogen 9-O-SAEs and viral 9-O-SAEs, such as those encoded by specific β-coronaviruses or influenza C.

Acknowledgements We thank Dr. Andrew Leech, University of York Bioscience Technology Facility for assistance with SEC-MALLS. We thank the Diamond Light Source for access to beamline I03 (proposal number mx-18598) that contributed to the results presented here and Dr. Johan Turkenburg and Sam Hart for coordinating data collection. G.J.D. thanks the Royal Society for the Ken Murray Research Professorship. We also acknowledge generous support from the BBSRC (BB/R001162/1 to G.J.D.).

References

Albers, M., Schröter, L., Belousov, S., Hartmann, M., Grove, M., Abeln, M. & Mühlenhoff, M. (2021). *Glycobiology*.
Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G. & Nielsen, H. (2019). *Nature Biotechnology* **37**, 420-423.
Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). *Journal of Molecular Biology* **215**, 403-410.
Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T. & Ben-Tal, N. (2016). *Nucleic Acids Research* **44**, W344-W350.
Bakkers, M. J. G., Lang, Y., Feitsma, L. J., Hulswit, R. J. G., de Poot, S. A. H., van Vliet, A. L. W., Margine, I., de Groot-Mijnes, J. D. F., van Kuppeveld, F. J. M., Langereis, M. A., Huizinga, E. G. & de Groot, R. J. (2017). *Cell Host & Microbe* **21**, 356-366.
Barnard, K. N., Alford-Lawrence, B. K., Buchholz, D. W., Wasik, B. R., LaClair, J. R., Yu, H., Honce, R., Ruhl, S., Pajic, P., Daugherity, E. K., Chen, X., Schultz-Cherry, S. L., Aguilar, H. C., Varki, A. & Parrish, C. R. (2020). *Journal of Virololgy* **94**.
Baumann, A.-M. T., Bakkers, M. J. G., Buettner, F. F. R., Hartmann, M., Grove, M., Langereis, M. A., de Groot, R. J. & Mühlenhoff, M. (2015). 9-O-Acetylation of sialic acids is catalysed by CASD1 via a covalent acetyl-enzyme intermediate, Vol. 6, Nat Commun, p. 7673.
Briliūtė, J., Urbanowicz, P. A., Luis, A. S., Baslé, A., Paterson, N., Rebello, O., Hendel, J., Ndeh, D. A., Lowe, E. C., Martens, E. C., Spencer, D. I. R., Bolam, D. N. & Crouch, L. I. (2019). *Nature Microbiology* **4**, 1571-1581.
Drula, E., Garron, M.-L., Dogan, S., Lombard, V., Henrissat, B. & Terrapon, N. (2022). *Nucleic Acids Research* **50**, D571-D577.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Crystallographica Section D, Structural Biology* **66**, 486-501.
Epstein, T. M., Samanta, U., Kirby, S. D., Cerasoli, D. M. & Bahnson, B. J. (2009). *Biochemistry* **48**, 3425-3435.
Evans, P. R. & Murshudov, G. N. (2013). *Acta Crystallographica Section D, Biological crystallography* **69**, 1204-1214.
Gorrec, F. (2009). *Journal of Applied Crystallography* **42**, 1035-1042.
Holm, L. (2020). *Protein Science* **29**, 128-140.
Huang, Y.-L., Chassard, C., Hausmann, M., von Itzstein, M. & Hennet, T. (2015). *Nat Commun* **6**, 8141.
Hurdiss, D. L., Drulyte, I., Lang, Y., Shamorkina, T. M., Pronker, M. F., van Kuppeveld, F. J. M., Snijder, J. & de Groot, R. J. (2020). *Nat Commun* **11**, 4646.
