1 **Bacterial outer-membrane polysaccharide export (OPX) proteins occupy three structural**
2 **classes with selective β-barrel porin requirements for polymer secretion**
3
4

5 Fares Saïdi[1,2†], Utkarsha Mahanta[3†], Adyasha Panda[3], Nicolas Y. Jolivet[1,2], Razieh Bitazar[1,2], Gavin
6 John[4], Matthew Martinez[5], Abdelkader Mellouk[1,2], Charles Calmettes[1,2], Yi-Wei Chang[5], Gaurav
7 Sharma[3*], Salim T. Islam[1,2*]

8 [1] Institut National de la Recherche Scientifique (INRS), Centre Armand-Frappier Santé
9 Biotechnologie, Université du Québec, Institut Pasteur International Network, Laval, QC, Canada

10 [2] PROTEO, the Quebec Network for Research on Protein Function, Engineering, and Applications,
11 Université Laval, Québec, QC, Canadas

12 [3] Institute of Bioinformatics and Applied Biotechnology (IBAB), Bengaluru, Karnataka, India

13 [4] Department of Pediatrics, Division of Infectious Diseases, Children's Hospital of Philadelphia,
14 Philadelphia, PA, U.S.A.

15 [5] Department of Biochemistry and Biophysics, Perelman School of Medicine, University of
16 Pennsylvania, Philadelphia, PA, U.S.A.

17 [†] co-1st authors
18
19 * corresponding authors:
20
21 Salim T. Islam   Gaurav Sharma
22 salim.islam@inrs.ca   gauravsharma@ibab.ac.in
23 ORCID ID: 0000-0001-6853-8446   ORCID ID: 0000-0002-2861-7446

24 **ABSTRACT**

25   Secretion of high-molecular-weight polysaccharides across the bacterial envelope is

26 ubiquitous as it enhances prokaryotic survival in (a)biotic settings. Such polymers are often

27 assembled by Wzx/Wzy- or ABC transporter-dependent schemes that implicate outer-membrane

28 (OM) polysaccharide export (OPX) proteins in polymer translocation to the cell surface. In the social

29 predatory bacterium *Myxococcus xanthus*, exopolysaccharide (EPS)-pathway WzaX, major spore

30 coat (MASC)-pathway WzaS, and biosurfactant polysaccharide-pathway WzaB were herein found to

31 be truncated OPX homologues of *Escherichia coli* Wza lacking OM-spanning α-helices.

32 Comparative genomics across all bacteria, complemented with cryo-electron tomography cell-

33 envelope analyses, revealed WzaX/S/B architecture to be the most common amongst three defined

34 OPX-protein structural classes independent of periplasmic thickness. Fold-recognition and deep-

35 learning analyses revealed the conserved *M. xanthus* proteins MXAN_7418/3226/1916 (encoded

36 adjacent to WzaX/S/B) to be integral OM β-barrels, with structural homology to the poly-*N*-acetyl-D-

37 glucosamine synthase-dependent pathway porin PgaA. Such porins were identified in bacteria near

38 numerous genes for all three OPX-protein classes. Interior MXAN_7418/3226/1916 β-barrel

39 electrostatics were found to match known properties of their associated polymers. With

40 MXAN_3226 essential for MASC export, and MXAN_7418 absence shown herein to compromise

41 EPS translocation, these data support a novel secretion paradigm for Wzx/Wzy-dependent pathways

42 in which those containing an OPX component that cannot span the OM instead utilize a β-barrel

43 porin to mediate polysaccharide transport across the OM.

INTRODUCTION

44

45   Diverse bacteria associated with biotic and abiotic settings secrete high-molecular-weight

46   (HMW) polysaccharides across the cell envelope to enhance their survival. Capsule polysaccharide

47   (CPS) chains are tightly bound to the cell surface and form hydrated exclusionary barriers to

48   molecule entry, whereas exopolysaccharide (EPS) polymers form a more loosely surface-associated

49   glycocalyx around cells that contributes to biofilm matrix formation in aggregates (Whitfield *et al.*,

50   2020). Certain HMW polysaccharides do not remain associated with the cell surface and are instead

51   secreted to the extracellular milieu where they can influence bacterial physiology (Saïdi *et al.*, 2021,

52   Islam *et al.*, 2020). In both monoderm (Gram-positive) and diderm (Gram-negative) bacteria,

53   multiple secreted polymers often act in concert to modulate complex physiology (Pérez-Burgos &

54   Søgaard-Andersen, 2020, Lavelle *et al.*, 2021, Franklin *et al.*, 2011).

55   *Myxococcus xanthus* is a Gram-negative bacterium that exhibits an intricate social

56   multicellular lifestyle (Muñoz-Dorado *et al.*, 2016). This motile soil bacterium (Faure *et al.*, 2016,

57   Islam & Mignot, 2015) is able to predate other bacteria (Seef *et al.*, 2021) and saprophytically feed

58   on the degradation products. Under nutrient deprivation, cells in a swarm biofilm form myxospore-

59   filled fruiting bodies through a developmental program resulting in a differentiated cell community

60   (Muñoz-Dorado *et al.*, 2016). This complex lifecycle is modulated by the secretion of three known

61   polysaccharides (Pérez-Burgos & Søgaard-Andersen, 2020). Cells constitutively produce EPS, a

62   specific surface-associated polymer that forms a glycocalyx surrounding the cell body (Saïdi *et al.*,

63   2021) and which constitutes the main matrix component in biofilms of this bacterium (Hu *et al.*,

64   2013, Smaldone *et al.*, 2014). A biosurfactant polysaccharide (BPS) is also synthesized, but is

65   instead secreted to the extracellular milieu (Islam *et al.*, 2020), where it functionally destabilizes the

66   EPS glycocalyx, leading to a range of fundamental behavioural and surface-property changes at the

67   single-cell level (Saïdi *et al.*, 2021). The synergy between EPS and BPS secretion as well as the

68   spatiospecific production patterns of the two polymers (Islam *et al.*, 2020) also impacts the internal

69   architecture of *M. xanthus* swarm biofilms, as well as their Type IV pilus (T4P)-dependent expansion

70   due to impacts on T4P production, stability, and positioning (Saïdi *et al.*, 2021). Finally, the major

71   spore coat (MASC) polymer is produced by spore-forming cells undergoing development, providing

72   a protective coat to cover nascent myxospores and provide resistance to environmental stresses

73   (Holkenbrink *et al.*, 2014, Wartel *et al.*, 2013).

74   Each *M. xanthus* polysaccharide is produced by a separate Wzx/Wzy-dependent assembly

75   pathway (Islam *et al.*, 2020, Pérez-Burgos *et al.*, 2020, Holkenbrink *et al.*, 2014), the components for

76   which have the suffixes X (e̲x̲opolysaccharide), B (b̲iosurfactant), or S (s̲pore coat). Therein, cluster-

77   specific glycosyltransferases synthesize polymer repeat units atop an undecaprenyl pyrophosphate

78   (UndPP) lipid anchor at the cytoplasmic leaflet of the inner membrane (IM). UndPP-linked repeats

79   are then translocated across the IM via the Wzx flippase (Islam *et al.*, 2012, Islam *et al.*, 2013a, Islam

80   *et al.*, 2010), followed by polymerization at the periplasmic leaflet of the IM by Wzy (Islam *et al.*,

81   2011, Islam *et al.*, 2013b) to modal lengths governed by the Wzc polysaccharide co-polymerase

82   (PCP) in non-O-antigen systems (**Fig. 1A**) (Islam & Lam, 2014, Whitfield *et al.*, 2020). In *M.*

83   *xanthus*, WzcB possesses an integrated cytosolic bacterial tyrosine autokinase (BYK) domain (PCP-

84   2A class), whereas WzcX and WzcS do not (PCP-2B class); the EPS and MASC pathways thus

85   encode a separate BYK (Wze) for association with the cognate PCP (Islam *et al.*, 2020). The Wzb

86   bacterial tyrosine phosphatase (BYP) protein in turn regulates the state of PCP-2A Wzc and PCP-2B-

87   associated Wze phosphorylation (Mori *et al.*, 2012). Wzb-mediated dephosphorylation of BYK

88   domains has been proposed to drive Wzc octamerization, in turn affecting Wzy-mediated

89   polymerization and interaction with the outer-membrane polysaccharide export (OPX) Wza

90   translocon needed for polymer transport through the periplasm and across the outer membrane (OM)

91   (Yang *et al.*, 2021) (**Fig. 1A**). Such pathways are one of the most widespread bacterial assembly

92   schemes for HMW polysaccharides, responsible for synthesizing diverse products such as Group 1

93   CPS and Group 4 CPS (i.e. "O antigen" capsule) as well as colanic acid polymers in enterobacteria

94   (Sande & Whitfield, 2021), in addition to holdfast polysaccharide in *Caulobacter crescentus*, (Toh *et*

95   *al.*, 2008), and xanthan in *Xanthomonas campestris* (Becker, 2015). In Group 1 CPS systems, the 18-

96   stranded integral OM β-barrel Wzi (internally occluded by an α-helical plug domain) is also

97   important as it displays lectin-like characteristics implicated in capsule structure organization

98   (Bushell *et al.*, 2013).

99   Secreted polymers can also be synthesized by an ABC transporter-dependent scheme in

100   which UndPP-linked polymers are built by sugar unit addition at the cytoplasmic leaflet of the IM,

101   with the polymer generated entirely in the cytoplasm. Subsequent ATP hydrolysis by the transporter

102   drives polymer transport across the IM, after which PCP and OPX proteins are needed for the

103   secretion of polymer through the periplasm and across the OM. These biosynthesis pathways are

104   implicated in the secretion of polysialic acid Group 2 CPS polymers from pathogenic extraintestinal

105 *E. coli*, as well as similar structures in *Neisseria meningitidis* and *Haemophilus influenza* (Willis &

106 Whitfield, 2013).

107   Alternatively, HMW polymers such as alginate in *Pseudomonas aeruginosa*, cellulose in

108 *Salmonella enterica*, and poly-*N*-acetyl-D-glucosamine (PNAG) in *Acinetobacter baumanii* are

109 produced via a synthase-dependent scheme in which the addition of a monosaccharide in the

110 cytoplasm by an integral IM synthase results in export of the polymer by a similar amount from the

111 cell surface. Polymer transport through the periplasm is mediated by a protein scaffold containing

112 TPR repeats followed by translocation across the OM through an integral OM β-barrel porin structure

113 (Whitney & Howell, 2013). While similarities exist between Wzx/Wzy- and ABC transporter-

114 dependent pathways (e.g. the presence of PCP and OPX proteins), no such schematic crossover with

115 proteins from synthase-dependent pathways has been identified.

116   In both Wzx/Wzy- and ABC transporter-dependent pathways, OPX-family proteins are

117 portrayed as forming an oligomeric channel of contiguous domains to allow polymer secretion

118 through the periplasm and across the OM (Whitfield *et al.*, 2020). All OPX proteins share a

119 conserved periplasmic N-terminal Poly_export domain (Pfam: PF02563, also called a PES

120 [polysaccharide export sequence] motif), followed by at least one copy of a soluble ligand-binding β-

121 grasp (SLBB) domain (Pfam: PF10531) (Sande *et al.*, 2019, Cuthbertson *et al.*, 2009), which is

122 predicted to interact with the sugar polymer in the periplasm (Burroughs *et al.*, 2007). OPX protein

123 domain architecture diverges at this point. In the prototypic OPX Wza from *E. coli* group 1 CPS

124 (Wza$_{Ec}$) — which is the only OPX protein with a solved 3D structure — after one Poly_export (D1)

125 and two SLBB domains (D2 & D3), the protein contains a Wza_C domain (Pfam: PF18412) at its C-

126 terminus, which forms a 35-residue amphipathic α-helical tract (D4) that crosses the OM (Dong *et*

127 *al.*, 2006) (**Fig. 1A**). Elucidation of the Wza$_{Ec}$ X-ray crystal structure (PDB: 2J58) was

128 revolutionary as it represented the first instance of such a fold in an integral OM protein. As part of

129 the Wza$_{Ec}$ oligomer, 8 copies of the α-helical Wza_C domain were shown to form a pore-like

130 structure, through which it is proposed that secreted polysaccharides exit the cell (Dong *et al.*, 2006,

131 Nickerson *et al.*, 2014) (**Fig. 1A**). Instead of a classical Wza_C domain, OPX proteins from ABC

132 transporter-dependent Group 2 CPS pathways usually contain a C-terminal Caps_synth_GfcC (Pfam:

133 PF06251, formerly DUF1017) module (Cuthbertson *et al.*, 2009), structurally similar to the stand-

134 alone GfcC protein (PDB: 3P42) from Group 4 Wzx/Wzy-dependent CPS pathways (Sande *et al.*,

135 2019). GfcC contains domains comparable to D2 and D3 from Wza$_{Ec}$, as well as a D4-like

136   amphipathic α-helix; however, the GfcC D4-like helix spanning the final 21 residues of the protein is

137   40% shorter than Wza$_{Ec}$ D4, bent at both ends, and structurally locked (Sathiyamoorthy *et al.*, 2011).

138   This overall OPX architecture is typified by the Group 2 CPS pathway protein KpsD from *E. coli*

139   (KpsD$_{Ec}$) (Sande *et al.*, 2019). Though it is uncertain if the C-terminal domain of KpsD$_{Ec}$ is able to

140   span the OM, KpsD$_{Ec}$ epitopes have previously been detected at the cell surface via anti-KpsD$_{Ec}$

141   antibody labelling (McNulty *et al.*, 2006). Numerous other annotated OPX proteins have been shown

142   to either (i) contain considerable-yet-uncharacterized protein sequences following their most C-

143   terminal identified domain, or (ii) be considerably shorter than either Wza$_{Ec}$ or KpsD$_{Ec}$, with

144   architecture beyond the Poly_export and SLBB domains largely absent (Cuthbertson *et al.*, 2009).

145   Ultimately, for OPX proteins that lack a canonical OM-spanning Wza_C domain, the manner by

146   which the respective secreted polymers traverse the asymmetric OM bilayer remains a fundamental

147   and pertinent question that has yet to be resolved.

148   Herein, we reveal that the WzaX, WzaB, and WzaS OPX proteins for the respective *M.*

149   *xanthus* EPS, BPS, and MASC pathways contain typical N-terminal "Poly_export–SLBB"

150   architecture but lack a C-terminal OM-spanning Wza_C domain. Comparative genomics analyses

151   reveal this architecture to be the most common amongst three distinct structural classes of OPX

152   proteins across all bacteria. However, in the *M. xanthus* EPS, BPS, and MASC biosynthetic clusters,

153   a conserved β-barrel protein (MXAN_7418/MXAN_1916/MXAN_3226) is encoded immediately

154   adjacent to the genes for WzaX/WzaB/WzaS (respectively). Fold-recognition and deep-learning

155   analyses reveal these adjacently-encoded proteins to be 18-stranded integral OM β-barrels with

156   structural homology to the barrel domain of the porin PgaA, required for PNAG secretion across the

157   OM by synthase-dependent pathways. In turn, PgaA-like β-barrel proteins are shown to be encoded

158   near numerous genes representing all three OPX structural classes in diverse Gram-negative bacteria.

159   The interior electrostatics of the *M. xanthus* β-barrels match known properties of their associated

160   polymers, and deletion of the MXAN_7418 β-barrel is shown to compromise EPS secretion.

161   Together with the known requirement for the MXAN_3226 β-barrel for MASC secretion

162   (Holkenbrink *et al.*, 2014), these data support a novel secretion paradigm for Wzx/Wzy-dependent

163   pathways in which those containing an OPX component that cannot span the OM instead utilize a β-

164   barrel porin to mediate translocation of HMW polymers across the OM.

165 **Figure 1. Wzx/Wzy-dependent polysaccharide assembly-and-secretion. A)** Pathway schematic.

