
Article:
Pearson, Caroline, Tindall, Sarah, Herman, Reyme et al. (5 more authors) (2020) Acetylation of surface carbohydrates in bacterial pathogens requires coordinated action of a two-domain membrane-bound acyltransferase. MBio. e01364-20. pp. 1-19. ISSN 2150-7511

https://doi.org/10.1128/mBio.01364-20

Acetylation of surface carbohydrates in bacterial pathogens requires coordinated action of a two-domain membrane-bound acyltransferase

Caroline R Pearson[a,b,1]; Sarah N Tindall [a,b,1]; Reyme Herman[b]; Huw T Jenkins[d]; Alex Bateman[e], Gavin H Thomas[a,b], Jennifer R Potts[b]; Marjan W Van der Woude[a,c]

Author Affiliation:

a York Biomedical Research Institute, University of York, York YO10 5DD, UK

b Department of Biology, University of York, York YO10 5DD, UK

c Hull York Medical School, University of York, York YO10 5DD, UK

d York Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5DD, UK

e European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK

Corresponding Author: M. W. Van der Woude; York Biomedical Research Institute and Hull York Medical School; Wentworth Way; University of York, Heslington, York YO10 5DD; United Kingdom

email: Marjan.vanderwoude@york.ac.uk

[1] Caroline R. Pearson and Sarah N. Tindall contributed equally to this work. Author order was determined alphabetically.

Abstract

Membrane bound Acyltransferase_3 (AT3) domain-containing proteins are implicated in a wide range of carbohydrate *O*-acyl modifications but their mechanism of action is largely unknown. O-antigen acetylation by AT3 domain-containing acetyltransferases of *Salmonella* spp. can generate a specific immune response upon infection and can influence bacteriophage interactions. This study integrates *in situ* and *in vitro* functional analysis of two of these proteins, OafA and OafB (formerly F2GtrC) which display an 'AT3-SGNH fused' domain architecture where an integral membrane AT3 domain is fused to an extra-cytoplasmic SGNH domain. An *in silico*-inspired mutagenesis approach of the AT3 domain identified seven residues which are fundamental for the mechanism of action of OafA, with a particularly conserved motif in TMH1 indicating a potential acyl donor interaction site. Genetic and *in vitro* evidence demonstrates that the SGNH domain is both necessary and sufficient for lipopolysaccharide acetylation. The structure of the periplasmic SGNH domain of OafB identified features not previously reported for SGNH proteins. In particular, the periplasmic portion of the inter-domain linking region is structured. Significantly, this region constrains acceptor substrate specificity, apparently by limiting access to the active site. Co-evolution analysis of the two domains suggests possible inter-domain interactions. Combining these data we propose a refined model of the AT3-SGNH proteins, with structurally constrained orientations of the two domains. These findings enhance our understanding of how cells can transfer acyl groups from the cytoplasm to specific extracellular carbohydrates.

Importance

Acyltransferase-3 (AT3) domain-containing membrane proteins are involved in O-acetylation of a diverse range of carbohydrates across all domains of life. In bacteria they are essential in processes including symbiosis, resistance to antimicrobials, and biosynthesis of antibiotics. Their mechanism of action, however, is poorly characterised. We analysed two acetyltransferases as models for this

important family of membrane proteins, that modify carbohydrates on the surface of the pathogen *Salmonella enterica*, affecting immunogenicity, virulence and bacteriophage resistance. We show that when these AT3 domains are fused to a periplasmic partner domain, both domains are required for substrate acetylation. The data shows conserved elements in the AT3 domain and unique structural features of the periplasmic domain. Our data provides a working model to probe the mechanism and function of the diverse and important members of the widespread AT3 protein family, which are required for biologically significant modifications of cell-surface carbohydrates.

Introduction

Salmonella infections are a considerable public health burden in both developing and developed countries. *Salmonella enterica* subspecies *enterica* serovar Typhimurium (STM) is estimated to cause more than 150,000 human deaths from gastroenteritis each year (1, 2). A sublineage of this serovar is the dominant cause of invasive non-Typhoidal *Salmonella* (iNTS) bloodstream infections in Africa (3). The Typhi serovar of this subspecies is the major cause of typhoid fever, resulting in over 200,000 deaths annually (2, 4). In the US, there are over 10,000 cases annually of these serovars combined (5, 6).

Cell surface lipopolysaccharide (LPS) is an important virulence factor. The O-antigen, the most distal and variable portion of LPS, is composed of repeating oligosaccharide units whose composition and structure varies between species and, in the case of *Salmonella* spp., between serovars. Modification of the O-antigen by alteration of sugar linkages or addition of moieties such as glucose or acetate (7, 8) can influence immunogenicity, virulence, and confer resistance to lytic phage infection (9–12).

Carbohydrates on the bacterial cell surface are frequently O-acetylated by acyltransferase proteins which contain a 10 transmembrane helix (TMH) Acyltransferase_3 (AT3, IPR002656, PF01757; also

known as Acyltransferase_3/Putative Acetyl-CoA Transporter, TC 9.B.97). This family of proteins is widespread in eukaryotes and prokaryotes and is involved in a range of acylation modifications. Examples of AT3-containing acetyltransferases from prokaryotes include those mediating peptidoglycan acetylation contributing to lysozyme resistance (13, 14), modification of root nodulation factors to initiate symbioses (15), and O-antigen acetylation (9, 16, 17). Despite the involvement of AT3-containing proteins in a wide range of reactions, their mechanism and structure are poorly characterised.

Among bacterial AT3 carbohydrate acetyltransferases, there are two known domain architectures; proteins consisting of an AT3 domain only (AT3-only) and an N-terminal AT3 domain linked to an extra-cytoplasmic domain, commonly an SGNH domain (AT3-SGNH fused). The SGNH domain is fused through addition of an 11th TMH and linking region. Oac (in *Shigella* spp.) is an example of an AT3-only protein that is essential for O-antigen acetylation (18) whereas OatA, the O-acetyltransferase of peptidoglycan in *Staphylococcus* spp., is an example of an AT3-SGNH fused protein (14). SGNH domains (InterPro IPR036514) are a large and diverse family of small catalytic domains of around 200 amino acids, originally characterised as a subgroup of the GDSL hydrolase family by their particular invariant residues, Ser, Gly, Asn, His - hence SGNH – which occur in four blocks of conserved sequence (19, 20). Members of this family that are active against carbohydrates are also classified as CE3 family proteins in CAZy (21). Subsequently, many more proteins have been found to belong to this diverse family and they no longer fully adhere to the original paradigm of SGNH. However, most members typically contain a catalytic triad of Ser, His, Asp and oxyanion hole residues within the four blocks of conserved sequence (22). It is not clear how the AT3 and SGNH domains function together in AT3-SGNH fused carbohydrate acetyltransferases, nor how the AT3-only proteins function independently of a linked periplasmic domain.

In *Salmonella* spp. there are two defined O-antigen acetyltransferases OafA and OafB (9, 10, 17, 23). Slauch *et al.* determined that the integral membrane protein OafA from STM (17), acetylates the 2-

hydroxyl group on the abequose moiety of the O-antigen unique to this serovar (24). This results in acquisition of the O:5 serotype (defined by the Kauffmann White Lee Minor scheme) (25, 26) which is required for production of protective antibodies against STM infection (24, 27). Multiple *Salmonella* serovars have a rhamnose moiety in the O-antigen that can be acetylated at the 2- and 3-hydroxyl groups by F2GtrC proteins (9, 10, 23). As it is clear that F2GtrC is an acetyltransferase with no functional relationship to the GtrABC glycosylating proteins, we propose to rename this and orthologous rhamnose acetyltransferases as OafB. The name reflects the protein architecture (O-antigen acetyltransferase fused B), similar to that we suggest for OafA (O-antigen acetyltransferase fused A).

In this work, using *in situ* and *in vitro* functional analysis of OafA and OafB O-antigen acetyltransferases, we address the following key questions to further our understanding of the mechanism of acetyl transport and transfer in AT3-SGNH fused acetyltransferases. (I) Are there essential residues in the membrane-bound AT3 domain that can give clues to their role in acetyl transfer? (II) Can we obtain insight into the architecture of these proteins by elucidating the structure of the SGNH domain and its N-terminal extension? (III) What is the function of the SGNH domain and can it function independently of the AT3 domain?

Results

In silico analysis identifies conserved features in the integral membrane domains of bacterial AT3 acetyltransferases

The STM O-antigen acetyltransferases OafA (17) and OafB (23) (formerly F2GtrC) are both predicted by InterPro to contain an N-terminal AT3 domain (IPR002656, PF01757) fused to an SGNH domain (IPR013830, PF14606 or PF13472) (28, 29) (Fig. 1*A*). The AT3 domain has 10 TMH and an additional

11th helix that is presumably required to localise the fused SGNH domain in the periplasm (Fig. 1*A*) (30); this prediction is supported by experimental topology analysis of OafB (9) and consistent with topology analysis of Oac (31) , a comparison enabled by our detailed alignments (see below). Reinforcing the widespread functions of these understudied proteins in bacteria, we identified in the literature 30 bacterial AT3 domain-containing proteins, with experimentally confirmed carbohydrate acetyltransferase activity (9, 14–17, 32–55). Of these 30 proteins, 19 contain just the AT3 domain, while 11, including OafA and OafB, have the fused AT3-SGNH architecture (*SI Appendix,* Table S1). Previous work showed that in OafA and OafB, the SGNH domain is essential for acetyltransferase activity (9, 56) and thus, we propose the following working model for the mechanism of action (Fig. 1*B*). In AT3-SGNH proteins, the AT3 domain passes an acetyl group from an unidentified donor in the cytoplasm to the periplasmic face of the inner membrane. This acetyl group is then transferred to the SGNH domain, which catalyses specific carbohydrate O-acetylation (Fig. 1*B*). To test this model, we first determined whether residues conserved between AT3-only and AT3-SGNH acetyltransferases are important for acetyltransferase activity.

Alignments of the 30 characterised AT3 acetyltransferases along with a *S. enterica* serovar Paratyphi A (SPA) OafB homologue revealed that only 4 amino acids are invariant across all 31 proteins, $OafA_{H25}$, $OafA_{F41,}$ $OafA_{G46}$ and $OafA_{G202}$ (Fig. 2*A*, *SI Appendix,* Fig. S1). $OafA_{F41}$ and $OafA_{G46}$ belong to the FFXISG motif previously identified in un-fused AT3 O-antigen acetyltransferases (*SI Appendix,* Fig. S1) (31). Two conserved residues are predicted in TMH1, separated by 10 amino acids, in an R/K-X_{10}-H motif (Fig. 2*A*, *SI Appendix,* Fig. S1). A previously identified RXXR motif ($OafA_{R69,R72}$) in loop 2-3 is essential for activity in *Shigella flexneri* Oac ($Oac_{R73,\ R75}$)(57), and OafB ($OafB_{R71,\ R73}$) (9). This motif is highly (but not absolutely) conserved across the 31 analysed acetyltransferases.

We next examined features unique to the AT3 domains of AT3-SGNH fused acetyltransferases; these 11 sequences derive from diverse Gram positive and Gram negative bacteria (Fig. 2*B*, *SI Appendix,* Table S1). The most striking shared feature of AT3-SGNH fused proteins is the highly conserved GG-

F/Y-XGV-D/P/V motif located at the periplasmic side of TMH2 ($OafA_{G33\text{-}D39}$), which replaces a longer and more divergent loop region between TMH1-2 in the non-fused AT3 proteins. Further conserved residues are seen in the periplasmic loop between TMH3-4, including $OafA_{S112}$, $OafA_{N113}$ and $OafA_{Y122}$. Together these observations suggest shared key residues in both AT3-only and AT3-SGNH fused proteins and possible adaption of AT3 domains in AT3-SGNH fused acetyltransferases towards their function together with the fused SGNH domain.

Site-directed mutagenesis combined with *in situ* functional analysis of OafA identifies functional residues within the AT3 domain

To determine the functional importance of conserved residues in $OafA_{STM}$, we developed an *in situ* functional assay using a double antibody LPS immunoblot. The assay quantifies both the level of acetylated abequose (O:5) and the amount of LPS based on the O-antigen core (Fig. 3). His-tagged OafA, or mutated versions thereof, were expressed *in trans* in a strain that lacks all O-antigen modification genes including *oafA* (strain 293) (Methods, *SI Appendix,* Table S2). Levels of abequose acetylation in these strains was determined by LPS immunoblot from the signal obtained with serotype antibody and protein expression was also confirmed (Fig. 3, *SI Appendix,* Fig. S2). We validated this approach by comparing abequose acetylation by the *in trans* system, to both chromosomal His-tagged OafA and wild type OafA. This showed that despite a higher level of protein in the *in trans* system (Fig. 3*B*), a comparable level of abequose acetylation was obtained in all strains (Fig. 3*A* and *SI Appendix,* Fig. S2).

Twenty positions in the membrane bound domain of OafA were individually engineered to replace the wild-type amino acid with alanine. The level of O-antigen acetylation *in situ* as a result of mutant protein expression is summarized in Table 1 and Fig. 4, and data are shown in *SI Appendix,* Fig. S2. Point mutants that gave <1% O-antigen acetylation signal in relation to wild type were considered to be inactive and those with <50% O-antigen acetylation signal were considered to have significantly

reduced activity. For all mutant proteins except G34A there was detectable full-length protein on the Western blot, sometimes in addition to degradation products (*SI Appendix,* Fig. S2). Assay validation experiments indicate that the levels of full length mutant protein is in excess of wild type levels and thus should be sufficient to confer detectable abequose O-acetylation.

