A novel fold for acyltransferase-3 (AT3) proteins provides a framework for transmembrane acyl-group transfer

Kahlan E. Newman[1**], Sarah N. Tindall[2**], Sophie L. Mader[3], Syma Khalid[3*], Gavin H. Thomas[2*], Marjan W. van der Woude[2*]

[1]School of Chemistry, University of Southampton, SO17 1BJ, UK
[2]Department of Biology and the York Biomedical Research Institute, University of York, UK
[3]Department of Biochemistry, University of Oxford, OX1 3QU, UK

*Correspondence:
Syma Khalid: syma.khalid@bioch.ox.ac.uk
Gavin H. Thomas: gavin.thomas@york.ac.uk
Marjan W. van der Woude: marjan.vanderwoude@york.ac.uk
** Joint first authors

Abstract

Acylation of diverse carbohydrates occurs across all domains of life and can be catalysed by proteins with a membrane bound acyltransferase-3 (AT3) domain (PF01757). In bacteria, these proteins are essential in processes including symbiosis, resistance to viruses and antimicrobials, and biosynthesis of antibiotics, yet their structure and mechanism is largely unknown. In this study, evolutionary co-variance analysis was used to build a computational model of the structure of a bacterial O-antigen modifying acetyltransferase, OafB. The resulting structure exhibited a novel fold for the AT3 domain, which molecular dynamics simulations demonstrated is stable in the membrane. The AT3 domain contains 10 transmembrane helices arranged to form a large cytoplasmic cavity lined by residues known to be essential for function. Further molecular dynamics simulations support a model where the acyl-coA donor spans the membrane through accessing a pore created by movement of an important loop capping the inner cavity, enabling OafB to present the acetyl group close to the likely catalytic resides on the extracytoplasmic surface. Limited but important interactions with the fused SGNH domain in OafB are identified and modelling suggests this domain is mobile and can both accept acyl-groups from the AT3 and then reach beyond the membrane to reach acceptor substrates. Together this new general model of AT3 function provides a

framework for the development of inhibitors that could abrogate critical functions of bacterial pathogens.

Introduction

Acyltransferase family 3 (*Acyl_transf_3,* AT3) domain (Interpro: IPR002656, PFAM: PF01757) containing proteins are found in all domains of life. These membrane-bound AT3 proteins are involved in acylation of a wide range of extracytoplasmic and surface bacterial polysaccharides (1), but are also important in Eukarya, for example in the regulation of lifespan in *Caenorhabditis elegans* (2) and in *Drosophila* development (3). In bacteria, where these proteins have been primarily studied, the resulting acylations have been shown to be involved in root nodulation (4), increase the efficacy of macrolide antibiotics (5), conferring resistance to lysozyme (6), influencing bacteriophage sensitivity (7), and altering antibody recognition (8–10). While AT3 domains most commonly exist as standalone proteins, there are many examples of AT3 domains fused with SGNH or alanine racemase domains, and fusions to other domains also exist in Eukarya (see Pfam architectures for PF10757). AT3 domains are predicted to have ten transmembrane helices (11–29); however, despite the wide-ranging functions of this family of proteins, there are currently no models for their overall structure and only limited information on mechanism.

The overall understanding of AT3 proteins is mainly derived from studies of these proteins in the context of bacterial virulence through changes in the cell surface (2,3). One example of this is the acetylation of lipopolysaccharide (LPS) and lipooligosaccharide (LOS) by AT3 domain containing proteins present in the inner (cytoplasmic) membrane of Gram-negative bacteria, including species of *Neisseria* (30), *Salmonella* (31,32), *Shigella* (33–38), *Haemophilus* (39,40), *Burkholderia* (41–44) and *Legionella* (27,45,46), that modify the O-antigen of the LPS during its synthesis and before LPS export to the outer membrane. Acetylation performed by these AT3 family proteins increases heterogeneity of the O-antigen, which can lead to an altered bacterial serotype and resistance to bacteriophage attack (31–36). Some of these reactions are carried out by AT3 only proteins, others by proteins with fused AT3 and SGNH domains. SGNH proteins are a family of proteins with hydrolase/transferase activity (47).

A second important process in bacteria that involves an AT3 protein is the modification of peptidoglycan that leads to resistance to both lysozyme and β-lactam antibiotics (29,48). One such protein, OatA, is an AT3 protein with attached SGNH domain (AT3-SGNH), that acetylates the MurNAc residue in peptidoglycan (29,48). Initially discovered in

70 *Staphylococcus aureus* (29), OatA homologues have since been identified in many Gram-
71 positive bacteria including *Listeria monocytogenes* (48), *Lactococcus lactis* (49,50) and
72 *Streptococcus pneumoniae* (14,51,52). Furthermore, standalone AT3 proteins contribute an
73 acyl group in the biosynthesis of macrolide antibiotics in *Streptomyces* species which
74 increases antibiotic efficacy (5,53–55). These selected examples (see (1) for a recent
75 review) illustrate that AT3 domain-containing proteins are involved in a range of highly
76 relevant processes for bacterial pathogens and may be relevant for biotechnological
77 applications.
78
79 Found in the cytoplasmic membrane, AT3 domains are highly hydrophobic integral
80 membrane proteins predicted to contain 10 transmembrane helices (11–29). While the
81 majority of AT3 domains consists of only an AT3 domain (standalone AT3), many are AT3
82 domains with an additional transmembrane helix (TMH) and C-terminal, periplasmic SGNH
83 domain attached via a periplasmic linking region (AT3-SGNH) (56). Thanweer *et al.* (26) and
84 Jones *et al*. (57) have performed PhoA-LacZα fusion analysis of the O-antigen
85 acetyltransferase Oac from *Shigella flexneri* (a standalone AT3) and OatA from
86 *Staphylococcus aureus* (an AT3-SGNH), respectively. Analysis by Thanweer *et al.*
87 suggested that Oac has 10 TMH, with the N- and C- termini in the cytoplasm (26). This is
88 consistent with fusion analysis of the O-antigen acetyltransferase OafB from *Salmonella* ser.
89 Typhimurium (20). The *Salmonella* O-antigen acetyltransferases OafB and OafA are both
90 AT3-SGNH fusion proteins and predicted to have an 11th TMH to allow the fused SGNH
91 domain to be located in the periplasm (20,58). In contrast, Jones *et al.* proposed that OatA
92 contain only 9 TMH with one large re-entrant loop (57); as an AT3-SGNH protein, OatA was
93 also previously predicted to contain 11 TMH, orientating the SGNH domain in the periplasm
94 (12,29). In addition, a large cytoplasmic loop was described between TMH 8-9 of OatA,
95 where Oac has only short cytoplasmic loops but a longer periplasmic loop between TMH3
96 and 4 (26,59). Currently, this is the extent of the knowledge of the architecture of AT3
97 domains and there are no published experimentally determined protein structures.
98
99 AT3 proteins are predicted to transport acyl groups, including acetyl groups, from the
100 cytoplasm across the cytoplasmic membrane to be transferred onto the receptor molecule
101 (6). However, the mechanism is currently largely unknown. The acetyl group has been
102 proposed to be donated by acetyl coenzyme A (acetyl-CoA)(24). Three arginine residues on
103 the cytoplasmic side of TMH1 and 3 have been identified as highly conserved in the PFAM
104 HMM logo, and in both standalone AT3 and AT3-SGNH proteins these residues are
105 essential for function (26,58). Arginine residues have been shown to be involved in binding
106 of the 3'-phosphate group of acetyl-CoA (60) suggesting that these conserved residues may

interact with the proposed acetyl group donor. Previous *in vitro* studies suggest that standalone AT3 proteins CmmA, MdmB and Asm19 are able to utilise acetyl-CoA as an acetyl donor (5,61,62). Importantly, Jones et al. (57) found OatA from *Staphylococcus aureus* (an AT3-SGNH protein) is able to hydrolyse the acetyl group from acetyl-CoA and transfer it to the a peptidoglycan-like acceptor substrate, effectively observing the entire catalytic cycle *in vitro* (57). Thus, the available data indicate that acetyl-CoA is the most likely acetyl group donor is acetyl-CoA, but how this interacts with the AT3 domain or how the acetyl group can cross the membrane, is as yet unknown.

In addition, site-directed mutagenesis identified three conserved tyrosine residues which are required for function (57). One of these Tyr residues is predicted to be located in the periplasm and is proposed to be involved in transfer of the acyl group from the AT3 domain to the SGNH domain in AT3-SGNH proteins (57). The catalytic triad of the SGNH domain of OafB from *Salmonella enterica* subsp. *enterica* serovar Typhimurium is known to be required for acetylation of the rhamnose moiety of the O-antigen of LPS (58), suggesting that in standalone AT3 proteins an equivalent, as yet unknown, partner protein may exist. If this were the case, it would suggest that the system may be similar to that of the PatA and PatB peptidoglycan acetyltransferase system. In this system it is hypothesised that the MBOAT (membrane-bound O-acyltransferase) protein PatA transports the acetyl group across the membrane where it is transferred to the peptidoglycan substrate by PatB (an SGNH domain containing protein) (63).

To further study the AT3-SGNH family of proteins, OafB from *Salmonella enterica* subsp. *enterica* was used as a model system. OafB acetylates the rhamnose moiety of the O-antigen repeating unit of LPS. Modification of the O-antigen has repeatedly been identified as important for virulence and persistence of bacteria (27,30,39,45). Similarly, O-acetylation by OafB has been acknowledged in the development of vaccines against *Salmonella* ser. Paratyphi A, where acetyl groups were shown to be required to elicit bactericidal antibodies (64,65). Furthermore, OafB-mediated acetylation in invasive non-typhoidal Salmonella (iNTS) alters susceptibility to bacteriophage (20). OafB consists of an AT3 domain linked to an SGNH domain via an 11th TMH and periplasmic linking region (AT3-SGNH) (58) (**Fig. 1A**). The structure of the SGNH domain and periplasmic linking region was previously solved using x-ray crystallography (58). The structure was found to be similar to that of other SGNH domains, but the periplasmic linking region formed a structured extension suggesting the AT3 and SGNH domains are likely to exist in close proximity and may even interact (58).

Herein we use RaptorX to determine a computationally derived structure for the transmembrane region, including the 10-TMH AT3 domain, of OafB. We show this novel structure to be stable under physiological conditions *via* molecular dynamics simulations. This model is integrated with the existing x-ray structure of the SGNH domain and published mutational data to allow additional structure-function analysis. Utilising both classical molecular dynamics simulations and quantum mechanical calculations, we consider the acetyl donor and acceptor substrates to generate a refined functional model for this important family of membrane proteins.

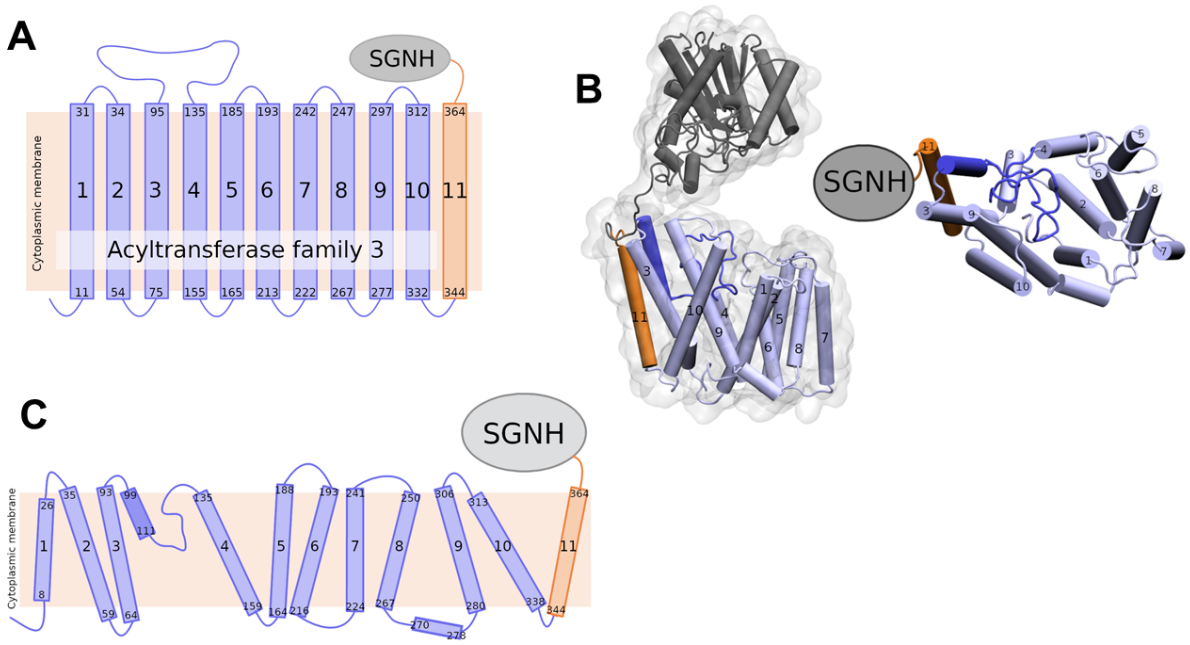


Figure 1: Topology and predicted structure of OafB. **(A)** TOPCONS topology prediction of OafB with N- and C-termini of each TMH indicated. Consistent with the RaptorX structure, TOPCONS predicts short cytoplasmic loops and a long periplasmic loop between TMH 3 and 4. **(B)** RaptorX predicted structure of OafB with AT3 domain and TMH 11 coloured light blue, periplasmic loop 3-4 coloured blue, SGNH domain coloured grey with the periplasmic linking region in orange and additional helix in teal. **(C)** Topology schematic of OafB based on the RaptorX predicted structure. The RaptorX structure has 11 TMH, with a long periplasmic loop between TMH 3 and 4 consisting of a short helix followed by an unstructured region. TMH1-10 form the AT3 domain and are coloured blue, TMH11 forms the linking region between the AT3 and SGNH domains and is coloured orange.