166 *Inset:* The Wza_{Ec} X-ray crystal structure octamer (PDB: 2J58) has been differentially coloured to

167 highlight the position of each chain in the structure. The D1 (Poly_export), D2 & D3 (both SLBB), and

168 D4 (Wza_C) domains have been indicated, with smooth loops. **B)** Tertiary structure models of *M.*

169 *xanthus* EPS-pathway WzaX (aa 51–212), MASC-pathway WzaS (aa 32–190), and BPS-pathway

170 WzaB (aa 38–202) based on the Wza_{Ec} structure (aa 22–376, depicted with a N-terminal lipid anchor).

171 Structures are displayed with smooth loops, highlighted β-sheets (*yellow*) and α-helices (*red*), with the
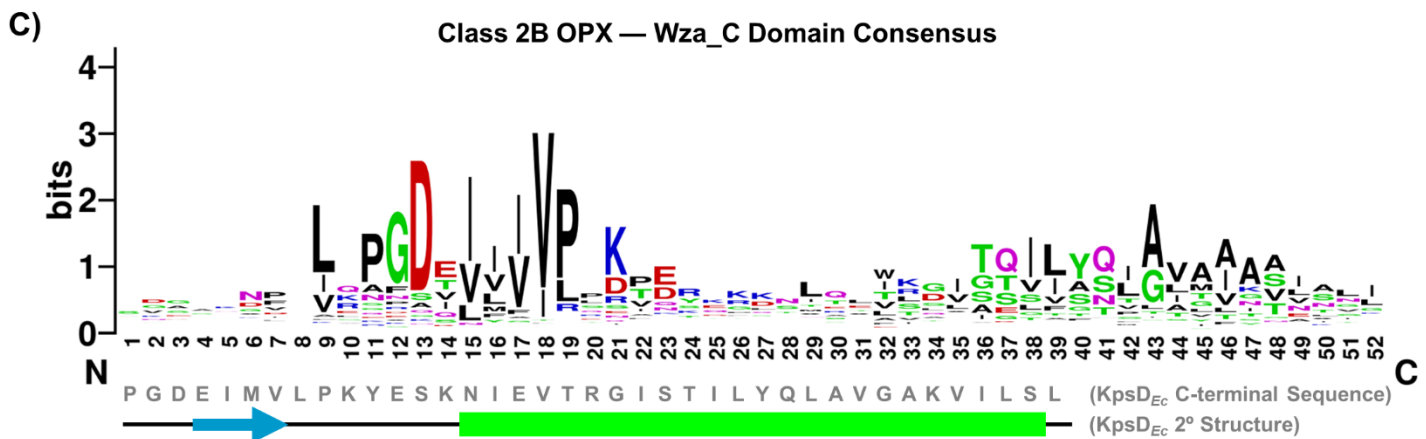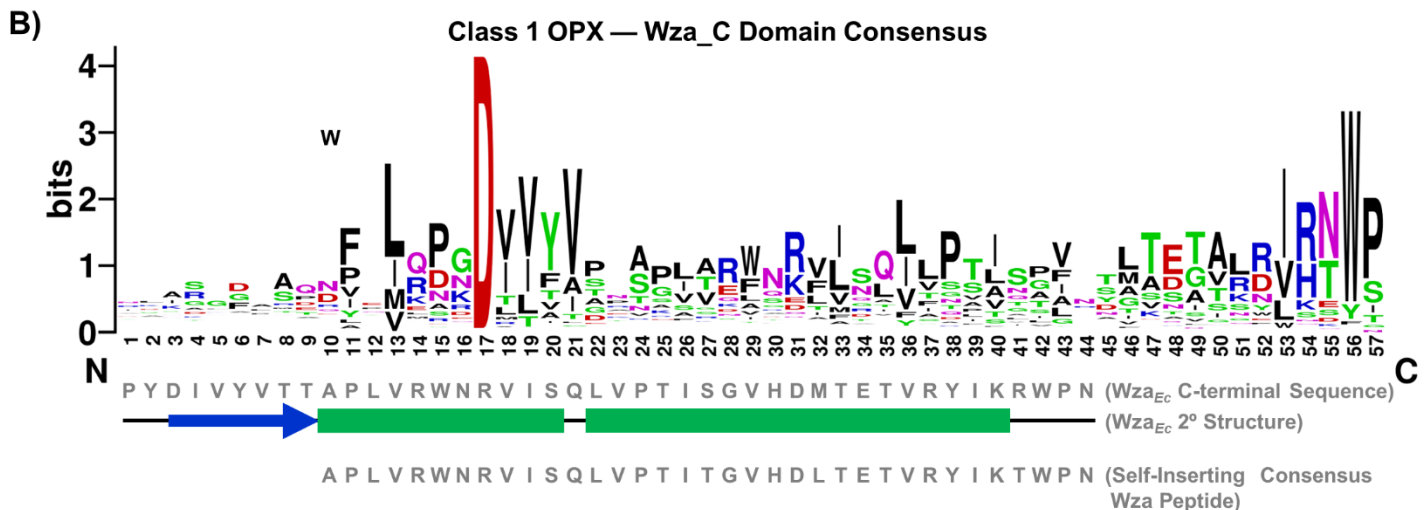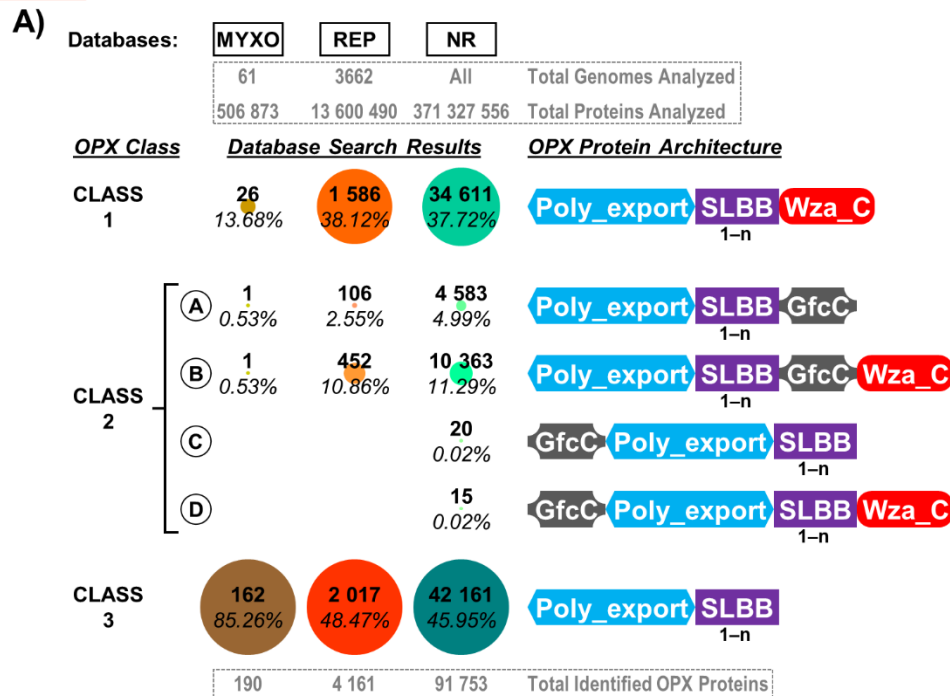
172 various N- and C- termini indicated.

FIGURE 1

A) Wzx/Wzy-Dependent Polymer Assembly & Secretion
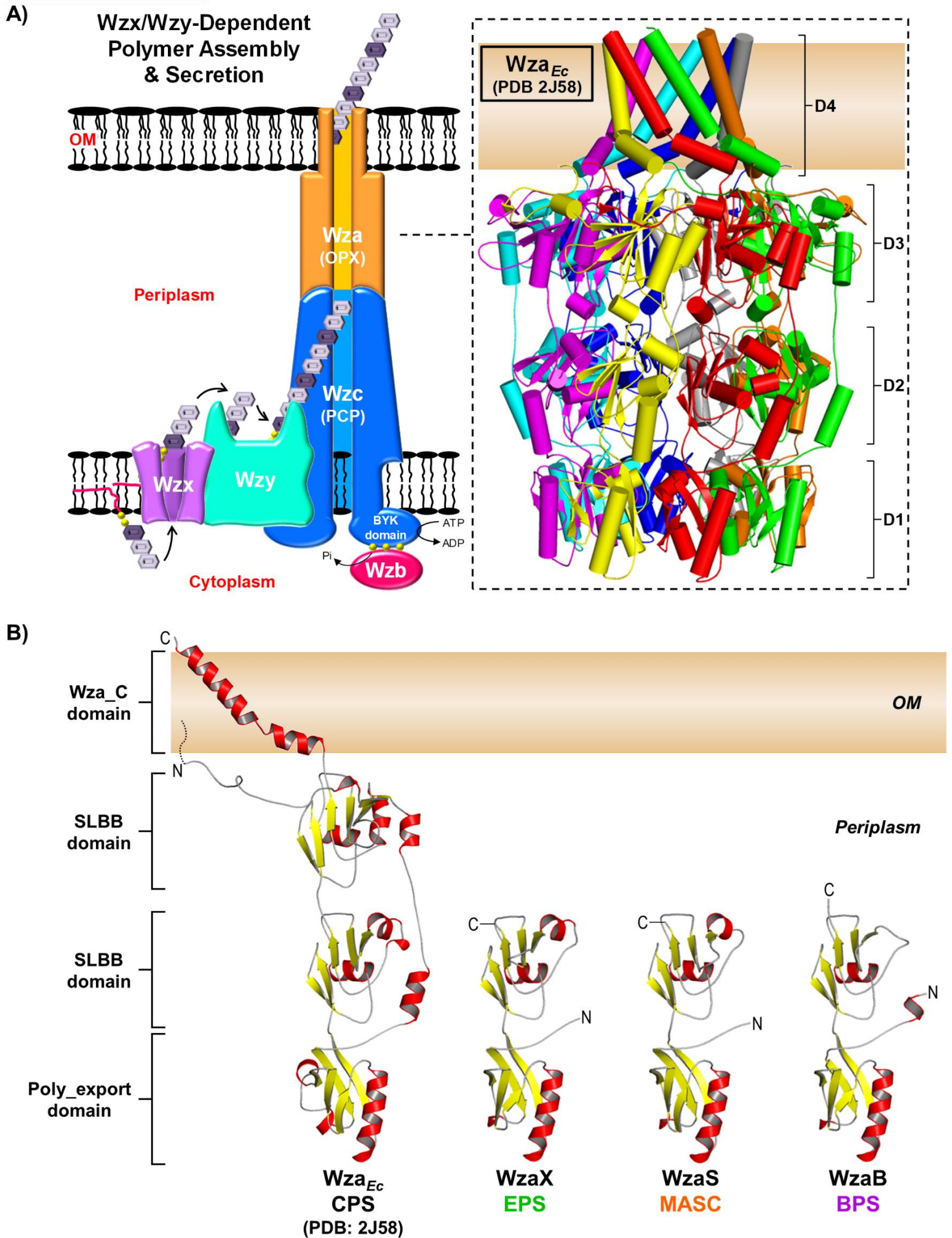
174   **RESULTS**

175   *The* M. xanthus *OPX proteins WzaX, WzaS, and WzaB lack an OM-spanning α-helix domain*

176   Each of the WzaX/S/B OPX proteins is essential for the secretion of its respective

177   EPS/MASC/BPS polymer in *M. xanthus* (Islam *et al.*, 2020, Holkenbrink *et al.*, 2014). However, as

178   each of WzaX, WzaS, and WzaB has a considerably shorter amino acid sequence than that of Wza_{Ec}

179   (44, 50, and 46% smaller, respectively), we sought to better understand the structural implications of

180   this difference. Fold-recognition analysis of each protein revealed 100%, 100%, and 99.9%

181   probability matches (respectively) to the N-terminal half of the high-resolution Wza_{Ec} X-ray crystal

182   structure. However, tertiary structure modelling of WzaX, WzaS, and WzaB against the Wza_{Ec} 3D

183   structure revealed WzaX, WzaS, and WzaB to be missing the 2nd SLBB domain (i.e. D3), and more

184   importantly the crucial OM-spanning α-helical Wza_C domain (i.e. D4), of Wza_{Ec} (**Fig. 1B**). The

185   absence of such an OM-spanning domain is consistent with the lack of WzaX/S/B detection in

186   proteomic analyses of OM vesicle (OMV) and biotinylated surface-protein samples (Kahnt *et al.*,

187   2010), despite the constitutive expression of the *wzaX/S/B* genes throughout the *M. xanthus* lifecycle

188   (Muñoz-Dorado *et al.*, 2019, Sharma *et al.*, 2021).

189   **Figure 2. Structural diversity of OPX proteins. A)** Domain organization and abundance of

190   bacterial OPX protein classes identified in the myxobacterial (MYXO, 61 genomes, 506 873 proteins

191   analyzed), representative (REP, 3662 genomes, 13 600 490 proteins analyzed), and NCBI non-

192   redundant (NR, 371 327 556 proteins analyzed) databases. The Poly_export (PF02563), SLBB

193   (PF10531), Wza_C (PF18412), and Caps_synth_GfcC (PF06251) Pfam domains were used to

194   query the various databases, followed by fold-recognition analysis using HHpred against the 3D X-

195   ray crystal structures of Wza$_{Ec}$ (PDB: 2J58) and GfcC (PDB: 3P42). The number of repeated copies

196   is indicated under each domain depiction. The number of OPX hits (*bold*) for a specific class is

197   indicated as well as the proportion of hits from each database (*italics*) represented by the hits.   **B)**

198   Sequence logo of the consensus amino acids constituting the OM-spanning α-helix based on a

199   multiple-sequence alignment of 1 586 Class 1 OPX proteins. The region of sequence alignment

200   with Wza$_{Ec}$ is indicated and depicted with the associated secondary structure from the Wza$_{Ec}$ X-ray

201   crystal structure (PDB: 2J58) (Dong *et al.*, 2006). Also depicted is the region of sequence alignment

202   with a previously-published optimized Wza_C synthetic peptide (based on 94 close Wza$_{Ec}$-related

203   homologues) capable of spontaneously inserting into lipid bilayers and self-assembling into stable α-

204   barrel pores (Mahendran *et al.*, 2017). **C)** Sequence logo of the consensus amino acids

205   constituting the putative OM-spanning α-helix based on a multiple-sequence alignment of 452 Class

206   2B OPX proteins. The region of sequence alignment with KpsD$_{Ec}$ is indicated, along with the

207   predicted KpsD$_{Ec}$ secondary structure. The position of observed (*dark-colored*) and predicted (*light-*

208   *colored*) α-helices (*boxes*) and β-strands (*arrows*) have been indicated.