The arginine and histidine residues in the R/K-X_{10}-H motif ($OafA_{R14}$ and $OafA_{H25}$) are essential for function. These residues are predicted to be on the same surface of the alpha helix with a spacing similar to the predicted distance between the 3' phosphate and the thioester bond of one coenzyme A molecule (~19 Å). Thus, we hypothesise these residues provide a potential Acetyl-CoA interaction site within the AT3 domain. The 100% conserved glycines ($OafA_{G46}$ and $OafA_{G202}$) could be replaced with alanine with no detriment. As expected, both arginines in the TMH3 RXXR motif ($OafA_{R69, R72}$) (9, 57) were essential for OafA function (Table 1, Fig. 4).

We next examined the unique aspects of the AT3 domains among the AT3-SGNH fused proteins. Of the conserved GG-F/Y-XGV-D/P/V motif and flanking residues, mutation of $OafA_{F35}$ and $OafA_{D39}$ caused significant reduction and complete loss of OafA activity, respectively. $OafA_{S112A}$ mutation also caused complete loss of OafA activity (Table 1, Fig. 4). AT3-only acetyltransferases do not contain an 11^{th} TMH but a glutamate residue after the C-terminal end of TMH10 ($OafA_{E325}$) is invariant across AT3-SGNH protein sequences; mutation of this residue ($OafA_{E325A}$) resulted in significant reduction in OafA activity (Table 1, Fig. 4). Thus, AT3-SGNH-specific conserved residues in the AT3 domain are inherently involved in the mechanism of action of OafA.

$OafB_{SPA}^{long}$ has an extended SGNH-like fold

To gain an understanding of the mechanism of OafA and OafB both domains must be analysed, thus *in vitro* analysis of the SGNH domain was conducted. Structural analysis of the SGNH domains and periplasmic linking regions of OafA and OafB were used to gain insight into the functional adaptations of an SGNH domain fused directly to an AT3 domain. We expressed and purified

residues 366 to 609 from OafA ($OafA_{STM}^{C\text{-}long}$) and residues 377 to 640 of SPA OafB from ($OafB_{SPA}^{C\text{-}long}$), which have 31% sequence identity (Fig. 1*A*). Although $OafB_{SPA}$ has not been experimentally characterised in the literature, SPA O-antigen rhamnose can be acetylated (58) and $OafB_{SPA}$ has 78% sequence identity to the experimentally characterised $OafB_{STM}$ rhamnose acetyltransferase (9).

No diffracting crystals of $OafA_{STM}^{C\text{-}long}$ could be obtained, however, crystals diffracting to 1.1 Å resolution were obtained for $OafB_{SPA}^{C\text{-}long}$, with a single molecule in the asymmetric unit. The structure could not be solved by molecular replacement using a number of known SGNH structures, but was solved using Fragon (59) with a 14 residue ideal polyalanine α-helix as the search model and refined to an R_{work}/R_{free} of 13.6/14.9% (*SI Appendix,* Table S3).

The core structure of $OafB_{SPA}^{C\text{-}long}$ resembles an SGNH domain, with an α/β/α hydrolase fold consisting of five central β-strands surrounded by six α-helices (Fig. 5*A*). Two disulfide bonds are seen in the structure (Fig. 5*A*) and were verified using mass spectrometry. The closest structural homologues to $OafB_{SPA}^{C\text{-}long}$, as identified by the DALI server, are carbohydrate esterases from *Talaromyces cellulolyticus* (5B5S) and *Clostridium thermocellum* (2VPT); each have an RMSD of 2.5 Å over 207 and 201 backbone residues, respectively.

The first clear difference between $OafB_{SPA}^{C\text{-}long}$ when compared to its closest structural homologues and the only other SGNH domain from a fused acyltransferase with a solved crystal structure, OatA-SGNH (5UFY) (60) is that the structure is significantly larger, at ~36k $Å^3$, compared to OatA-SGNH at ~23k $Å^3$, which is more similar to the size of the two most closely related structures of the carbohydrate esterases (2VPT is ~26k $Å^3$ and 5B5S is 27k $Å^3$). This additional volume in the fold is contributed by two separate non-contiguous parts of the structure, the first being helix α8, which comprises 10% of the SGNH domain volume (Fig. 5). A structure-based alignment of related SGNH domains indicated that the sequence forming this additional helix is only present in AT3-SGNH domains involved in acetylation of LPS O-antigens (Fig. 6A*, SI Appendix,* Fig. S3) and so is missing on OatA. Secondly, and most significantly the region that connects the end of TM11 and the start of the

sequence of other known SGNH domains (residues 377-421) is clearly structured and forms a long extension of the SGNH domain that we now term the SGNH extension ($SGNH_{ext}$). The $SGNH_{ext}$ interacts extensively with the SGNH domain covering 1500 $Å^2$ of the SGNH domain, including interactions with helix α8; 38 amino acids of the SGNH domain interact with 32 (of 48) residues in the extension. Removal of the most N-terminal half of the $SGNH_{ext}$ ($OafA_{STM}^{C\text{-}short}$ and $OafB_{SPA}^{C\text{-}short}$ (Fig. 1A)), results in a decrease in melting temperature of 5.7 ^{o}C in OafA and 8.9 ^{o}C in OafB suggesting that the $SGNH_{ext}$ has a stabilising effect on the SGNH domain (*SI Appendix,* Fig. S4). These observations show that $OafB_{SPA}^{C\text{-}long}$ forms an extended SGNH-like fold with an additional helix, and the periplasmic portion of the linking region is structured and interacts with the SGNH domain.

Catalytic residues of $OafB_{SPA}^{long}$ resemble a typical SGNH domain with an atypical oxyanion hole

SGNH domains are usually characterised by the presence of four blocks of sequence, containing conserved residues: block I – GDS, block II – G, block III – GxND and block V – DxxH (where x is any non-proline residue) (22). The structure-based sequence alignment was used to identify conserved residues in the SGNH domain of fused acyltransferases (Fig. 6*B, SI Appendix,* Fig. S3). The typical SGNH catalytic triad, consisting of serine (block I), aspartic acid and histidine (block V), is conserved in the sequence of both OafA and OafB. *In situ* functional analysis of catalytic triad mutants $OafA_{S412A}$ and $OafA_{H590A}$ showed almost complete loss of function, whereas $OafA_{D587A}$ showed reduced activity (Table 2, Fig. S2). This is consistent with analyses of typical catalytic triad activity in other SGNH proteins (61, 62).

While the catalytic triad is conserved in both proteins, the oxyanion hole residues, glycine (block II) and asparagine (block III), are not (Fig. 6*B*). Analysis of the structure-based alignment of the block II region (Fig. 6*B, SI Appendix,* Fig. S3) reveals the conserved glycine is replaced by an asparagine in OafB ($OafB_{N459}$). The structure of $OafB_{SPA}^{C\text{-}long}$ shows $OafB_{N459}$ to be within hydrogen bonding distance of a co-crystallised sulfate ion (Fig. 6*C*) suggesting that $OafB_{N459}$ could interact with bound substrate

and participate in oxyanion hole formation. Homology modelling of $OafA_{STM}^{C\text{-}long}$ based on the structure of $OafB_{SPA}^{C\text{-}long}$ (*SI Appendix,* Fig. S5) suggests that the $OafA_{S437}$ side chain or $OafA_{L438}$ are most likely to replace the block II glycine in the oxyanion hole. This was supported by the *in situ* abequose acetylation assay which shows $OafA_{S437A}$ has significantly reduced activity in comparison to wild type OafA (Table 2, Fig. S2) consistent with the decrease in activity seen on mutation of the oxyanion hole residues in other SGNH domains (60, 61, 63)**.**

The GxND motif (block III), where Asn is typically involved in oxyanion hole formation (20), is not evident in OafA or OafB in the structure-based alignment (Fig. 6*B*). $OafB_{SPA}^{C\text{-}long}$ contains a GTNG motif ($OafB_{G502\text{-}G505}$) close to sequence block III (Fig. 6*B*)**,** but the side chains of these residues are oriented away from the catalytic triad (Fig. 6*C*). These observations suggest that**,** although OafA and OafB display the typical catalytic triad of an SGNH domain, their oxyanion hole arrangement is atypical.

The $SGNH_{ext}$ confers acceptor specificity

The structured region that extends the OafB SGNH domain ($SGNH_{ext}$) appears to occlude the active site and results in significantly lower solvent accessible surface area (SASA) of the catalytic triad residues (40 Å) than in OatA, 2VPT and 5B5S (132 Å, 110 Å and 126 Å, respectively) (Fig. 5*C*). Removing the 22 most N-terminal residues from the structure of $OafB_{SPA}^{C\text{-}long}$ ($OafB_{SPA}^{C\text{-}short}$, Fig. 1*A*) increases the SASA of the catalytic triad residues of OafB to 107.9 Å.

To assess the potential consequences of an occluded active site for substrate specificity, assays were carried out for OafA and OafB containing the full $SGNH_{ext}$ ($OafA_{STM}^{C\text{-}long}$ and $OafB_{SPA}^{C\text{-}long}$) and those with half the $SGNH_{ext}$ ($OafA_{STM}^{C\text{-}short}$ and $OafB_{SPA}^{C\text{-}short}$) (Fig. 1*A*). *In vitro* catalytic activity was first confirmed for all constructs via their ability to hydrolyse the ester substrate p-nitrophenyl acetate (pNP-Ac) (*SI Appendix,* Fig. S6), an assay commonly used to test SGNH domain function (64, 65). This

activity suggests that all four proteins are correctly folded and catalytically active regardless of the presence or absence of the $SGNH_{ext}$ residues covering the active site (*SI Appendix,* Fig. S6).

To assess whether $SGNH_{ext}$ affects the *in vitro* acceptor substrate specificity of $OafA_{STM}^{C\text{-}term}$ and $OafB_{SPA}^{C\text{-}term}$ proteins, purified proteins were incubated with pNP-Ac (acetyl group donor) and unmodified STM LPS (Path993, *SI Appendix,* Table S2) as the acceptor substrate, and O:5 antibodies were used to probe for O-antigen abequose acetylation. Abequose is the native acceptor sugar for OafA whereas OafB acetylates rhamnose *in situ*. A positive signal for O:5 antibody binding is gained after incubation with $OafA_{STM}^{C\text{-}long}$ and $OafA_{STM}^{C\text{-}short}$ (Fig. 7). Thus, $OafA_{STM}^{C\text{-}long}$ and $OafA_{STM}^{C\text{-}short}$ are able to acetylate their native substrate in solution. In contrast, acetylation of the non-native acceptor substrate by OafB occurs only in the absence of the OafB $SGNH_{ext}$ ($OafB_{SPA}^{C\text{-}short}$) (Fig. 7). Firstly, these results support our working model that the SGNH domain performs the last step in the transferase reaction; the transfer of the acetyl moiety to the acceptor carbohydrate. Furthermore, these results strongly indicate that the acceptor substrate specificity of this SGNH domain is constrained by the cognate, structured $SGNH_{ext}$.

Evolutionary support for an interaction between the AT3 domain and the SGNH domain

The discovery that the 'linker' region that is present between the more clearly defined AT3 and SGNH domains is in fact a long structured component of the SGNH domain, means that the SGNH is much more constrained and proximal to the membrane that initially proposed if this region was a long flexible linker. The discovery that there are residues in the AT3 loop between TMH3-4 that are only conserved in the AT3-SGNH fused proteins, suggests potential protein-protein contacts between the two domains during catalysis. To test this hypothesis we used a co-evolution analysis of the OafA-B type acetyltransferases to assess whether there was any evidence for correlated changes in the two domains consistent with a physiological interaction (Fig. S7a). While there are many correlated changes within the two separate domains, a significant correlated changes was observed

between residues 95 and 97, located in the periplasmic loop between TMH3-4 of the AT3 domain (Fig. 8) and residues 542 and 545-546, which form a surface-accessible patch (Fig. S7b) on the additional helix (α8) of the SGNH domain (Fig. 8). This predicted interaction further informs our refined topological model of these AT3-SGNH acetyltransferases (Fig. 8).

Discussion

AT3 domain-containing proteins (PF01757) are a ubiquitous family of proteins involved in diverse carbohydrate modifications across the domains of life. Prokaryotic members of this family play roles in modification of antibiotics and antitumor drugs, as well as initiation of microbial symbioses with plants (15, 66, 67)(*SI Appendix,* Table S1). In bacterial pathogens, such as *Salmonella enterica*, *Listeria monocytogenes*, *Haemophilus influenzae* and *Streptococcus pneumonia*, these proteins are implicated in acetylation of extra-cytoplasmic polysaccharides which can have significance for interactions with phage and hosts and can affect virulence and antibiotic resistance (24, 32, 36, 38). The current experimentally characterised AT3 domain-containing carbohydrate O-acetyltransferases display SGNH-fused or AT3-only domain architecture. Although both AT3 and SGNH domains display broad substrate ranges in diverse biological systems, the mechanism of action of both SGNH-fused and AT3-only acetyltransferases is largely unknown.