Kim, M. H., Kang, B. S., Kim, S., Kim, K.-J., Lee, C. H., Oh, B.-C., Park, S.-C. & Oh, T.-K. (2008). *Proteins: Structure, Function, and Bioinformatics* **70**, 578-583.
Krissinel, E. & Henrick, K. (2007). *Journal of Molecular Biology* **372**, 774-797.
Langereis, M. A., Zeng, Q., Heesters, B. A., Huizinga, E. G. & de Groot, R. J. (2012). *PLoS Pathogens* **8**, e1002492.
Lenfant, N., Hotelier, T., Velluet, E., Bourne, Y., Marchot, P. & Chatonnet, A. (2012). *Nucleic Acids Research* **41**, D423-D429.
Long, F., Nicholls, R. A., Emsley, P., Gražulis, S., Merkys, A., Vaitkus, A. & Murshudov, G. N. (2017). *Acta Crystallographica Section D Structural Biology* **73**, 112-122.
Mather, R. L., Loveson, K. F. & Fillmore, H. L. (2019). *Scientific Reports* **9**, 8609.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *Journal of Applied Crystallography* **40**, 658-674.
Mitchell, T. J., Kahya, H. F., Andrew, P. W. & Yesilkaya, H. (2017). *PLOS Pathogens* **13**, e1006263.
Mølgaard, A., Kauppinen, S. & Larsen, S. (2000). *Structure* **8**, 373-383.
Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S. & Olson, A. J. (2009). *Journal of Computational Chemistry* **30**, 2785-2791.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Crystallographica Section D, Structural Biology* **67**, 355-367.
Nakayama-Imaohji, H., Ichimura, M., Iwasa, T., Okada, N., Ohnishi, Y. & Kuwahara, T. (2012). *The Journal of Medical Investigation* **59**, 79-94.
Orizio, F., Damiati, E., Giacopuzzi, E., Benaglia, G., Pianta, S., Schauer, R., Schwartz-Albiez, R., Borsani, G., Bresciani, R. & Monti, E. (2015). *Glycobiology* **25**, 992-1006.
Park (2019). *Vaccines* **7**.
Rangarajan, E. S., Ruane, K. M., Proteau, A., Schrag, J. D., Valladares, R., Gonzalez, C. F., Gilbert, M., Yakunin, A. F. & Cygler, M. (2011). *Protein Science* **20**, 1208-1219.
Robinson, L. S., Lewis, W. G. & Lewis, A. L. (2017). *J Biol Chem* **292**, 11861-11872.
Rosenthal, P. B., Zhang, X., Formanowski, F., Fitz, W., Wong, C. H., Meier-Ewert, H., Skehel, J. J. & Wiley, D. C. (1998). *Nature* **396**, 92-96.
Studier, F. W. (2005). *Protein Expression and Purification* **41**, 207-234.
Tortorici, M. A., Walls, A. C., Lang, Y., Wang, C., Li, Z., Koerhuis, D., Boons, G.-J., Bosch, B.-J., Rey, F. A., de Groot, R. J. & Veesler, D. (2019). *Nature Structural & Molecular Biology* **26**, 481-489.
Trott, O. & Olson, A. J. (2009). *Journal of Computational Chemistry*, NA-NA.
Upton, C. & Buckley, J. T. (1995). *Trends in Biochemical Sciences* **20**, 178-179.
Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). *Acta Crystallographica. Section D, Biological crystallography* **67**, 235-242.
Zeng, Q., Langereis, M. A., van Vliet, A. L. W., Huizinga, E. G. & de Groot, R. J. (2008). *Proc Natl Acad Sci U S A* **105**, 9065-9069.
Zimm, B. H. (1945). *The Journal of Chemical Physics* **13**, 141-145.