FIGURE 2

A)

Databases: MYXO | REP | NR

61	3662	All	Total Genomes Analyzed
506 873	13 600 490	371 327 556	Total Proteins Analyzed

OPX Class — **Database Search Results** — **OPX Protein Architecture**

CLASS 1: 26 (13.68%), 1 586 (38.12%), 34 611 (37.72%) — Poly_export SLBB Wza_C (1–n)

CLASS 2:
- (A) 1 (0.53%), 106 (2.55%), 4 583 (4.99%) — Poly_export SLBB GfcC (1–n)
- (B) 1 (0.53%), 452 (10.86%), 10 363 (11.29%) — Poly_export SLBB GfcC Wza_C (1–n)
- (C) 20 (0.02%) — GfcC Poly_export SLBB (1–n)
- (D) 15 (0.02%) — GfcC Poly_export SLBB Wza_C (1–n)

CLASS 3: 162 (85.26%), 2 017 (48.47%), 42 161 (45.95%) — Poly_export SLBB (1–n)

| 190 | 4 161 | 91 753 | Total Identified OPX Proteins |

B)

Class 1 OPX — Wza_C Domain Consensus

P Y D I V Y V T T A P L V R W N R V I S Q L V P T I S G V H D M T E T V R Y I K R W P N (Wza_{Ec} C-terminal Sequence)

(Wza_{Ec} 2° Structure)

A P L V R W N R V I S Q L V P T I T G V H D L T E T V R Y I K T W P N (Self-Inserting Consensus Wza Peptide)

C)

Class 2B OPX — Wza_C Domain Consensus

P G D E I M V L P K Y E S K N I E V T R G I S T I L Y Q L A V G A K V I L S L ($KpsD_{Ec}$ C-terminal Sequence)

($KpsD_{Ec}$ 2° Structure)

OPX proteins constitute three distinct structural classes

211   To determine if the absence of the Wza_C domain was an aberration confined to the subset of

212   OPX proteins from *M. xanthus* under study, we first performed a comparative genomics analysis

213   using profile-based homology searches across three different datasets: (i) 61 myxobacterial genomes

214   (MYXO) (**Supplementary Table S1**), (ii) 3662 representative genomes (REP) (**Supplementary**

215   **Table S2**), and (iii) the non-redundant (NR) NCBI database (371 327 556 proteins at 100% identity

216   as of June 10, 2021) (**Supplementary Table S3**), to identify encoded OPX proteins, using

217   PF02563 [Poly_export], PF10531 [SLBB], PF18412 [Wza_C], and PF06251 [Caps_synth_GfcC;

218   used here as "GfcC"] as our query domains. These profile-based analyses identified diverse putative

219   OPX homologues that we divided into three distinct classes according to their domain architecture

220   (**Fig. 2A**). The first set of OPX proteins was found to contain Poly_export–SLBB$_{(1-14)}$ architecture

221   ending with a C-terminal OM-spanning Wza_C domain, similar to Wza$_{Ec}$, and was assigned the

222   designation "Class 1" (**Fig. 2A**). A second set of OPX proteins was found to possess Poly_export–

223   SLBB$_{(1-6)}$–GfcC architecture similar to KpsD$_{Ec}$, ending with or without a C-terminal OM-spanning

224   Wza_C domain, and was assigned the designation "Class 2". However, most OPX proteins were

225   found to contain only Poly_export–SLBB$_{(1-7)}$ architecture lacking either a Wza_C or GfcC domain;

226   these hits were designated "Class 3"; however, many of these initial hits were found to contain

227   additional amino acids that may have remained uncharacterized following sequence-based domain

228   detection. Therefore, to probe these partially-characterized hits in more detail, we subjected all

229   identified OPX proteins to fold-recognition analysis using HHpred to identify matches with more

230   remote sequence homology but conserved structural properties. These analyses resulted in

231   reclassification of several original Class 3 hits to either Class 1 or Class 2

232   Incidentally for Class-1 OPX proteins, while the secondary structure was conserved,

233   considerable sequence variation was detected within certain regions of the putative OM-spanning

234   Wza_C domains, with this domain extending up to 48 residues in length (compared to 35 residues in

235   Wza$_{Ec}$) (**Fig. 2B**). As per the MYXO/REP/NR databases, 13.7/38.1/37.7% of OPX proteins possess

236   Class 1 Wza$_{Ec}$-like organization with a putative OM-spanning C-terminal α-helix (**Fig. 2A,**

237   **Supplementary Tables S1, S2, S3**). Class 1 OPX proteins were found to have a median length

238   of 378 amino acids and most (1233/1586, ~78%) were predicted to be lipoproteins with Sec/SPII

239   signal sequences.

240   These proteins were largely confined to the phylum Proteobacteria (837/1632 genomes;

241   ~51%) with a predominance in classes Gammaproteobacteria (430/760 genomes; ~56%),

242   Alphaproteobacteria (232/410 genomes; ~57%), and Betaproteobacteria (133/274 genomes; ~49%),

243   and representation also in phyla Bacteroidetes (67/283 genomes; ~24%), Planctomycetes (56/62

244   genomes; ~90%), and Cyanobacteria (46/55 genomes; ~84%) (**Supplementary Table S2**). Based

245   on species-level PSORTdb classification, the REP database contains 698 Gram-positive and 1381

246   Gram-negative organisms. Our analysis revealed that Class 1 OPX proteins are encoded by many

247   Gram-negative bacteria (639/1381 genomes, ~46%), whereas these proteins were completely absent

248   in Gram-positive species (**Supplementary Table S2**).

249   Our MYXO/REP/NR database comparative genomic analysis revealed that Class 2 OPX

250   proteins can be further divided into four subclasses. Proteins belonging to Class 2A contain

251   Poly_export–SLBB$_{(1-n)}$–GfcC architecture, whereas those assigned to Class 2B possess Poly_export–

252   SLBB$_{(1-n)}$–GfcC–Wza_C architecture ending with an OM-spanning α-helical domain. Classes 2C

253   and 2D are variations of Classes 2A and 2B (respectively) where the Poly_export domain is preceded

254   by a GfcC domain; however, only 20 Class 2C and 15 Class 2D proteins were identified across the

255   entire NR database.

256   Class 2A OPX proteins constitute 0.5/2.6/5.0% of all OPX proteins identified in the

257   MYXO/REP/NR databases, with a median length of 605 amino acids. These proteins were found to

258   be encoded mainly in Proteobacteria (77/1632 genomes, ~5%), Bacteroidetes (9/283 genomes, ~3%),

259   and Acidobacteria (3/14 genomes, ~21%). Taxonomy orders Alteromonadales (14/83 genomes,

260   ~17%), Campylobacterales (10/88 genomes, ~11%), Burkholderiales (9/175 genomes, ~5%), and

261   Oceanospirillales (9/43 genomes, ~21%) display the maximum representation for Class 2A. In

262   addition, Class 2A OPX proteins are encoded in only ~5% (66/1381 genomes) of Gram-negative

263   bacteria and are completely absent among Gram-positive species (**Supplementary Table S4**).

264   Class 2B OPX proteins represent 0.5/10.9/11.3% of all OPX proteins identified in the

265   MYXO/REP/NR databases (**Fig. 2A, Supplementary Table S1, S2, S3**). The median length of

266   Class 2B proteins was found to be 824 amino acids, with most (375/452, ~83%) found to possess

267   standard Sec/SPI secretory signal peptides. These proteins were largely encoded by Proteobacteria

268   (226/1632 genomes, ~14%), Bacteroidetes (107/283 genomes, ~38%) and Cyanobacteria (14/55

269   genomes, ~25%). At the level of Order, Alteromonadales (66/83 genomes, ~80%), Bacteroidales

270   (38/52 genomes, ~73%), Cytophagales (23/50 genomes, 46%), and Vibrionales (20/46 genomes,

271   ~43%) were found to contain the most organisms encoding Class 2B OPX architecture. Class 2B

272   OPX proteins have representation in only ~18% (243/1381 genomes) of Gram-negative organisms

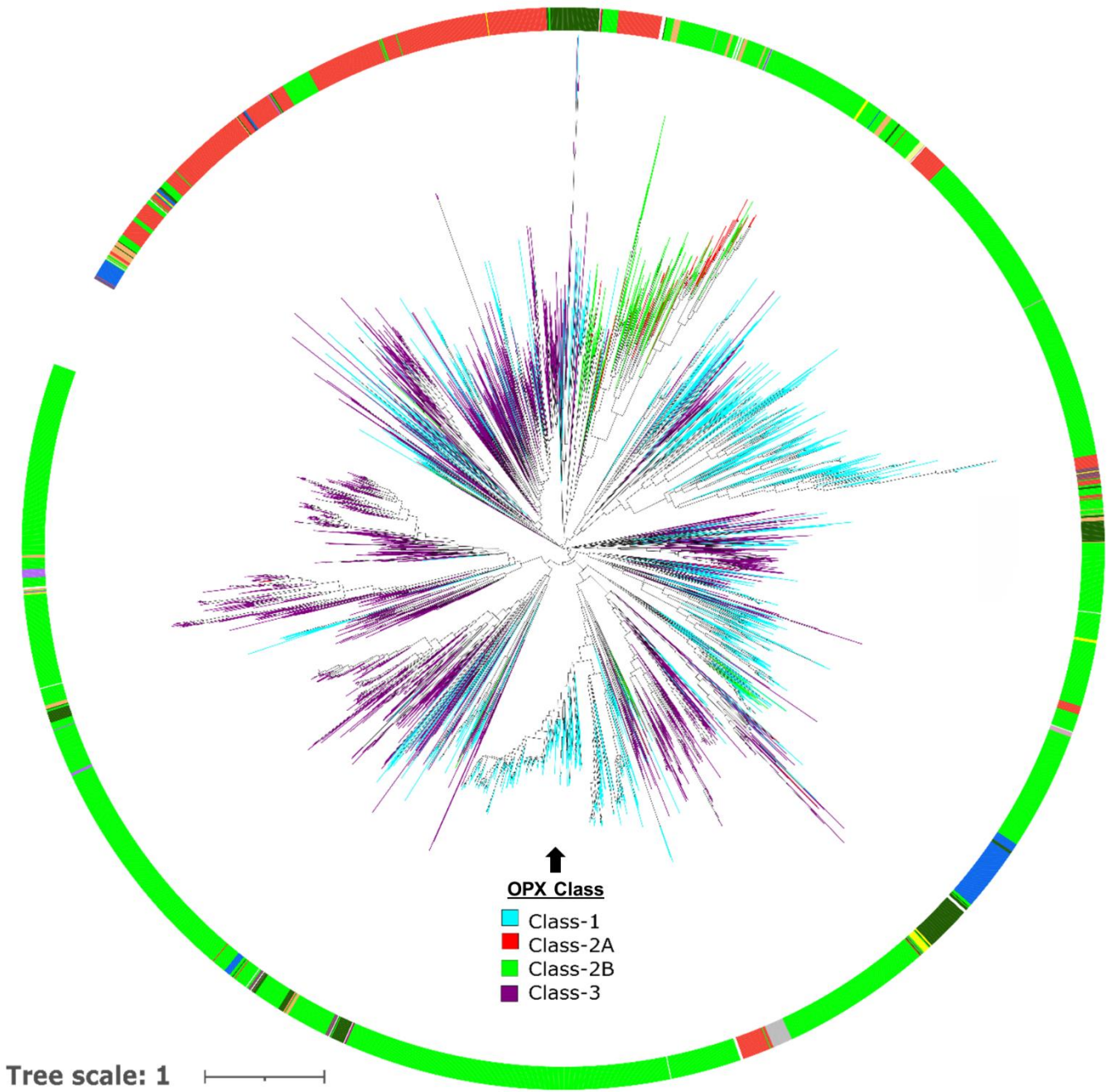273   and are absent in Gram-positive bacteria (**Supplementary Table S4**).

274   Class 2B architecture is typified by $KpsD_{Ec}$. Consistent with a previous report (Sande *et al.*,

275   2019), fold-recognition analysis of $KpsD_{Ec}$ revealed that most of its N-terminus is structurally

276   homologous to Wza_{Ec}, while the bulk of its C-terminus is a structural match to the standalone GfcC

277   protein (**Supplementary Fig. S1A**). However, the extreme C-terminus of $KpsD_{Ec}$ — i.e. the

278   portion of $KpsD_{Ec}$ surpassing the end of structural homology with the GfcC D4 α-helix — was found

279   to have considerable structural homology with the most C-terminal region of Wza_{Ec}, including a 25-

280   residue tract with α-helical propensity matched to the OM-spanning α-helical tract of Wza_{Ec}

281   (**Supplementary Fig. S1B,C**). A similarly-extended C-terminal α-helix was found throughout the

282   Class 2B OPX hits identified herein, with considerable variation in certain regions of its sequence,

283   and extending to 38 residues (compared to 25 residues in $KpsD_{Ec}$) (**Fig. 2C**). This observation

284   supports the notion that a part of $KpsD_{Ec}$ (and by extension Class 2B OPX proteins) may indeed be

285   able to span the OM and access the cell surface.

286   Finally, Class 3 OPX proteins with Poly_export–SLBB$_{(1–n)}$ architecture (but no appreciable

287   peptide sequence following their respective C-terminal-most SLBB domain), with a median length of

288   256 amino acids, represent a plurality (~85/49/46%) of OPX proteins identified across the

289   MYXO/REP/NR databases (**Fig. 2A, Supplementary Table S1, S2, S3**). Almost 50% are

290   predicted lipoproteins (Sec/SPII signal sequences) while ~30% are secreted proteins (Sec/SPI signal

291   sequences). Such OPX proteins (i.e. those lacking a Wza_C or GfcC) domain were found across

292   multiple bacterial classes such as Alphaproteobacteria (274/410 genomes, ~67%),

293   Gammaproteobacteria (247/760 genomes, ~33%), Betaproteobacteria (100/274 genomes, ~37%),

294   Flavobacteria (116/133 genomes, ~87%), and Deltaproteobacteria (66/82 genomes, ~80%).

295   Expectedly, our analysis detected Class 3 OPX proteins in Gram-negative bacteria (600/1382

296   genomes, ~43%), but also intriguingly in several Gram-positive organisms (52/699 genomes, ~7%)

297   (**Supplementary Table S4**). Of note, the proportions of each Class of OPX protein detected in the

298   REP database were highly reflective of those found in the NR database (**Fig. 2A**), reinforcing the

299   utility and applicability of the REP database.

300   Within the MXYO dataset, we identified 26 Class 1, two Class 2, and 162 Class 3 OPX

301   proteins (**Fig. 2A**, **Supplementary Table S1**). Class 3 OPX proteins were encoded in all 61

302   myxobacterial organisms, without any exception, in the range of 1-4 proteins. Class 1 OPX proteins

303   were encoded by several members of the suborder Cystobacterineae such as *Anaeromyxobacter*,

304   *Cystobacter*, *Corallococcus*, *Pyridicoccus*, and *Simulacricoccus*. However, Class 1 OPX proteins

305   were not represented within any species of the genus *Myxococcus*. Class 2 OPX proteins were only

306   present in two myxobacteria, namely *Sandaracinus amylolyticus* (Class 2A OPX) and *Haliangium*

307   *ochraceum* (Class 2B OPX). All proteins in the MYXO dataset belonging to the three OPX classes

308   possess similar median lengths to those described for all OPX proteins in the REP database (i.e.

309   MYXO Class 1, 373 amino acids; MYXO Class 2, 550 amino acids; MYXO Class 3, 205 amino

310   acids).

311 **Figure 3. Phylogenetic tree of OPX proteins.** Sequence alignment of all "Poly_export" domains
312 as identified in 4 161 OPX proteins in the REP dataset was used to generate a maximum-likelihood
313 phylogenetic tree. The classes of OPX proteins (*inner tree*) and their respective phyla (*outer ring*)
314 have been coloured accordingly for effective visualization.