Previous understanding of AT3-SGNH fused acetyltransferases was obtained by *in situ* functional assays, and structure-function assessment of the SGNH domain (9, 17, 60). Here, expanded bioinformatic analysis with a set of 30 experimentally characterised bacterial AT3 acetyltransferases, including AT3-only and AT3-SGNH fused protein sequences, which perform a range of biological functions (*SI Appendix,* Table S1), revealed commonalities and key differences. For example, an R/K-X_{10}-H motif in TMH1 is shared across all the bacterial AT3 acetyltransferases studied (Fig. 2) and is also highly conserved across all AT3 domain proteins in the Pfam database (29), strongly suggesting that these are critical catalytic residues relevant to the whole protein family.

313 OafA$_{R14}$ and OafA$_{H25}$ within this motif were essential for activity (Table 1, Fig. 4) and are predicted to
314 be at opposite ends, but on the same surface, of the TMH1 helix (arginine towards the cytoplasmic
315 side) providing a potential interaction site for the proposed acetyl group donor acetyl-CoA. Although
316 cytoplasmic acetyl-CoA has not been confirmed as the donor for O-antigen acetylation, it occupies a
317 central role in bacterial metabolism and is a prominent source of acetate in bacterial cells (68, 69).
318 Arginine residues have previously been implicated in binding of the 3' phosphate of acetyl-CoA in
319 other acetyltransferase proteins (70) and conserved histidine residues in the soluble mitochondrial
320 carnitine O-acyltransferase co-ordinate the thioester bond of acyl-CoA with the carnitine acceptor to
321 catalyse the acyl-transfer reaction (71). Significantly the equivalent residue was discovered as a
322 natural histidine to tyrosine point mutation that decreased function of the *Streptococcus pneumonia*
323 capsule acetylation protein WcjE in clinical isolates (72).

324 A similar role for a conserved intermembrane histidine residue has also been suggested for
325 membrane bound O-acyltransferases containing an MBOAT (IPR004299) rather than AT3 domain
326 (73). These observations support a role of the R/K-X$_{10}$H motif in coordinating a cytoplasmic derived
327 acetyl-CoA molecule within the membrane bound AT3 domain for transfer of the acetyl group to the
328 SGNH domain, consistent with our model (Fig. 1). AT3 domain-containing proteins are implicated in
329 transferring a wide range of acyl groups such as succinate, isovalerate, and propionate (67, 74, 75);
330 these can all be carried by Coenzyme-A. The proposed mechanism of acetyl donor interaction would
331 provide a potential conserved mechanism for transfer of any of these acyl substituents, supporting
332 the idea that the TMH1 arginine and histidine are fundamentally important for the mechanism of all
333 AT3 domain-containing acyltransferases.

334 Residues specifically conserved in the AT3 domains of AT3-SGNH fused proteins (OafA$_{F35}$ and OafA$_{D39}$
335 in TMH2 and OafA$_{S112}$ between TMH3-4) are located towards the periplasmic side of the AT3 domain
336 (Fig. 2); we suggest these are likely to be important for interaction with the O-antigen substrate or
337 SGNH domain for acetyl group transfer. In contrast to the essential nature of OafA$_{S112}$ in the

periplasmic loop between TMH3-4, no functional residues have been identified in the equivalent region of *S. flexneri* Oac (an AT3-only O-antigen acetyltransferase) (57). Conversely, the invariant glycine residue $OafA_{G46}$, which was critical in *S. flexneri* Oac (Oac_{G53}) (*SI Appendix,* Figure S1) (31, 57), could be replaced by alanine without affecting the function of OafA. These observations imply a divergence between AT3-only and AT3-SGNH fused proteins. The location of critical residues specific to the AT3-SGNH fused proteins, further suggest that this divergence occurs at the point of acetyl group transfer to the acceptor substrate.

This study demonstrates that the SGNH domain of OafA is able to acetylate the abequose of the O-antigen of *Salmonella in vitro* without the presence of its cognate fused AT3 domain. This supports the predicted role for SGNH in the final step of acetyl group transfer to the acceptor substrate in fused acetyltransferases (Fig. 1*B*). In agreement with this, in the two component PatA/PatB peptidoglycan acetyltransferase system, PatB, a soluble SGNH protein, is responsible for transfer of the acetyl group onto the peptidoglycan substrate (62). Moynihan and Clarke *et al.* hypothesised that PatA (an MBOAT protein not an AT3) is responsible for transporting the acetyl group across the membrane where it is transferred to the acceptor by the soluble PatB protein (62). The membrane bound PatA MBOAT protein in this system is interchangeable with WecH, an AT3-only acetyltransferase protein (52, 76), giving an example of direct transfer of acetate between a membrane bound AT3 domain and soluble SGNH domain protein. This supports the mechanistic model of the AT3 domain delivering the acetyl group to the SGNH domain for transfer onto the acceptor substrate in AT3-SGNH fused proteins (Fig. 1*B*).

Our data demonstrated, for the first time in a fused system, the necessity for the fused SGNH domain in glycan carbohydrate acetylation. However, this poses the conundrum that other closely related systems, such as OacA from *Shigella* that O-acetylates rhamnose in the O-antigen (57), lack either a fused or genetically linked partner SGNH domain. Consequently, either the AT3 domain functions differently, or there is a currently undiscovered partner protein.

363 This study elucidates the structure of the $SGNH_{ext}$ in $OafB_{SPA}^{C\text{-}long}$ and shows that removal of this
364 region results in promiscuity of carbohydrate modification in *in vitro* acetyltransferase reactions (Fig.
365 7). These findings suggest that the $SGNH_{ext}$ plays a role in determining the specificity of the O-
366 antigen residue to be acetylated. Closer examination of the structure reveals that two tyrosines,
367 Tyr289 and Tyr394, in the $SGNH_{ext}$ sit closely to the active site and could potentially be involved in a
368 mechanism to limit off-target acetylation. Inadvertent acetylation of complex carbohydrates could
369 potentially have diverse and undesired biological effects due to the variation of cellular processes
370 that can be affected by acetylation (9, 33, 39, 77–79). Whether this also implies that AT3 proteins all
371 need a partner domain or protein for substrate specific transferase activity remains to be
372 determined.

373 Co-evolution analysis predicts interaction between periplasmic loops of the AT3 domain and the
374 SGNH domain of OafB. This is similar to the arrangement of domains seen in PglB, an oligosaccharide
375 transferase from *Campylobacter lari* (80), with 13 TMH and a periplasmic domain. In PglB the
376 periplasmic domain interacts via periplasmic loops in the transmembrane domain and both domains
377 are hypothesised to interact with the peptide substrate (80). In our model, the co-evolution analysis
378 positions the periplasmic loops of the AT3 domain close to α8 helix in the SGNH domain allowing for
379 a interaction with each other and with the acceptor substrate (Fig. 8).

380 AT3 domain-containing proteins are involved in the modification of a wide range of polysaccharides
381 and influence many host-pathogen interactions. These structural and functional insights can be
382 applied to the well-studied and biotechnologically relevant AT3 proteins, including Nod factor
383 modifications important for plant microbe symbiosis, and anti-tumour and antibiotic modifying
384 proteins. Furthermore, this work can inform future studies in eukaryotic systems where AT3 domain-
385 containing proteins are involved in regulation of the lifespan of *Caenorhabditis elegans* (81) and in
386 *Drosophila* development (82).

387

Methods

Bacterial strains, plasmids and culture conditions:

Escherichia coli and STM strains and plasmids are listed in *SI Appendix,* Table S2. Strains were cultured in Lennox broth (LB; Fisher Scientific) at 37 °C with appropriate antibiotic selection unless otherwise stated.

In silico analysis of bacterial AT3 domains to identify conserved residues

A survey of the literature identified 30 experimentally-characterised bacterial carbohydrate acetyltransferases, these sequences were aligned along with OafB from *Salmonella* ser. Paratyphi A, using TCoffee (83). Protein accession numbers are in Fig. S1. TCoffee was also used to align $OafA_{STM}$, $OafB_{STM}$ and $OafB_{SPA}$ protein sequences for direct comparison.

Structure based sequence alignments using PROMALS-3D with default settings were carried out with the two closest structural homologues identified using the DALI server, and a selection of typical SGNH domains for which structural information is available: $OafB_{SPA}$, 1IVN, 4K40, 1DEX, 5UFY, 5B5S and 2VPT. Five further representative sequences of OafA, OafB, and OatA were included (A0A0H2WM30, STMMW_03911, Q8ZNJ3, NTHI0512, Q2FV54).

Co-evolution analysis

A multiple sequence alignment of AT3 SGNH domain fused proteins was constructed using the MUSCLE alignment tool based on 1,188 full length sequences from the UniProt Reference Proteomes. This alignment was used to construct a profile-HMM to detect further homologues in the UniProt Reference Proteome set as well as within the MGnify protein sequence set. We required that all matches to this profile-HMM had a sequence and domain threshold of 27 bits. We also

required that the sequence matched > 90% of the HMM match states to ensure that homologues with only one of the two domains were not included in the alignment.

2,713 homologues were identified from the UniProt Reference Proteome set and 9,757 homologues were identified from the MGnify metagenomics sequences. A large sequence alignment was constructed using OafB as the master with no indels with all the sequence matches aligned to it using the hmmalign package and a custom Perl script to format the alignment for contact prediction. The alignment was submitted to the RaptorX contact prediction server (84).

Molecular Biology

Primers (Sigma-Aldrich) are listed in *SI Appendix,* Table S2. Mutations were introduced into the OafA sequence (pMV433 as template), which had been cloned into pBADcLIC using blunt end ligation, placing the gene under control of an arabinose inducible promoter. Mutants were confirmed by sequencing. Plasmids were electroporated into STM strain 293 (*SI Appendix,* Table S2) for analysis of activity.

All $oafA_{STM}$ and $oafB_{SPA}$ sequences for protein expression were cloned into pETFPP_2 (Technology Facility, University of York) using in-fusion cloning (Clontech) to add a 3C-protease cleavable N-terminal His-MBP tag. Plasmid pMV433 (*SI Appendix,* Table S2) was used as the template for creation of expression plasmids encoding the protein sequence for $OafA_{STM}^{C\text{-}long}$ (residues 366-609) and $OafA_{STM}^{C\text{-}short}$ (residues 379-609). $oafB_{SPA}$ (A0A0H2WM30) amino acid residues 377 to 640 for $OafB_{SPA}^{C\text{-}long}$, was codon-optimised for *E. coli* and synthesised by Genewiz in a pUC57-Kan vector. This vector was then used as a template for the sequence encoding $OafB_{SPA}^{C\text{-}short}$ (residues 399-640); see *SI Appendix,* Table S2 for primers used.

In situ functional analysis of OafA variants

All *in situ* functional analyses of OafA variants cloned into pBADcLIC were carried out in strain Path293 (23) (*SI Appendix,* Table S2). Strains for the *in situ* functional analysis were cultured at pH 7.0 in 100 mM sodium phosphate-buffered LB at 37 °C in a baffled conical flask with shaking at 200 rpm. Overnight cultures were diluted 100-fold and grown for 16 hr. Samples were normalised to (OD_{600}) of 3.0 per ml for LPS and protein extraction.

Crude LPS sample preparation

The method was adapted from Davies *et al.* 2013 (23). 1 ml of OD-normalised (OD_{600} 3.0) overnight culture was pelleted for 5 min at 16,000xg. Cell pellets were re-suspended in 100 µl LPS sample buffer (60 mM Tris-HCL, 1mM EDTA, pH 6.8) containing 2% (w/v) SDS then boiled at 100 °C for 5 min. 400 µl of LPS buffer was then used to dilute the solution before RNAse (Roche) and DNAse (Sigma) treatment at 37 ˚C for 16 hours. Samples were then treated with 100 µg proteinase K for 16 hours at 50 ˚C. 7.5 µl of crude LPS extracts were run on 1.0 mm Tricine SDS - Poly Acrylamide Gel Electrophoresis TSDS-PAGE gel for analysis by immunoblotting.

Detection of OafA protein expression for *in situ* assays

1ml of OD-normalised culture was pelleted for 5 min at 16,000xg. Soluble and insoluble fractions were isolated from cell pellets using Bug Buster™ solution (Novagen) following manufacturer's instructions for soluble protein extraction. The insoluble pellet was resuspended in 75 µl of sample buffer (10% (v/v) Glycerol, 1% (w/v) SDS, 10mM Tris-HCL, pH 7.2, 0.06% (w/v) Bromophenol Blue, 3% (v/v) β-mercaptoethanol), heated to 60˚C for 10 min and centrifuged for 10 min at 16,000xg. 10 µl of insoluble fraction samples were loaded onto a 12% acrylamide 1.0 mm SDS-PAGE gel for analysis.