Results & Discussion

RaptorX model of OafB supports specific topology predictions and identifies new features

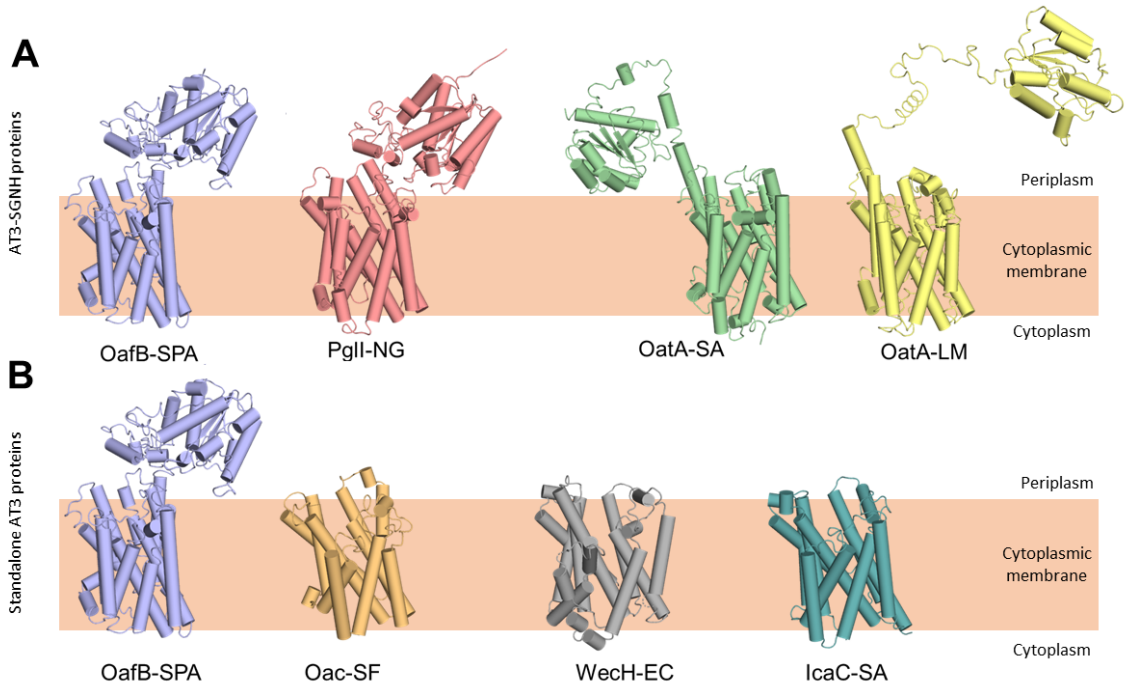
The structure of OafB was predicted by RaptorX (66–68) using the protein sequence of a *Salmonella* rhamose O-acetyltransferase (OafB). This method has been successfully validated with numerous proteins (69–73). The structure of OafB consists of two key

168 domains: the AT3 domain and SGNH domain (**Fig. 1B**). The transmembrane region has 13
169 helices, of which 11 completely span the membrane (**Fig. 1B, C**). The AT3 domain (residues
170 1-338) has 10 TMH and an additional 11th TMH (residues 339-377) fused to the periplasmic
171 linking region (containing SGNH-ext, residues 377-421), which facilitates a periplasmic
172 location of the fused SGNH domain (residues 422-640) (58), as predicted from simple
173 TOPCONS topological analysis (74,75). The topological analysis also predicted a loop
174 region between helices 3 and 4 (**Fig. 1A**); in the RaptorX model this region is clearly
175 suggested to be a short re-entrant loop consisting of a short, well-structured helix (residue
176 99-111) followed by a ~24 residue-long unstructured region **(Fig. 1B,C**). This means that
177 there are no long hydrophilic loops in the structure on either side of the membrane (58).
178
179 The SGNH domain and periplasmic linking region of OafB (residues 377-640) was
180 previously solved using x-ray crystallography (58). The RaptorX structure of the same region
181 showed remarkable similarity, with a root-mean-square deviation (RMSD) of 2.8 Å. As seen
182 in the crystal structure, the SGNH domain from OafB resembles a typical SGNH domain but
183 with an additional helix and structured extension. Neither of these regions are seen in the
184 structures of other SGNH domains in the RCSB PDB; despite this, RaptorX predicts these
185 regions with high accuracy. There are very few residues between the C-terminus of TMH11
186 (residue 367) and the N-terminus of the structured extension (residue 377), suggesting that
187 the AT3 and SGNH domains are likely in close proximity and may interact (58). However, in
188 the RaptorX structure of OafB, the two domains show limited interaction and are not in close
189 proximity (**Fig. 1B**).
190
191 Similar to RaptorX, AlphaFold (76) predicted 11 transmembrane helices, with the SGNH
192 domain in the periplasm. The structures predicted for residues 1-338 (TMH1-10 of the
193 transmembrane domain, the AT3 domain) showed 97% coverage (RMSD of 3.04 Å) and a
194 template matching score of 0.85, indicating the same structure across the two models
195 (**Supp. Fig. S1**).
196

197 **Diverse bacterial AT3 domain-containing proteins share a common 10 TMH**
198 **structure**

199 Having built a model for our primary experimentally characterised protein, we expanded the
200 analysis to include other important AT3 domain-containing proteins, including both additional
201 AT3-SGNH fusion proteins and standalone AT3 proteins. The RaptorX generated structures
202 of the AT3-SGNH proteins OatA from *Staphylococcus aureus* (OatA-SA), OatA from *Listeria*
203 *monocytogenes* (OatA-LM) and PglI from *Neisseria gonorrhoeae* (PglI-NG) all closely

204 resembled the OafB (**Fig. 2A**), with an RMSD of less than 3 Å for the AT3-domain over the
205 10 TMHs. Importantly, as they all contain a C-terminal SGNH domain that functions extra-
206 cytoplasmically, they all contain an 11th TMH to enable correct localisation of the fused
207 domain (**Fig 2A**) (23,58,77,78). In addition, the core of the SGNH domains are also similar
208 (79) with variation occurring in the length of the linking regions that connect the AT3 to the
209 SGNH. While the PglI-NG protein closely resembles OafB with a short, structured linker (1),
210 OatA proteins show a more extended, and potentially more mobile and flexible, structure,
211 locating the SGNH domain further from the AT3 domain (**Fig. 2A**).
212
213 Expanding the analysis to standalone AT3 proteins, the RaptorX outputs from Oac from
214 *Shigella flexneri* (Oac-SF), WecH from *Escherichia coli* (WecH-EC) and IcaC from
215 *Staphylococcus aureus* (IcaC-SA) were analysed (**Fig. 2B**). While the RMSD of these
216 compared to the AT3-domain of OafB was more variable (RMSD of 4.5 - 6.7 Å when
217 compared to residues 1-376 of OafB) the overall 10 TMH features were conserved and are
218 consistent with the experimentally-determined topology of Oac-SF (10,26). However, a
219 recent analysis of *Staphylococcus aureus* OatA combining LacZ-PhoA fusion data with *in*
220 *silico* predictions (56), led to an alternative model for the topology of the AT3 domain with 9
221 TMH, a further re-entrant helix and a long cytoplasmic loop preceding the final TMH that
222 leads to the SGNH domain in the periplasm. This topology is not seen in any of our models,
223 including similar models generated in AlphaFold (data not shown) and these conflicts will be
224 discussed later.



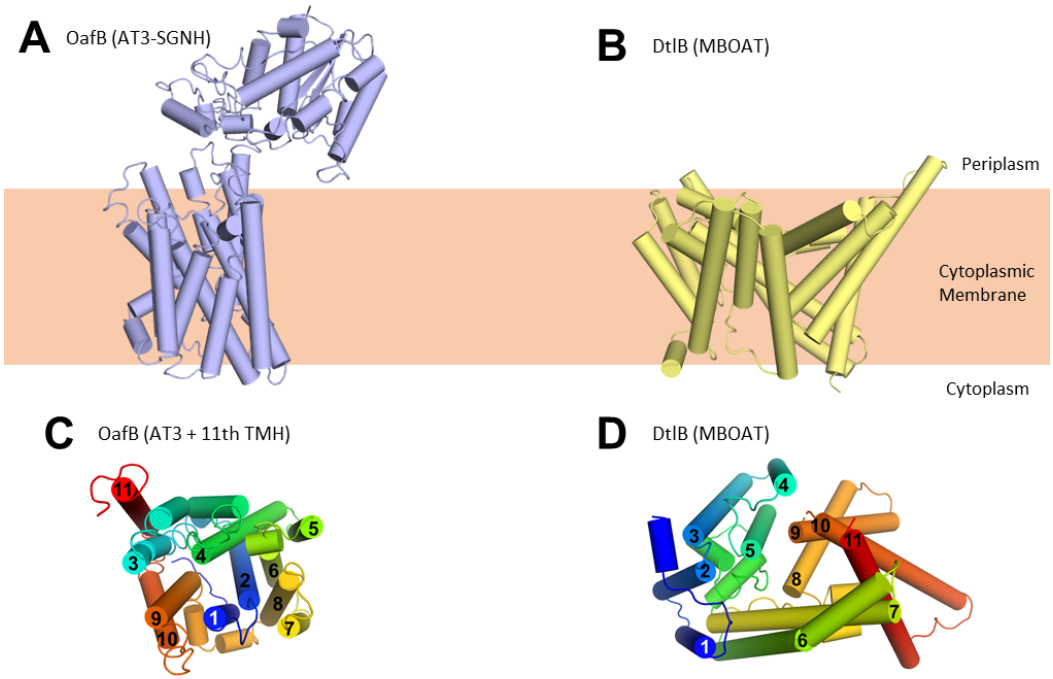
225
226 **Figure 2:** RaptorX predicted structures of AT3 domain containing proteins. **(A)** Fused AT3-SGNH proteins with
227 OafB for comparison. Left to right: OafB from *Salmonella enterica* subsp. *enterica* ser. Paratyphi A (OafB); PglI

228 from *Neisseria gonorrhoeae* (PglI-NG); OatA from *Staphylococcus aureus* (OatA-SA); OatA from *Listeria*
229 *monocytogenes* (OatA-LM). **(B)** Standalone AT3 proteins with OafB for comparison. Left to right: OafB-SPA; Oac
230 from *Shigella flexneri* (Oac-SF); WecH from *Escherichia coli* (WecH-EC); IcaC from *Staphylococcus aureus*
231 (IcaC-SA). The structure of the AT3 domain is largely conserved between these proteins.
232

233 **The AT3 domain defines a novel and stable membrane protein fold**

234 Using our new structural models for AT3 proteins, we assessed their similarity to known
235 structures using the DALI server (80). The closest structural homologues consisted of small
236 helical bundles that were fragments of larger proteins and no significant matches to full
237 length proteins were identified. Looking specifically at other membrane proteins with 10-11
238 TMH (3M73, 4J72, 4KJR, 1RH5) we could not see similarities with the AT3 structure
239 (RMSDs between 12-24 Å) and the arrangement of TMH is very different. Finally, we
240 compared our AT3 structure to that of the other known family of membrane-bound
241 acyltransferases, the MBOAT proteins. These also function in the acylation of complex
242 extracytoplasmic carbohydrates (81) and previous analysis has shown an MBOAT protein to
243 be functionally interchangeable with a standalone AT3 protein (19,82) (**Fig. 3B**), again
244 suggesting that these proteins have closely related functions. The structure of DltB (an
245 MBOAT protein from *Streptococcus thermophilus*) (81) similarly contains 11 TMH, however,
246 the helical arrangement is entirely different (**Fig. 3C,D**). Consistent with a core 10 TMH fold
247 in the AT3-SGNH protein, the 11th TMH present sits on the outside of the bundle as would
248 be expected for a non-essential feature of the AT3 domain, in contrast to the 11th TMH in
249 DltB (**Fig. 3C**).
250

251 To investigate the biological plausibility of the proposed AT3 fold, we applied Molecular
252 Dynamics (MD) techniques to emulate the behaviour of the protein in a physiological
253 environment, and to extract meaningful data on the dynamics and structure-function
254 relationships of the system. The RaptorX model of the transmembrane domain (TMH1-11 of
255 OafB, residues 1-377) was embedded in a model *Escherichia coli* inner membrane in 150
256 mM KCl solution. When subjected to MD simulations of 50 and 100 ns at 320 and 303 K,
257 respectively, the protein model remained stable within the membrane. There was no
258 significant unfolding observed; secondary structure analysis indicates that the alpha-helical
259 segments of the transmembrane domain are maintained throughout and the RMSD of the
260 protein backbone remained below 0.5 nm at both temperatures (**Supp. Fig. S2A**). The root-
261 mean-square fluctuation (RMSF) by residue, as expected, is highest for the termini and for
262 residues in unstructured loops, and lower for the transmembrane helices (**Supp. Fig. S2B).**
263 Together these data suggest that the computational approach described a novel, stable
264 membrane protein fold for AT3 proteins.