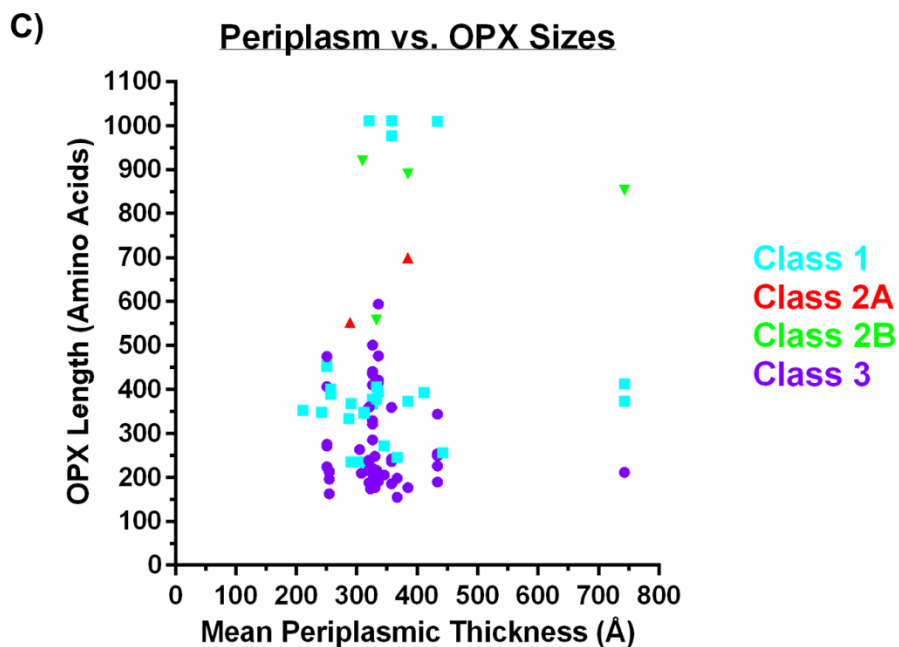**FIGURE 3**

Colored Ranges (Phylum-Level Taxonomy)

- Proteobacteria
- Actinobacteria
- Firmicutes
- Cyanobacteria
- Planctomycetes
- Bacteroidetes
- Verrucomicrobia
- Chlorobi
- Nitrospirae
- Chlamydiae
- Thermotogae
- Acidobacteria

OPX Class
- Class-1
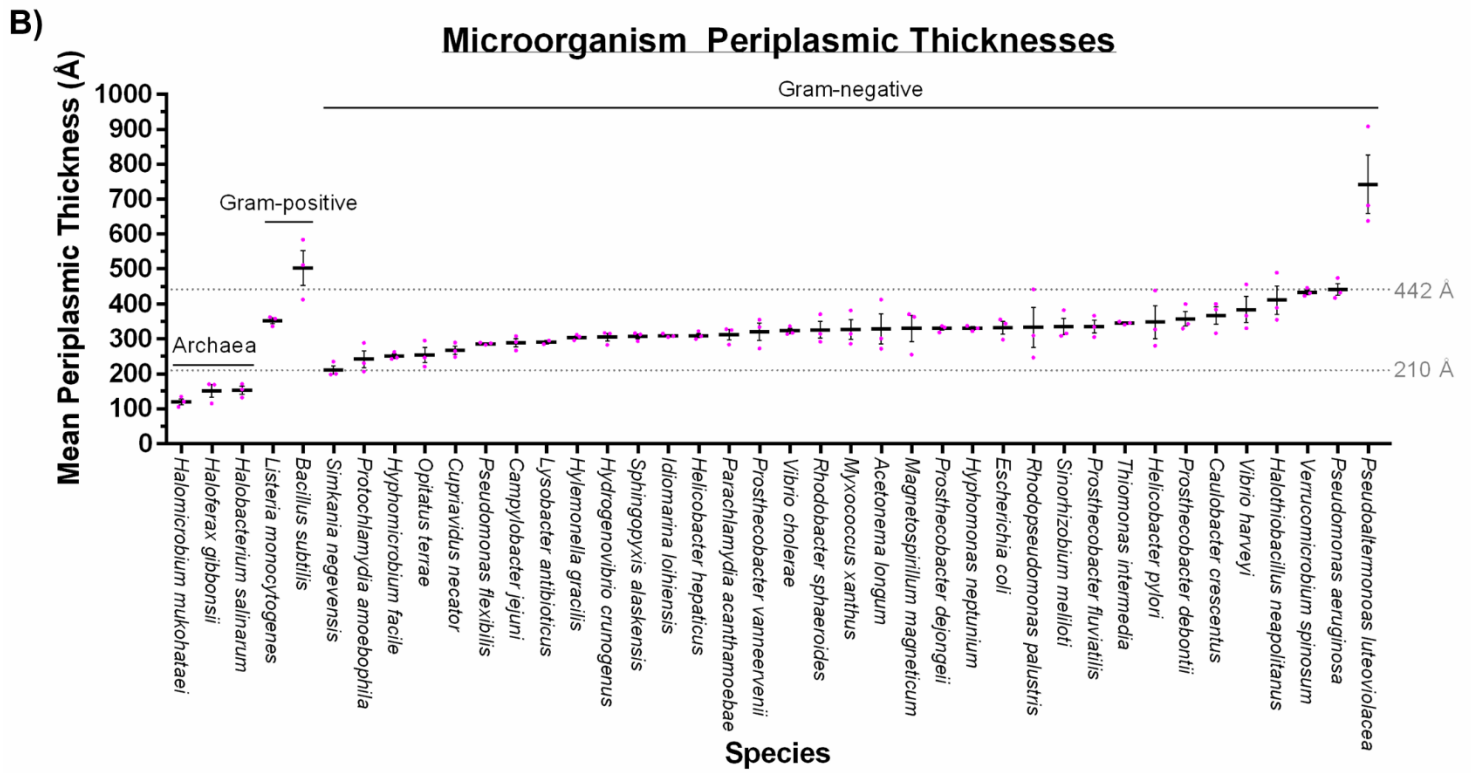- Class-2A
- Class-2B
- Class-3

Tree scale: 1

316   *Molecular phylogenetics suggests the coevolution of three OPX protein Classes*

317   Taken together, these analyses have identified three principal structural classes of OPX

318   proteins, namely: (i) Class 1 (i.e. Wza$_{Ec}$-like with a canonical OM-spanning α-helix), (ii) Class 2 (i.e.

319   KpsD$_{Ec}$-like with a potentially OM-spanning α-helix at the terminus of the GfcC domain), and (iii)

320   the majority Class 3 (i.e. those with structural homology to the Wza$_{Ec}$ N-terminus, but with no

321   discernible OM-spanning domain). By extension, WzaX/S/B from *M. xanthus* (**Fig. 1B**) are thus

322   Class 3 OPX proteins that share the domain architecture of most other OPX proteins encoded by

323   bacteria (**Fig. 2A**). Given that all OPX proteins among the three different classes have a conserved

324   "Poly_export" domain, this can be utilized as a phylogenetic marker. Therefore, based on the

325   hmmscan results, we extracted the location of the "Poly_export" domain from REP dataset hits,

326   aligned those sequences using MUSCLE, and generated a maximum-likelihood phylogeny. The

327   generated phylogeny (**Fig. 3**) revealed that Class 1 and Class 3 OPX proteins are interspersed with

328   each other in all taxonomic clades, suggesting that proteins from these two classes have co-evolved

329   by losing or gaining the Wza_C segment in closely-related organisms. However, Class 2A and Class

330   2B are both present in nearby sister clades and away from Class 1 and Class 3. This denotes that

331   Class 2 OPX proteins, similar to proteins from Class 1 and Class 3, are highly similar to each other

332   and have coevolved by losing or gaining their respective Wza_C segments.

333 **Figure 4. Cell envelope ultrastructure in Gram-negative bacteria. A)** Cryo-electron microscopy

334 tomogram slice of a *M. xanthus* cell, showing the IM, OM, and intervening periplasmic space and

335 their respective measured thicknesses. **B)** Comparison of means (*black bar*) for periplasmic

336 distances in 40 microbial species (± SEM). Individual replicate measurements are indicated (*red*

337 *dots*). Data from organisms with increasing mean periplasmic thickness values have been depicted

338 from left to right, grouped according to Gram-negative, Gram-positive, or Archaea organism

339 designation. **C)** Scatterplot of mean periplasmic thickness values plotted against the length of each

340 OPX protein from the same organism. Data points have been coloured according to the Class of

341 OPX protein assigned herein. No correlation was detected between periplasmic thickness and (i)

342 overall OPX protein length (Pearson coefficient: 0.1450; Spearman coefficient: 0.05884, calculated

343 over 92 data pairs), (ii) Class 1 OPX hits (Pearson coefficient: 0.07250; Spearman coefficient:

344 0.2318; calculated over 33 data pairs), or (iii) Class 3 OPX hits (Pearson coefficient: -0.1127;

345 Spearman coefficient: -0.1242; calculated over 53 data pairs).

FIGURE 4

Figure 4. A) OM, IM membrane image with measurements 70 Å, 327 Å, 62 Å. B) Microorganism Periplasmic Thicknesses. C) Periplasm vs. OPX Sizes.

347   *OPX protein lengths and classes in Gram-negative bacteria are not linked to periplasmic distance*

348   In *E. coli*, the integral OM Class 1 Wza$_{Ec}$ OPX protein is proposed to form a complex with

349   the integral IM Wzc PCP protein that creates a contiguous periplasm-spanning channel for polymer

350   export (Collins *et al.*, 2007) (**Fig. 1A**). To gain an understanding of the relationship between the

351   subcellular architecture of *M. xanthus* and the role of the WzaX, WzaS, and WzaB Class 3 OPX

352   proteins, we therefore compared the sizes of various cellular compartments and structures from cryo-

353   electron tomography projections of the *M. xanthus* envelope. This revealed the *M. xanthus* OM to

354   have an average thickness of 69.8 ± 1.8 Å, compared to the average thickness of 62 ± 1.6 Å for its

355   IM, with a mean inter-membrane periplasmic thickness of 327 ± 28.4 Å (**Fig. 4A**).

356   Given the enrichment of Class 1 OPX proteins in certain bacterial genera and different

357   median OPX sizes for each OPX protein class (**Supplementary Table S4**), we examined whether

358   the specific size of an OPX protein identified in a given bacterium was associated with the thickness

359   of the periplasm in that organism. We first measured the distance between the IM and OM at lateral

360   positions in cryo-electron tomography projections of cells from an additional 34 species of Gram-

361   negative bacteria. For reference, we also analyzed IM–peptidoglycan and IM–S-layer thickness in

362   several Gram-positive bacteria and Archaea (respectively) (**Fig. 4B**). This analysis revealed a range

363   of Gram-negative periplasmic thicknesses confined between the lower and upper thresholds of 210

364   and 442 Å (respectively), with the lone exception being *Campylobacter jejuni*, displaying a mean

365   periplasmic thickness of 743 Å (**Fig. 4B**). For any of these species in which OPX proteins were

366   herein identified (**Supplementary Tables S1, S2, S3**), we next compared the average measured

367   periplasmic thickness with the length of the OPX protein(s) in each system. However, no overall

368   correlation between the two variables was detected across all OPX proteins in this analysis, nor

369   specifically within Class 1 or Class 3 OPX hits (**Fig. 4C**).

WzaX, WzaS, and WzaB are genomically paired with 18-stranded β-barrel proteins

370

371   Given the lack of identifiable OM-spanning domains in WzaX/S/B (**Fig. 1B**), we sought to

372 identify candidate proteins that could permit export of synthesized EPS, MASC, and BPS polymers

373 across the *M. xanthus* OM. Through our previous analyses of the EPS, MASC, and BPS biosynthesis

374 clusters, we demonstrated that WzaX (MXAN_7417), WzaS (MXAN_3225), and WzaB

375 (MXAN_1916) were encoded immediately adjacent to MXAN_7418, MXAN_3226, and

376 MXAN_1916 (respectively), with this synteny conserved for the majority (115/162, ~71%) of Class

377 3 OPX proteins in myxobacterial genomes (Islam *et al.*, 2020), supporting the notion that the latter

378 three proteins are important for each respective polymer synthesis pathway. To analyze the specific

379 structural potential for each protein, MXAN_7418, 3226, and 1916 were first subjected to

380 evolutionary coupling analysis, revealing the predicted presence of 18 principal β-strands for

381 MXAN_7418, MXAN_3226, and MXAN_1916 (**Supplementary Figs. 2A,3A,4A**).

382   Fold-recognition analyses revealed structural homology of the C-terminal 76-86% of

383 MXAN_7418/3226/1916 to the complete C-terminal integral OM β-barrel module (residues 513-807)

384 of PgaA (PgaA$_{βb}$, PDB: 4Y25) (Wang *et al.*, 2016), at 98.9%, 99.4%, and 99.2% probability

385 (respectively). PgaA is the OM porin responsible for secretion of PNAG polymer (which is heavily

386 implicated in biofilm integrity) in Gram-negative bacteria; it contains multiple periplasmic

387 tetratricopeptide repeats at its N-terminus, followed by a 16-stranded integral OM β-barrel domain

388 closed by four extracellular loops (Wang *et al.*, 2016). Intriguingly, PNAG is produced by a

389 synthase-dependent pathway (Whitney & Howell, 2013). In each of MXAN_7418/3226/1916, the

390 PgaA$_{βb}$-like module is extended by two integral OM β-strands at the N-terminus of each protein,

391 suggesting that MXAN_7418, 3226, and 1916 do indeed have the propensity to form 18-stranded β-

392 barrels (**Supplementary Figs. 2B,3B,4B**).

393   For the 162 Class 3 OPX proteins identified across 61 myxobacterial genomes (with

394 Poly_export—SLBB$_{1-2}$ architecture), most (115/162, ~71%) were found to be encoded near an

395 extended PgaA$_{βb}$-like protein, whereas the 26 Class 1 (with Poly_export—SLBB$_{1-2}$—Wza_C

396 organization) and two Class 2 OPX proteins were not encoded near any such β-barrel protein

397 (**Supplementary Table S1**). We again expanded our analysis beyond *M. xanthus* to determine

398 whether the presence of a β-barrel porin was a common occurrence in pathways containing an OPX

399 protein. Intriguingly in *E. coli*, the *gfcABCDE-etp-etk* cluster needed for Wzx/Wzy-dependent Group

400   4 CPS production, encodes the OPX protein GfcE (formerly YccZ/Wza$_{22min}$) as well as the protein

401   GfcD (formerly YmcA) (Peleg *et al.*, 2005). The separate *yjbEFGH* (paralogous to *gfcABCD*)

402   operon implicated in polysaccharide secretion encodes the GfcD-like protein YjbH (Ferrières *et al.*,

403   2007). Both GfcD and YjbH are β-barrel proteins, which were recently identified to be part of a

404   novel class of OM proteins (lacking published structures) with two separate β-barrels predicted to be

405   formed by the same polypeptide chain (Solan *et al.*, 2021). Herein, fold-recognition analysis

406   revealed the N-terminal halves to be matches to the β-barrel amyloid transporter FapF from

407   *Pseudomonas*, whereas the C-terminal halves (GfcD$_{Cterβb}$ and YjbH$_{Cterβb}$, respectively) possessed

408   structural homology to the PNAG PgaA$_{βb}$ module described above (**Supplementary Fig. 5A**).

409   This double-barrel arrangement was supported by AlphaFold2-generated deep-learning structure

410   models for both full-length GfcD and YjbH (**Supplementary Fig. 5B**), with the larger barrel

411   portions displaying considerable sequence homology to PgaA$_{βb}$ (**Supplementary Fig. 5C**).

412   To probe for the presence of similar β-barrels encoded near other OPX proteins, we used

413   sequence homology searches (BLAST and HMMER) to examine the genomic context (up to 10

414   genes upstream and downstream) of the various OPX proteins we identified in the REP dataset,

415   beginning with the β-barrel sequences of MXAN_7418, MXAN_3226, and MXAN_1916. Given the

416   homology of the above proteins to PgaA, we added the PgaA$_{βb}$, GfcD$_{Cterβb}$, and YjbH$_{Cterβb}$ sequences

417   as well. In addition, the β-barrel sequences of BcsC (BcsC$_{βb}$, from PDB: 6TZK) (Acheson *et al.*,

418   2019) and AlgE (from PDB: 4AFK) (Tan *et al.*, 2014) were also included, given their porin functions

419   in the synthase-dependent cellulose and alginate production pathways, respectively (Whitney &

420   Howell, 2013). Finally, we also included the sequence of Wzi (PDB: 2YNK) lacking the plug

421   domain (Wzi$_{βb}$), as this is an 18-stranded β-barrel known to be linked with polysaccharide

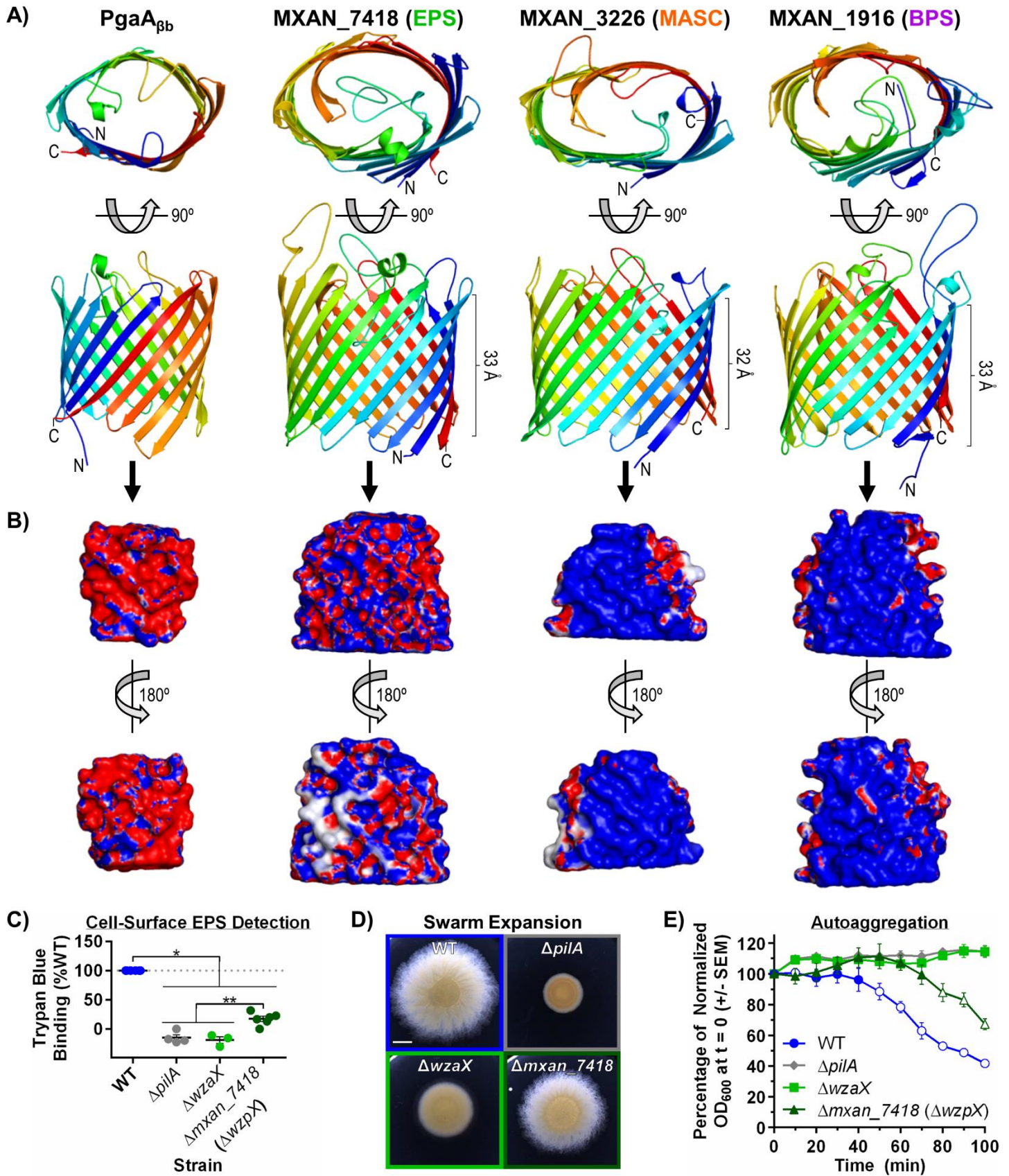422   biosynthesis clusters (Bushell *et al.*, 2013).