Immunoblotting

7.5 µl of crude LPS extracts were run on 1.0 mm Tricine SDS - Polyacrylamide Gel Electrophoresis (TSDS-PAGE) gel for analysis by immunoblotting. The TSDS-PAGE-separated LPS samples and SDS-

PAGE separated protein samples were transferred onto Immobilon-P PVDF membrane (Merck-Millipore). For His-tagged protein detection, the primary antibody was Tetra·His Antibody (1:1000) (Qiagen; in 3% (w/v) BSA TBS) and the secondary antibody was goat anti-mouse IgG-HRP (1:10000) (Sigma-Aldrich; in 5% (w/v) Milk TBS). The blot was developed using Luminata Classico Western HRP substrate (Merck-Millipore). For LPS detection O:5 serotyping antibody (1:10000) (Statens Serum Institute; 40272) and *Salmonella* core antigen (1:200) (Insight Biotechnology; 5D12A) were used as the primary antibodies and Goat Anti-Rabbit IgG StarBright Blue700 (1:5000) (Bio-Rad) and Goat anti-mouse IgG (H+L) DyLight 800 (1:5000) as the respective secondary antibodies. LPS antibodies were diluted in 5% Milk PBS-T. ChemiDoc MP Imaging System (Bio-Rad) and Image Lab™ (Bio-Rad) were used for image capture and analysis. The *in situ* activity of OafA mutant relative to wild type protein was derived from quantification of the O:5 signal in each lane, standardised to the intensity of the single O-antigen repeat band for the *Salmonella* core signal on LPS immunoblots. Assay validation demonstrated that <1% O:5 signal with respect to wild type was within the background variation. Variation increased significantly for signal intensities in the higher range, therefore O:5 signal recoded between 50 and 100% relative to wild type was not interpreted further.

Expression and purification of $OafA_{STM}^{C\text{-}term}$ and $OafB_{SPA}^{C\text{-}term}$

pETFPP_2 vectors containing the inserted $OafA_{STM}^{C\text{-}term}$ and $OafB_{SPA}^{C\text{-}term}$ constructs (Fig. 1) were transformed into Origami (Novagen) *E. coli* for protein expression. Protein expression was carried out as described by Gruszka et al. 2015 (85) without the addition of protease inhibitor. The proteins were purified using immobilised metal affinity chromatography with a HisTrap FF column (GE Healthcare) utilising a His-tag, followed by size exclusion chromatography after His-tag removal, as described by Gruszka et al. 2012 (86); purified protein was eluted in 20 mM Tris-HCl pH 7.5, 100 mM NaCl.

Melting temperature of OafA and OafB SGNH domains

The melting temperature of SGNH domains was determined using NanoDSF with a protein concentration of 1 mg/mL in 20 mM TrisHCl pH 7.5, 100 mM NaCl. Proteins were heated from 20 ^{o}C to 95 ^{o}C with a heating rate of 2 ^{o}C / min. The fluorescence at 330 and 350 nm was measured every 0.05 ^{o}C.

In vitro acetylesterase activity assay

The catalytic activity of OafA and OafB C-terminal constructs was confirmed by acetyl esterase activity using pNP-Ac as a substrate. 100 µl of enzyme solution (10 µM $OafA_{STM}^{C\text{-}term}$, 40 µM $OafB_{STM}^{C\text{-}term}$ or 0.04 U/ml Acetylxylanesterase) or appropriate control buffers were added to relevant wells of a 96 well plate and incubated at 37 ˚C for 10 min prior to addition of pNP-Ac. 100 µl of 1 mM pNPA in the corresponding buffer was then added to matching sample and control wells and immediately placed into a plate reader incubated at 37 ˚C. Absorbance at 405 nm was measured at T=0, and then at 5 min intervals.

In vitro abequose acetyltransferase activity assay

Crude LPS extracted from OafA-negative STM LT2 strain (Path993) was heated at 100°C for 20 min to inactivate the proteinase K (see above). Heat-treated LPS was mixed 1:1 with KPi buffer (200 mM NaCl, 50 mM Potassium Phosphate buffer pH 7.8). 10 µM $OafA_{STM}^{C\text{-}term}$ and 20 µM $OafB_{SPA}^{C\text{-}term}$ were incubated at 4 °C in LPS-KPi mixture with 4 mM pNP-Ac dissolved in ethanol (4% (v/v) final concentration in reaction). Samples of the reaction mix were taken after specified time points and inactivated by boiling for 10 min.

5 µl of LPS reaction samples were loaded onto methanol-activated PVDF membrane using a BioRad Bio-Dot® microfiltration apparatus. The protocol for LPS detection with O:5 serotyping antibodies

and *Salmonella* core antigen was followed as per immunoblotting, following removal of the membrane from the apparatus after sample loading.

Protein structure analysis

To crystallise $OafB_{SPA}^{C\text{-}long}$, a hanging-drop vapour diffusion method was used with 20 mg/mL $OafB_{SPA}^{C\text{-}long}$ in a drop ratio of 1:1 protein:reservoir solution. After incubation for 24 hours at 20^{o}C crystals grown in 100 mM BisTris pH 5.5, 0.25 M lithium sulfate, 25% PEG 3350 were cryoprotected by addition of glycerol to a final concentration of 20% and vitrified in liquid nitrogen.

X-ray diffraction data for crystals of $OafB_{SPA}^{C\text{-}long}$ were collected on beamline I04-1 (Diamond Light Source, UK) at a wavelength of 0.9282 Å using a Pilatus 6M-F detector. Data were integrated with XDS (87), and scaled and merged with AIMLESS (88) via the Xia2 pipeline (89). Fragon molecular replacement (59) used Phaser (90) to place an ideal poly-alanine helix of 14 amino acids in length followed by density modification with ACORN (91). ARP-wARP (92) was used for automated chain tracing, and the model was refined using REFMAC5 (93–98). Manual manipulation of the model between refinement cycles was performed using Coot (99, 100). The final model was evaluated using MolProbity (101) and PDB validate, secondary structure shown in Fig. 5A was annotated using STRIDE (102).

A homology model of $OafA_{STM}^{C\text{-}long}$ was produced using SwissModel with the structure of $OafB_{SPA}^{C\text{-}long}$ as a template (103–107).

Data Availability

The atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB ID code 6SE1).

Acknowledgments

CP and ST were supported by a PhD studentship from the Biotechnology and Biological Sciences Research Council White Rose Doctoral Training Programme (BB/M011151/1): Mechanistic Biology and its Strategic Application. The authors thank the support of the University of York Technology Facility, Steinar Mannsverk for technical assistance and Jean Whittingham for crystallography support. The authors thank Diamond Light Source for access to beamline I04-1 (under proposal DLS-MX-13587).

References

1. Majowicz SE, Musto J, Scallan E, Angulo FJ, Kirk M, O'Brien SJ, Jones TF, Fazil A, Hoekstra RM. 2010. The Global Burden of Nontyphoidal Salmonella Gastroenteritis. Clin Infect Dis 50:882–889.

2. Hiyoshi H, Tiffany CR, Bronner DN, Bäumler AJ. 2018. Typhoidal Salmonella serovars: ecological opportunity and the evolution of a new pathovar. FEMS Microbiol Rev. 42:527-541.

3. Reddy EA, Shaw A V, Crump JA. 2010. Community-acquired bloodstream infections in Africa: a systematic review and meta-analysis. Lancet Infect Dis 10:417–32.

4. Crump JA, Luby SP, Mintz ED. 2004. The global burden of typhoid fever. Bull World Health Organ 82:346–353.

5. Lynch MF, Blanton EM, Bulens S, Polyak C, Vojdani J, Stevenson J, Medalla F, Barzilay E, Joyce K, Barrett T, Mintz ED. 2009. Typhoid Fever in the United States, 1999-2006. JAMA 302:859.

6. Boore AL, Hoekstra RM, Iwamoto M, Fields PI, Bishop RD, Swerdlow DL. 2015. Salmonella enterica Infections in the United States and Assessment of Coefficients of Variation: A Novel Approach to Identify Epidemiologic Characteristics of Individual Serotypes, 1996-2011. PLoS One 10:e0145416.

7. Liu B, Knirel YA, Feng L, Perepelov A V., Senchenkova SN, Reeves PR, Wang L. 2014. Structural diversity in Salmonella O antigens and its genetic basis. FEMS Microbiol Rev 38:56–89.

8. Raetz C, Whitfield C. 2002. Lipopolysaccharide endotoxins. Annu Rev Biochem 71:635–700.

9. Kintz E, Davies MR, Hammarlöf DL, Canals R, Hinton JCDD, van der Woude MW. 2015. A BTP1 prophage gene present in invasive non-typhoidal S almonella determines composition and length of the O-antigen of the lipopolysaccharide. Mol Microbiol 96:263–275.

10. Kintz E, Heiss C, Black I, Donohue N, Brown N, Davies MR, Azadi P, Baker S, Kaye PM, van der Woude M. 2017. Salmonella Typhi Lipopolysaccharide O-antigen Modifications Impact on Serum Resistance and Antibody Recognition. Infect Immun 85:e01021-16.

11. Kim ML, Slauch JM. 1999. Effect of acetylation (O-factor 5) on the polyclonal antibody response to Salmonella typhimurium O-antigen. FEMS Immunol Med Microbiol 26:83–92.

12. Fierer J, Guiney DG. 2001. Diverse virulence traits underlying different clinical outcomes of

555 Salmonella infection. J Clin Invest 107:775–80.

556 13. Moynihan PJ, Clarke AJ. 2011. O-Acetylated peptidoglycan: Controlling the activity of
557 bacterial autolysins and lytic enzymes of innate immune systems. Int J Biochem Cell Biol
558 43:1655–1659.

559 14. Bera A, Herbert S, Jakob A, Vollmer W, Götz F. 2005. Why are pathogenic staphylococci so
560 lysozyme resistant? The peptidoglycan O-acetyltransferase OatA is the major determinant for
561 lysozyme resistance of Staphylococcus aureus. Mol Microbiol 55:778–787.

562 15. Davis EO, Evans IJ, Johnston AWB. 1988. Identification of nodX, a gene that allows Rhizobium
563 leguminosarum biovar viciae strain TOM to nodulate Afghanistan peas. MGG Mol Gen Genet
564 212:531–535.

565 16. Verma NK, Brandt JM, Verma DJ, Lindberg AA. 1991. Molecular characterization of the O-
566 acetyl transferase gene of converting bacteriophage SF6 that adds group antigen 6 to Shigella
567 flexneri. Mol Microbiol 5:71–75.

568 17. Slauch JM, Lee AA, Mahan MJ, Mekalanos JJ. 1996. Molecular characterization of the oafA
569 locus responsible for acetylation of Salmonella typhimurium O-antigen: OafA is a member of
570 a family of integral membrane trans-acylases. J Bacteriol 178:5904–5909.

571 18. Clark CA, Beltrame J, Manning PA. 1991. The oac gene encoding a lipopolysaccharide O-
572 antigen acetylase maps adjacent to the integrase-encoding gene on the genome of Shigella
573 flexneri bacteriophage Sf6. Gene 107:43–52.

574 19. Dalrymple BP, Cybinski DH, Layton I, McSweeney CS, Xue GP, Swadling YJ, Lowry JB. 1997.
575 Three Neocallimastix patriciarum esterases associated with the degradation of complex
576 polysaccharides are members of a new family of hydrolases. Microbiology 143:2605–2614.

577 20. Mølgaard A, Kauppinen S, Larsen S. 2000. Rhamnogalacturonan acetylesterase elucidates the
578 structure and function of a new family of hydrolases. Structure 8:373–383.

579 21. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The
580 carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490-495.

581 22. Akoh CC, Lee G-C, Liaw Y-C, Huang T-H, Shaw J-F. 2004. GDSL family of serine
582 esterases/lipases. Prog Lipid Res 43:534–52.

583 23. Davies MR, Broadbent SE, Harris SR, Thomson NR, van der Woude MW. 2013. Horizontally

584 Acquired Glycosyltransferase Operons Drive Salmonellae Lipopolysaccharide Diversity. PLoS
585 Genet 9:e1003568.

586 24. Slauch JM, Mahan MJ, Michetti P, Neutra MR, Mekalanos JJ. 1995. Acetylation (O-factor 5)
587 affects the structural and immunological properties of Salmonella typhimurium
588 lipopolysaccharide O antigen. Infect Immun 63:437–41.

589 25. Grimont P, Weill F-X. 2008. Antigenic formulae of the Salmonella servovars. WHO Collab
590 Cent Ref Res Salmonella 1–167.

591 26. Issenhuth-Jeanjean S, Roggentin P, Mikoleit M, Guibourdenche M, de Pinna E, Nair S, Fields
592 PI, Weill FX. 2014. Supplement 2008-2010 (no. 48) to the White-Kauffmann-Le Minor scheme.
593 Res Microbiol 165:526–530.

594 27. Lanzilao L, Stefanetti G, Saul A, MacLennan CA, Micoli F, Rondini S. 2015. Strain selection for
595 generation of O-antigen-based glycoconjugate vaccines against invasive nontyphoidal
596 Salmonella disease. PLoS One 10:e0139847.

597 28. Mitchell A, Chang H-Y, Daugherty L, Fraser M, Hunter S, Lopez R, McAnulla C, McMenamin C,
598 Nuka G, Pesseat S, Sangrador-Vegas A, Scheremetjew M, Rato C, Yong S-Y, Bateman A, Punta
599 M, Attwood TK, Sigrist CJA, Redaschi N, Rivoire C, Xenarios I, Kahn D, Guyot D, Bork P, Letunic
600 I, Gough J, Oates M, Haft D, Huang H, Natale DA, Wu CH, Orengo C, Sillitoe I, Mi H, Thomas
601 PD, Finn RD. 2015. The InterPro protein families database: the classification resource after 15
602 years. Nucleic Acids Res 43:D213-21.