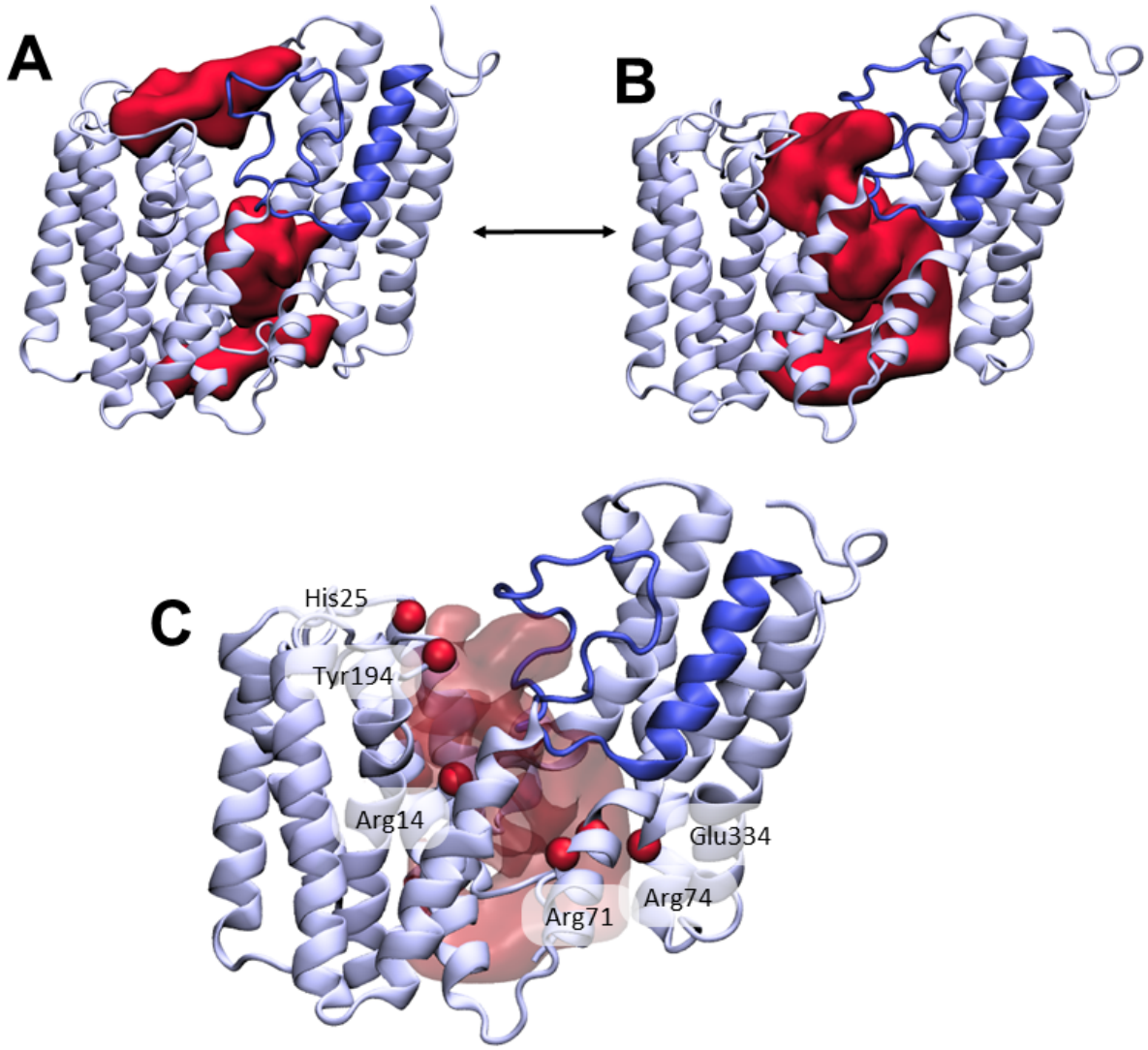
265
266 **Figure 3:** Structure of OafB, an AT3-SGNH protein (left, panels **A** and **C**), compared to DltB from *Streptococcus*
267 *thermophilus*, an MBOAT (Membrane Bound O-Acyltransferase) protein (right, panels **B** and **D**). Both OafB and
268 DltB have 11 TMH **(C, D)** however the arrangements differ considerably. **(C)** OafB and **(D)** DltB viewed from
269 periplasm, coloured blue to red, N- to C- termini with transmembrane helices numbered.
270

271 **The AT3 domain from OafB has a cavity lined with essential residues forming a**
272 **putative acetyl-CoA binding site**

273 Having established that the AT3 domain constitutes a novel stable fold in the membrane we
274 then wished to use this structural model to try and understand the mechanism of this
275 membrane-bound enzyme. A large cavity with a volume of ~3,620 $Å^3$ can be observed in
276 the AT3 protein on the cytoplasmic side, which is enclosed by enclosed by TMH 1, 2, 9, and
277 the unstructured loop between TMH 5 and 6 (**Fig. 4A**). The top end of this cavity is blocked
278 by the novel re-entrant loop between THM 3-4, which in contrast to TOPCONS predictions
279 (**Fig. 1A**) is not in the periplasm but resides within the transmembrane domain (**Fig. 4A**).
280 Examination of the positioning of core amino acids known to be important and/or essential
281 for function of different AT3 proteins onto the predicted structure reveals that many of these
282 in fact line this cytoplasmic cavity in the protein (**Fig. 4C**).
283

284 Two conserved motifs have been identified in AT3 domains: an R-X_{10}-H motif located in
285 TMH1 (Arg14 and His25 in OafB, **Fig. 4C**) (58); and an RXXR motif located on the
286 cytoplasmic face of TMH3 (Arg71 and Arg74) (17,20,24,83). The four residues found in
287 these motifs have all been identified in experimental studies as being essential for function

288 (20,25,57,58,84). Both arginine and histidine residues have previously been implicated in
289 other acetyltransferase proteins as important for binding of acetyl-CoA (60,81,85). Perhaps
290 significantly, all three Arg residues are in close proximity in positions approximately in the
291 middle of the membrane, a usually unfavourable location, suggesting a key function in the
292 mechanism of the enzyme.
293



294
295 **Figure 4:** The loop between TMH3 and 4 controls the formation of a transmembrane channel lined with essential
296 residues. **(A)** Initial equilibrated structure of the transmembrane domain (residues 1-377) of OafB. Residues 1-94
297 and 136-377 in pale blue, residues 95-135 in blue. The two largest cavities identified by the ProPores2 server are
298 shown as red surfaces. Initially, the loop between TMH3 and 4 occludes the central pore in the AT3 domain. **(B)**
299 Structure of the transmembrane domain after 50 ns equilibrium MD at 320 K. The loop between TMH3 and 4 is
300 dynamic – this snapshot from the end of the simulation shows the loop to have moved away from the center of
301 the cavity to allow a channel to form. This channel could be occupied by the acetyl group donor. **(C)** Same
302 snapshot as (B), but pore shown as a transparent red surface. Important conserved residues shown as red
303 spheres: these residues line the pore in the center of the AT3 domain.

304 In addition, Jones et al. performed site-directed mutagenesis on a number of residues in
305 OatA from *Staphylococcus aureus* (57). These identified, amongst others, three conserved
306 tyrosine residues which are required for function (57). It is thought that Tyr206, located in the
307 periplasm (equivalent residue in OafB Tyr194, **Fig. 4A**), is involved in transfer of the acetyl
308 group from the AT3 domain to the SGNH domain (57). Similarly, a catalytic triad consisting
309 of this aforesaid tyrosine, plus a glutamic acid and a histidine in the AT3 domain was
310 proposed to remove the acetyl group from acetyl-CoA (57), allowing transport across the
311 membrane (the three equivalent residues are highlighted in **Figure 4A** – Tyr194, His25 and
312 Glu334). In the RaptorX structure of OafB, Tyr194 and His25 are in close proximity (C_α
313 separated by 8 Å) and both sit on the periplasmic face of the membrane. However, Glu334 is
314 located on the cytoplasmic side very distant from His25 and Tyr194, suggesting that it is not
315 part of a catalytic triad.
316
317 Together this mapping suggests two important sites in the protein, a mid-membrane region
318 of positive charge created by 3 conserved Arg residues and a potential catalytic site on the
319 periplasmic surface comprising at least His25 and Tyr194. In the RaptorX model these two
320 features of the protein are physically separated by the loop 3-4 region (**Fig. 4A**). During
321 equilibrium MD simulations this loop is in fact dynamic (**Fig. 4B**). Structures taken from later
322 in the simulations indicate that the loop can move outwards (displacement of residues on the
323 order of 3-4 Å) from the centre of the cavity towards TMH2, allowing the potential formation
324 of a pore between the cytoplasmic and periplasmic surfaces of the transmembrane domain,
325 which could be a critical stage in the mechanism of the acyltransferase in the trans-
326 membrane movement of acyl-groups (**Fig. 4B**).
327

328 The AT3 domain facilitates presentation of acyl-CoA molecules to the 329 periplasm

330 Any model for an acyl-coA dependent acylation of an extracytoplasmic acceptor sugar
331 requires the transfer of the cytoplasmically-located acyl group across the membrane. The
332 inner cavity and dynamic conversion of this to a pore, could provide the route for acyl-CoA
333 molecules to enter the protein in a way to present the acyl group for use in the catalytic
334 process (**Fig. 5A, left**). OafB would use acetyl-CoA as the substrate, which is an extended
335 molecule with a volume of ~630 $Å^3$. To test this hypothesis, one molecule of acetyl-CoA was
336 pulled into this channel (entrance delimited by the N-terminus and TMH9 and 10) from the
337 cytoplasmic entrance of the AT3 domain of a hybrid full-length OafB model (assembly
338 discussed in the following section) using steered MD (SMD). It was found that this channel
339 could indeed accommodate acetyl-CoA (**Fig. 5A, right**) and, significantly, the thioester bond

340 was positioned close to the essential His25 residue within the membrane. The conserved
341 Arg14 was also positioned to help coordinate the 3'-phosphate of acetyl-CoA, as previously
342 predicted (58) (**Fig. 5A**).
343
344 The acetyl-CoA molecule was then allowed to equilibrate within the transmembrane domain
345 using equilibrium MD. Three replicates of this system were generated and simulated with
346 altered starting conformations for the acetyl-CoA within the protein. In the unrestrained final
347 20 ns of these simulations, acetyl-CoA was observed to hydrogen bond to several basic
348 residues within the pore (**Fig. 5B**). In addition to the Arg14 coordinating the phosphate of 3'-
349 phosphate acetyl-CoA, we noted that the conserved Arg74 (of the RXXR motif of TMH3)
350 along with Lys279 and Arg338 formed a small pocket capable of orienting such that they can
351 all simultaneously hydrogen bond to the 3'-phosphate group (**Fig. 5B**).
352
353 Further to this, energy decomposition analysis using ONETEP (86,87), indicates strong,
354 attractive interactions between the AT3 domain and acetyl-CoA. The transmembrane
355 domain-acetyl-CoA complex was optimised, and a single point energy calculated. This
356 energy was then decomposed into its contributing components. Accounting for frozen
357 density, polarisation, charge transfer, and dispersion components, the overall energy of their
358 interactions was calculated to be -1,035.943 kcal mol^{-1}. Strong charge transfer interactions
359 were identified between the protein and small molecule. The most prominent of these were
360 between the 3'-phosphate of acetyl-CoA and nearby basic residues in the AT3 domain
361 (Arg14, Arg74, Arg338, Lys279); these interactions are shown in **Figure 5C** and are in
362 strong agreement with the above MD simulations.
363
364 We calculated the electrostatic potential of the transmembrane domain and acetyl-CoA, and
365 found them to be complementary: the transmembrane domain displayed positive potentials,
366 while the acetyl-CoA displayed negative potentials, of similar magnitudes (**Fig. 5D,E**). The
367 greatest positive potentials are found towards the cytoplasmic side of the transmembrane
368 domain (in agreement with the positive inside rule (88)), and the greatest negative potentials
369 are found at the phosphate groups of acetyl-CoA (which are positioned towards the
370 cytoplasmic side of the AT3 domain). Together these data are strongly supportive of a model
371 whereby acetyl-CoA from the cytoplasmic side is able to penetrate the enzyme to bring the
372 acetyl group into range for use in catalytic transfer to the receptor sugar on the periplasmic
373 face of the membrane.
374
375
376

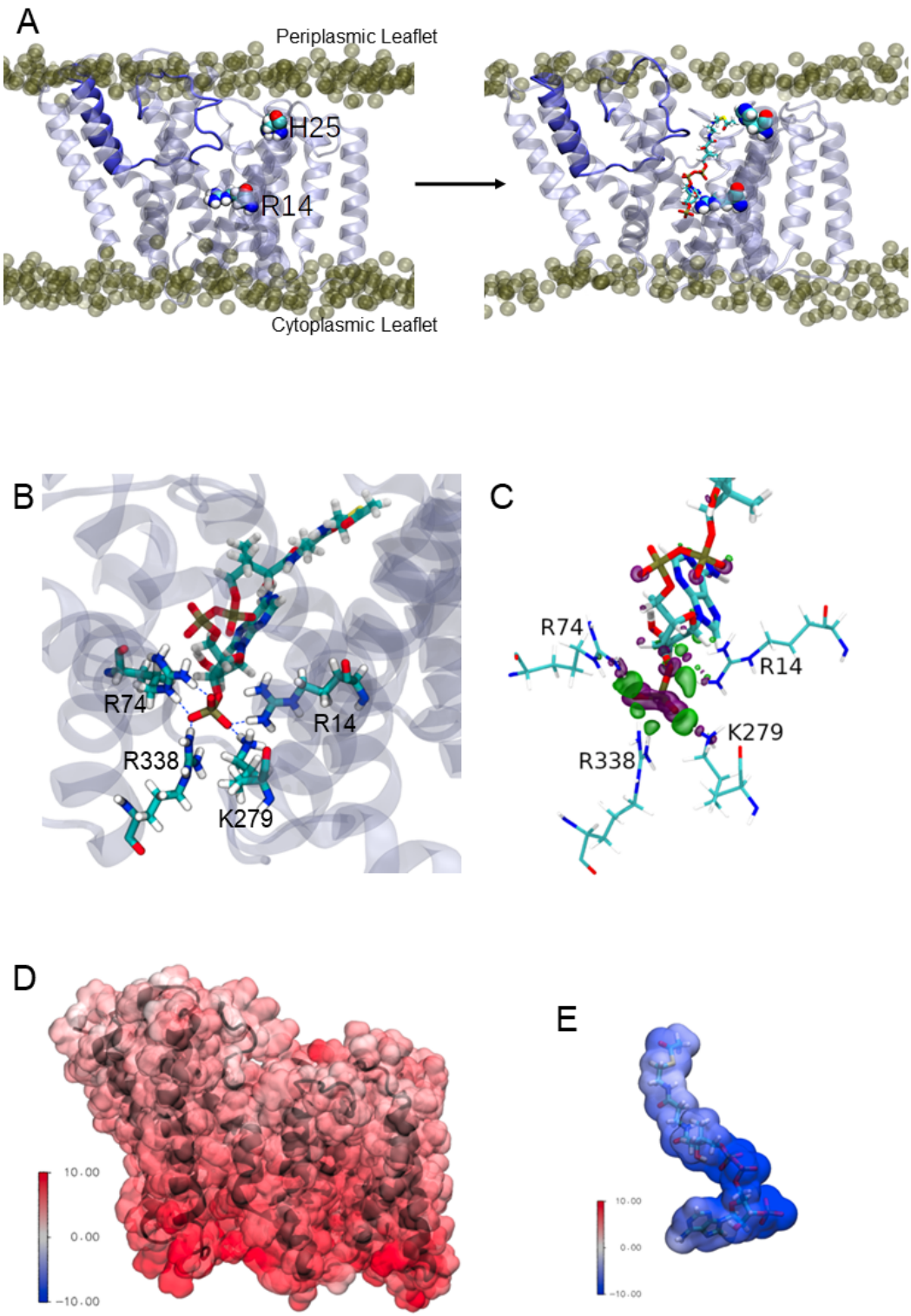


Figure 5: Interactions between the Transmembrane domain of OafB and the putative acetyl donor molecule, acetyl coezyme-A. Note these structures are taken from simulations in which we use the full OafB protein, but only the transmembrane domain (residues 1-377) is shown here for clarity. **(A)** Left: Initial structure of the OafB transmembrane domain. Essential residues H25 and R14 shown as spheres, the loop between helices 3 and 4 is shown in dark blue. Phosphate headgroups of the phospholipids shown as tan spheres. The loop initially occludes the pore in the AT3 domain. Right: Structure from the end of the steered MD simulation in which acetyl