423   Altogether, this analysis detected 365 β-barrel query homologues encoded near 344 OPX

424   proteins of all three classes (**Table 1**). Of the 156 matches to the Wzi$_{βb}$ query, 128 were of

425   comparable size and conserved alignment to full-length Wzi (including the α-helical plug domain),

426   consistent with these hits not being likely porin candidates. However, 28 homologues were a match

427   to only Wzi$_{βb}$ (i.e. no plug domain), reinforcing their candidacies as trans-OM export β-barrels;

428   intriguingly, HHpred analysis of these 28 hits also revealed many with strong similarity to DUF6029,

429   ascribing a potential polysaccharide secretion role to this heretofore uncharacterized protein domain.

430   The PgaA$_{βb}$ query detected 12 homologues, 2 of which were uniquely-matched to only the β-barrel

431 domain, whereas the remaining 10 aligned with full-length PgaA, consistent with these latter hits

432 possessing the numerous N-terminal TPR domains as well as the C-terminal PgaA$_{\beta b}$ porin domain.

433 These results were similar for BcsC$_{\beta b}$, i.e. that of the 5 homologues detected near OPX proteins, each

434 was a match to full-length BcsC indicating a conservation of N-terminal TPR and C-terminal porin

435 domains (with no matches confined to BcsC$_{\beta b}$ detected). Fifteen homologues to synthase-dependent

436 pathway AlgE were also detected. Profile matches to GfcD$_{Cter\beta b}$ and YjbH$_{Cter\beta b}$ queries (68 and 84,

437 respectively) were largely matched via length and conserved alignment to full-length GfcD and

438 YjbH, suggesting that these homologues contain the N-terminal FapF-like β-barrel domain as well as

439 the C-terminal PgaA$_{\beta b}$-like polymer-secretion domains; only two GfcD$_{Cter\beta b}$ and three YjbH$_{Cter\beta b}$ hits

440 were identified with homology to only the C-terminal barrels of each protein. However, the

441 GfcD$_{Cter\beta b}$ homologue detected in *Aquifex aeolicus* (WP_164930611.1), as well as the hits to

442 YjbH$_{Cter\beta b}$ detected in *A. aeolicus* (WP_010880290.1) and *Sulfitobacter pseudonitzschiae*

443 (WP_174861591.1), were only matched to the respective polysaccharide-secretion modules,

444 indicating the potential for stand-alone export of sugar chains in these systems. Finally, homologues

445 to MXAN_7418, MXAN_3226, and MXAN_1916 were only found to be encoded in myxobacterial

446 genomes. Importantly, the presence of these extended PgaA$_{\beta b}$-like 18-stranded β-barrels in

447 myxobacteria was linked to nearby Class 3 OPX proteins. However, unlike OPX hits from the

448 MYXO database, only 77/2017 Class 3 OPX proteins from the REP dataset were identified to be

449 encoded near MXAN_7418/3226/1916/ PgaA$_{\beta b}$/GfcD$_{Cter\beta b}$/YjbH$_{Cter\beta b}$/BcsC$_{\beta b}$/AlgE/Wzi$_{\beta b}$ stand-alone

450 porin homologues.

451   Together, these data reveal intriguing architectural similarities between β-barrel porin

452 modules from synthase-dependent polymer export pathways and those implicated in myxobacterial

453 Wzx/Wzy-dependent secretion, as well as analogous or ABC transporter-dependent pathways in

454 diverse bacteria, all previously unreported associations.

455 **Figure 5. OPX-companion β-barrel structures.** **A)** Tertiary structure models (top and front views)
456 for MXAN_7418, MXAN_3226, and MXAN_1916, as generated using deep learning via AlphaFold2,
457 as well as the PgaA C-terminal domain (aa 513–807) X-ray crystal structure (PDB: 4Y25) (Wang *et*
458 *al.*, 2016). Structures are coloured with a spectrum, from the N-terminus (*blue*) to the C-terminus
459 (*red*), and depicted with smooth loops. **B)** Front and back views of the interior spaces of the β-
460 barrels depicted in Panel A overlaid with the electrostatic character of the residues contacting the
461 lumenal volume, as generated via HOLLOW (Ho & Gruswitz, 2008). Surfaces have been colored
462 according to charge, from blue (positive, +5 kT/e) to white (uncharged/hydrophobic), to red
463 (negative, −5 kT/e). **C)** Trypan Blue dye-binding for *M. xanthus* DZ2 WT (n = 6), Δ*pilA* (n = 4),
464 Δ*wzaX* (n = 3), and Δ*mxan_7418* (i.e. Δ*wzpX,* n = 4) to probe cell-surface EPS levels. Mean values
465 are indicated (+/− SEM), with each biological replicate data point indicated. Means of all mutants
466 were significantly lower than WT, while that of Δ*mxan_7418* was significantly higher than either
467 Δ*pilA* or Δ*wzaX*, as calculated via Student's T-test ($p < 0.05$). **D)** T4P-dependent swarm expansion
468 of strains tested in Panel C. Scale bar: 4 mm. **E)** Auto-aggregation profiles of strains tested in Panel
469 C for cells resuspended in CYE rich medium at an initial OD_{600} of 1.0. Mean values (n = 3) are
470 indicated +/− SEM. Open plot points: no statistically significant difference in mean relative to Δ*wzaX*
471 at a given time point. Closed plot points: statistically significant difference of means relative to
472 Δ*wzaX* at a given time point. Significance was evaluated via Student's T-test ($p \leq 0.05$).

FIGURE 5

A) PgaA$_{\beta b}$ | MXAN_7418 (EPS) | MXAN_3226 (MASC) | MXAN_1916 (BPS)

B)

C) Cell-Surface EPS Detection

D) Swarm Expansion

E) Autoaggregation

474   ***WzpX, WzpS, and WzpB are (respectively) integral OM β-barrel EPS-, MASC-, and BPS-pathway***
475   ***porins***

476   To examine the structural suitability of the WzaX/S/B co-occurring β-barrels for the
477   respective translocation of EPS/MASC/BPS, and since no full-length template structure could be
478   identified, we employed the deep-learning approach provided by AlphaFold2 to generate a tertiary
479   structure model. AlphaFold2 employs information from evolutionarily-coupled amino acids as well
480   as templates with structural homology to fold a polypeptide sequence using an iterative process
481   (Jumper *et al.*, 2021). Consistent with the above-described data (**Supplementary Figs. 2,3,4**),
482   MXAN_7418, 3226, and 1916 were all predicted to form 18-stranded β-barrels with sizeable central
483   cavities, with respective barrel heights of 33, 32, and 33 Å (**Fig. 5A**). As molecular dynamics
484   simulations typically calculate the hydrophobic thickness of asymmetric OM bilayers to be ~40% of
485   their total solvated thickness (Pavlova *et al.*, 2016), based on our measured *M. xanthus* OM thickness
486   of 69.8 ± 1.8 Å (**Fig. 4A**), an approximated hydrophobic thickness of ~28 Å would indeed by
487   traversable by the proposed MXAN_7418, MXAN_3226, and MXAN_1916 tertiary structures.

488   We subsequently used HOLLOW to probe the lumenal volume of the EPS/MASC/BPS-
489   cluster β-barrels via filling of the internal space with dummy atoms to generate a cast of the void
490   space, after which the electrostatic potential of the contacting β-barrel surface was overlaid. To
491   validate this approach, we first probed the internal electrostatics of the PgaA$_{\beta b}$ template structure,
492   revealing a highly anionic interior (**Fig. 5B**), consistent with passage of the cationic PNAG polymer
493   through the lumen of the barrel. BPS was previously discovered to be a randomly-acetylated anionic
494   repeating tetrasaccharide, with the distal three sugars of each repeat constituted by
495   mannosaminuronic acid (ManNAcA) units (Islam *et al.*, 2020). Therefore, the cationic charge
496   character of the MXAN_1916 lumen (**Fig. 5B**) is indeed suitable for passage of its associated HMW
497   BPS polymer. While the chemical structures or exact compositions of MASC or EPS are not known,
498   isolated spore coat material was found to contain GalNAc chains with potential glucose (Glc) and
499   glycine decorations (Holkenbrink *et al.*, 2014). As the interior of the MXAN_3226 β-barrel is
500   cationic (**Fig. 5B**), this suggests that MASC may have a net-anionic charge character, as contributed
501   via as-yet-unidentified sugars and/or chemical modifications. EPS composition has been probed
502   across four investigations (Islam *et al.*, 2020, Behmlander & Dworkin, 1994, Gibiansky *et al.*, 2013,
503   Sutherland & Thomson, 1975), with Ara, Gal, GalNAc, Glc, GlcN, GlcNAc, Man, ManNAc, Rha,
504   and Xyl having been identified (depending on the publication); however, none of these sugars are

505   highly charged, which is consistent with the more neutral character of the MXAN_7418 interior

506   (compared to that of either MXAN_1916 or MXAN_3226) (**Fig. 5B**).

507   To probe the implication of Wzx/Wzy-dependent pathway-associated β-barrels in polymer

508   secretion, we next set out to better understand their physiological contexts. RNAseq analysis

509   previously detected the transcripts encoding MXAN_7418 and MXAN_1916 in vegetative cells, as

510   well as MXAN_3226 in developmental cells, indicating that all three β-barrels are indeed expressed

511   over the course of the *M. xanthus* lifecycle (Muñoz-Dorado *et al.*, 2019, Sharma *et al.*, 2021).

512   Furthermore, MXAN_1916 was detected in proteomic screens of biotinylated surface-exposed

513   proteins, and MXAN_1916 and MXAN_3226 were both detected in OMV samples from vegetative

514   cells (Kahnt *et al.*, 2010). Importantly, the MXAN_3226 β-barrel was already shown to be an

515   essential part of the MASC pathway as its respective *M. xanthus* deletion-mutant strain was found to

516   be deficient in sporulation and MASC production (Holkenbrink *et al.*, 2014). To examine effects of

517   β-barrel deletion on EPS levels in vegetative cells, we first generated a Δ*mxan_7418* chromosomal

518   deletion mutant strain. Since retention of Trypan Blue has become a well-established indicator for

519   the presence of EPS on the surface of *M. xanthus* cells, we next compared the dye-binding capacity

520   of Δ*mxan_7418* cells versus EPS-pathway OPX⁻ (Δ*wzaX*) and T4P⁻ (Δ*pilA*) cells, both known to be

521   defective in EPS production. Relative to WT cells, absence of the EPS-pathway β-barrel resulted in

522   an 83% loss of Trypan Blue retention by Δ*mxan_7418* cells (**Fig. 5C**), indicating a severe reduction

523   in the amount of cell-surface EPS in the β-barrel mutant, consistent with cell-surface EPS

524   deficiencies previously probed in Δ*wzxX,* Δ*wzyX,* Δ*wzcX,* Δ*wzeX,* and Δ*wzaX* EPS-pathway mutants

525   (Islam *et al.*, 2020).

526   Compared to the baseline readings in the Δ*wzaX* EPS-pathway OPX⁻ mutant strain,

527   Δ*mxan_7418* cells displayed marginally higher levels of Trypan Blue binding (**Fig. 5C**). These

528   results are consistent with EPS-pathway β-barrel deficiency principally impacting polymer export to

529   the cell surface, as opposed to both polymer assembly and export being compromised in the absence

530   of EPS-pathway OPX proteins. To probe whether residual quantities of EPS were indeed present on

531   the surface of Δ*mxan_7418* cells, we compared swarm expansion on solid medium as well as auto-

532   aggregation in liquid medium; though both phenotypes are multifactorial, each requires T4P

533   engagement with cell-surface EPS. Relative to WT, Δ*mxan_7418* swarm expansion was reduced,

534   with this phenotype even more pronounced in Δ*wzaX* swarms (**Fig. 5D**). Similarly for auto-

535   aggregation testing in rich medium, consistent with previous findings, WT cells steadily clumped

536   together and sedimented in the cuvette, whereas both Δ*pilA* and Δ*wzaX* cells did not (**Fig. 5E**). Cells

537   of the Δ*mxan_7418* mutant remained in suspension analogous to Δ*wzaX* cells for ~75% of the assay,

538   after which they began to slowly sediment (**Fig. 5E**), suggesting that cell-surface EPS had eventually

539   accumulated to a sufficient threshold to support T4P-mediated clumping in liquid. Taken together,

540   these data indicate that while minimal amounts of EPS can reach the cell surface through as-yet-

541   undetermined means (see Discussion for further comment), MXAN_7418 serves as the principal

542   trans-OM conduit for EPS export in *M. xanthus*.

543   Ultimately, the findings detailed in this investigation support a model for polysaccharide

544   export in Wzx/Wzy-dependent pathways lacking an integral OM OPX protein — as represented by

545   the independent EPS, BPS and MASC pathways in *M. xanthus* — in which the final polymer must

546   pass through an integral OM β-barrel porin for efficient secretion to the cell exterior. For these

547   reasons, as well as the long-established naming convention for Wzx/Wzy-dependent pathways in

548   bacteria (Reeves *et al.*, 1996), we propose the designation Wzp (i.e. W̲z̲ porin) for the newly-

549   identified component of these secretion systems (**Fig. 6**).

550 **Figure 6. Wzx/Wzy-dependent polysaccharide assembly-and-secretion pathways in *M.***

551 ***xanthus.*** In these schematics, the WzaX/S/B proteins have been depicted in a linking capacity

552 between the apex of the WzcX/S/B PCP periplasmic domains and the periplasmic opening of the

553 integral OM WzpX/S/B β-barrel porins identified herein. However, (i) the exposure of the

554 EPS/MASC/BPS polymers to the periplasmic space as each transits between the IM and OM

555 components of each system, and (ii) the exact role(s) of the WzaX/S/B proteins in *M. xanthus*

556 polymer translocation (Supplementary Fig. S6), remain open questions for each pathway. To

557 denote these uncertainties, this stage of the transport cycle has been marked with an asterisk (*).