603 29. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K,
604 Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M. 2014. Pfam: the protein families
605 database. Nucleic Acids Res 42:D222-30.

606 30. Krogh A, Larsson B, Von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein
607 topology with a hidden Markov model: Application to complete genomes. J Mol Biol
608 305:567–580.

609 31. Thanweer F, Verma NK. 2012. Identification of critical residues of the serotype modifying O-
610 acetyltransferase of Shigella flexneri. BMC Biochem 13:13.

611 32. Crisóstomo MI, Vollmer W, Kharat AS, Inhülsen S, Gehre F, Buckenmaier S, Tomasz A. 2006.
612 Attenuation of penicillin resistance in a peptidoglycan O-acetyl transferase mutant of
613 Streptococcus pneumoniae. Mol Microbiol 61:1497–1509.

614 33. Laaberki MH, Pfeffer J, Clarke AJ, Dworkin J. 2011. O-acetylation of peptidoglycan is required
615 for proper cell separation and S-layer anchoring in Bacillus anthracis. J Biol Chem 286:5278–
616 5288.

617 34. Buendia AM, Enenkel B, Köplin R, Niehaus K, Arnold W, Pünier A. 1991. The Rhizobium
618 meliloti exoZl exoB fragment of megaplasmid 2: ExoB functions as a UDP-glucose 4-epimerase
619 and ExoZ shows homology to NodX of Rhizobium leguminosarum biovar viciae strain TOM.
620 Mol Microbiol. 5:1519-1530.

621 35. Katzen F, Ferreiro DU, Oddo CG, Ielmini M V, Becker A, Pühler A, Ielpi L. 1998. Xanthomonas
622 campestris pv. campestris gum mutants: effects on xanthan biosynthesis and plant virulence.
623 J Bacteriol 180:1607–17.

624 36. Fox KL, Yildirim HH, Deadman ME, Schweda EKH, Moxon ER, Hood DW. 2005. Novel
625 lipopolysaccharide biosynthetic genes containing tetranucleotide repeats in Haemophilus
626 influenzae, identification of a gene for adding O-acetyl groups. Mol Microbiol 58:207–216.

627 37. Zou CH, Knirel YA, Helbig JH, Zähringer U, Mintz CS. 1999. Molecular cloning and
628 characterization of a locus responsible for O acetylation of the O polysaccharide of Legionella
629 pneumophila serogroup 1 lipopolysaccharide. J Bacteriol 181:4137–41.

630 38. Aubry C, Goulard C, Nahori MA, Cayet N, Decalf J, Sachse M, Boneca IG, Cossart P, Dussurget
631 O. 2011. OatA, a peptidoglycan O-acetyltransferase involved in Listeria monocytogenes
632 immune escape, is critical for virulence. J Infect Dis 204:731–740.

633 39. Kahler CM, Lyons-Schindler S, Choudhury B, Glushka J, Carlson RW, Stephens DS. 2006. O-
634 acetylation of the terminal N-acetylglucosamine of the lipooligosaccharide inner core in
635 Neisseria meningitidis: Influence on inner core structure and assembly. J Biol Chem
636 281:19939–19948.

637 40. Bernard E, Rolain T, Courtin P, Guillot A, Langella P, Hols P, Chapot-Chartier M-P. 2011.
638 Characterization of *O* -Acetylation of *N* -Acetylglucosamine. J Biol Chem 286:23950–23958.

639 41. Pacios Bras C, Jordá MA, Wijfjes a H, Harteveld M, Stuurman N, Thomas-Oates JE, Spaink HP.
640 2000. A Lotus japonicus nodulation system based on heterologous expression of the fucosyl
641 transferase NodZ and the acetyl transferase NoIL in Rhizobium leguminosarum. Mol Plant
642 Microbe Interact 13:475–479.

643 42. Brett PJ, Burtnick MN, Heiss C, Azadi P, DeShazer D, Woods DE, Gherardini FC. 2011.

644 Burkholderia thailandensis oacA mutants facilitate the expression of Burkholderia mallei-like
645 O polysaccharides. Infect Immun 79:961–9.

646 43. Wang J, Knirel YA, Lan R, Senchenkova SN, Luo X, Perepelov A V, Wang Y, Shashkov AS, Xu J,
647 Sun Q. 2014. Identification of an O-acyltransferase gene (oacB) that mediates 3- and 4-O-
648 acetylation of rhamnose III in Shigella flexneri O antigens. J Bacteriol 196:1525–31.

649 44. Knirel YA, Wang J, Luo X, Senchenkova SN, Lan R, Shpirt AM, Du P, Shashkov AS, Zhang N, Xu
650 J, Sun Q. 2014. Genetic and structural identification of an O-acyltransferase gene (oacC)
651 responsible for the 3/4-O-acetylation on rhamnose III in Shigella flexneri serotype 6. BMC
652 Microbiol 14:266.

653 45. Sun Q, Knirel YA, Wang J, Luo X, Senchenkova SN, Lan R, Shashkov AS, Xu J. 2014. Serotype-
654 converting bacteriophage SfII encodes an acyltransferase protein that mediates 6-O-
655 acetylation of GlcNAc in Shigella flexneri O-antigens, conferring on the host a novel O-antigen
656 epitope. J Bacteriol 196:3656–3666.

657 46. Cogez V, Gak E, Puskas A, Kaplan S, Bohin JP. 2002. The opgGIH and opgC genes of
658 Rhodobacter sphaeroides form an operon that controls backbone synthesis and succinylation
659 of osmoregulated periplasmic glucans. Eur J Biochem 269:2473–2484.

660 47. Bontemps-Gallo S, Madec E, Robbe-Masselot C, Souche E, Dondeyne J, Lacroix J-M. 2016. The
661 opgC gene is required for OPGs succinylation and is osmoregulated through RcsCDB and
662 EnvZ/OmpR in the phytopathogen Dickeya dadantii. Sci Rep 6:19619.

663 48. Hong Y, Duda KA, Cunneen MM, Holst O, Reeves PR. 2013. The WbaK acetyltransferase of
664 Salmonella enterica group E gives insights into O antigen evolution. Microbiol (United
665 Kingdom) 159:2316–2322.

666 49. Brett PJ, Burtnick M, Woods D. 2003. The wbiA locus is required for the 2-O-acetylation of
667 lipopolysaccharides expressed by Burkholderia pseudomallei and Burkholderia thailandensis.
668 FEMS Microbiol Lett. 218:323-328.

669 50. Geno KA, Saad JS, Nahm MH. 2017. Discovery of novel pneumococcal serotype 35D, a natural
670 WciG-deficient variant of serotype 35B. J Clin Microbiol 55:1416–1425.

671 51. Calix JJ, Nahm MH. 2010. A new pneumococcal serotype, 11E, has a variably inactivated wcjE
672 gene. J Infect Dis 202:29–38.

673 52. Kajimura J, Rahman A, Hsu J, Evans MR, Gardner KH, Rick PD. 2006. O Acetylation of the

674 Enterobacterial Common Antigen Polysaccharide Is Catalyzed by the Product of the yiaH
675 Gene of Escherichia coli K-12. J Bacteriol 188:7542–7550.

676 53. Veiga P, Bulbarela-Sampieri C, Furlan S, Maisons A, Chapot-Chartier M-P, Erkelenz M,
677 Mervelet P, Noirot P, Frees D, Kuipers OP, Kok J, Gruss A, Buist G, Kulakauskas S. 2007. SpxB
678 Regulates O-Acetylation-dependent Resistance of Lactococcus lactis Peptidoglycan to
679 Hydrolysis. J Biol Chem 282:19342–19354.

680 54. Menéndez N, Nur-e-Alam M, Braña AF, Rohr J, Salas JA, Méndez C. 2004. Biosynthesis of the
681 Antitumor Chromomycin A3 in Streptomyces griseus: Analysis of the Gene Cluster and
682 Rational Design of Novel Chromomycin Analogs. Chem Biol 11:21–32.

683 55. Warren MJ, Roddam LF, Power PM, Terry TD, Jennings MP. 2004. Analysis of the role of pglI
684 in pilin glycosylation of Neisseria meningitidis. FEMS Immunol Med Microbiol 41:43–50.

685 56. Hauser E, Junker E, Helmuth R, Malorny B. 2011. Different mutations in the oafA gene lead to
686 loss of O5-antigen expression in Salmonella enterica serovar Typhimurium. J Appl Microbiol
687 110:248–253.

688 57. Thanweer F, Tahiliani V, Korres H, Verma NK. 2008. Topology and identification of critical
689 residues of the O-acetyltransferase of serotype-converting bacteriophage, SF6, of Shigella
690 flexneri. Biochem Biophys Res Commun 375:581–585.

691 58. Ravenscroft N, Cescutti P, Gavini M, Stefanetti G, Maclennan CA, Martin LB, Micoli F. 2015.
692 Structural analysis of the O-acetylated O-polysaccharide isolated from Salmonella paratyphi A
693 and used for vaccine preparation. Carbohydr Res 404:108–116.

694 59. Jenkins HT. 2018. Fragon: rapid high-resolution structure determination from ideal protein
695 fragments. Acta Crystallogr Sect D Struct Biol 74:205–214.

696 60. Sychantha D, Jones CS, Little DJ, Moynihan PJ, Robinson H, Galley NF, Roper DI, Dowson CG,
697 Howell PL, Clarke AJ. 2017. In vitro characterization of the antivirulence target of Gram-
698 positive pathogens, peptidoglycan O-acetyltransferase A (OatA). PLoS Pathog 13:e1006667.

699 61. Lee L-C, Lee Y-L, Leu R-J, Shaw J-F. 2006. Functional role of catalytic triad and oxyanion hole-
700 forming residues on enzyme activity of Escherichia coli thioesterase I/protease
701 I/phospholipase L1. Biochem J 397:69–76.

702 62. Moynihan PJ, Clarke AJ. 2014. Substrate specificity and kinetic characterization of
703 peptidoglycan O-acetyltransferase B from Neisseria gonorrhoeae. J Biol Chem 289:16748–

16760.

63. Pfeffer JM, Weadge JT, Clarke AJ. 2013. Mechanism of action of neisseria gonorrhoeae O-acetylpeptidoglycan esterase, an SGNH serine esterase. J Biol Chem 288:2605–2613.

64. Baker P, Ricer T, Moynihan PJ, Kitova EN, Walvoort MTC, Little DJ, Whitney JC, Dawson K, Weadge JT, Robinson H, Ohman DE, Codée JDC, Klassen JS, Clarke AJ, Howell PL. 2014. P. aeruginosa SGNH Hydrolase-Like Proteins AlgJ and AlgX Have Similar Topology but Separate and Distinct Roles in Alginate Acetylation. PLoS Pathog 10: e1004334.

65. Moynihan PJ, Clarke AJ. 2013. Assay for peptidoglycan O-acetyltransferase: A potential new antibacterial target. Anal Biochem 439:73–79.

66. Menéndez N, Nur-e-Alam M, Braña AF, Rohr J, Salas JA, Méndez C. 2004. Tailoring modification of deoxysugars during biosynthesis of the antitumour drug chromomycin A3 by Streptomyces griseus ssp. griseus. Mol Microbiol 53:903–915.

67. Arisawa A, Kawamura N, Tsunekawa H, Okamura K, Tone H, Okamoto R. 1993. Cloning and nucleotide sequences of two genes involved in the 4''-O-acylation of macrolide antibiotics from Streptomyces thermotolerans. Biosci Biotechnol Biochem 57:2020–5.

68. Takamura Y NG. 1988. Changes in the intracellular concentration of acetyl-CoA and malonyl-CoA in relation to the carbon and energy metabolism of Escherichia coli K12. J Gen Microbiol 134:224:2249–2253.

69. Krivoruchko A, Zhang Y, Siewers V, Chen Y, Nielsen J. 2015. Microbial acetyl-CoA metabolism and metabolic engineering. Metab Eng 28:28–42.

70. Wu D, Hersh LB. 1995. Identification of an active site arginine in rat choline acetyltransferase by alanine scanning mutagenesis. J Biol Chem 270:29111–6.

71. JOGL G, HSIAO Y-S, TONG L. 2004. Structure and Function of Carnitine Acyltransferases. Ann N Y Acad Sci 1033:17–29.

72. Calix JJ, Oliver MB, Sherwood LK, Beall BW, Hollingshead SK, Nahm MH. 2011. Streptococcus pneumoniae Serotype 9A Isolates Contain Diverse Mutations to wcjE That Result in Variable Expression of Serotype 9V-specific Epitope. J Infect Dis 204:1585–1595.

73. Ma D, Wang Z, Merrikh CN, Lang KS, Lu P, Li X, Merrikh H, Rao Z, Xu W. 2018. Crystal structure of a membrane-bound O-acyltransferase. Nature 562:286-290.

733 74. Roset MS, Ciocchini AE, Ugalde RA, Inon de Iannino N. 2006. The Brucella abortus Cyclic -1,2-
734 Glucan Virulence Factor Is Substituted with O-Ester-Linked Succinyl Residues. J Bacteriol
735 188:5003–5013.