Supporting information

Table S1 Structurally homologous enzymes to the PvSAE Apo I structure, identified using the DALI server (Holm, 2020).

Description	Source	PBD	Z score	RMSD / Å	L-align	Identity / %
Putative platelet activating factor	*Streptococcus pneumonia*	2hsj	27.4	1.7	194	28
GDSL-like lipase	*Parabacteroides distasonis*	3p94	24.7	2.0	183	25
Putative acylhydrolase	*Bacteroides fragilis*	4ppy	24.2	2.4	190	26
GDSL-like protein	*Bacteroides uniformis*	4iyj	24.2	2.4	190	26
Brain platelet-activating factor acetylhydrolase	*Bos taurus*	3dt9	22.3	2.1	185	20

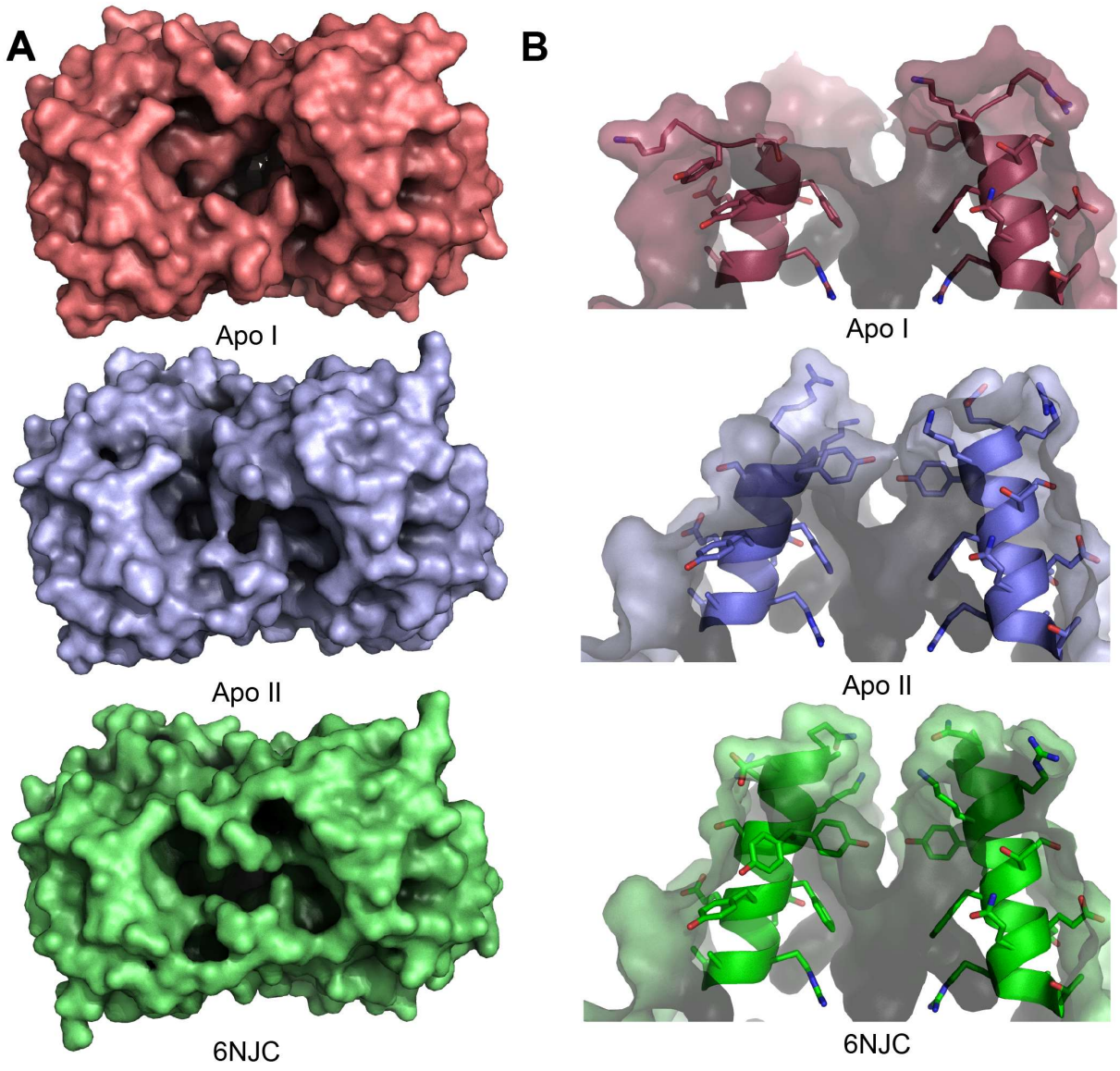

Figure S1 Conformational changes of the N-terminal α-helix. **A**) Surface representation of the Apo I, Apo II and 6njc structures as viewed from above the active site. **B**) The side view of the N-terminal helix from the same structures with residues present in the helix shown in stick representation.