FIGURE 6
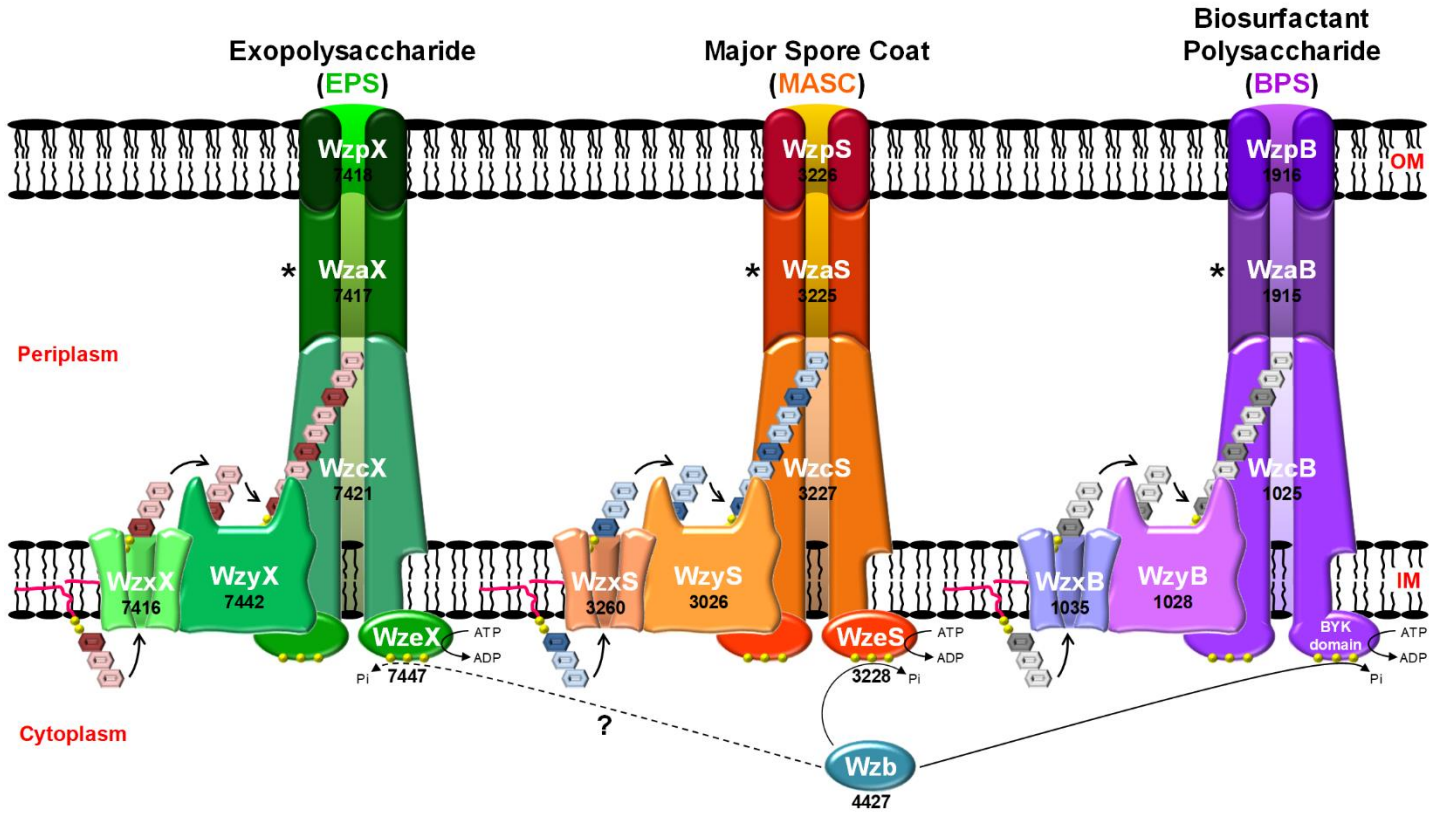
559   **DISCUSSION**

560   Knowledge of the terminal component through which secreted polysaccharides exit a

561   bacterial cell is crucial for the development of targeted antimicrobial agents that could be used to

562   inhibit this process (Kong *et al.*, 2013). From the data presented herein, we have provided evidence

563   that bacterial OPX proteins fall into one of three distinct structural classes, the first two of which

564   have either the demonstrated (Class 1) or predicted (Class 2) capacity to span the OM, whereas the

565   third (Class 3) is missing any such domains. Instead, as demonstrated by the Class 3 OPX WzaX/S/B

566   proteins from *M. xanthus*, these Wzx/Wzy-dependent pathway proteins are genomically and

567   functionally paired with a complementary integral OM-spanning β-barrel porin similar to that

568   required for PNAG export in synthase-dependent pathways.

569   The Class 1 OPX protein Wza$_{Ec}$ is the most extensively-characterized OPX protein with

570   respect to structure–function relationships. The orientation of Wza$_{Ec}$ in the OM was established

571   through introduction of a FLAG tag on the Wza$_{Ec}$ C-terminus, allowing for a FLAG epitope to be

572   detected on the cell surface with α-FLAG antibodies (Dong *et al.*, 2006). Similar introduction of a C-

573   terminal His$_6$ epitope tag resulted in a partially functional Wza$_{Ec}$-His$_6$ construct that was able to

574   restore K30 Group 1 CPS biosynthesis to ~20% of the level restored by an untagged Wza$_{Ec}$ construct

575   following expression *in trans* (Nesper *et al.*, 2003). Various C-terminal α-helix truncations also

576   abolished Wza$_{Ec}$ function (Ford *et al.*, 2009). This indicates that the OM-spanning domain of Wza$_{Ec}$

577   can be functionally-sensitive to structural perturbation. In lipid bilayers, purified Wza$_{Ec}$-His$_6$ was

578   able to form 2D octameric ring-like crystal arrays (Beis *et al.*, 2004), suggestive of channel

579   formation, with this arrangement confirmed by the 3D X-ray crystal structure (Dong *et al.*, 2006).

580   Through introduction of the photo-crosslinkable synthetic amino acid *p*-benzoyl-L-phenylalanine

581   (*p*Bpa) at various sites in Wza$_{Ec}$, Nickerson and colleagues elegantly demonstrated that K30 CPS

582   polymers could be trapped in the lumen of the translocon (Nickerson *et al.*, 2014), confirming that

583   polysaccharides do indeed pass through Wza$_{Ec}$ during secretion. Specifically, substitutions at certain

584   sites within periplasmic D2 (**Fig. 1A**) were able to maintain translocon functionality as well as form

585   crosslinks with polymers. Site-specific *p*Bpa substitutions within periplasmic D1 rendered Wza$_{Ec}$

586   non-functional. Conversely, *p*Bpa-substituted positions in periplasmic D3, and more importantly at

587   the kink in the OM-traversing α-helical D4, maintained Wza$_{Ec}$ functionality, but were unable to form

588   detectable intermolecular crosslinks upon UV exposure (Nickerson *et al.*, 2014). As such, the transit

589   of secreted polymer through the α-helical D4 Wza_C domain pore could not be demonstrated via this

590   technique. However, purified Wza_{Ec} was shown to stably insert into planar lipid bilayers and form

591   electro-conductive channels, with site-specific amino acids substitutions confirming that ion flow

592   does indeed proceed via the D4 pore formed by the OM-spanning α-helix (Kong *et al.*, 2013).

593   Synthetic peptides corresponding to the native Wza_{Ec} D4 α-helix sequence were also able to

594   spontaneously insert into such bilayers, but formed unstable pores; however, modification of the

595   native D4 sequence through consensus generation (based on 94 closely-related sequences) yielded an

596   optimized peptide (**Fig. 2B**) that could spontaneously insert into bilayers and form stable pores

597   (Mahendran *et al.*, 2017). Given the primary structure diversity amongst OM-spanning α-helices

598   uncovered herein for both Class 1 and Class 2B OPX proteins (**Fig. 2B,C**), additional optimization

599   of synthetic peptide sequences should be possible to further improve spontaneous membrane

600   insertion, self-assembly, and α-barrel pore stability.

601   The Class 2B OPX protein $KpsD_{Ec}$ has long been known to be essential for Group 2 CPS

602   export. When expressed in isolation, $KpsD_{Ec}$ was shown to reside in the periplasm (Silver *et al.*,

603   1987). However, upon expression of $KpsD_{Ec}$ along with its cognate IM-localized PCP KpsE (which

604   extends into the periplasm), $KpsD_{Ec}$ was also detected in IM and OM fractions of lysed cells

605   (Arrecubieta *et al.*, 2001). Intriguingly, OM-localized $KpsD_{Ec}$ is detected as a multimer, whereas the

606   minimally-detected IM-localized $KpsD_{Ec}$ is present as a monomer, consistent with the adoption of

607   quaternary structure by $KpsD_{Ec}$ at the site of polysaccharide egress from the cell (Sande *et al.*, 2019).

608   Furthermore, immunolabelling of intact *E. coli* cells using α-$KpsD_{Ec}$ antiserum resulted in the

609   detection of $KpsD_{Ec}$ epitopes at the cell surface (McNulty *et al.*, 2006), suggesting that a portion of

610   $KpsD_{Ec}$ was indeed surface-accessible. The detection herein of structural homology of the extreme

611   $KpsD_{Ec}$ C-terminus to the OM-spanning domain of Wza_{Ec} further supports the contention that a part

612   of $KpsD_{Ec}$ is able to interact with, and span, the OM bilayer, albeit in a conditional manner. The 24-

613   residue length of the $KpsD_{Ec}$ C-terminal α-helix is well within the threshold of 20 amino acids

614   required to span the hydrophobic core of a membrane bilayer for α-helical integral membrane

615   domains (Baeza-Delgado *et al.*, 2013). Analogous to Wza_{Ec}, truncation of the $KpsD_{Ec}$ C-terminus by

616   11 amino acids abrogated protein function (Wunder *et al.*, 1994). A potential OM-spanning α-helix

617   is indeed a conserved property of the Class 2B OPX proteins identified in this study. Despite the

618   detectability of $KpsD_{Ec}$ at the cell surface with α-$KpsD_{Ec}$ antibodies, α-His-tag antibodies were

619   unable to label the surface of cells expressing a $KpsD_{Ec}$ variant encoding a C-terminal His_6 affinity

620   tag (McNulty *et al.*, 2006); in this instance, the highly-cationic nature of the affinity tag may have

621   impeded its translocation across the hydrophobic OM bilayer, thus inhibiting immunodetection.

622   Given that KpsD$_{Ec}$ by itself does not intrinsically associate with the OM (Silver *et al.*, 1987), this

623   may suggest that the protein can become directly inserted into the OM (as opposed to requiring OM

624   insertion via the Bam/Tam or Lol machinery). In this manner, KpsD$_{Ec}$ could indeed function as the

625   terminal piece of the Group 2 capsule secretion machinery.

626   Historically, Group 2 CPS secretion across the OM had been suggested to implicate integral

627   OM β-barrel porins (Whitfield & Valvano, 1993, Bliss & Silver, 1996), however such a model has

628   fallen out of favour given that no such β-barrels have ever been detected in or near related Kps

629   synthesis clusters. This would be consistent with an ability of KpsD$_{Ec}$-like Class 2B OPX proteins to

630   traverse the OM, thus not requiring any downstream piece of export machinery in certain organisms.

631   However, this scenario may not be absolute, as numerous instances of β-barrel proteins encoded near

632   both Class 2A and Class 2B (as well as Class 1 and Class 3) OPX genes were uncovered herein,

633   suggesting that integral OM porins may play an important role in non-synthase-dependent secretion

634   in diverse organisms. Though not all Class 2 OPX proteins identified in our investigation were

635   matched with a nearby β-barrel, our synteny analysis window was limited to $+/-$ 10 genes from each

636   OPX gene and would thus not have captured candidate porins elsewhere in the genome. As a case-

637   in-point, the *M. xanthus* EPS, MASC, and BPS clusters contain respective insertions of >18 kbp,

638   >223 kbp, and > 1 Mbp that separate constituent members of the same cluster, which in the latter

639   results in extraordinary genomic distance between *wzaB-wzpB* and the remainder of the BPS

640   assembly genes (Islam *et al.*, 2020). Moreover, as our synteny analyses were limited to 9 query

641   sequences (8 with similarity to PgaA$_{\beta b}$), this does not preclude the presence of other/more distantly-

642   related β-barrels near "unmatched" OPX proteins. However, in the absence of specific templates

643   with which to search, such an analysis was beyond the scope of the current investigation.

644   The identification of "OPX" proteins (an obvious misnomer) in Gram-positive bacteria, but

645   specifically those of Class 3 architecture, point to a broadly-conserved periplasmic function for these

646   proteins, likely through interfacing with the periplasmic domains of PCP proteins in most organisms.

647   However, as dedicated peptidoglycan-spanning polysaccharide export channels have yet to be

648   identified in Gram-positive bacteria, any role for Class 3 OPX proteins in these systems with regards

649   to interactions with a secretion pore of some sort would be unfounded speculation. In *M. xanthus*

650   cells, WzaX/S/B Class 3 OPX protein deficiency does not lead to visible accumulations of polymeric

651   material in the periplasm (Saïdi *et al.*, 2021) suggesting that EPS/MASC/BPS polymer assembly via

652   Wzx/Wzy-dependent pathways does not indiscriminately continue in these mutant backgrounds; this

653   is a similar observation to that for Wza$_{Ec}$ Class 1 OPX deficiency in *E. coli* cells (Nesper *et al.*,

654   2003). Such material from ABC transporter-dependent synthesis does however accumulate in the

655   periplasm of KpsD$_{Ec}$-deficient Class 2B OPX-mutant cells (Wunder *et al.*, 1994, Bliss & Silver,

656   1996).

657   For myxobacterial Class 3 OPX proteins, a clear partnership has now been demonstrated

658   between these secretion-pathway components and integral OM β-barrel porins illustrating the

659   requirement of the latter for HMW polysaccharide export across the OM. The WzpS (MXAN_3226)

660   β-barrel was already shown to be essential for MASC secretion, but its function in the MASC

661   transport cycle was not known at the time (Holkenbrink *et al.*, 2014). In the current investigation, we

662   have shown that the properties of WzpB (MXAN_1916) make it suitable for BPS translocation

663   across the OM. Furthermore, we have herein demonstrated that WzpX (MXAN_7418) functions as

664   the principal export conduit for EPS across the OM to the cell surface. Intriguingly, EPS still appears

665   to be assembled to a certain degree in mutant cells lacking WzpX (given the residual amount detected

666   on the cell surface), as opposed to cells lacking WzaX which manifest an even more robust EPS$^-$

667   phenotype. This may indicate that while lack of the OPX component in the EPS pathway results in a

668   severe reduction (or shutdown) of EPS production, in cells lacking WzpX, other β-barrel porins (e.g.

669   WzpS and/or WzpB) may be able to inefficiently cross-complement the deficiency in the EPS-

670   pathway machinery, resulting in residual EPS localization to the cell surface. Part of this inefficiency

671   could arise from the cationic natures of the WzpS and WzpB barrel interiors not serving as suitable

672   conduits for a more neutral EPS polymer.

673   The presence of a Wza_C domain in a Class 1 or 2B OPX protein is not mutually exclusive to

674   the presence of an integral OM β-barrel protein encoded nearby, considering that such barrels were

675   found to be encoded near genes representing all three OPX protein classes across a range of bacterial

676   genomes. The GfcD protein is particularly intriguing given the GfcE protein also encoded by the

677   Group 4 CPS export locus (Peleg *et al.*, 2005). While GfcD contains a PgaA$_{\beta b}$-like polysaccharide

678   secretion component with an attached SLBB-like domain (as part of the GfcD C-terminal module)

679   (**Supplementary Fig. 5B**), GfcE is a functional Class 1 OPX paralogue of Wza$_{Ec}$; this was

680   evidenced by the ability of GfcE (formerly YccZ/Wza$_{22min}$) expressed *in trans* to partially restore *E.*

681   *coli* K30 CPS production in a mutant lacking Wza$_{Ec}$ (Drummelsmith & Whitfield, 2000). GfcE

682   possesses a complete Wza_C domain, and as such would be expected to span the OM in a Wza$_{Ec}$-like

683   manner. One possibility could be that the Wza_C domain of the GfcE OPX protein may be required

684 to properly interact/organize around the GfcD$_{\text{Cter}\beta\text{b}}$ polysaccharide secretion module. Furthermore,

685 the presence of a putative FapF amyloid secretion β-barrel fused to the same polypeptide as that of a

686 PgaA$_{\beta\text{b}}$-like module typically associated with polysaccharide secretion raises an interesting

687 possibility. Amyloid proteins are frequently secreted by bacteria in order to stabilize biofilm

688 matrices composed largely of secreted polysaccharides (Erskine *et al.*, 2018). Thus, in Group 4

689 capsules, secretion of amyloidogenic polypeptides (via the GfcD FapF-like N-terminal module) could

690 help to stabilize the polysaccharide component of the capsule structure and/or anchor the CPS to the

691 cell surface.