736 75. Cong L, Piepersberg W. 2007. Cloning and characterization of genes encoded in dTDP-D-
737 mycaminose biosynthetic pathway from a midecamycin-producing strain, streptomyces
738 mycarofaciens. Acta Biochim Biophys Sin (Shanghai) 39:187–193.

739 76. Moynihan PJ, Clarke AJ. 2010. O-acetylation of peptidoglycan in gram-negative bacteria:
740 Identification and characterization of peptidoglycan O-acetyltransferase in Neisseria
741 gonorrhoeae. J Biol Chem 285:13264–13273.

742 77. Bernard E, Rolain T, David B, André G, Dupres V, Dufrêne YF, Hallet B, Chapot-Chartier M-PP,
743 Hols P. 2012. Dual Role for the O-Acetyltransferase OatA in Peptidoglycan Modification and
744 Control of Cell Septation in Lactobacillus plantarum PLoS One 7:e47893.

745 78. Baranwal G, Mohammad M, Jarneborn A, Reddy BR, Golla A, Chakravarty S, Biswas L, Götz F,
746 Shankarappa S, Jin T, Biswas R. 2017. Impact of cell wall peptidoglycan O-acetylation on the
747 pathogenesis of Staphylococcus aureus in septic arthritis. Int J Med Microbiol 307:388–397.

748 79. Knirel YA, Prokhorov NS, Shashkov AS, Ovchinnikova OG, Zdorovenko EL, Liu B, Kostryukova
749 ES, Larin AK, Golomidova AK, Letarov A V. 2015. Variations in O-antigen biosynthesis and O-
750 acetylation associated with altered phage sensitivity in Escherichia coli 4s. J Bacteriol
751 197:905–912.

752 80. Lizak C, Gerber S, Numao S, Aebi M, Locher KP. 2011. X-ray structure of a bacterial
753 oligosaccharyltransferase. Nature 474: 350-355.

754 81. Vora M, Shah M, Ostafi S, Onken B, Xue J, Ni JZ, Gu S, Driscoll M. 2013. Deletion of microRNA-
755 80 Activates Dietary Restriction to Extend C. elegans Healthspan and Lifespan. PLoS Genet
756 9:e1003737.

757 82. Dzitoyeva S, Dimitrijevic N, Manev H. 2003. Identification of a novel Drosophila gene, beltless
758 , using injectable embryonic and adult RNA interference (RNAi). BMC Genomics 4:33.

759 83. Notredame C, Higgins DG, Heringa J. 2000. T-coffee: A novel method for fast and accurate
760 multiple sequence alignment. J Mol Biol 302:205–217.

761 84. Källberg M, Margaryan G, Wang S, Ma J, Xu J. 2014. Raptorx server: A resource for template-
762 based protein structure modeling. Methods Mol Biol 1137:17–27.

763 85. Gruszka DT, Whelan F, Farrance OE, Fung HKH, Paci E, Jeffries CM, Svergun DI, Baldock C,
764 Baumann CG, Brockwell DJ, Potts JR, Clarke J. 2015. Cooperative folding of intrinsically
765 disordered domains drives assembly of a strong elongated protein. Nat Commun 6:7271.

766 86. Wojdyla JA, Gruszka DT, Foster TJ, Geoghegan JA, Manfield IW, Bingham RJ, Potts JR,
767 Turkenburg JP, Leech AP, Clarke J, Steward A. 2012. Staphylococcal biofilm-forming protein
768 has a contiguous rod-like structure. Proc Natl Acad Sci 109:E1011–E1018.

769 87. Kabsch W, IUCr. 2010. *XDS*. Acta Crystallogr Sect D Biol Crystallogr 66:125–132.

770 88. Evans PR, Murshudov GN. 2013. How good are my data and what is the resolution? Acta
771 Crystallogr D Biol Crystallogr 69:1204–14.

772 89. Winter G. 2010. xia2: an expert system for macromolecular crystallography data reduction. J
773 Appl Cryst 43:186–190.

774 90. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ, IUCr. 2007.
775 *Phaser* crystallographic software. J Appl Crystallogr 40:658–674.

776 91. Jia-Xing Y, Woolfson MM, Wilson KS, Dodson EJ. 2005. A modified ACORN to solve protein
777 structures at resolutions of 1.7 A or better. Acta Crystallogr Sect D Biol Crystallogr 61:1465–
778 1475.

779 92. Perrakis A, Morris R, Lamzin VS. 1999. Automated protein model building combined with
780 iterative structure refinement. Nat Struct Biol 6:458–463.

781 93. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F,
782 Vagin AA, IUCr. 2011. *REFMAC* 5 for the refinement of macromolecular crystal structures.
783 Acta Crystallogr Sect D Biol Crystallogr 67:355–367.

784 94. Vagin AA, Steiner RA, Lebedev AA, Potterton L, McNicholas S, Long F, Murshudov GN. 2004.
785 *REFMAC* 5 dictionary: organization of prior chemical knowledge and guidelines for its use.
786 Acta Crystallogr Sect D Biol Crystallogr 60:2184–2195.

787 95. Nicholls RA, Long F, Murshudov GN, IUCr. 2012. Low-resolution refinement tools in *REFMAC*
788 5. Acta Crystallogr Sect D Biol Crystallogr 68:404–417.

789 96. Murshudov GN, Vagin AA, Lebedev A, Wilson KS, Dodson EJ. 1999. Efficient anisotropic
790 refinement of macromolecular structures using FFT. Acta Crystallogr Sect D Biol Crystallogr
791 55:247–255.

97. Winn MD, Murshudov GN, Papiz MZ. 2003. Macromolecular TLS Refinement in REFMAC at Moderate Resolutions, p. 300–321. *In* .

98. Garib N. Murshudov; Alexei A. Vagin; Eleanor J. Dodson. 1997. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. Acta Crystallogr Sect D Biol Crystallogr D53:240–255.

99. Emsley P, Cowtan K, IUCr. 2004. *Coot* : model-building tools for molecular graphics. Acta Crystallogr Sect D Biol Crystallogr 60:2126–2132.

100. Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. Acta Crystallogr Sect D Biol Crystallogr 66:486–501.

101. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC, Richardson DC. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 66:12–21.

102. Heinig M, Frishman D. 2004. STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. Nucleic Acids Res 32:W500–W502.

103. Bertoni M, Kiefer F, Biasini M, Bordoli L, Schwede T. 2017. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. Sci Rep 7.

104. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 46:W296–W303.

105. Bienert S, Waterhouse A, de Beer TAP, Tauriello G, Studer G, Bordoli L, Schwede T. 2017. The SWISS-MODEL Repository—new features and functionality. Nucleic Acids Res 45:D313–D319.

106. Benkert P, Biasini M, Schwede T. 2011. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics 27:343–350.

107. Guex N, Peitsch MC, Schwede T. 2009. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. Electrophoresis 30:S162–S173.

Figure Legends

Fig. 1. OafA and OafB are membrane bound O-acetyltransferases that acetylate the O-antigen of *Salmonella*. (*A*) Schematic representation of OafA and OafB functional (coloured) and transmembrane (shaded) domains predicted by InterPro and TMHMM respectively. C-term constructs used for *in vitro* characterisation are indicated below the protein. (*B*) Proposed mechanism of action of O-antigen acetyltransferases during maturation of the LPS in the periplasm using OafA as an example. AT3 = IPR002656, SGNH= IPR013830

Fig. 2. Conservation in transmembrane domains of experimentally characterised bacterial AT3 carbohydrate acetyltransferases. 100% identical residues are coloured orange, similar residues in > 90% sequences are coloured blue, conserved small hydrophobic residues in transmembrane helices were not coloured. (*A*) Conserved residues across all 30 currently known experimentally characterised proteins and $OafB_{SPA}$. (*B*) Conservation in only AT3-SGNH fused proteins in the alignment. See Table S1 for details of aligned sequences and Fig. S1 for full alignment.

Fig. 3. O-antigen acetylation and OafA expression from plasmid and chromosomally expressed protein. (*A*) LPS immunoblot with crude LPS extracts from *Salmonella* ser. Typhimurium: LT2 basal O-antigen strain expressing OafA from pBADcLIC plasmid (pWT OafA-His), LT2 WT O-antigen strain with a C-terminal Deca-His tag added to the chromosomal copy of OafA (cWT OafA-His), the same strain with unmodified OafA (cWT OafA), and the LT2 basal O-antigen strain with an empty pBADcLIC plasmid (Empty Vector). O:5 antibody binding (Blue) shows abequose acetylation, *Salmonella* LPS core antibody binding (Green) acts as a loading control. (*B*) Corresponding anti-His Western blot of insoluble protein fraction for detection of His tagged OafA. Arrow indicates full length OafA protein.

Fig. 4. Summary of mutagenesis analysis of STM OafA. A diamond shape indicates residues that were mutated, cysteine residues were mutated to serine and all other residues were mutated to alanine. Results relate to % O-antigen acetylation compared to wild type, mutants that caused loss of protein expression are diamond shaped but not coloured (G34A).

Fig. 5. Analysis of the crystal structure of $OafB_{SPA}^{C\text{-}long}$. (*A*) Cartoon representation of $OafB_{SPA}^{C\text{-}long}$ with helices and sheets numbered, with the additional helix (α8) coloured teal and SGNH-extension coloured orange. Catalytic residues and disulfide bonds are shown as sticks and labelled. (*B*) Surface representation of $OafB_{SPA}^{C\text{-}long}$ with colouring as above and catalytic triad coloured red. (*C*) Surface representation of $OafB_{SPA}^{long}$, 5UFY, 5B5S and 2VPT.

Fig. 6. Analysis of additional helix and catalytic triad residues (A*, B*) Structure based sequence alignments of additional helix (A), indicated by a line above the sequence, and blocks I-V (B) with residues conserved in > 50% of sequences highlighted blue, catalytic and oxyanion hole residues are indicated by an arrow. Abbreviations and details of sequences used in methods section. (*C*) Catalytic triad and potential oxyanion hole residues shown as sticks, hydrogen bonds to co-crystallised sulfate ion shown as dashed black lines.

Fig. 7. Effect of $SGNH_{ext}$ length on substrate specificity of C-terminal OafA and OafB. Dot blot for acetylated abequose (αO:5 – blue) on basal *Salmonella* ser. Typhimurium LPS after incubation with purified the C-terminal OafA and OafB and pNPA as an acetyl group donor. 10 µM OafA and 20 µM OafB were used in these reactions. αCore antibody (green) is used as a loading control. WT acetylated LPS is used as a positive control. (+) = Active protein, (-) = Heat treated protein. Representative of N=3 repeats. 'C-Long' constructs comprise the SGNH domain with full $SGNH_{ext,}$ 'C-Short' constructs comprise the SGNH domain with fewer $SGNH_{ext}$ residues to expose the SGNH domain active site. See Figure 1 for details of the C-terminal OafA and OafB constructs

Fig. 8. Refined model of AT3-SGNH fused O-antigen acetyltransferases. Periplasmic SGNHext (Orange) is structured, therefore positioning the SGNH domain (Grey) close to the AT3 domain (Purple), this orients the additional helix (Teal) in close proximity to the AT3 domain with interactions between the two domains as proposed by the co-evolution analysis. These observations result in the current hypothesis: 1) Cytoplasmic acetyl group donor interacts with conserved Arg in TMH1, the acetyl group is processed and transferred to the periplasmic side of the inner membrane

and this process involves catalytic His residue of TMH1. 2) Conserved Asp and Ser mediate transfer of acetate to the SGNH domain. 3) SGNH domain catalyses addition of the acetate to specific O-antigen monosaccharide. The active site of the SGNH domain is highlighted by an asterisk and interaction site highlighted by a "+".

Table Legends

Table 1. Summary of site directed mutagenesis analysis of the transmembrane domain of OafA

Table 2. Summary of site directed mutagenesis analysis of the periplasmic domain of OafA

Supplementary Figure Legends

Fig. S1. Alignment of characterised AT3 acetyltransferases. Protein sequences are in the same order as Table S1 after *Salmonella* ser. Paratyphi A OafB WP_00400612. SGNH fused acetyltransferases are indicated by a grey box. Asterisks mark residues selected for mutation from this alignment.

Fig. S2. Functional analysis of OafA membrane bound domain point mutants *in situ*. Left panel shows LPS western blot with crude LPS extracts from Salmonella ser. Typhimurium basal O-antigen strain expressing OafA point mutant variants in (A) the membrane domain and (B) the periplasmic domain. O:5 antibody binding (Blue) shows abequose acetylation and

Salmonella LPS core antibody binding (Green) acts as a loading control. Right panel shows corresponding anti-His western blot for expression of His tagged OafA. Arrow indicates full length OafA protein.

Fig. S3. Structure based sequence alignment of OafB, OafA and closest structural homologues. Residues conserved in >50% highlighted blue, catalytic and oxyanion hole residues are indicated by an arrow. Abbreviations and details of sequences used in methods section.

892 **Fig. S4.** Melting curves of $OafA_{STM}^{C\text{-}long}$, $OafA_{STM}^{C\text{-}short}$, $OafB_{SPA}^{C\text{-}long}$ and $OafB_{SPA}^{C\text{-}short}$, with melting
893 temperatures of $OafA_{STM}^{C\text{-}long}$ = 63.8 ^{o}C, $OafA_{STM}^{C\text{-}short}$ = 58.1 ^{o}C, $OafB_{SPA}^{C\text{-}long}$ = 58.9 ^{o}C, $OafB_{SPA}^{C\text{-}short}$ =
894 50.0 ^{o}C.