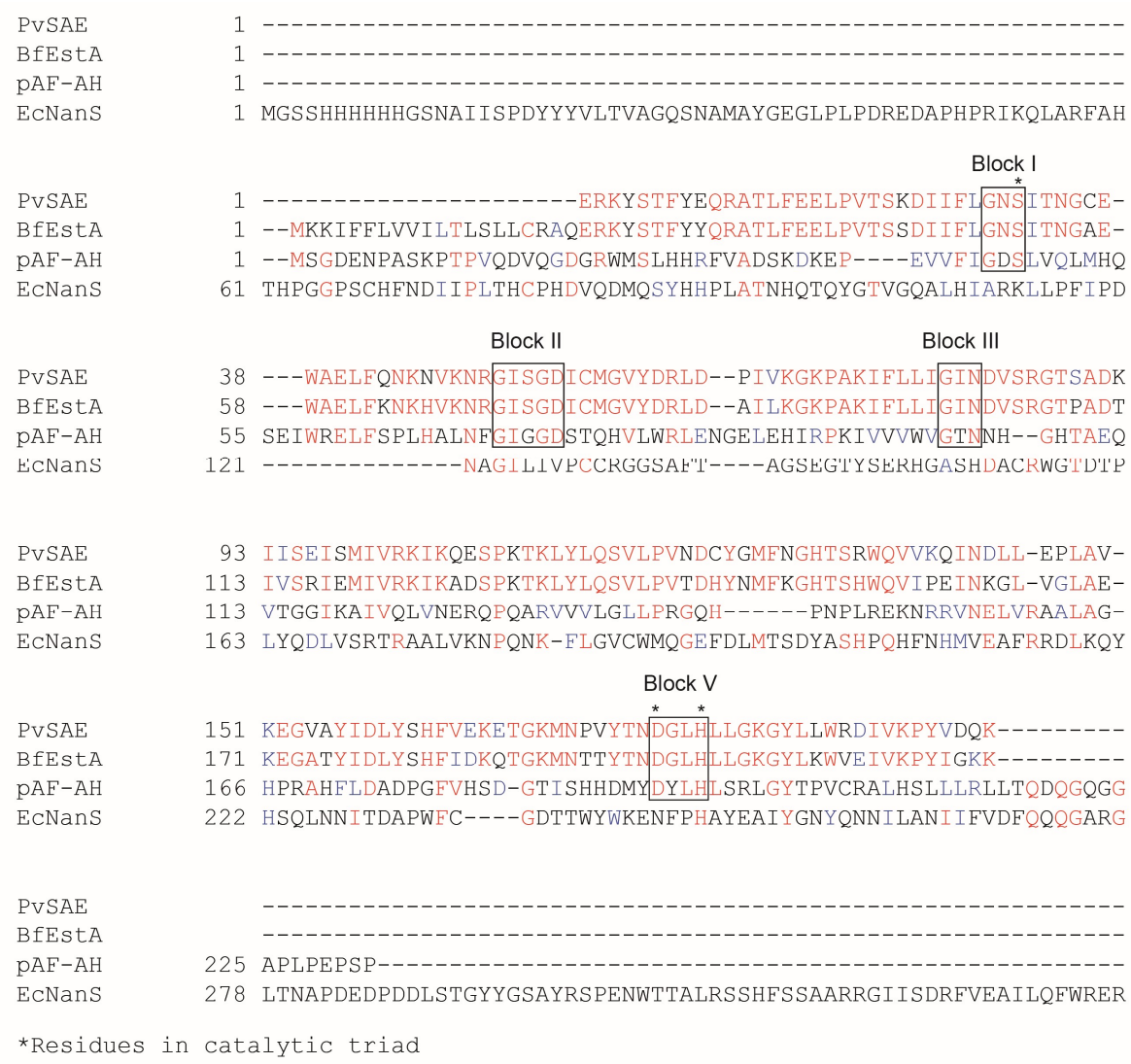

Figure S2 Amino acid sequence alignments for a range of 9-O-SAEs. Alignments include *Phocaeicola vulgatus* PvSAE (ABR41743.1), *Bacteroides fragilis* EstA (WP_100721746.11), *Bos taurus* pAF-AH (PDBID: 3DT6) and *E. coli* NanS (PDBID: 3PT5). Residues forming the catalytic Ser-His-Asp triad are indicated with a star above the amino acid.