692

693 ***Ideas and Speculation***

694 For Class 3 OPX proteins, the lack of OM-spanning domains and the (relatively) small size of

695 these proteins (compared to other OPX classes), present a dilemma regarding the mechanism of

696 transit across the periplasm for polymers produced by these systems (**Supplementary Fig. 6**). As

697 a case-in-point, the thickness of the *M. xanthus* periplasm was measured to be 327 Å (**Fig. 4A,B**),

698 while the periplasmic domain of the BPS-pathway WzcB PCP-2B protein was estimated to extend up

699 to ~165 Å into the periplasmic space. Coupled with the maximum possible height of ~62 Å for one

700 unit of OPX protein WzaB, together this only accounts for ~227 Å of periplasmic thickness, leaving

701 ~100 Å of periplasmic height unaccounted for (**Supplementary Fig. 6**), compared to standard

702 models of polysaccharide secretion (**Fig. 1A**). Furthermore, high-confidence co-evolving amino

703 acids between WzcB and WzaB localize to the apex of the PCP and the base of the Poly_export

704 domain of the Class 3 OPX protein (respectively) (**Supplementary Fig. 6, Supplementary**

705 **Table S5**), heavily implying the presence of a conserved interaction interface between the two

706 proteins. With the assumption of pathway specificity for each *M. xanthus* Class 3 OPX protein — as

707 evidenced by lack of cross-complementation in single-OPX-knockout strains (Islam *et al.*, 2020) —

708 and using components of the BPS pathway as examples (**Fig. 6**), several potential models for trans-

709 envelope transit can thus be proposed:

710 (i) Model 1: As the polymer exits the WzcB cavity in the periplasm (following WzyB-

711 mediated polymerization), it passes through a periplasmic WzcB-associated single-layer WzaB

712 oligomer. Once past this point, the polymer would have to independently locate its cognate integral

713 OM WzpB β-barrel porin in order to reach the outside of the cell. This presumes the polymer is

714   exposed to the periplasm for a substantial portion of its trip between the IM and OM. In such a

715   model, polymer export might still be possible in the absence of the cognate OPX (as long as this

716   absence does not impact polymer assembly). However, since this is not the case in *M. xanthus*, it

717   would argue against this model.

718   (ii) Model 2: This is a variation of Model 1 in which single-layer periplasmic oligomers of

719   WzaB are located both on the apical point of the PCP octamer as well as the proximal face of the

720   integral OM WzpB β-barrel porin. In this manner, *M. xanthus* OPX-mediated substrate specificity

721   would exist at both ends of the transport process, but again the polymer could be exposed to the

722   periplasm during stages of the transport cycle bridging the OPX proteins. Nonetheless, this model

723   provides a solution to the question of "targeting" of the nascent polymer to the proper OM-spanning

724   apparatus. However, this model also presumes a constant presence of Class 3 OPX proteins

725   associated with the OM, as well as specific interactions between the periplasmic OPX WzaB and the

726   integral OM porin WzpB. At the moment, this is not bolstered by high-confidence evolutionary

727   couplings data between the two proteins (**Supplementary Table S5**), but this may be partially due

728   to an insufficient number of barrel homologues with which to fully probe coevolution. However, as

729   OPX proteins are not detected in surface-biotinylated or OMV samples from *M. xanthus*, supporting

730   evidence for this concept is lacking.

731   (iii) Model 3: The periplasmic distance of 327 Å could be approximately accounted for by

732   ~165 Å of WzcB PCP-2B periplasmic domain height, followed by OPX oligomers of WzaB stacked

733   in duplicate (~62 Å × 2 = ~124 Å) or triplicate (~62 Å × 3 = ~186 Å), depending on the packing

734   arrangement of the oligomers. Along with the PCP channel, such architecture could be envisaged to

735   form a protected channel lumen spanning from the IM to the OM, precluding exposure of the

736   polymer to the periplasm. This would also abrogate any "targeting" issues of the polymer to the

737   WzpB β-barrel secretin. In this case, HMW oligomers of *M. xanthus* OPX proteins would be

738   expected in the periplasm, which could possibly co-precipitate with IM and/or OM fractions. Once

739   again, data to support this contention awaits further experimentation.

740   (iv) Model 4: As the nascent polymer emerges from the PCP-2B opening, the polymer

741   interacts with copies of WzaB at the apex of the PCP-2B periplasmic domain, allowing WzaB to bind

742   the polymer and detach from the PCP. As additional polymer elongation occurs, more units of WzaB

743   are able to bind further down the polymer. In this manner, WzaB bound to the polymer would serve

744   as a type of targeting chaperone which preferentially directs the periplasmic polymer to its cognate

745   WzpB β-barrel porin in the OM. Once a given WzaB has reached the porin, it would disengage from

746   the polymer and be able to undergo subsequent rounds of binding in the periplasm to the nascent

747   polymer. Such a mechanism would afford a certain degree of protection to the translocating polymer

748   against the periplasmic environment, as well as provide a mechanism for targeting of the polymer to

749   the specific β-barrel machinery needed for transport across the OM. This would also explain why the

750   Class 3 OPX proteins WzaX, WzaS, and WzaB were not detected in screens of surface-biotinylated

751   proteins or those identified in OMV fractions (from vegetative and developmental cells), i.e. that OM

752   association of WzaX/S/B (via interaction with WzpX/WzpS/WzpB) is of a more transient nature.

753   Exposure of a sugar polymer to the periplasm as it transits between the IM and OM is a

754   common occurrence in synthase-dependent systems such as those involved in alginate, cellulose, and

755   PNAG assembly-and-export (Whitney & Howell, 2013). Recently, Group 2 CPS was also suggested

756   to be exposed to the periplasm at some point during its secretion across the cell envelope (Liston *et*

757   *al.*, 2018). In these systems, periplasmic enzymes are able to access the transiting polymer and

758   introduce various modifications including sugar epimerization and/or (de)acetylation. Such a

759   mechanism may be in place for the *M. xanthus* BPS polymer, for which the secreted form displays

760   random acetylation (Islam *et al.*, 2020). In synthase-dependent pathways, TPR domains (either

761   standalone or attached) extend into the periplasm from the β-barrel porin and are proposed to interact

762   with these chain-modifying enzymes (Whitney & Howell, 2013). Of note, numerous OPX proteins

763   were found to be encoded near β-barrel genes for full-length PgaA and BcsC homologues complete

764   with the respective TPR architectures (**Table 1**). In these systems, questions arise as to the functional

765   relationships between periplasmic OPX proteins and integral OM β-barrel porins with TPR domains

766   potentially occluding access of the OPX protein to the periplasmic face of the porin.

767   Ultimately, this study reveals OPX-protein complexities in diverse organisms that differ from

768   the well-studied Wza_{Ec} protein, which will further our understanding of the mechanism of sugar

769   polymer export in bacterial cells. Moreover, updated genomic, structural, and functional knowledge

770   of the terminal step in the polysaccharide secretion pathway will enable researchers to selectively

771   develop novel antimicrobial compounds targeted to blocking bacterial polymer secretion from the

772   outside, thus bypassing any requirements for access to the cell interior to compromise the viability of

773   a bacterial cell.

774 **MATERIALS AND METHODS**

775 *Protein structure analysis & modelling*

776   Given the high fold-recognition equivalence between WzaX/S/B and Wza$_{Ec}$, the tertiary

777 structure of each *M. xanthus* OPX protein was modelled against the Wza$_{Ec}$ template (PDB: 2J58)

778 using MODELLER. For β-barrel proteins MXAN_7418/3226/1918 that lacked a suitable full-length

779 structural template, protein structure models were computed using deep learning and artificial

780 intelligence via AlphaFold2 (Jumper *et al.*, 2021). Multiple sequence alignment entries were

781 generated using 606, 251 and 101 unique sequences, respectively for MXAN_1916, MXAN_3226

782 and MXAN_7418, with the program run for 5 independent prediction models, leading to convergence

783 after 3 recycling iterations. For proteins GfcD and YjbH, deposited AlphaFold2-generated structures

784 were mined from the UniProt entries for P75882 and P32689, respectively. HOLLOW (Ho &

785 Gruswitz, 2008) was used to generate internal volume casts of the MXAN_7418/3226/1918 β-

786 barrels, followed by overlaying of the solvent-accessible electrostatic potential contributed by amino

787 acids in contact with the internal volume, as calculated using PDB2PQR and APBS (Propka pH 7.0,

788 Swanson force field, ±5 kT/e). All protein structures were visualized and rendered in PyMol.

789 Evolutionarily-coupled amino acids within the same protein were analyzed using RaptorX-Contact

790 (Wang *et al.*, 2017), while evolutionarily-coupled residues between two proteins were determined

791 using RaptorX-ComplexContact (Zeng *et al.*, 2018) (http://raptorx.uchicago.edu/). Protein contact

792 maps were displayed in GraphPad.

793

794 *Identification of OPX proteins*

795   Three types of datasets i.e., 61 order Myxococcales genomes (MYXO), 3662 reference and

796 representative bacterial genomes (REP; downloaded on Dec 7, 2021), and non-redundant NCBI

797 database (NR) (371 327 556 proteins at 100% identify as of June 10, 2021) were downloaded from

798 NCBI. Pfam domains attributed to PF02563 [Poly_export], PF10531 [SLBB], PF18412 [Wza_C],

799 and PF06251 [Caps_synth_GfcC; used here as GfcC] were extracted from the Pfam-A v34.0

800 database (Mistry *et al.*, 2021) (downloaded: March 24, 2021) and a reduced combined profile

801 database was created. These functional domains were identified by scanning all three types of

802 datasets (MYXO, REP, and NR) using offline hmmscan (Potter *et al.*, 2018) against the created

803 database with an E-value cutoff of 1×10^{-5}. The resultant files were parsed using hmmscan-parser.sh,

804 sorted and arranged in the form of protein architecture using in-house scripts. Based on the identified

805 domains per protein per dataset, three primary clusters were curated. All identified proteins within

806   these primary clusters were subjected to fold-recognition analysis using HHpred (Söding *et al.*, 2005)

807   against the database of two proteins (PDB: 2J58 [i.e. Wza$_{Ec}$] and PDB: 3P42 [i.e. GfcC]; extracted

808   from PDBmmCIF70 downloaded on Nov 19, 2021) using "-p 5 -Z 500 -loc -z 1 -b 1 -B 500 -all -id

809   35 -ssm 2 -sc 1 -seq 1 -dbstrlen 10000 -norealign -maxres 32000" parameters. HHpred raw data was

810   parsed using in-house scripts to generate the architecture of each protein in terms of non-overlapping

811   homologous regions to 2J58 and 3P42. To identify the type of signal peptide and the cleavage site

812   location, each OPX protein was subject to SignalP 6.0 analysis (Teufel *et al.*, 2022) and to predict the

813   membrane topology of the proteins, TMHMM (Server v. 2.0) (Krogh *et al.*, 2001) was used.

814   Based on HHpred analysis, we also investigated the secondary structure-based homology of

815   primary cluster proteins with a Wza_C segment (aa 326 – 359; 34 amino acid length) in PDB 2J58. If

816   a protein was showing secondary structural homology with at least 10/34 amino acids of the Wza_C

817   segment (aa 326 – 359) in 2J58, we considered it as a true Wza_C segment in the respective protein.

818   Proteins with both 2J58- and 3P42-homologous non-overlapping regions were classified as Class 2.

819   Proteins having only 2J58-homologous regions along with a Wza_C segment were classified as Class

820   1. Other proteins with only 2J58-homologous regions and no Wza_C segment were classified as

821   Class 3. We also generated sequence logos using WebLogo (Crooks *et al.*, 2004) (v.2.8) for 2º

822   structure-based homologues of Wza_C segments as identified in Class 1 and Class 2B OPX proteins.

823

824   *Synteny analysis of β-barrel query proteins with identified OPX proteins*

825   Along with the identification of the above-mentioned domains and classification of all OPX

826   proteins into three classes, we identified the β-barrel homologues encoded in the vicinity (± 10 genes)

827   of our OPX genes in genomes from both the MYXO and REP databases. We analyzed the various

828   predicted proteomes of both datasets with BLASTp and hmmscan using protein sequences from *M.*

829   *xanthus* DZ2 (MXAN_7418 [aa 24 – 415], MXAN_3226 [aa 23 – 381], and MXAN_1916 [aa 26 –

830   421]), *E. coli* K12 (PgaA [aa 511 – 807], GfcD [aa 425 – 698], YjbH [aa 423 – 698], BcsC [aa 785 –

831   1157], and Wzi [aa 92 – 479], as well as *P. aeruginosa* PAO1 AlgE [aa 33 – 490]. Detected

832   homologues for each query profile were aligned against the truncated and full-length sequences for

833   each query protein using Clustal Omega to probe for the lone presence of a putative β-barrel

834   polysaccharide secretion porin domain versus its presence as part of a multi-domain polypeptide

835   comparable to the native sequence. Hits resembling truncated versions of various β-barrel queries

836   were individually profiled via fold-recognition using the online HHpred bioinformatics toolkit

837   (Zimmermann *et al.*, 2018) (https://toolkit.tuebingen.mpg.de/tools/hhpred). Depicted sequence

838   alignments were displayed in GeneDoc with residues colored according to conservation score (out of

839   10) as indicated in JalView (Waterhouse *et al.*, 2009).

840

Phylogenetic analysis of OPX proteins

842   The PF02563 [Poly_export] domain of OPX proteins was used as a phylogenetic marker. The

843   locations of all identified 'Poly_export' domains were first extracted, after which extracted sequences

844   were aligned using MUSCLE (with 10 iterations). The resultant alignment was analyzed via

845   FastTree 2.1.10 to generate a maximum likelihood tree of OPX proteins. Tree visualization as well as

846   mapping of the OPX protein classification and taxonomy of each branch were performed via the

847   iTOL web server Version 6.4.3 (Letunic & Bork, 2021).

848

Bacterial membrane modeling and intermembrane distance measurements

850   Desired prokaryotic tomograms were downloaded from the Caltech Electron Tomography

851   Database (https://etdb.caltech.edu/). Forty species were analyzed, each via three separate tomograms.

852   Tomogram inspection and modeling were performed using the IMOD software package (Kremer *et*

853   *al.*, 1996). Tomograms were first oriented in 3D using the IMOD "Slicer" window to identify the

854   central slice through each bacterium. To enhance contrast, 5 layers of voxels were averaged around

855   the section of interest. Model points were then placed along corresponding regions of the OM and IM

856   for a total distance of ~100 nm. A custom Python script was subsequently used to calculate the

857   intermembrane distance every 0.1 nm along the modeled stretch of membranes. GraphPad was used

858   to prepare plots and carry out correlation analyses.

859

Bacterial cell culture

861   Information on wild-type *M. xanthus* DZ2 (Campos & Zusman, 1975) and isogenic mutant

862   strains analyzed herein can be found in Table 2. Strains were grown and maintained at 32 °C on

863   Casitone-yeast extract (CYE) (1% casitone, 0.5% yeast extract, 10 mM MOPS [pH 7.5], 4 mM

864    $MgSO_4$) 1.5% agar (BD Difco) plates or in CYE liquid medium at 32 °C on a rotary shaker at 220

865   rpm.

866

Mutant construction

868   As previously described (Islam *et al.*, 2020), to generate a *M. xanthus* deletion-mutant strain,

869   500 bp upstream and 500 bp downstream of the target gene were amplified and fused via PCR,

870 subjected to double digestion with restriction enzymes, then ligated into the pBJ114 plasmid.

871 Products were used to transform chemically-competent *E. coli* DH10B via heatshock, after which

872 cells were plated on LB agar supplemented with kanamycin (50 µg/mL) to select for drug-resistant

873 colonies. Successful clones were verified via sequencing. Resultant plasmids were then introduced

874 into *M. xanthus* DZ2 via electroporation. Mutants resulting from homologous recombination of

875 deletion alleles were obtained by selection on CYE agar plates containing kanamycin (100 µg/mL).