384 Coenzyme-A was pulled into the central channel within the AT3 domain. The loop between helices 3 and 4 has
385 moved away from the center of the pore towards TMH2 (to the left) sufficiently to allow passage of acetyl-coa into
386 the transmembrane domain. **(B)** A pocket of basic residues is observed in the AT3 domain, complementary to the
387 3'-phosphate of acetyl coenzyme-A. Several high occupancy hydrogen bonds are observed between the 3'-
388 phosphate and transmembrane domain residues R14, R74, R338, and K279. **(C)** Charge transfer interactions
389 between acetyl coenzyme-A and the transmembrane domain of OafB, identified via ONETEP Energy
390 Decomposition Analysis. Loss of electron density is depicted as a purple surface and gain of electron density in a
391 green surface. Significant charge transfer interactions were identified between the 3'-phosphate of acetyl
392 Coenzyme-A (losing electron density) and surrounding basic residues R14, R74, R338, and K279 (gaining
393 electron density) **(D)** Electrostatic potential of the OafB transmembrane domain. Protein in black New Cartoon
394 representation; calculated electrostatic potential overlayed in Surface representation. Scale from -10 V (blue) to
395 +10 V (red). **(E)** Electrostatic potential of acetyl Coenzyme-A. Molecule in Licorice representation; calculated
396 electrostatic potential overlayed in Surface representation. Using the same scale as the AT3 domain, it is clear
397 that acetyl coenzyme-A and the AT3 domain are complementary.
398

399 Modelling suggests the two domains in OafB could function in a pedal bin 400 mechanism

401 To learn more about the function of the model OafB proteins, which requires both the AT3
402 domain and its linked SGNH domain to function, we first constructed a hybrid structure that
403 combined the RaptorX predicted structure of the transmembrane domain (residues 1-377)
404 with the x-ray crystal structure (380-640) of the SGNH-ext (380-421) and the SGNH domain
405 (422-640). Residues 378-379 were added to the N-terminus of the SGNH domain using
406 MODELLER 10.0 (89), and a peptide bond was generated between residues 377 and 378 in
407 ChimeraX (90) with varying C-N-C_α-C dihedral angles (80, 100, 120, 140°) to generate
408 proteins with differing relative conformations of the AT3 and SGNH domains (henceforth
409 referred to as $OafB_{80}$, $OafB_{100}$, $OafB_{120}$, and $OafB_{140}$). To investigate the flexibility of the
410 hybrid structure we used equilibrium MD simulations (see Methods).
411
412 The x-ray crystal structure of the SGNH-ext region of OafB (377-421) (58) shows it to be
413 structured and interact substantially with the SGNH domain. The biological relevance of this
414 close interaction is supported by the equilibrium MD simulations of $OafB_x$: the SGNH-ext
415 remains structured and packed close to the SGNH domain throughout the 4 x 250 ns
416 simulations, indicating this fold is stable under physiological conditions. Unfolding of the
417 SGNH-ext to generate a longer flexible linker - such as that seen in the RaptorX predicted
418 structures of OatA-LM and OatA-SA - is not observed.
419
420 A short stretch of only around 10 residues (~370-380) remains that could form a flexible
421 linker between the transmembrane and periplasmic domains. Indeed, in the MD simulations,
422 the periplasmic domain exhibits a wide range of orientations relative to the transmembrane

423 domain. Principal component analysis (PCA) of the backbone atoms was used to extract the
424 first 2 major motions (accounting for 77% or more of the variance) of the protein in the four
425 equilibrium replicates. The first of these motions was a pedal bin-like action, where the
426 periplasmic domain 'lid' opens and closes relative to the transmembrane domain, hinged at
427 the flexible linker. The second motion was the rotation of the SGNH domain within the
428 periplasmic space about the linker. These two motions can be seen in **Videos 1 and 2**.
429
430 Despite these motions and the range of relative orientations, the SGNH and AT3 domains
431 were not observed to spontaneously interact in the equilibrium simulations. This does not
432 necessarily mean that the domains do not interact *in vivo*; given the short timescales of MD
433 simulations and the vast conformational space available to the protein, sampling all possible
434 conformations is an intractable problem. To direct computational efforts to the
435 conformational space of interest, elastic network bonds were applied to pull the
436 transmembrane and periplasmic domains together.
437
438 Elastic networks are a set of artificial harmonic 'bonds' added to a molecular model. These
439 are commonly used in coarse-grained models to maintain the secondary structure of
440 molecules (91); here, we use them to induce and then maintain a change in the tertiary
441 structure of $OafB_x$ by pulling the periplasmic and transmembrane domains towards each
442 other. The 'bonds' were added between residues likely to be proximal to one another as
443 identified through coevolution analysis (see Methods). Around 75% of co-evolving residues
444 have heavy atoms (carbon, oxygen, nitrogen, sulphur) within 5 Å of one another (92). The
445 side chains of these residues were generally 3-5 Å in length; to avoid artificial re-orientation
446 of the side chains but still bring the residues close enough to interact, elastic bonds of length
447 10 Å were added between the alpha carbons of the residue pairs. Applying elastic network
448 bonds between the transmembrane and periplasmic domains of the $OafB_x$ proteins resulted
449 in the rapid and irreversible 'closing' of the protein (the closed state), for the duration of the
450 simulation time that the elastic network bonds were applied.
451
452 The elastic network bonds were then removed from these closed state structures, and over
453 the subsequent 100 ns equilibrium MD simulations, all four replicates remained closed
454 throughout. In this configuration, inter-domain interactions, sufficiently stable to maintain the
455 closed state without the elastic network, are identified. Hydrogen bond analysis (distance
456 and angle cutoffs of 3 Å and 20°, respectively) of the subsequent equilibrium simulations
457 revealed a minimum of 35 unique hydrogen bonds between the periplasmic and
458 transmembrane domains across each simulation. When averaged across all four
459 simulations, the number of hydrogen bonds between the two domains at any given time was

3.16 ± 1.67 (**Fig. 6A**). Three of the higher occupancy hydrogen bonds found in all four replicates reside close to the linker region, as shown in **Figure 6B**. Specifically, these are Lys132-Asp387 (all >51% occupancy), Thr386-Asp98 (>25% occupancy), and Ser129-Asp387 (>70% occupancy in 3 of the 4 simulations, 23% in the fourth). Additional hydrogen bonds were observed further from the linker in each system, on the other side of the catalytic triad: for example, Leu408-Glu189 (34% occupancy) and Ser474-Glu189 (21%) are observed in the $OafB_{100}$ system (**Fig. 6C**), but do not form simultaneously (**Supp. Fig. S4**). Taken together, these findings from equilibrium MD simulations identify interactions between the residues in the AT3 membrane domain and the periplasmic domains, supporting the hypothesis that the two domains may cooperate.

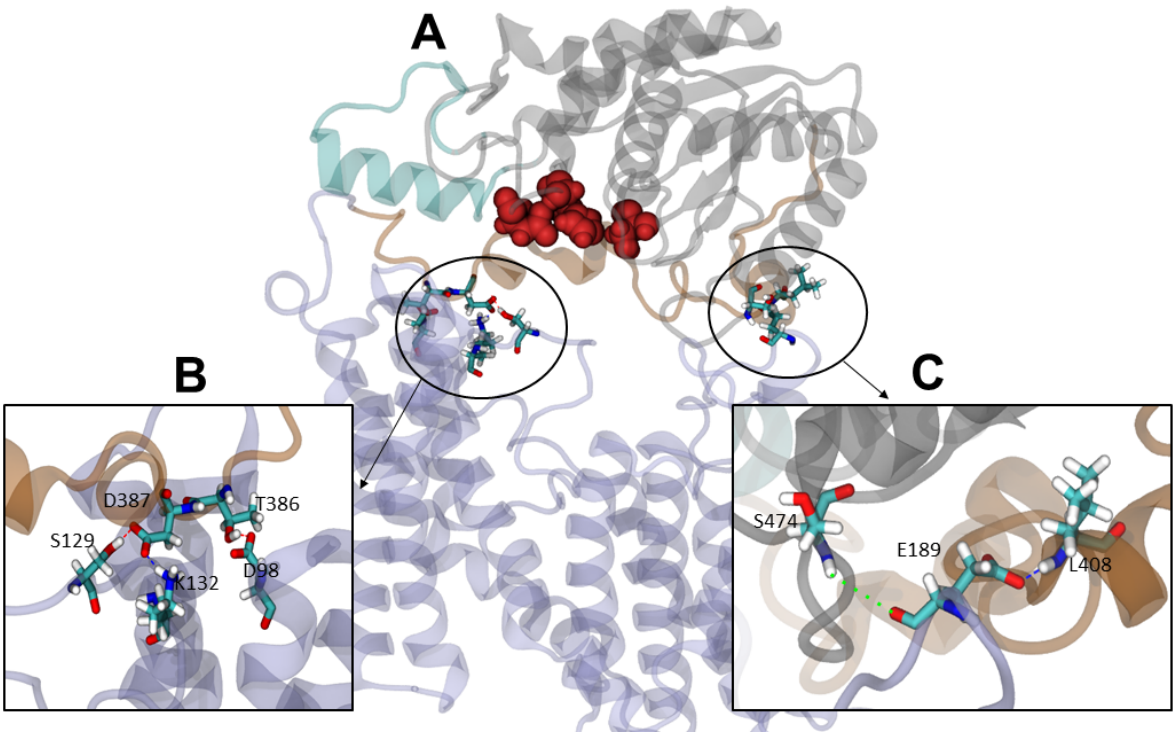
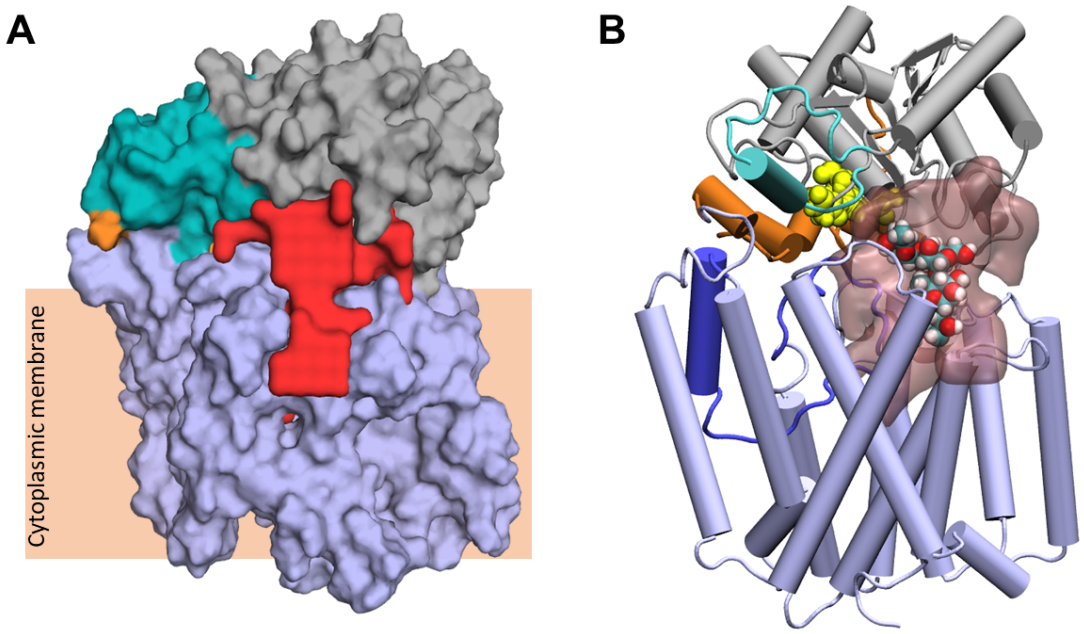


Figure 6: Interactions between the transmembrane and periplasmic domains in the closed state of OafB. **(A)** Closed OafB structure in New Cartoon representation (TM domain in lilac, SGNH-ext in orange, α8 in teal, SGNH domain in grey). Catalytic triad (D618, H621, S430) as red spheres. Residues identified in high-occupancy or important hydrogen bonding interactions between the SGNH and AT3 domains in Licorice representation. **(B)** Hydrogen bonding observed between SGNH-ext and AT3 domains in all equilibrium simulations of the closed OafB structure. High-occupancy hydrogen bonding between side chain: carboxylate of E387 and hydroxyl of S129, amine of K132; carboxylate of E98 and hydroxyl of T386. **(C)** Hydrogen bonding between the SGNH and AT3 domains observed in the $OafB_{100}$ simulations. E189 can hydrogen bond to S474 via its backbone carbonyl (green dashed line), or to L408 via its carboxylate side chain (blue dashed line) but cannot form these interactions simultaneously. Timeseries of the separation of these residues shown in Supplementary Figure S4.

483 These fully closed structures were submitted to the ProPores2 webserver (93) to identify and
484 parameterise cavities within the protein. The closing of the pedal bin converts the outer
485 cavity from the transmembrane domain into a larger enclosed cavity at the interface of the
486 AT3 and SNGH domains (**Figure 7A**) This pore has a volume of ~4,780 $Å^3$ and includes
487 both the SGNH catalytic triad comprising residues Asp618, His621 and Ser430, and at the
488 membrane surface the proposed catalytic residues of the AT3 domain, namely, His25 and
489 Tyr194. This raises the possibility that the transfer of the acyl-group to the acceptor molecule
490 could occur in this region of the protein.