895 **Fig. S5.** Comparison of potential oxyanion hole residues in OafA and OafB. A. Homology model of
896 $OafA_{STM}^{C\text{-}long}$ (yellow) modelled on the structure of OafB (grey, extension in orange and additional
897 helix in teal). Catalytic triad and potential oxyanion hole residues shown as sticks. Residues indicated
898 with OafA first. Both Ser 437 side chain and Leu 438 backbone amide are in close proximity to
899 catalytic triad and active site sulfate. B. Sequence alignment of OafA from Salmonella ser.
900 Tyhimurium and OafB from Typhimurium and Paratyphi A serovars of Salmonella from OafA residues
901 410-450. Alignments were carried out using Tcoffee with default settings. Red Box highlights
902 predicted replacements for catalytic block II glycine.

903 **Fig. S6.** In vitro acetyl-esterase activity of C-terminal OafA and OafB assessed by hydrolysis of
904 pNitrophenyl acetate (pNPA). Solid line = Active protein, dashed line = Heat treated protein. Error
905 bars = SEM, N=3. Some error bars are obscured by point markers. 'C-Long' constructs comprise the
906 SGNH domain with full SGNHext, 'C-Short' constructs comprise the SGNH domain with fewer $SGNH_{ext}$
907 residues to expose the SGNH domain active site. See Figure 1 for details of the C-terminal OafA and
908 OafB constructs.

909 **Fig. S7.** A) Predicted contact map for OafB based on a correlated mutation analysis using the RaptorX
910 webserver. The horizontal/vertical line marks residue 377, which forms the boundary at the end of
911 the AT3 domain. High confidence interactions within the AT3 domain (top left) and the SGNH
912 domain (bottom right), while a single high scoring interaction between the AT3 (93-97) and SGNH
913 (524-546) is marked (bottom left). B) Structure of $OafB_{SPA}^{C\text{-}long}$ with residues (542-546) predicted to
914 interact with the acyltransferase domain coloured blue. The extension is coloured orange, the
915 additional helix coloured teal, and catalytic triad coloured red.

Supplementary Table Legends

Table S1. Experimentally characterised bacterial AT3 acetyltransferases

Table S2. Molecular biology materials. Bacterial strains and primers used in this study. Primers for cloning of OafA and OafB constructs and creation of OafA point mutant variants on the pBADcLIC_WT-OafA plasmid. Amp = Ampicillin 100 µg/ml, Kan = Kanamycin 50 µg/ml.

Table S3. X-ray crystallography data and statistics for the structure of $OafB_{SPA}^{C\text{-}long}$. Values in parenthesis correspond to the highest resolution shell unless otherwise stated.

Table 1. Summary of site directed mutagenesis analysis of the transmembrane domain of OafA

Mutant	O:5 signal intensity compared to wild type % (± SEM)	Position	Reason for mutation
R14A	0.07 ± 0.04	TMH1	Specifically conserved in AT3-SGNH proteins
H25A	0.33 ± 0.18		Conserved in TMH1 across all aligned proteins
S32A	105.25 ± 30.89	Periplasmic loop & TMH2	XGG-F/Y-XGV-D/P/V-X motif found to be conserved in AT3-SGNH fused acyltransferases. In the first periplasmic loop between TMH1-2
G33A	119.17 ± 18.72		
G34A	1.36 ± 0.88*		
F35A	19.24 ± 2.70		
I36A	101.47 ± 22.72		
G37A	118.13 ± 22.11		
V38A	86.38 ± 12.73		
D39A	0.31 ± 0.07		
V40A	121.28 ± 23.82		
S45A	98.18 ± 24.30	TMH2	Conserved in SG in TMH2
G46A	99.59 ± 22.01		
R69A	0.10 ± 0.04	TMH3	RXXR motif previously identified as critical for function
R72A	0.07 ± 0.02		
S112A	0.24 ± 0.09	TMH3-4 Periplasmic loop	Conserved in periplasmic loop between TMH3-4 in AT3-SGNH fused proteins
N113A	93.79 ± 14.92		
Y122A	85.76 ± 7.58		
G202A	74.14 ± 10.70	TMH6	Conserved trans membrane glycine
E325A (Linker)	4.84 ± 1.13	TMH10-11 Cytoplasmic loop	Conserved after TMH10 in all AT3-SGNH fused proteins

Dark Grey= Point mutants with <1% O:5 signal intensity, Light Grey = Point mutants with <50% O:5 signal intensity, * = No OafA protein expression detected. Values represent the average of 2 biological repeats with 3 technical replicates.

Table 2. Summary of site directed mutagenesis analysis of the periplasmic domain of OafA

Mutant	O:5 signal intensity compared to wild type % (± SEM)	Reason for mutation
C383,397S (Linker)	107.40 ± 26.80	Conserved disulphide bonding pairs
C439,453S	185.06 ± 54.63	
C567,572S	49.98 ± 4.33	
S437A	45.59 ± 3.42	Potential oxyanion hole residue
E569A	99.87 ± 7.01	Conserved between most C-term Cys pair
S412A	0.36 ± 0.26	SGNH domain catalytic triad residues
D587A	10.13 ± 1.70	
H590A	0.87 ± 0.62	

Dark Grey= Point mutants with <1% O:5 signal intensity, Light Grey = Point mutants with <50% O:5 signal intensity. Values represent the average of 2 biological repeats with 3 technical replicates.

Figure 1

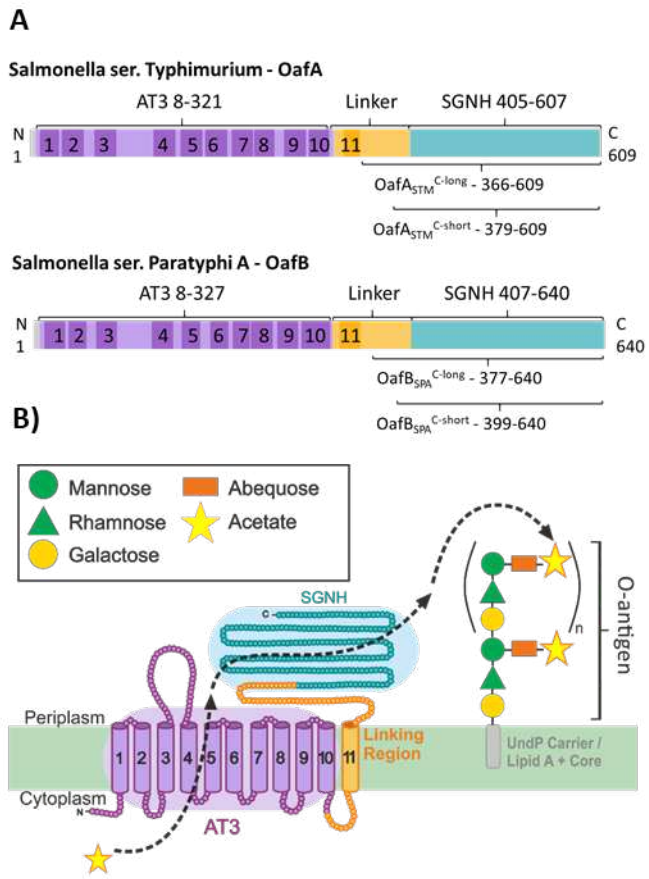

A
Salmonella ser. Typhimurium - OafA
AT3 8-321
Linker
SGNH 405-607
N
1
1 2 3 4 5 6 7 8 9 10 11
C
609
OafA STM C-long - 366-609
OafA STM C-short - 379-609
Salmonella ser. Paratyphi A - OafB
AT3 8-327
Linker
SGNH 407-640
N
1
1 2 3 4 5 6 7 8 9 10 11
C
640
OafB SPA C-long - 377-640
OafB SPA C-short - 399-640
B)
Mannose
Rhamnose
Galactose
Abequose
Acetate
SGNH
O-antigen
n
Periplasm
Cytoplasm
1 2 3 4 5 6 7 8 9 10 11
Linking Region
UndP Carrier / Lipid A + Core
AT3

Figure 2

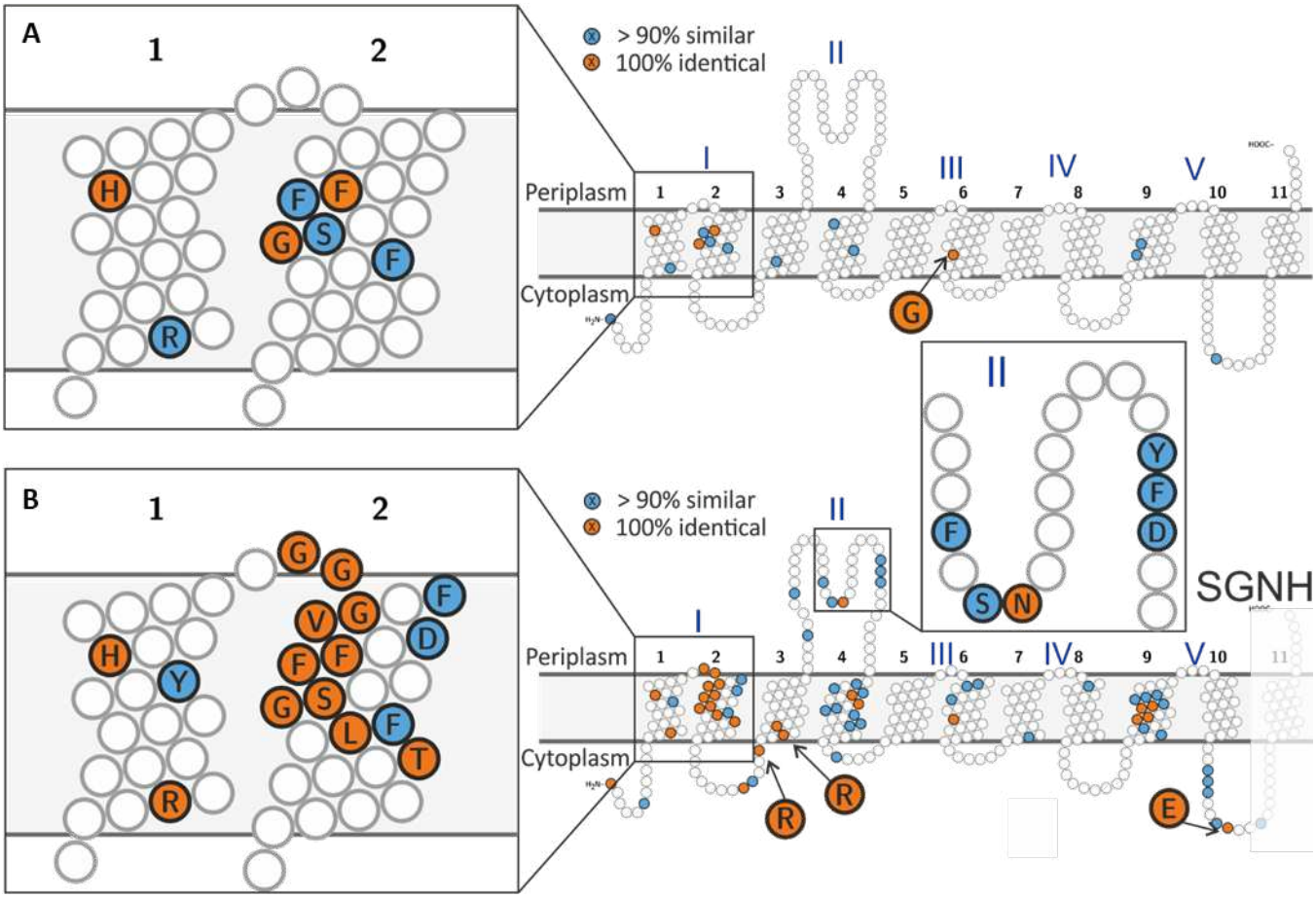

A
1
2
H
F
F
G
S
F
R
> 90% similar
100% identical
II
I
III
IV
V
Periplasm
1
2
3
4
5
6
7
8
9
10
11
Cytoplasm
G
II
Y
F
F
D
S
N
SGNH
B
1
2
G
G
F
V
G
D
H
F
F
Y
G
S
L
F
T
R
> 90% similar
100% identical
II
I
Periplasm
Cytoplasm
R
R
E