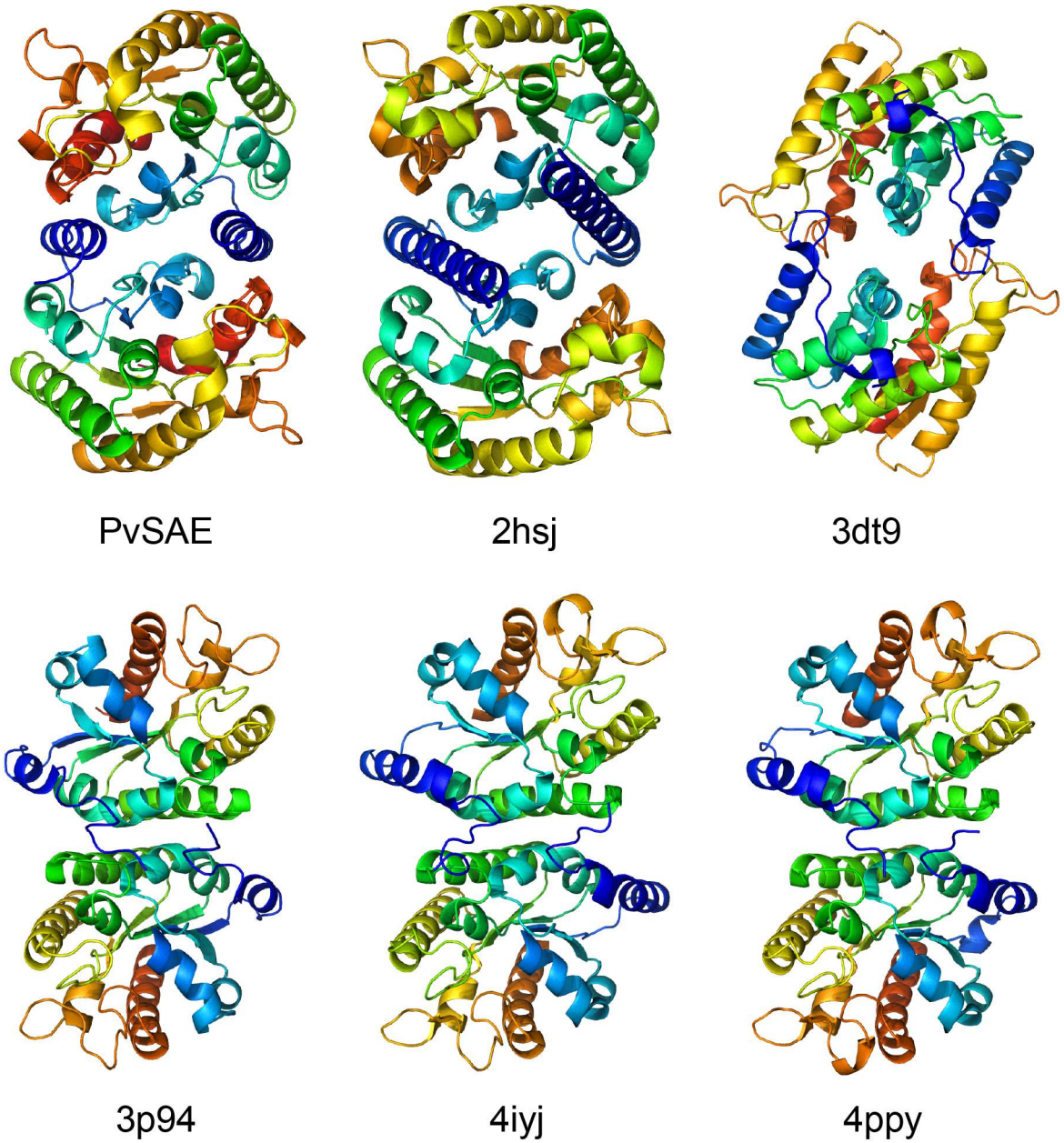

Figure S3 Dimer interfaces of PvSAE and homologs. Dimers are shown in cartoon representation and colored from the N-terminus (purple) to the C-terminus (red).