491 **Figure 7:** Largest cavity in the closed state of OafB. **(A)** Largest cavity identified by the ProPores2 webserver in
492 the closed OafB structure shown as a red surface. The cavity extends down into the transmembrane domain. **(B).**
493 Closed OafB in Cartoon representation: transmembrane domain coloured light blue, periplasmic loop between
494 TMH 3-4 coloured dark blue, SGNH domain coloured grey with the periplasmic linking region in orange, and
495 additional helix in the SGNH domain in teal. The SGNH catalytic triad (Ser430, Asp618, His621) is shown as
496 yellow spheres. Largest pore identified by the ProPores2 Webserver shown as a transparent pink surface. A
497 single O-antigen unit is shown inside the pore in van der Waals representation, coloured by atom name (cyan for
498 carbon; red for oxygen; white for hydrogen). The single O-antigen unit can be comfortably accommodated in this
499 space in several orientations.
500

501 **Fully closed structure exhibits cavity able to accommodate the LPS O-antigen**

502 Finally, we considered the potential mechanisms by which the acceptor molecule(s) would
503 interact with open and closed 'pedal bin' model of OafB, working with two alternative
504 hypotheses that the acceptor is presented either as a single or multiple O-antigen repeat
505 unit(s) anchored to a lipid carrier or later after transfer to the lipidA core molecule.
506

507 The closed form of the pedal bin model presents a possible enclosed catalytic site for OafB,
508 which is known to acetylate the rhamnose residue in the O-antigen repeating unit of the LPS
509 (10). This cavity resides directly below the catalytic triad (Ser430, His621, Asp618), and
510 extends down into the AT3 domain (**Fig. 7A**). The measured volume of ~4,780 $Å^3$ is an order
511 of magnitude greater than the volume of the LPS O-antigen unit (volume of ~502 $Å^3$), hence
512 one or more units could be manually placed within the cavity (**Fig. 7B**). Importantly the
513 enclosed cavity on the periplasmic surface of the membrane does extend into the
514 membrane, the lower portion residing below the plane of the upper leaflet headgroups and
515 hence is exposed to the membrane core (**Fig. 7A**), consistent with a ligand delivered on a
516 lipid anchor.
517
518 *S. enterica* serovars Typhimurium and Paratyphi use the prevalent Wzx/Wzy-dependent
519 pathway for O-antigen assembly, in which a single O-antigen unit is assembled in the
520 cytoplasm on undecaprenyl phosphate (Und-PP) in the membrane, which is then flipped
521 placing the O-antigen unit in the periplasm. Lipid linked O-antigen units are assembled to
522 the final length in the periplasm before ligation to lipid A-core. Subsequently, the completed
523 LPS molecule is transported to the outer membrane (94). Thus, the question remains, does
524 OafB acetylate the rhamnose moiety of a single O-antigen unit bound to Und-PP or after O-
525 antigen polymerisation? If the latter, does acetylation occur with the polymerised OAg
526 unit(s) attached to the Und-PP or after transfer to the lipidA-core molecule?
527
528 To contextualise the OafB models with biologically relevant molecules, a system containing
529 one molecule of the lipid carrier undecaprenyl phosphate (Und-PP) attached to an O-antigen
530 unit was placed next to the OafB model (**Fig 8A**). A distance of ~9 Å separates O3 of the
531 rhamnose from the periplasmic surface of the membrane, which is similar to the distance of
532 the SGNH active site (Ser430) from the membrane (which was ~12 Å calculated with
533 $OafB_{100}$ after 100 ns with elastic network bonds applied), meaning it is possible that the
534 substrate diffuses laterally into the cavity, either with the pedal bin open or closed.
535
536 The analysis above suggests that acetylation could occur using single O-antigen units,
537 before O-antigen polymerisation. Alternatively, acetylation occurs during or after
538 polymerisation on the Und-PP carrier before transfer to the lipidA-core (not shown). The
539 distance of multimers of 9 Å (single O-antigen unit) stretching away from the membrane
540 could possibly be accommodated by the open state of the pedal bin with the SGNH domain
541 extending into the periplasm.
542
543

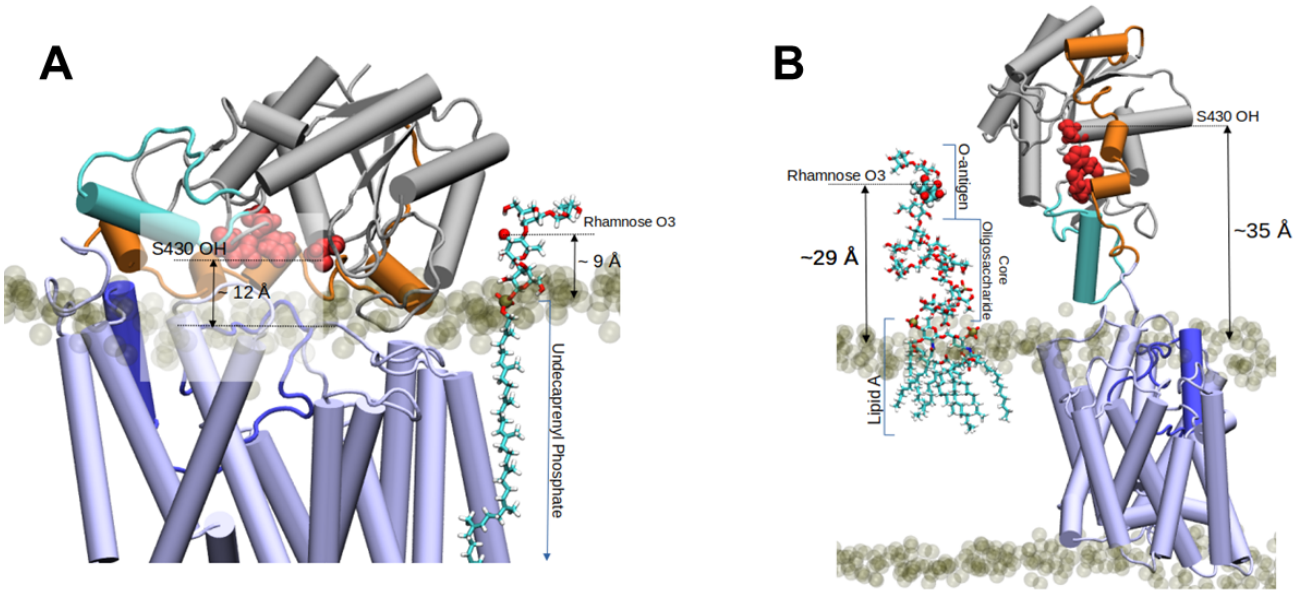


Figure 8: Two proposed models for the O-antigen rhamnose moiety acetylation by OafB. In both panels, OafB is shown in a cartoon representation with the AT3 domain coloured light blue, periplasmic loop between TMH3-4 coloured blue, SGNH domain coloured grey, the periplasmic linking region in orange, and additional helix in the SGNH domain in teal. Catalytic triad of the SGNH domain (Ser430, Asp618, His621) shown as red spheres. POPE and POPG phosphate headgroups shown as transparent tan spheres. **(A)** Undecaprenyl phosphate carrier lipid with a single O-antigen unit in Licorice representation. O3 of the rhamnose moiety (marked by a red sphere) has a vertical separation from the periplasmic leaflet surface of ~9 Å. The OH of S430 has a vertical separation from the bulk membrane of ~12 Å. **(B)** OafB-LPS system after equilibration steps - one LPS molecule with a single O-antigen unit shown in Licorice representation, and its rhamnose moiety shown in van der Waals representation. OafB remains in an open conformation. After relaxation, the oligosaccharide of LPS has kinked, with the distance between the lipid headgroups of the periplasmic leaflet surface and the acetylation site reduced to ~29 Å (from ~42 Å, as shown in Figure S5). This is comparable to the ~35 Å separation between the OH of S430 and the periplasmic leaflet surface

To investigate the final model of acetylation after assembly onto the lipidA core, we built a model with LPS (with one O-antigen unit) and OafB-SPA embedded in the inner membrane to assess the distances defined between the periplasmic leaflet surface and various groups within the system relevant to the acetylation (**Fig. 8B**). One molecule of *Salmonella* LPS was generated in CHARMM-GUI's LPS Modeler and inserted into the periplasmic leaflet of the $OafB_{100}$ system (replacing 2 POPE and 2 POPG molecules). Initially the modelled LPS adopts a quasi-linear conformation; the z-separation between the rhamnose moiety and the periplasmic membrane surface was found to be ~42 Å, as shown in **Supplementary Figure S5**. During minimisation and equilibration, the oligosaccharide was allowed to relax; this resulted in a kinked structure, where the distance between the acetylation site and the periplasmic leaflet surface was reduced to around 29 Å (**Fig. 8B**). By comparison, after equilibration, the z-separation of the hydroxyl of Ser430 in the catalytic triad of $OafB_{100}$ and the periplasmic leaflet is ~35 Å, hence it is not inconceivable that the open pedal bin could

571 function to deliver acetyl groups at a distance from the membrane. However, we note that in
572 reality the full-length O-antigen will be attached to the lipidA core, which would require
573 significant, entropically unfavourable conformational rearrangement of the O-antigen for
574 OafB to catalyse significant overall levels of O-acetylation across the whole polymer.

575 Steered molecular dynamics (SMD) was used to pull the rhamnose moiety of LPS and Ser
576 430 of OafB towards one another, and subsequent restrained MD simulations held the
577 protein and LPS close to one another to assess possible interactions. Many hydrogen bonds
578 were observed between the O-antigen and the periplasmic domain of OafB over the two
579 replicates. Several of these were with residues from the SGNH-ext (galactose-Asp387;
580 galactose-Tyr389; rhamnose-Tyr394). Tyr389 and Tyr394 are both residues which were
581 previously identified by FTMap *in silico* docking analysis as involved in binding to rhamnose
582 sugars. Hydrogen bonding was also observed between the 3' and 2' hydroxy groups of the
583 rhamnose moiety and Ser430. Even when all position restraints were removed in the MD
584 simulations, this interaction persisted for a further 10 ns.

585 Together these analyses indicate that O-acetylation of the O-antigen may occur either when
586 a single O-antigen unit attached to a lipid carrier is presented in the periplasm, or when the
587 O-antigen is polymerized, either on the lipid carrier, or when attached to the lipid A-core.
588

589 Ideas and Speculation

590 The realisation that diverse bacteria use AT3 proteins in a plethora of different biological
591 processes for the modification of extracytoplasmically localised glycans, directed our efforts
592 into describing the fundamental mechanism of these proteins (1). While the role of AT3
593 proteins in these diverse processes is often implicated from genetics only, the OafB protein
594 has been studied in much more detail, along with other selected AT3 proteins including the
595 OatA protein from *S. aureus* (10,57,58). These systems are examples of AT3 proteins
596 working with SGNH domains to acylate LPS and peptidoglycan, respectively, which are key
597 components of bacterial envelopes.
598

599 Integration of experimental and computational data has provided detailed insight into the
600 structure and potential mechanism of action of AT3-SGNH proteins. Our structural model,
601 combined with molecular dynamics simulations, suggests first that the transmembrane
602 domain of OafB has a novel fold. Importantly this is distinct from the MBOAT family of
603 membrane-bound acyltransferases, suggesting that nature has evolved two independent
604 routes to solve this problem. The discovery of a large cytoplasmic cavity, which can open
605 dynamically to accommodate an acyl-CoA molecule that spans the membrane, provides a

key mechanistic breakthrough in our understanding of the protein function, which in OafB is mediated by a loop region which would need to block the pore in the absence of donor/acceptor to prevent leakage of protons.

The putative catalytic residues His25 and Tyr194 are close enough to the acetyl-group (~ 5 Å) to conceivably be involved in catalysis, although the MD simulations cannot capture bond making or breaking (nor proton transfer). The loop between TMH 5 and 6, in which Tyr194 resides, shows significant mobility and the side chain of Tyr194 is observed to point inwards, towards the centre of the pore (and therefore towards acetyl-CoA). Given the conformational mobility of the 4'-phosphopantetheine segment of the coenzyme, and of the loop between TMH 5 and 6, it is not unreasonable to suggest that the thioester bond is accessible to both His25 and Tyr194 to aid the acetyl transfer in a mechanism such as that suggested by Jones et al (57). Addressing the role of Glu334, which Jones et al (56) proposed was the third residue in the catalytic site (Glu357 in OatA), our model places this distantly from the His25 and Tyr194, but yet in OafB we also know this residue is important for function (79). The residue sits close to the cytoplasmic side and the internal cavity and perhaps has a role in the initial recognition of the acyl-CoA. Other conserved residues, including Phe42, Tyr124, Trp138 and Glu143 in OafB, which were identified by mutagenesis as being important for both the OafB-related protein OafA and OatA, cluster on the extracytoplasmic surface and could have related roles in acyl-CoA binding, catalysis or acceptor recognition.

While the function of the AT3 to bind and deliver acyl-groups to a periplasmic catalytic site is well supported by our model, and must be a conserved feature of all AT3 proteins, the further discrete role of the SGNH domain is still not fully understood, although is essential for transferase function in both OafB and OatA (57,58). The linker region, while short and structured in OafB suggests that the domain can function in a 'pedal bin' like model, opening and closing to define a large cavity on the extracytoplasmic surface, which we demonstrate could then capture a lipid-anchored O-antigen repeat structure. For OatA and other related proteins, this linker can be much longer (**Fig. 2**) and whether this then allows the pedal bin lid to close and accept the acyl-group and then open and extend away from the membrane to reach the acceptor substrate needs further investigation, but the MD simulations suggest it is possible. However, for the OafB system we favour the model of lid opening and closing to help trap the substrate in a membrane-surface-located activity site, similar to the function of the Sus protein in capturing ligand in the outer membrane of *Bacteroidetes* bacteria (95). For LPS O-acetylation by OafB and similar proteins we favour the model where the substrate is the single O-antigen repeat bound by the Und-PP lipid anchor before later polymerisation. Given that overall levels of OafB-dependent O-acetylation vary between 40-70% (10,96), the

643 stochastic interaction of OafB with the Und-PP anchored O-antigen repeat appears a more
644 parsimonious mechanism than relying on the enzyme trying to reach rhamnose sugars in a
645 fully assembled O-antigen.
646
647 Beyond LPS and peptidoglycan, AT3 domain-containing proteins acylate a diverse range of
648 complex carbohydrates in bacteria (1). Each receptor substrate may present its own unique
649 challenges for the AT3 domain. In both the LPS and peptidoglycan examples, additional
650 SGNH domains are required for function, however, AT3 domains can function as standalone
651 proteins. One example of this is IcaC, which is involved in the O-succinylation of poly N-
652 acetylglucosamine (PNAG) biofilm component of *Staphylococci.* Here, the acceptor is not
653 delivered on a lipid anchor, but rather through a continuous synthesis and extrusion
654 mechanism catalysed by IcaAD, a 'synthase' mechanism (94). Perhaps a close physical
655 association of IcaC with IcaAD removes the need for the additional SGNH domain and
656 allows direct transfer of the acyl group onto the emerging PNAG polymer giving stochastic
657 levels of O-succinylation of about 40% (97,98). Hence, the differences between AT3 and
658 AT3 with additional domains such as SGNH, perhaps could relate to the route of
659 presentation of the acceptor molecule within the context of different biosynthetic pathways
660 for extracytoplasmic glycans.
661

662 # Conclusions

663 Here we provide strong evidence that the integral membrane AT3 domain (PF01757),
664 mediating acylation of a diverse range of complex carbohydrates, has a novel fold. The
665 stability of this structure was confirmed using molecular dynamics simulations, which also
666 enabled insights into the mechanism of action of AT3 domains. Importantly, for the first time
667 our model provides a solution to the problem of acetylation occurring in the periplasm using
668 the cytoplasmic acetyl donor acetyl-CoA. We show a membrane-spanning pore occurs
669 transiently in the AT3 domain which can accommodate acetyl-CoA, presenting the acyl
670 group to the periplasmic side. This part of our model is supported by quantum-level
671 calculations to probe the hypothesised protein-acetyl-CoA interactions, and the presence of
672 key, conserved and essential residues that line the pore. The second cavity at the
673 periplasmic side in OafB is able to accommodate the receptor substrate O-antigen.
674 Together, our data give important new insights into the family of AT3 domain containing
675 proteins. Our well-supported model can support and drive targeted research to gain a full
676 understanding of the mechanism of action of this family or membrane bound acyltransferase
677 proteins that mediate acylation of complex carbohydrates across the domains of life.
678

Methods

Co-evolution and protein structure predictions

OafB Protein sequence (Uniprot accession A0A0H2WM30) was submitted to either the RaptorX contact prediction server (66–68) or AlphaFold server (76) and analysis was run with default parameters. Predicted structures using RCSB PDB Pairwise Structure Alignment webserver, employing the flexible jFATCAT alignment algorithm (99).