Figure 3

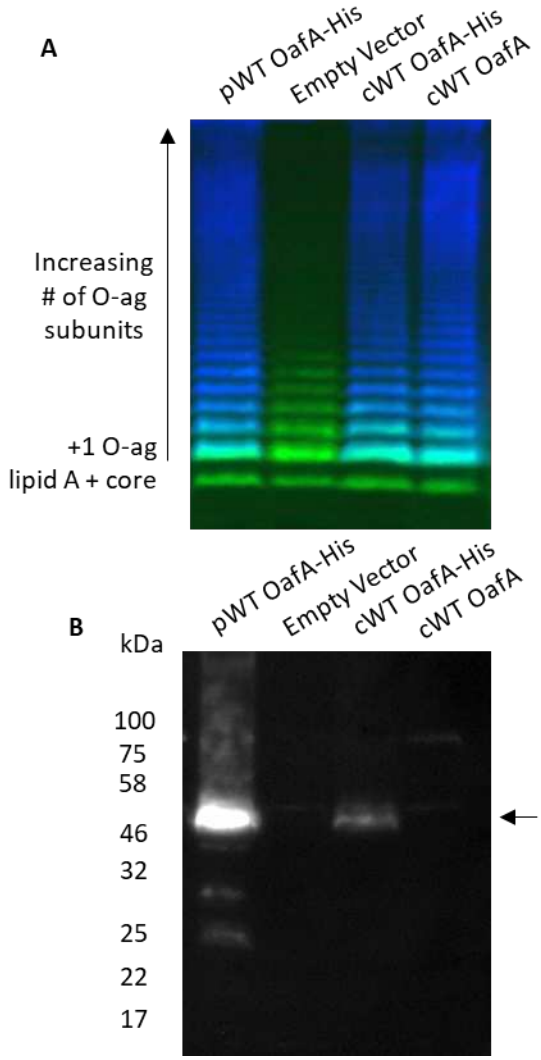

A
pWT OafA-His
Empty Vector
cWT OafA-His
cWT OafA
Increasing
# of O-ag
subunits
+1 O-ag
lipid A + core
B
pWT OafA-His
Empty Vector
cWT OafA-His
cWT OafA
kDa
100
75
58
46
32
25
22
17

Figure 4

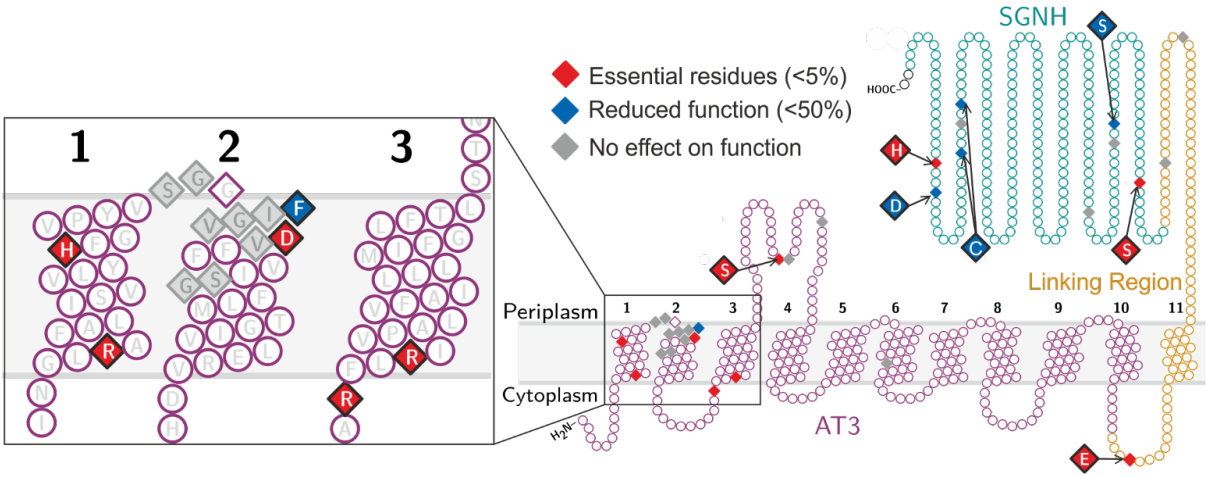

SGNH
Essential residues (<5%)
Reduced function (<50%)
No effect on function
1
2
3
HOOC
H
D
C
S
S
S
Linking Region
Periplasm
Cytoplasm
1
2
3
4
5
6
7
8
9
10
11
AT3
H2N
E

Figure 5

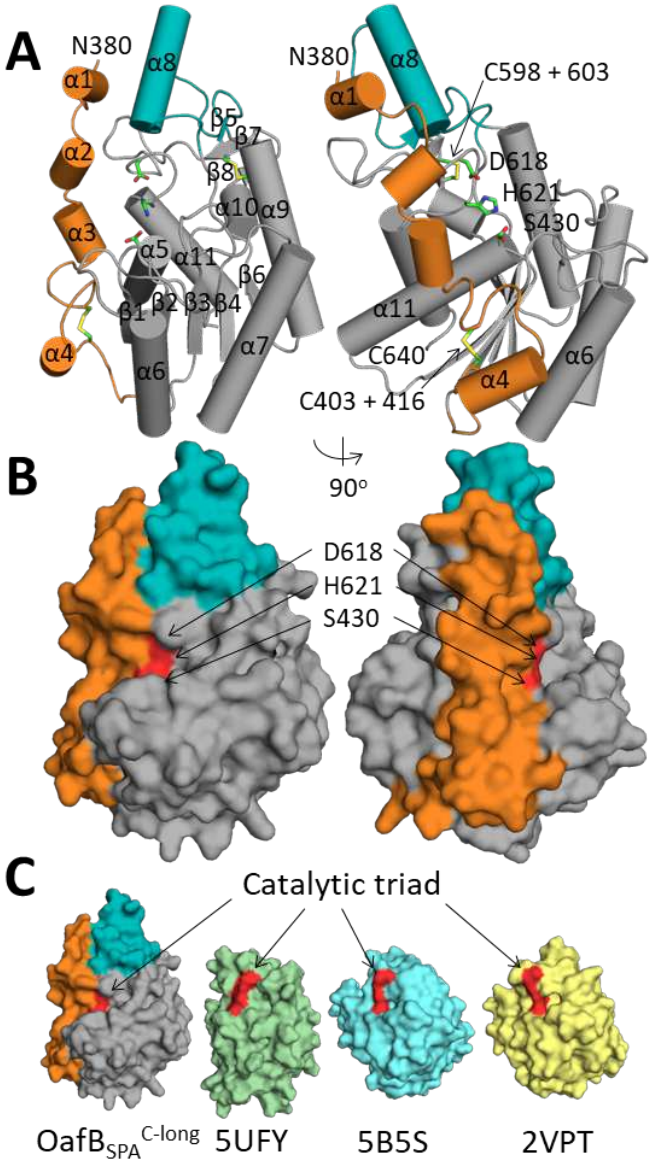
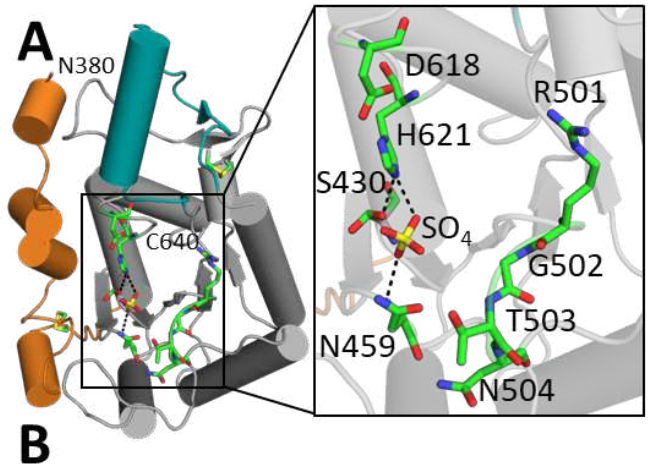

A
N380
α8
α1
α2
β5
β7
β8
α10
α9
α3
α5
α11
β6
β1
β2
β3
β4
α7
α4
α6
N380
α8
α1
C598 + 603
D618
H621
S430
α11
C640
α6
α4
C403 + 416
90°
B
D618
H621
S430
C
Catalytic triad
OafB$_{SPA}^{C\text{-}long}$
5UFY
5B5S
2VPT

Figure 6



Additional helix

OafB-S.PA	535	.....WNANLVKIISNYLSEFKKTPPLYMTYGLNSEI	575
OafB-S.Tym	535	.....WNANLVKVISNYTSEFKKTPPIYMSYGLNDEI	575
OafA-S.Tym	505	.....KKTMIDTIEDMGINSGRTVPWSM.TDETRNL	544
OafA-H.inf	514	.....SPL...RGYLLENYGLEKYLTPI....HRMGDI	551
OatA-S.pne	516	.....DKT.....KETYAIV	534
OatA-S.aur	530	.....DYE	542
Ape1-N.men	285	TLGVCGTRPV.....RL	308
TAP1-E.col	131	.....GRR.....YN	147
RGAE-A.acu	131	.....ETGTFVN.....SP	151
5B5S-T.cel	127	.....DATI	142
2VPT-C.the	129	.....AI	141

C

Block I

OafB-S.PA	425	FIIGDSYAAA	434
OafB-S.Tym	425	FIIGDSYAAA	434
OafA-S.Tym	407	VVWGDSHAAH	416
OafA-H.inf	405	IILGDSHSSH	414
OatA-S.pne	433	MLIGDSVALR	442
OatA-S.aur	448	LLIGDSVMVD	457
Ape1-N.men	57	LQIGDSHTAG	66
TAP1-E.col	31	LILGDSLSAG	40
RGAE-A.acu	4	YLAGDSTMAK	13
5B5S-T.cel	5	MLLGDSITEI	14
2VPT-C.the	7	MPVGDSCTEG	16

Block II

OafB-S.PA	455	MTD.....GNAPPLFV	465
OafB-S.Tym	455	MTD.....GNAPPLFV	465
OafA-S.Tym	434	RTASLCPPIIGLQKDD	449
OafA-H.inf	435	DKFECSFIVN.EQYQL	449
OatA-S.pne	457	NAQ.....VS.....	461
OatA-S.aur	472	DGK.....VG.....	476
Ape1-N.men	214	MGI.....NG.....	218
TAP1-E.col	66	ASI.....SG.....	70
RGAE-A.acu	38	DAV.....AG.....	42
5B5S-T.cel	58	EGH.....SG.....	62
2VPT-C.the	63	EGH.....SG.....	67

Block III

OafB-S.PA	498	WSV..RGTNG...VHD	508
OafB-S.Tym	498	WSV..RGSNG...VHD	508
OafA-S.Tym	476	ALWPVY.....	481
OafA-H.inf	476	MGGQPVPRFRPETFIE	491
OatA-S.pne	488	TGVNNPE.....	494
OatA-S.aur	504	LGINGAFTK.....	512
Ape1-N.men	247	YGTNEAFNNNID....	258
TAP1-E.col	96	LGGNDGLRG....FQP	107
RGAE-A.acu	71	FGHNDGGSLS...TDN	83
5B5S-T.cel	89	LGTNDVNIGH...RNA	101
2VPT-C.the	93	IGGNDLLLNG.....	102

Block V

OafB-S.PA	615	TAVDWGHL..	622
OafB-S.Tym	615	TAVDWGHL..	622
OafA-S.Tym	585	.QYDNAHL..	591
OafA-H.inf	587	.YGDQDHL..	593
OatA-S.pne	565	AGTDQVHFGS	574
OatA-S.aur	573	.AYDGIHL..	579
Ape1-N.men	346	.AKDGVHF..	352
TAP1-E.col	178	.QDDGIHP..	184
RGAE-A.acu	190	.PIDHTHT..	196
5B5S-T.cel	177	.RDDGVHP..	183
2VPT-C.the	176	.SWDGLHL..	182

Figure 7

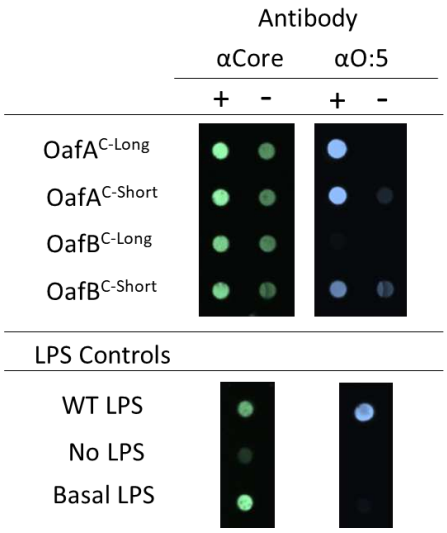
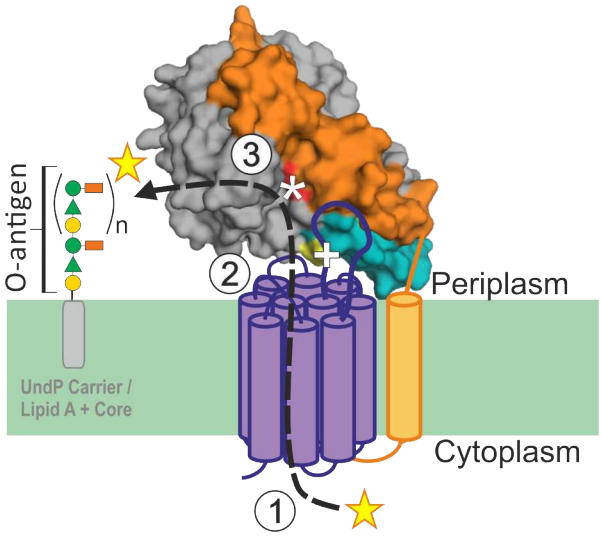

Antibody
αCore
αO:5
+ - + -
OafA C-Long
OafA C-Short
OafB C-Long
OafB C-Short
LPS Controls
WT LPS
No LPS
Basal LPS

Figure 8


O-antigen
n
3
2
1
Periplasm
Cytoplasm
UndP Carrier /
Lipid A + Core