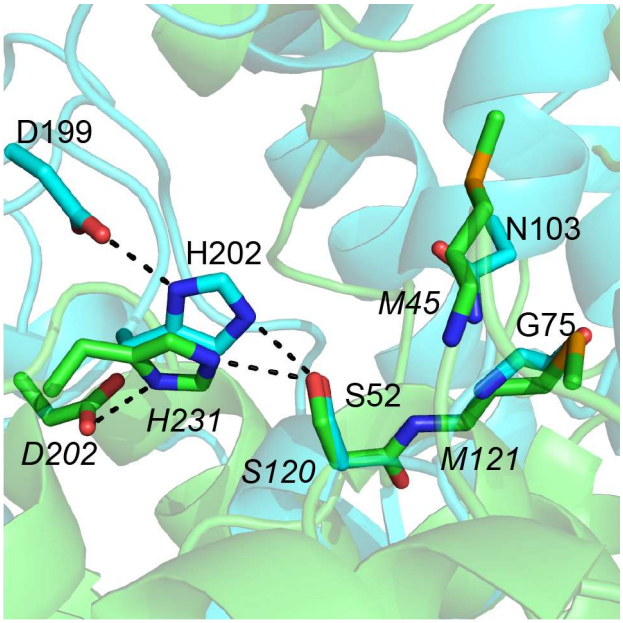

Figure S4 Active site overlay of PvSAE and *S. pneumonia* EstA (PDB: 2uz0)(Kim *et al.*, 2008) with active site residues shown as sticks. *S. pneumonia* EstA residues are labelled in italics and carbons are colored in green, while PvSAE carbons are cyan. Hydrogen bonds between the catalytic triad are shown as black dashed lines.

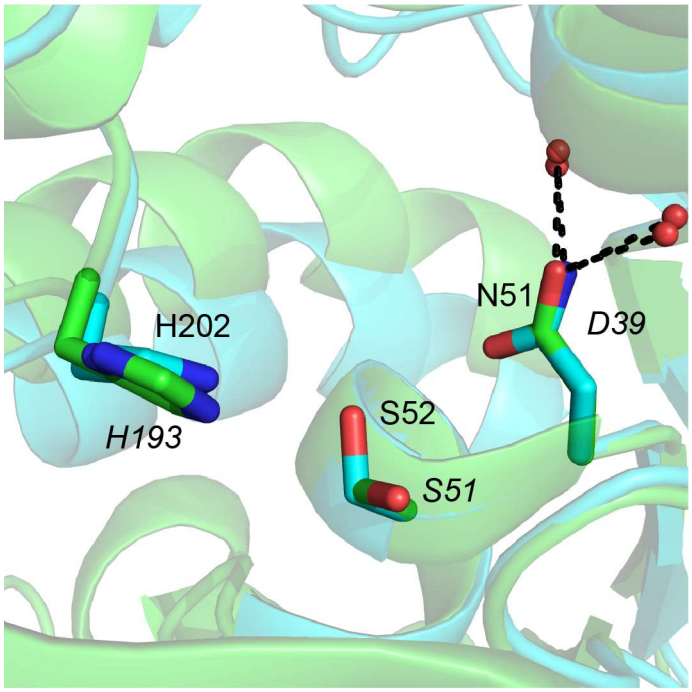

Figure S5 Active site overlay of PvSAE and *S. pneumonia* putative platelet activating factor (PDB-ID: 2hsj) with active site residues shown as sticks. *S. pneumonia* putative platelet activating factor residues are labelled in italics and carbons are colored in green, while PvSAE carbons are cyan. Hydrogen bonds between the catalytic triad are shown as black dashed lines.

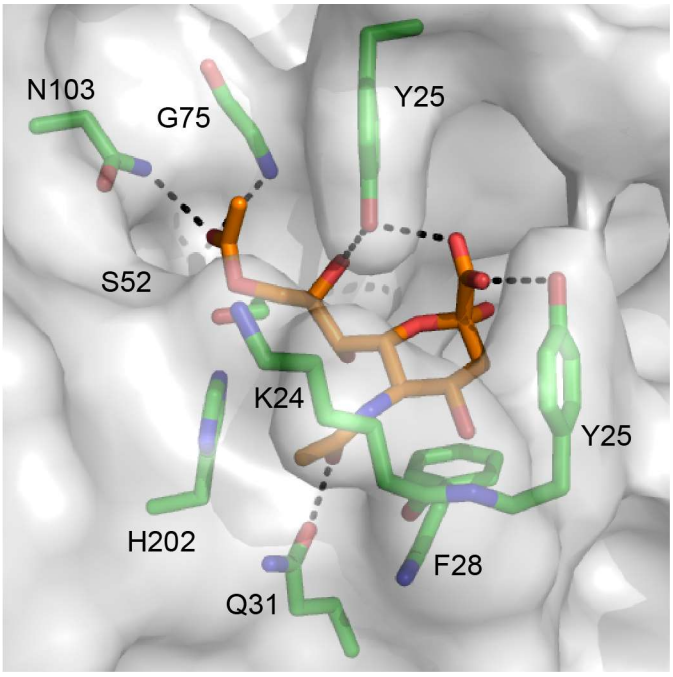

Figure S6 Docking of 9-O-acetylsialic acid into the active site of the Apo II structure. Polar interactions within 3.2 Å are shown as black dashed lines.