Molecular dynamics simulations

All simulations used the GROMACS simulation package (version 2020.3) (100). Results were analysed using GROMACS tools and in-house python scripts utilising MDAnalysis (101–104). Molecular graphics were generated using VMD 1.9.4 (105).

Atomistic simulations

Atomistic simulations used the CHARMM36m forcefield (106) with the TIP3P water model (107). A cut-off of 1.2 nm was applied to Lennard-Jones interactions and short-range electrostatics using the potential shift Verlet scheme. Long-range electrostatics were treated using the particle mesh-Ewald (PME) method (108). Atoms were constrained using the LINCS algorithm to allow the use of a 2 fs timestep (109). For production simulations, temperatures were maintained using the Nosé-Hoover thermostat (110,111) (1.0 ps coupling constant), and pressure was maintained at 1 bar using the Parrinello-Rahman semi-isotropic barostat (112) (5.0 ps coupling constant).

Transmembrane Domain Model

A model of the transmembrane domain (residues 1-376, i.e. the 10 TMH of the AT3 domain and the 11th TMH) was generated using the RaptorX webserver (66,67). Using the CHARMM-GUI membrane builder module (113), this structure was embedded within a model *Escherichia coli* inner membrane: a symmetric bilayer of 18:1:1 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE), 1-Palmitoyl-2-oleoyl-sn-glycero-3-phosphoglycerol (POPG), and 1’,3’-bis[1-palmitoyl-2-oleoyl-sn-glycero-3-phospho]glycerol (cardiolipin). The bilayer system was subsequently solvated in 0.15 M KCl. The system was energy minimised in 5000 steps using the steepest descent method (114). The subsequent structure was equilibrated in six phases in which the protein and lipid head groups were subjected to position restraints with varying force constants. The full equilibration protocol is detailed below. Two short, independent equilibrium molecular dynamics simulations were carried out on the equilibrated system: one simulation of 100 ns at 303.15 K, and one of 50 ns at 320 K.

715 **Equilibration Protocol**

716 Each system was equilibrated using two NVT phases followed by four NPT phases. NVT
717 phases used a timestep of 1 fs and each lasted 0.125 ns, and the NPT stages used a
718 timestep of 2 fs and each lasted 0.5 ns. The velocity-rescaling thermostat (115) was applied
719 at all stages to bring the system to 303.15 K with a coupling constant of 1.0 ps. Semi-
720 isotropic Berendsen pressure coupling (116) was applied to the NPT phases to equilibrate
721 with a pressure bath of 1 bar (τ_P = 5.0 ps, compressibility of 4.5×10^{-5} bar^{-1}).

722

723 **Table 1:** Position restraints used in the equilibration steps for all atomistic systems

Equilibration Phase	Position restraint / kJ mol^{-1} nm^{-2}			
	Protein backbone	**Protein Sidechains**	**Lipid head groups**	**Dihedrals**
NVT1	4000	2000	1000	1000
NVT2	2000	1000	400	400
NPT1	1000	500	400	200
NPT2	500	200	200	200
NPT3	200	50	40	100
NPT4	50	-	-	-

724

725 **Full OafB model**

726 The RaptorX transmembrane domain model and periplasmic domain crystal structure (PDB
727 ID: 6SE1) were used to build a full OafB model. Missing linker residues (residues 377-380)
728 were modelled into the N-terminus of the SGNH domain using MODELLER 10.0 (89). A
729 peptide bond was generated between the N-terminus of the resulting periplasmic domain
730 (K377) and the C-terminus of the transmembrane domain (N376) using UCSF ChimeraX
731 (90). Four structures were generated, varying the C-N-C_α-C dihedral angle; values of 80,
732 100, 120, and 140° were used to give structures with different relative orientations of the two
733 domains. The C_α-C-N-C_α dihedral was maintained at 170°. These proteins will henceforth be
734 referred to as $OafB_x$, where x is the C-N-C_α-C dihedral angle (80, 100, 120, or 140°).

735

736 The transmembrane domain of each model was embedded in an *Escherichia coli* inner
737 membrane model and solvated in the CHARMM-GUI membrane builder as described for the
738 RaptorX model. Systems were energy minimised in 5000 steps using steepest descent
739 algorithm, and subsequently equilibrated using the protocol described above.

740

741 Duplicates of each equilibrated system were generated. The first replicate of each system
742 was simulated for 250 ns at 303.15 K. The topologies of the second replicates were modified
743 to add elastic network bonds that would bring the periplasmic and transmembrane domains

744 together. Elastic bands (of length 1 nm and strength 100 kJ mol^{-1} nm^{-2}) were generated
745 between the alpha carbons of 10 residue pairs, identified through coevolution analysis, to be
746 more than 50% likely to be proximal to one another.
747
748 **Table 2:** Residues identified through co-evolution as likely to be proximal to one another

Residue 1 (TM domain)	Residue 2 (periplasmic domain)	% Probability
94	545	56.1
95	546	80.5
96	546	57.2
97	546	70.5
98	546	58.4
123	458	53.3
125	403	59.2
126	404	55.9
194	627	50.9
242	503	51.3

749
750 These four replicates were simulated for 100 ns with the elastic network bonds applied. A
751 subsequent 100 ns simulation was undertaken on each system with the network removed.
752

753 **Full OafB Model: AlphaFold**

754 For comparison to the hybrid models built above, the OafB amino acid sequence was
755 submitted to AlphaFold (76) to generate a structure prediction for the full protein. The RCSB
756 PDB Pairwise Structure Alignment web service (99) was used to compare our structures to
757 the AlphaFold prediction. Residues 1-327 of the RaptorX model, residues 328-406 of
758 OafB80, and residues 407-640 of the periplasmic domain crystal structure (PDB ID: 6SE1)
759 were aligned to the equivalent residues in the AlphaFold model using the flexible jFATCAT
760 algorithm.
761

762 **Acetyl Coenzyme A**

763 One molecule of acetyl coenzyme A (acetyl-coA) was generated in CHARMM-GUI's Ligand
764 Reader & Modeler (117). Using VMD as a visual tool to guide positioning, the acetyl group of
765 acetyl-coA was placed beneath the gap between transmembrane helices 9 and 10, and the
766 N-terminus of the $OafB_{100}$ protein. The protein was embedded in the *Escherichia coli* inner
767 membrane model in CHARMM-GUI and solvated in 0.15 M KCl. The system was energy

768 minimised using the steepest descent algorithm, and subsequently equilibrated using the
769 protocol described above.
770
771 A short steered molecular dynamics (SMD) simulation was used to insert acetyl-coA into the
772 pore (delimited by transmembrane helices 1, 2, and 9, and the periplasmic loops between
773 helices 5 and 6, and 3 and 4). Pulling at a rate of 0.5 nm ns^{-1} and with a harmonic force of
774 1000 kJ mol^{-1} nm^{-2}, the sulphur of the acetyl-coA molecule was pulled upwards, towards the
775 centre of mass of the alpha carbons of residues E243, D126, E189 (all in loops at the
776 periplasmic surface of the transmembrane domain, surrounding the pore) until the z-
777 separation of the two groups was 0 (7.5 ns).
778
779 Three variations on the final frame of the SMD were generated in VMD by rotating and
780 translating acetyl-coA within this pore to use as starting conformations for equilibrium MD.
781 These systems were energy minimised using the steepest descent algorithm to resolve
782 steric clashes, and equilibrated as described above, with additional position restraints (**Table**
783 **3**) on the acetyl-coA molecule. Each system was subjected to three consecutive MD
784 simulations of 20 ns at 303.15 K, with decreasing position restraint strength on acetyl-coA to
785 allow the protein to relax around the substrate.
786
787 **Table 3:** Position restraints used in the equilibration and simulation of the OafB-acetyl CoA
788 systems.

Equilibration Phase	**Position restraint / kJ mol^{-1} nm^{-2}**				
	Protein backbone	**Protein Sidechains**	**Lipid head groups**	**Dihedrals**	**Acetyl-coA**
NVT1	4000	2000	1000	1000	4000
NVT2	2000	1000	400	400	4000
NPT1	1000	500	400	200	4000
NPT2	500	200	200	200	3000
NPT3	200	50	40	100	3000
NPT4	50	-	-	-	2000
MD1	-	-	-	-	1000
MD2	-	-	-	-	500
MD3	-	-	-	-	-

789
790
791
792 **LPS in the Periplasmic Leaflet of the Inner Membrane**
793 One molecule of *Salmonella spp.* LPS was generated using the CHARMM-GUI LPS
794 modeller (O9-4 core with one O-antigen unit). Using VMD as a tool to guide positioning, 2

795 POPE and 2 POPG molecules were removed from the periplasmic leaflet of the $OafB_{100}$
796 system and the LPS lipid A moiety inserted in their place. 8 water molecules were replaced
797 by potassium ions to neutralise the additional negative charge of the LPS. The system was
798 energy minimised in 5000 steps via the steepest descent algorithm to resolve steric clashes.
799 The system was equilibrated in 6 phases, as described for the other atomistic systems.
800
801 Steered MD was used to guide the rhamnose moiety of the LPS towards the Serine 430
802 residue (2 replicates, one with a pull rate of 0.5 nm ns^{-1}, the other with a pull rate of 1 nm ns^{-}
803 ^{1}; both with a pull force of 500 kJ mol^{-1}). The final frame of each of these trajectories (13.1 ns
804 and 6.5 ns, respectively) was used as the starting conformation for 3 consecutive MD
805 simulations at 303 K of 20 ns each, with varying position restraints on the protein backbone
806 (250, 0, 0 kJ mol^{-1}) and LPS molecule (1000, 500, 0 kJ mol^{-1}). Hydrogen bond analysis was
807 undertaken in VMD (distance cutoff = 3 Å, angle cutoff = 20°) to assess the interactions
808 between the LPS and the SGNH domain.
809
810 **DFT Calculations**
811 The electrostatic potentials of the transmembrane domain of OafB (residues 1-370) and the
812 acetyl-coA molecule were calculated using the linear-scaling density functional theory
813 package, ONETEP (86,87), using the PBE exchange-correlation functional (118),
814 augmented with Grimme's D2 dispersion correction (119). Open boundary conditions *via*
815 real-space solution of the electrostatics were used in a simulation cell with dimensions 9 nm
816 × 7 nm × 8 nm. Norm-conserving pseudopotentials were used for the core electrons, and the
817 psinc basis set, equivalent to a plane wave basis set with a kinetic energy cut-off of 800 eV,
818 was employed. 8.0 Bohr localisation radii were used for the nonorthogonal generalised
819 Wannier functions (NGWFs). The energy decomposition analysis (EDA) implemented in
820 ONETEP (120,121) was used to extract polarisation and/or charge transfer contributions to
821 the interaction energies between Acetyl-CoA and the AT3 domain.
822

823 # Acknowledgements

824 ● The authors acknowledge access to the following High Performance Computing
825 resources: Iridis 5 at the University of Southampton and the JADE Tier 2 facility
826 (EPSRC grant no. EP/T022205/1) to which access was granted via HECBioSim, the
827 UK High-End Computing Consortium for Biomolecular Simulation (EPSRC grant no.
828 EP/R029407/1)
829 ● The Chris Skylaris Group at the University of Southampton for help with ONETEP

- K.E.N. was supported by a Ph.D. Studentship from EPSRC (Project Number: 2446840).
- S.N.T. was supported by a Ph.D. studentship from the Biotechnology and Biological Sciences Research Council White Rose Doctoral Training Program (BB/M011151/1), “Mechanistic Biology and its Strategic Application.”
- SLM acknowledges the support of the Federation of European Biochemical Societies (FEBS) through a long-term fellowship.

References

1. Pearson CR, Tindall SN, Potts JR, Thomas GH, Van der Woude MW. Diverse functions for acyltransferase-3 proteins in the modification of bacterial cell surfaces. Microbiology. 2022
2. Vora M, Shah M, Ostafi S, Onken B, Xue J, Ni JZ, et al. Deletion of microRNA-80 activates dietary restriction to extend C. elegans healthspan and lifespan. PLoS Genet. 2013;9(8):e1003737.
3. Dzitoyeva S, Dimitrijevic N, Manev H. Identification of a novel Drosophila gene, beltless, using injectable embryonic and adult RNA interference (RNAi). BMC Genomics. 2003;4(1):33.
4. Davis EO, Evans IJ, Johnston AWB. Identification of nodX, a gene that allows Rhizobium leguminosarum biovar viciae strain TOM to nodulate Afghanistan peas. Mol Gen Genet MGG. 1988;212(3):531–5.
5. Hara O, Hutchinson CR. A macrolide 3-O-acyltransferase gene from the midecamycin-producing species Streptomyces mycarofaciens. J Bacteriol. 1992;174(15):5141–4.
6. Laaberki MH, Pfeffer J, Clarke AJ, Dworkin J. O-Acetylation of Peptidoglycan Is Required for Proper Cell Separation and S-layer Anchoring in Bacillus anthracis. J Biol Chem. 2011; 18;286(7):5278–88.
7. Teh MY, Furevi A, Widmalm G, Morona R. Influence of Shigella flexneri 2a O Antigen Acetylation on Its Bacteriophage Sf6 Receptor Activity and Bacterial Interaction with Human Cells. J Bacteriol. 2020;202(24):e00363-20.
8. Davies MR, Broadbent SE, Harris SR, Thomson NR, Woude MW van der. Horizontally Acquired Glycosyltransferase Operons Drive Salmonellae Lipopolysaccharide Diversity. PLOS Genet. 2013;9(6):e1003568.
9. Broadbent SE, Davies MR, van der Woude MW. Phase variation controls expression of Salmonella lipopolysaccharide modification genes by a DNA methylation-dependent mechanism. Mol Microbiol. 2010;77(2):337–53.
10. Kintz E, Heiss C, Black I, Donohue N, Brown N, Davies MR, et al. Salmonella enterica Serovar Typhi Lipopolysaccharide O-Antigen Modification Impact on Serum Resistance and Antibody Recognition. Infect Immun. 2017;85(4):e01021-16.
11. Allison GE, Verma NK. Serotype-converting bacteriophages and O-antigen modification in Shigella flexneri. Trends Microbiol. 2000;8(1):17–23.
12. Bernard E, Rolain T, Courtin P, Guillot A, Langella P, Hols P, et al. Characterization of O-acetylation of N-acetylglucosamine: a novel structural variation of bacterial peptidoglycan. J Biol Chem. 2011;286(27):23950–8.
13. Bohin JP. Osmoregulated periplasmic glucans in Proteobacteria. FEMS Microbiol Lett. 2000;186(1):11–9.
14. Bonnet J, Durmort C, Jacq M, Mortier-Barrière I, Campo N, VanNieuwenhze MS, et al. Peptidoglycan O-acetylation is functionally related to cell wall biosynthesis and cell division in Streptococcus pneumoniae. Mol Microbiol. 2017;106(5):832–46.
15. Bontemps-Gallo S, Madec E, Robbe-Masselot C, Souche E, Dondeyne J, Lacroix JM.

879 The opgC gene is required for OPGs succinylation and is osmoregulated through
880 RcsCDB and EnvZ/OmpR in the phytopathogen Dickeya dadantii. Sci Rep.
881 2016;6(1):19619.
882 16. Cogez V, Gak E, Puskas A, Kaplan S, Bohin JP. The opgGIH and opgC genes of
883 Rhodobacter sphaeroides form an operon that controls backbone synthesis and
884 succinylation of osmoregulated periplasmic glucans. Eur J Biochem.
885 2002;269(10):2473–84.
886 17. Corvera A, Promé D, Promé JC, Martínez-Romero E, Romero D. The nolL gene from
887 Rhizobium etli determines nodulation efficiency by mediating the acetylation of the
888 fucosyl residue in the nodulation factor. Mol Plant-Microbe Interact MPMI.
889 1999;12(3):236–46.
890 18. Geno KA, Bush CA, Wang M, Jin C, Nahm MH, Yang J. WciG O-Acetyltransferase
891 Functionality Differentiates Pneumococcal Serotypes 35C and 42. J Clin Microbiol.
892 2017;55(9):2775–84.
893 19. Kajimura J, Rahman A, Hsu J, Evans MR, Gardner KH, Rick PD. O Acetylation of the
894 Enterobacterial Common Antigen Polysaccharide Is Catalyzed by the Product of the
895 yiaH Gene of Escherichia coli K-12. J Bacteriol. 2006;188(21):7542–50.
896 20. Kintz E, Davies MR, Hammarlöf DL, Canals R, Hinton JCD, van der Woude MW. A
897 BTP1 prophage gene present in invasive non-typhoidal Salmonella determines
898 composition and length of the O-antigen of the lipopolysaccharide. Mol Microbiol.
899 2015;96(2):263–75.
900 21. Lacroix JM, Lanfroy E, Cogez V, Lequette Y, Bohin A, Bohin JP. The mdoC gene of
901 Escherichia coli encodes a membrane protein that is required for succinylation of
902 osmoregulated periplasmic glucans. J Bacteriol. 1999;181(12):3626–31.
903 22. Menéndez N, Nur-e-Alam M, Braña AF, Rohr J, Salas JA, Méndez C. Biosynthesis of
904 the antitumor chromomycin A3 in Streptomyces griseus: analysis of the gene cluster
905 and rational design of novel chromomycin analogs. Chem Biol. 2004;11(1):21–32.
906 23. Moynihan PJ, Clarke AJ. O-Acetylated peptidoglycan: controlling the activity of
907 bacterial autolysins and lytic enzymes of innate immune systems. Int J Biochem Cell
908 Biol. 2011;43(12):1655–9.
909 24. Slauch JM, Lee AA, Mahan MJ, Mekalanos JJ. Molecular characterization of the oafA
910 locus responsible for acetylation of Salmonella typhimurium O-antigen: OafA is a
911 member of a family of integral membrane trans-acylases. J Bacteriol.
912 1996;178(20):5904–9.
913 25. Spencer BL, Shenoy AT, Orihuela CJ, Nahm MH. The Pneumococcal Serotype 15C
914 Capsule Is Partially O-Acetylated and Allows for Limited Evasion of 23-Valent
915 Pneumococcal Polysaccharide Vaccine-Elicited Anti-Serotype 15B Antibodies. Clin
916 Vaccine Immunol CVI. 2017;24(8):e00099-17.
917 26. Thanweer F, Tahiliani V, Korres H, Verma NK. Topology and identification of critical
918 residues of the O-acetyltransferase of serotype-converting bacteriophage, SF6, of
919 Shigella flexneri. Biochem Biophys Res Commun. 2008;375(4):581–5.
920 27. Zou CH, Knirel YA, Helbig JH, Zähringer U, Mintz CS. Molecular cloning and
921 characterization of a locus responsible for O acetylation of the O polysaccharide of
922 Legionella pneumophila serogroup 1 lipopolysaccharide. J Bacteriol.
923 1999;181(13):4137–41.
924 28. Buendia AM, Enenkel B, Köplin R, Niehaus K, Arnold W, Pünier A. The Rhizobium
925 meliloti exoZl exoB fragment of megaplasmid 2: ExoB functions as a UDP-glucose 4-
926 epimerase and ExoZ shows homology to NodX of Rhizobium leguminosarum biovar
927 viciae strain TOM. Mol Microbiol. 1991;5(6):1519–30.
928 29. Bera A, Herbert S, Jakob A, Vollmer W, Götz F. Why are pathogenic staphylococci so
929 lysozyme resistant? The peptidoglycan O-acetyltransferase OatA is the major
930 determinant for lysozyme resistance of Staphylococcus aureus. Mol Microbiol.
931 2005;55(3):778–87.
932 30. Kahler CM, Lyons-Schindler S, Choudhury B, Glushka J, Carlson RW, Stephens DS.
933 O-Acetylation of the terminal N-acetylglucosamine of the lipooligosaccharide inner core

934 in Neisseria meningitidis. Influence on inner core structure and assembly. J Biol Chem.
935 2006;281(29):19939–48.
936 31. Hellerqvist CG, Lindberg B, Svensson S, Holme T, Lindberg AA. Structural studies on
937 the O-specific side-chains of the cell-wall lipopolysaccharide from Salmonella
938 typhimurium LT2. Carbohydr Res. 1969;9(2):237–41.
939 32. Hellerqvist CG, Lindberg B, Samuelsson K, Lindberg AA. Structural studies on the O-
940 specific side-chains of the cell-wall lipopolysaccharide from Salmonella parathyphi A
941 var. durazzo. Acta Chem Scand. 1971;25(3):955–61.
942 33. Clark CA, Beltrame J, Manning PA. The oac gene encoding a lipopolysaccharide O-
943 antigen acetylase maps adjacent to the integrase-encoding gene on the genome of
944 Shigella flexneri bacteriophage Sf6. Gene. 1991;107(1):43–52.
945 34. Sun Q, Lan R, Wang Y, Wang J, Xia S, Wang Y, et al. Identification of a divergent O-
946 acetyltransferase gene oac1b from Shigella flexneri serotype 1b strains. Emerg
947 Microbes Infect. 2012;1(9):e21.
948 35. Verma NK, Brandt JM, Verma DJ, Lindberg AA. Molecular characterization of the O-
949 acetyl transferase gene of converting bacteriophage SF6 that adds group antigen 6 to
950 Shigella flexneri. Mol Microbiol. 1991;5(1):71–5.
951 36. Knirel YA, Wang J, Luo X, Senchenkova SN, Lan R, Shpirt AM, et al. Genetic and
952 structural identification of an O-acyltransferase gene (oacC) responsible for the 3/4-O-
953 acetylation on rhamnose III in Shigella flexneri serotype 6. BMC Microbiol.
954 2014;14(1):266.
955 37. Wang J, Lan R, Knirel YA, Luo X, Senchenkova SN, Shashkov AS, et al. Serological
956 Identification and Prevalence of a Novel O-Antigen Epitope Linked to 3- and 4-O-
957 Acetylated Rhamnose III of Lipopolysaccharide in Shigella flexneri. J Clin Microbiol.
958 2014;52(6):2033–8.
959 38. Sun Q, Knirel YA, Wang J, Luo X, Senchenkova SN, Lan R, et al. Serotype-converting
960 bacteriophage SfII encodes an acyltransferase protein that mediates 6-O-acetylation of
961 GlcNAc in Shigella flexneri O-antigens, conferring on the host a novel O-antigen
962 epitope. J Bacteriol. 2014;196(20):3656–66.
963 39. Fox KL, Yildirim HH, Deadman ME, Schweda EKH, Moxon ER, Hood DW. Novel
964 lipopolysaccharide biosynthetic genes containing tetranucleotide repeats in
965 Haemophilus influenzae, identification of a gene for adding O-acetyl groups. Mol
966 Microbiol. 2005;58(1):207–16.
967 40. Yildirim HH, Li J, Richards JC, Hood DW, Moxon ER, Schweda EKH. Complex O-
968 acetylation in non-typeable Haemophilus influenzae lipopolysaccharide: evidence for a
969 novel site of O-acetylation. Carbohydr Res. 2005;12;340(17):2598–611.
970 41. Brett PJ, Burtnick MN, Heiss C, Azadi P, DeShazer D, Woods DE, et al. Burkholderia
971 thailandensis oacA Mutants Facilitate the Expression of Burkholderia mallei-Like O
972 Polysaccharides. Infect Immun. 2011;79(2):961–9.
973 42. Brett PJ, Burtnick MN, Woods DE. The wbiA locus is required for the 2-O-acetylation of
974 lipopolysaccharides expressed by Burkholderia pseudomallei and Burkholderia
975 thailandensis. FEMS Microbiol Lett. 2003;218(2):323–8.
976 43. Heiss C, Burtnick MN, Black I, Azadi P, Brett PJ. Detailed structural analysis of the O-
977 polysaccharide expressed by Burkholderia thailandensis E264. Carbohydr Res.
978 2012;363:23–8.
979 44. Heiss C, Burtnick MN, Roberts RA, Black I, Azadi P, Brett PJ. Revised structures for
980 the predominant O-polysaccharides expressed by Burkholderia pseudomallei and
981 Burkholderia mallei. Carbohydr Res. 2013;381:6–11.
982 45. Kooistra O, Lüneberg E, Lindner B, Knirel YA, Frosch M, Zähringer U. Complex O-
983 acetylation in Legionella pneumophila serogroup 1 lipopolysaccharide. Evidence for
984 two genes involved in 8-O-acetylation of legionaminic acid. Biochemistry.
985 2001;40(25):7630–40.
986 46. Wagner C, Krönert C, Lück PC, Jacobs E, Cianciotto NP, Helbig JH. Random
987 mutagenesis of Legionella pneumophila reveals genes associated with
988 lipopolysaccharide synthesis and recognition by typing monoclonal antibodies. J Appl

989 Microbiol. 2007;103(5):1975–82.
990 47. Akoh CC, Lee GC, Liaw YC, Huang TH, Shaw JF. GDSL family of serine
991 esterases/lipases. Prog Lipid Res. 2004;43(6):534–52.
992 48. Aubry C, Goulard C, Nahori MA, Cayet N, Decalf J, Sachse M, et al. OatA, a
993 Peptidoglycan O-Acetyltransferase Involved in Listeria monocytogenes Immune
994 Escape, Is Critical for Virulence. J Infect Dis. 2011;204(5):731–40.
995 49. Cao L, Liang D, Hao P, Song Q, Xue E, Caiyin Q, et al. The increase of O-acetylation
996 and N-deacetylation in cell wall promotes acid resistance and nisin production through
997 improving cell wall integrity in Lactococcus lactis. J Ind Microbiol Biotechnol.
998 2018;45(9):813–25.
999 50. Veiga P, Bulbarela-Sampieri C, Furlan S, Maisons A, Chapot-Chartier MP, Erkelenz M,
1000 et al. SpxB regulates O-acetylation-dependent resistance of Lactococcus lactis
1001 peptidoglycan to hydrolysis. J Biol Chem. 2007;282(27):19342–54.
1002 51. Crisóstomo MI, Vollmer W, Kharat AS, Inhülsen S, Gehre F, Buckenmaier S, et al.
1003 Attenuation of penicillin resistance in a peptidoglycan O-acetyl transferase mutant of
1004 Streptococcus pneumoniae. Mol Microbiol. 2006;61(6):1497–509.
1005 52. Davis KM, Akinbi HT, Standish AJ, Weiser JN. Resistance to Mucosal Lysozyme
1006 Compensates for the Fitness Deficit of Peptidoglycan Modifications by Streptococcus
1007 pneumoniae. PLOS Pathog. 2008;4(12):e1000241.
1008 53. Arisawa A, Tsunekawa H, Okamura K, Okamoto R. Nucleotide sequence analysis of
1009 the carbomycin biosynthetic genes including the 3-O-acyltransferase gene from
1010 Streptomyces thermotolerans. Biosci Biotechnol Biochem. 1995;59(4):582–8.
1011 54. Arisawa A, Kawamura N, Takeda K, Tsunekawa H, Okamura K, Okamoto R. Cloning of
1012 the macrolide antibiotic biosynthesis gene acyA, which encodes 3-O-acyltransferase,
1013 from Streptomyces thermotolerans and its use for direct fermentative production of a
1014 hybrid macrolide antibiotic. Appl Environ Microbiol. 1994;60(7):2657–60.
1015 55. Epp JK, Huber ML, Turner JR, Goodson T, Schoner BE. Production of a hybrid
1016 macrolide antibiotic in Streptomyces ambofaciens and Streptomyces lividans by
1017 introduction of a cloned carbomycin biosynthetic gene from Streptomyces
1018 thermotolerans. Gene. 1989;85(2):293–301.
1019 56. Meziane-Cherif D, Stogios PJ, Evdokimova E, Egorova O, Savchenko A, Courvalin P.
1020 Structural and Functional Adaptation of Vancomycin Resistance VanT Serine
1021 Racemases. mBio. 2015;6(4):e00806.
1022 57. Jones CS, Anderson AC, Clarke AJ. Mechanism of Staphylococcus aureus
1023 peptidoglycan O-acetyltransferase A as an O-acyltransferase. Proc Natl Acad Sci.
1024 2021;118(36).
1025 58. Pearson CR, Tindall SN, Herman R, Jenkins HT, Bateman A, Thomas GH, et al.
1026 Acetylation of Surface Carbohydrates in Bacterial Pathogens Requires Coordinated
1027 Action of a Two-Domain Membrane-Bound Acyltransferase. mBio. 2020;11(4):e01364-
1028 20.
1029 59. Rajput MI, Verma NK. Identification of critical residues of O-antigen-modifying O-
1030 acetyltransferase B (OacB) of Shigella flexneri. BMC Mol Cell Biol. 2022; 24;23(1):16.
1031 60. Wu D, Hersh LB. Identification of an active site arginine in rat choline acetyltransferase
1032 by alanine scanning mutagenesis. J Biol Chem. 1995;270(49):29111–6.
1033 61. Moss SJ, Bai L, Toelzer S, Carroll BJ, Mahmud T, Yu TW, et al. Identification of asm19
1034 as an acyltransferase attaching the biologically essential ester side chain of
1035 ansamitocins using N-desmethyl-4,5-desepoxymaytansinol, not maytansinol, as its
1036 substrate. J Am Chem Soc. 2002;124(23):6544–5.
1037 62. García B, González-Sabín J, Menéndez N, Braña AF, Núñez LE, Morís F, et al. The
1038 chromomycin CmmA acetyltransferase: a membrane-bound enzyme as a tool for
1039 increasing structural diversity of the antitumour mithramycin. Microb Biotechnol.
1040 2011;4(2):226–38.
1041 63. Moynihan PJ, Clarke AJ. Mechanism of action of peptidoglycan O-acetyltransferase B
1042 involves a Ser-His-Asp catalytic triad. Biochemistry. 2014;53(39):6243–51.
1043 64. Konadu E, Shiloach J, Bryla DA, Robbins JB, Szu SC. Synthesis, characterization, and

1044 immunological properties in mice of conjugates composed of detoxified
1045 lipopolysaccharide of Salmonella paratyphi A bound to tetanus toxoid with emphasis on
1046 the role of O acetyls. Infect Immun. 1996;64(7):2709–15.
1047 65. Ravenscroft N, Cescutti P, Gavini M, Stefanetti G, MacLennan CA, Martin LB, et al.
1048 Structural analysis of the O-acetylated O-polysaccharide isolated from Salmonella
1049 paratyphi A and used for vaccine preparation. Carbohydr Res. 2015;404:108–16.
1050 66. Wang S, Li W, Zhang R, Liu S, Xu J. CoinFold: a web server for protein contact
1051 prediction and contact-assisted protein folding. Nucleic Acids Res. 2016;44(W1):W361-
1052 366.
1053 67. Källberg M, Margaryan G, Wang S, Ma J, Xu J. RaptorX server: a resource for
1054 template-based protein structure modeling. Methods Mol Biol Clifton NJ.
1055 2014;1137:17–27.
1056 68. Ma J, Wang S, Wang Z, Xu J. Protein contact prediction by integrating joint
1057 evolutionary coupling analysis and supervised learning. Bioinformatics.
1058 2015;31(21):3506–13.
1059 69. Sharma C, Nigam A, Singh R. Computational-approach understanding the structure-
1060 function prophecy of Fibrinolytic Protease RFEA1 from Bacillus cereus RSA1. PeerJ.
1061 2021;9:e11570.
1062 70. Mariani V, Kiefer F, Schmidt T, Haas J, Schwede T. Assessment of template based
1063 protein structure predictions in CASP9. Proteins Struct Funct Bioinforma.
1064 2011;79(S10):37–58.
1065 71. Lopes-Rodrigues M, Zanuy D, Alemán C, Michaux C, Perpète EA. 3D structure of a
1066 Brucella melitensis porin: molecular modelling in lipid membranes. J Biomol Struct Dyn.
1067 2019;37(15):3923–35.
1068 72. Kaplan-Türköz B, Bakar B. Structural Modelling and Structure-Function Analysis of
1069 Zymomonas mobilis Levansucrase. Süleyman Demirel Üniversitesi Fen Bilim Enstitüsü
1070 Derg. 2017;21(1):279–85.
1071 73. Xu J, Wang S. Analysis of distance-based protein structure prediction by deep learning
1072 in CASP13. Proteins Struct Funct Bioinforma. 2019;87(12):1069–81.
1073 74. Bernsel A, Viklund H, Hennerdal A, Elofsson A. TOPCONS: consensus prediction of
1074 membrane protein topology. Nucleic Acids Res. 2009;37(Web Server issue):W465-468.
1075 75. Tsirigos KD, Peters C, Shu N, Käll L, Elofsson A. The TOPCONS web server for
1076 consensus prediction of membrane protein topology and signal peptides. Nucleic Acids
1077 Res. 2015;43(W1):W401-407.
1078 76. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly
1079 accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583–9.
1080 77. Bonnet J, Durmort C, Jacq M, Mortier-Barrière I, Campo N, VanNieuwenhze MS, et al.
1081 Peptidoglycan O-acetylation is functionally related to cell wall biosynthesis and cell
1082 division in Streptococcus pneumoniae. Mol Microbiol. 2017;106(5):832–46.
1083 78. Bernard E, Rolain T, David B, André G, Dupres V, Dufrêne YF, et al. Dual Role for the
1084 O-Acetyltransferase OatA in Peptidoglycan Modification and Control of Cell Septation
1085 in Lactobacillus plantarum. PLoS ONE. 2012;7(10):e47893.
1086 79. Pearson CR, Tindall SN, Herman R, Jenkins HT, Bateman A, Thomas GH, et al.
1087 Acetylation of Surface Carbohydrates in Bacterial Pathogens Requires Coordinated
1088 Action of a Two-Domain Membrane-Bound Acyltransferase. Clarke AJ, Pier GB,
1089 editors. mBio. 2020;11(4).
1090 80. Holm L. Using Dali for Protein Structure Comparison. In: Gáspári Z, editor. Structural
1091 Bioinformatics: Methods and Protocols. New York, NY: Springer US; 2020. p. 29–42.
1092 (Methods in Molecular Biology).
1093 81. Ma D, Wang Z, Merrikh CN, Lang KS, Lu P, Li X, et al. Crystal structure of a
1094 membrane-bound O-acyltransferase. Nature. 2018;562(7726):286–90.
1095 82. Moynihan PJ, Clarke AJ. O-acetylation of peptidoglycan in gram-negative bacteria:
1096 identification and characterization of peptidoglycan O-acetyltransferase in Neisseria
1097 gonorrhoeae. J Biol Chem. 2010;285(17):13264–73.
1098 83. Thanweer F, Verma NK. Identification of critical residues of the serotype modifying O-

acetyltransferase of Shigella flexneri. BMC Biochem. 2012;13:13.
84. Calix JJ, Oliver MB, Sherwood LK, Beall BW, Hollingshead SK, Nahm MH. Streptococcus pneumoniae Serotype 9A Isolates Contain Diverse Mutations to wcjE That Result in Variable Expression of Serotype 9V-specific Epitope. J Infect Dis. 2011;204(10):1585–95.
85. Jogl G, Hsiao YS, Tong L. Structure and function of carnitine acyltransferases. Ann N Y Acad Sci. 2004;1033:17–29.
86. Skylaris CK, Haynes PD, Mostofi AA, Payne MC. Introducing ONETEP: Linear-scaling density functional simulations on parallel computers. J Chem Phys. 2005 ;122(8):084119.
87. Prentice JCA, Aarons J, Womack JC, Allen AEA, Andrinopoulos L, Anton L, et al. The ONETEP linear-scaling density functional theory program. J Chem Phys. 2020;152(17):174111.
88. Heijne G. The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. EMBO J. 1986;5(11):3021–7.
89. Webb B, Sali A. Comparative Protein Structure Modeling Using MODELLER. Curr Protoc Bioinforma. 2016;54:5.6.1-5.6.37.
90. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci Publ Protein Soc. 2021;30(1):70–82.
91. Periole X, Cavalli M, Marrink SJ, Ceruso MA. Combining an Elastic Network with a Coarse-Grained Molecular Force Field: Structure, Dynamics, and Intermolecular Recognition. J Chem Theory Comput. 2009;5(9):2531–43.
92. Anishchenko I, Ovchinnikov S, Kamisetty H, Baker D. Origins of coevolution between residues distant in protein 3D structures. Proc Natl Acad Sci. 2017;114(34):9122–7.
93. Hollander M, Rasp D, Aziz M, Helms V. ProPores2: Web Service and Stand-Alone Tool for Identifying, Manipulating, and Visualizing Pores in Protein Structures. J Chem Inf Model. 2021;61(4):1555–9.
94. Whitney JC, Howell PL. Synthase-dependent exopolysaccharide secretion in Gram-negative bacteria. Trends Microbiol. 2013;21(2):63–72.
95. Glenwright AJ, Pothula KR, Bhamidimarri SP, Chorev DS, Baslé A, Firbank SJ, et al. Structural basis for nutrient acquisition by dominant members of the human gut microbiota. Nature. 2017;541(7637):407–11.
96. Micoli F, Ravenscroft N, Cescutti P, Stefanetti G, Londero S, Rondini S, et al. Structural analysis of O-polysaccharide chains extracted from different Salmonella Typhimurium strains. Carbohydr Res. 2014;385:1–8.
97. Sadovskaya I, Vinogradov E, Flahaut S, Kogan G, Jabbouri S. Extracellular carbohydrate-containing polymers of a model biofilm-producing strain, Staphylococcus epidermidis RP62A. Infect Immun. 2005;73(5):3007–17.
98. Joyce JG, Abeygunawardana C, Xu Q, Cook JC, Hepler R, Przysiecki CT, et al. Isolation, structural characterization, and immunological evaluation of a high-molecular-weight exopolysaccharide from Staphylococcus aureus. Carbohydr Res. 2003;338(9):903–22.
99. RCSB PDB. Pairwise Structure Alignment Guide [Internet]. https://www.rcsb.org/docs/tools/pairwise-structure-alignment. [cited 2021 Nov 30]. Available from: https://www.rcsb.org/docs/tools/pairwise-structure-alignment
100. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: Fast, flexible, and free. J Comput Chem. 2005;26(16):1701–18.
101. Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. J Comput Chem. 2011;32(10):2319–27.
102. R. J. Gowers, M. Linke, J. Barnoud, T. J. E. Reddy, M. N. Melo, S. L. Seyler, D. L. Dotson, J. Domanski, S. Buchoux, I. M. Kenney, and O. Beckstein. MDAnalysis: A Python package for the rapid analysis of molecular dynamics simulations. In S. Benthall and S. Rostrup, editors, Proceedings of the 15th Python in Science Conference, pages 98-105, Austin, TX, 2016. SciPy, doi:10.25080/majora-629e541a-00e

103. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature. 2020;585(7825):357–62.
104. McKinney W. pandas: a Foundational Python Library for Data Analysis and Statistics. Python High Perform Sci Comput. 2011;9.
105. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. J Mol Graph. 1996;14(1):33–8.
106. Huang J, MacKerell AD. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. J Comput Chem. 2013;34(25):2135–45.
107. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 1983;79(2):926–35.
108. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. J Chem Phys. 1995;103(19):8577–93.
109. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. J Comput Chem. 1997;18(12):1463–72.
110. Nosé S. A molecular dynamics method for simulations in the canonical ensemble. Mol Phys. 1984;52(2):255–68.
111. Hoover WG. Canonical dynamics: Equilibrium phase-space distributions. Phys Rev A. 1985;31(3):1695–7.
112. Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. J Appl Phys. 1981;52(12):7182–90.
113. Lee J, Patel DS, Ståhle J, Park SJ, Kern NR, Kim S, et al. CHARMM-GUI *Membrane Builder* for Complex Biological Membrane Simulations with Glycolipids and Lipoglycans. J Chem Theory Comput. 2019;15(1):775–86.
114. Goldstein AA. On Steepest Descent. J Soc Ind Appl Math Ser Control. 1965;3(1):147–51.
115. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. J Chem Phys. 2007;126(1):014101.
116. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys. 1984;81(8):3684–90.
117. Kim S, Lee J, Jo S, Brooks III CL, Lee HS, Im W. CHARMM-GUI ligand reader and modeler for CHARMM force field generation of small molecules. J Comput Chem. 2017;38(21):1879–86.
118. Perdew JP, Burke K, Ernzerhof M. Generalized Gradient Approximation Made Simple. Phys Rev Lett. 1996;77(18):3865–8.
119. Grimme S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. J Comput Chem. 2006;27(15):1787–99.
120. Phipps MJS, Fox T, Tautermann CS, Skylaris CK. Energy decomposition analysis approaches and their evaluation on prototypical protein–drug interaction patterns. Chem Soc Rev. 2015;44(10):3177–211.
121. Phipps MJS, Fox T, Tautermann CS, Skylaris CK. Energy Decomposition Analysis Based on Absolutely Localized Molecular Orbitals for Large-Scale Density Functional Theory Calculations in Drug Design. J Chem Theory Comput. 2016;12(7):3135–48.