876 A second selection was then made on CYE agar plates containing galactose (2.5%) to obtain the final

877 deletion strain, with mutants verified via PCR amplification using flanking primers.

878

879 ***Trypan blue dye retention***

880 Retention of the dye Trypan Blue was carried out as previously described (Islam *et al.*, 2020).

881 In brief, cells from overnight CYE cultures were first resuspended to OD_{600} 1.0 in TPM buffer.

882 Resuspended cells (or a cell-free blank) (900 µL) were then mixed with Trypan Blue solution (100

883 µL, 100 µg/mL stock concentration) in a microfuge tube and briefly pulsed (1 s) via vortex mixer.

884 Samples were incubated at room temperature, in a tube rack covered with aluminum foil, atop a

885 rocker platform (1 h) to facilitate dye binding by the cells. Samples were subsequently sedimented

886 ($16\,000 \times g$, 5 min), after which the top 900 µL of blank or clarified supernatant was transferred to a

887 disposable spectrophotometer cuvette. The cell-free "TPM + Trypan Blue" sample was used to blank

888 the spectrophotometer at 585 nm. The absorbance at 585 nm (A_{585}) was then determined for each

889 clarified supernatant. Finally, absorbance values were normalized to the respective A_{585} value for the

890 WT of each biological replicate. Sub-zero final values are due to trace amounts of cell debris

891 detected at 585 nm in individual samples in which absolutely no binding of Trypan Blue occurred.

892

893 ***Phenotypic analyses***

894 Cells from exponentially-growing cultures were harvested and resuspended in TPM buffer (10

895 mM Tris-HCl, pH 7.6, 8 mM $MgSO_4$ and 1 mM KH_2PO_4) at a final concentration of OD_{600} 5.0. To

896 study T4P-dependent swarm expansion, this cell suspension (5 µL) was spotted onto CYE 0.5% agar.

897 Plates were incubated at 32 °C (72 h), then imaged with an Olympus SZX16 stereoscope with UC90

898 4K camera. For T4P-dependent motility, swarm images were captured using the 0.5× objective at 1×

899 zoom, using linear color and darkfield illumination.

900

901

902   *Auto-aggregation testing*

903   The protocol followed has been previously detailed (Saïdi *et al.*, 2021). In brief, the turbidity

904   (OD_{600}) of *M. xanthus* CYE cultures (12.5 mL) grown overnight was determined via

905   spectrophotometer. Specific culture volumes were aspirated, then sedimented via microfuge (4000 $\times$

906   *g*, 5 min) so pellet resuspension in 1 mL CYE broth would yield a final OD_{600} of either 1.0. Broth-

907   resuspended cells were then transferred to a polystyrene spectrophotometer cuvette. Resuspensions

908   were strongly aspirated and ejected in the cuvette (10 s) via p200 micropipette, then immediate read

909   for OD_{600} ($t = 0$) in a spectrophotometer. Time-course readings of OD_{600} were taken every 10 min up

910   to 100 min of monitoring. In between readings, cuvettes were left covered and undisturbed on the

911   benchtop in a cuvette box. All OD_{600} readings were normalized to the OD_{600} determined at $t = 0$ for

912   each sample.

ACKNOWLEDGEMENTS

927 **COMPETING INTERESTS**

928   The authors declare no financial or non-financial competing interests.

AUTHOR CONTRIBUTIONS

929

930   STI and GS conceived of and planned the study.

931   UM, AP, and GS performed comparative genomics studies.

932   FS and RB generated mutant strains.

933   FS performed phenotypic, dye-binding, and auto-aggregation analyses.

934   FS and NYJ performed HOLLOW and electrostatics analyses.

935   STI and AM carried out protein modelling.

936   MM and GJ designed the periplasmic analysis workflow, with analysis by FS.

937   STI, GS, and FS wrote the manuscript.

938   STI and GS generated figures.

939   STI, GS, CC, and YWC contributed personnel and funding support.

REFERENCES

Acheson, J.F., Derewenda, Z.S., and Zimmer, J. (2019) Architecture of the cellulose synthase outer membrane channel and its association with the periplasmic TPR domain. *Structure* **27**: 1855-1861.e1853.

Arrecubieta, C., Hammarton, T.C., Barrett, B., Chareonsudjai, S., Hodson, N., Rainey, D., and Roberts, I.S. (2001) The transport of Group 2 capsular polysaccharides across the periplasmic space in *Escherichia coli*: roles for the KpsE and KpsD proteins*. *J. Biol. Chem.* **276**: 4245-4250.

Baeza-Delgado, C., Marti-Renom, M.A., and Mingarro, I. (2013) Structure-based statistical analysis of transmembrane helices. *Eur. Biophys. J.* **42**: 199-207.

Becker, A. (2015) Challenges and perspectives in combinatorial assembly of novel exopolysaccharide biosynthesis pathways. *Front. Microbiol.* **6**.

Behmlander, R.M., and Dworkin, M. (1994) Biochemical and structural analyses of the extracellular matrix fibrils of *Myxococcus xanthus*. *J. Bacteriol.* **176**: 6295-6303.

Beis, K., Collins, R.F., Ford, R.C., Kamis, A.B., Whitfield, C., and Naismith, J.H. (2004) Three-dimensional structure of Wza, the protein required for translocation of Group 1 capsular polysaccharide across the outer membrane of *Escherichia coli**. *J. Biol. Chem.* **279**: 28227-28232.

Bliss, J.M., and Silver, R.P. (1996) Coating the surface: a model for expression of capsular polysialic acid in *Escherichia coli* K1. *Mol. Microbiol.* **21**: 221-231.

Burroughs, A.M., Balaji, S., Iyer, L.M., and Aravind, L. (2007) A novel superfamily containing the β-grasp fold involved in binding diverse soluble ligands. *Biol. Direct* **2**: 4.

Bushell, Simon R., Mainprize, Iain L., Wear, Martin A., Lou, H., Whitfield, C., and Naismith, James H. (2013) Wzi Is an outer membrane lectin that underpins group 1 capsule assembly in *Escherichia coli*. *Structure* **21**: 844-853.

Campos, J.M., and Zusman, D.R. (1975) Regulation of development in *Myxococcus xanthus*: effect of 3':5'-cyclic AMP, ADP, and nutrition. *Proc. Natl. Acad. Sci. U. S. A.* **72**: 518-522.

Collins, R.F., Beis, K., Dong, C., Botting, C.H., McDonnell, C., Ford, R.C., Clarke, B.R., Whitfield, C., and Naismith, J.H. (2007) The 3D structure of a periplasm-spanning platform required for assembly of group 1 capsular polysaccharides in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **104**: 2390-2395.

Crooks, G.E., Hon, G., Chandonia, J.-M., and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.* **14**: 1188-1190.

Cuthbertson, L., Mainprize, I.L., Naismith, J.H., and Whitfield, C. (2009) Pivotal roles of the outer membrane polysaccharide export and polysaccharide copolymerase protein families in export of extracellular polysaccharides in Gram-negative bacteria. *Microbiol. Mol. Biol. Rev.* **73**: 155-177.

Dong, C., Beis, K., Nesper, J., Brunkan-LaMontagne, A.L., Clarke, B.R., Whitfield, C., and Naismith, J.H. (2006) Wza the translocon for *E. coli* capsular polysaccharides defines a new class of membrane protein. *Nature* **444**: 226.

980 Drummelsmith, J., and Whitfield, C. (2000) Translocation of group 1 capsular polysaccharide to the
981   surface of *Escherichia coli* requires a multimeric complex in the outer membrane. *EMBO J.*
982   **19**: 57-66.

983 Ducret, A., Valignat, M.-P., Mouhamar, F., Mignot, T., and Theodoly, O. (2012) Wet-surface-
984   enhanced ellipsometric contrast microscopy identifies slime as a major adhesion factor during
985   bacterial surface motility. *Proc. Natl. Acad. Sci. USA* **109**: 10036-10041.

986 Erskine, E., MacPhee, C.E., and Stanley-Wall, N.R. (2018) Functional amyloid and other protein
987   fibers in the biofilm matrix. *J. Mol. Biol.* **430**: 3642-3656.

988 Faure, L.M., Fiche, J.-B., Espinosa, L., Ducret, A., Anantharaman, V., Luciano, J., Lhospice, S.,
989   Islam, S.T., Tréguier, J., Sotes, M., Kuru, E., Van Nieuwenhze, M.S., Brun, Y., Théodoly, O.,
990   L, A., Nollmann, M., and Mignot, T. (2016) The mechanism of force transmission at bacterial
991   focal adhesion complexes. *Nature* **539**: 530-535.

992 Ferrières, L., Aslam, S.N., Cooper, R.M., and Clarke, D.J. (2007) The *yjbEFGH* locus in *Escherichia*
993   *coli* K-12 is an operon encoding proteins involved in exopolysaccharide production.
994   *Microbiology* **153**: 1070-1080.

995 Ford, R.C., Brunkan-LaMontagne, A.L., Collins, R.F., Clarke, B.R., Harris, R., Naismith, J.H., and
996   Whitfield, C. (2009) Structure-function relationships of the outer membrane translocon Wza
997   investigated by cryo-electron microscopy and mutagenesis. *J. Struct. Biol.* **166**: 172-182.

998 Franklin, M.J., Nivens, D.E., Weadge, J.T., and Howell, P.L. (2011) Biosynthesis of the
999   *Pseudomonas aeruginosa* extracellular polysaccharides, alginate, Pel, and Psl. *Front.*
1000   *Microbiol.* **2**: 167.

1001 Gibiansky, M.L., Hu, W., Dahmen, K.A., Shi, W., and Wong, G.C.L. (2013) Earthquake-like
1002   dynamics in *Myxococcus xanthus* social motility. *Proc. Natl. Acad. Sci. USA* **110**: 2330-2335.

1003 Ho, B.K., and Gruswitz, F. (2008) HOLLOW: generating accurate representations of channel and
1004   interior surfaces in molecular structures. *BMC Struct. Biol.* **8**: 49.

1005 Holkenbrink, C., Hoiczyk, E., Kahnt, J., and Higgs, P.I. (2014) Synthesis and assembly of a novel
1006   glycan layer in *Myxococcus xanthus* spores. *J. Biol. Chem.* **289**: 32364-32378.

1007 Hu, W., Lux, R., and Shi, W., (2013) Analysis of exopolysaccharides in *Myxococcus xanthus* using
1008   confocal laser scanning microscopy. In: Bacterial Cell Surfaces: Methods and Protocols. A.H.
1009   Delcour (ed). Totowa, NJ: Humana Press, pp. 121-131.

1010 Islam, S.T., Eckford, P.D.W., Jones, M.L., Nugent, T., Bear, C.E., Vogel, C., and Lam, J.S. (2013a)
1011   Proton-dependent gating and proton uptake by Wzx support O-antigen-subunit antiport across
1012   the bacterial inner membrane. *mBio* **4**: e00678-00613.

1013 Islam, S.T., Fieldhouse, R.J., Anderson, E.M., Taylor, V.L., Keates, R.A.B., Ford, R.C., and Lam,
1014   J.S. (2012) A cationic lumen in the Wzx flippase mediates anionic O-antigen subunit
1015   translocation in *Pseudomonas aeruginosa* PAO1. *Mol. Microbiol.* **84**: 1165-1176.

1016 Islam, S.T., Gold, A.C., Taylor, V.L., Anderson, E.M., Ford, R.C., and Lam, J.S. (2011) Dual
1017   conserved periplasmic loops possess essential charge characteristics that support a catch-and-
1018   release mechanism of O-antigen polymerization by Wzy in *Pseudomonas aeruginosa* PAO1.
1019   *J. Biol. Chem.* **286**: 20600-20605.

1020   Islam, S.T., Huszczynski, S.M., Nugent, T., Gold, A.C., and Lam, J.S. (2013b) Conserved-residue
1021   mutations in Wzy affect O-antigen polymerization and Wzz-mediated chain-length regulation
1022   in *Pseudomonas aeruginosa* PAO1. *Sci. Rep.* **3**: 3441.

1023   Islam, S.T., and Lam, J.S. (2014) Synthesis of bacterial polysaccharides via the Wzx/Wzy-dependent
1024   pathway. *Can. J. Microbiol.* **60**: 697-716.

1025   Islam, S.T., and Mignot, T. (2015) The mysterious nature of bacterial surface (gliding) motility: a
1026   focal adhesion-based mechanism in *Myxococcus xanthus*. *Semin. Cell Dev. Biol.* **46**: 143-154.

1027   Islam, S.T., Taylor, V.L., Qi, M., and Lam, J.S. (2010) Membrane topology mapping of the O-
1028   antigen flippase (Wzx), polymerase (Wzy), and ligase (WaaL) from *Pseudomonas aeruginosa*
1029   PAO1 reveals novel domain architectures. *mBio* **1**: e00189-00110.

1030   Islam, S.T., Vergara Alvarez, I., Saïdi, F., Guiseppi, A., Vinogradov, E., Sharma, G., Espinosa, L.,
1031   Morrone, C., Brasseur, G., Guillemot, J.-F., Benarouche, A., Bridot, J.-L., Ravicoularamin,
1032   G., Cagna, A., Gauthier, C., Singer, M., Fierobe, H.-P., Mignot, T., and Mauriello, E.M.F.
1033   (2020) Modulation of bacterial multicellularity via spatio-specific polysaccharide secretion.
1034   *PLOS Biol.* **18**: e3000728.

1035   Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K.,
1036   Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J.,
1037   Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S.,
1038   Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T.,
1039   Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., and
1040   Hassabis, D. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**:
1041   583-589.

1042   Kahnt, J.r., Aguiluz, K., Koch, J.r., Treuner-Lange, A., Konovalova, A., Huntley, S., Hoppert, M.,
1043   Søgaard-Andersen, L., and Hedderich, R. (2010) Profiling the outer membrane proteome
1044   during growth and development of the social bacterium *Myxococcus xanthus* by selective
1045   biotinylation and analyses of outer membrane vesicles. *J. Proteome Res.* **9**: 5197-5208.

1046   Kong, L., Harrington, L., Li, Q., Cheley, S., Davis, B.G., and Bayley, H. (2013) Single-molecule
1047   interrogation of a bacterial sugar transporter allows the discovery of an extracellular inhibitor.
1048   *Nat. Chem.* **5**: 651-659.

1049   Kremer, J.R., Mastronarde, D.N., and McIntosh, J.R. (1996) Computer visualization of three-
1050   dimensional image data using IMOD. *J. Struct. Biol.* **116**: 71-76.

1051   Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L.L. (2001) Predicting transmembrane
1052   protein topology with a hidden markov model: application to complete genomes. *J. Mol. Biol.*
1053   **305**: 567-580.

1054   Lavelle, K., Sinderen, D.v., and Mahony, J. (2021) Cell wall polysaccharides of Gram positive
1055   ovococcoid bacteria and their role as bacteriophage receptors. *Comput. Struct. Biotechnol. J.*
1056   **19**: 4018-4031.

1057   Letunic, I., and Bork, P. (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic
1058   tree display and annotation. *Nucleic Acids Res.* **49**: W293-W296.

1059   Liston, S.D., McMahon, S.A., Le Bas, A., Suits, M.D.L., Naismith, J.H., and Whitfield, C. (2018)
1060   Periplasmic depolymerase provides insight into ABC transporter-dependent secretion of
1061   bacterial capsular polysaccharides. *Proc. Natl. Acad. Sci. U. S. A.* **115**: E4870-E4879.

1062 Mahendran, K.R., Niitsu, A., Kong, L., Thomson, A.R., Sessions, R.B., Woolfson, D.N., and Bayley,
1063   H. (2017) A monodisperse transmembrane α-helical peptide barrel. *Nat. Chem.* **9**: 411-419.

1064 Mathavan, I., Zirah, S., Mehmood, S., Choudhury, H.G., Goulard, C., Li, Y., Robinson, C.V.,
1065   Rebuffat, S., and Beis, K. (2014) Structural basis for hijacking siderophore receptors by
1066   antimicrobial lasso peptides. *Nat. Chem. Biol.* **10**: 340-342.

1067 McNulty, C., Thompson, J., Barrett, B., Lord, L., Andersen, C., and Roberts, I.S. (2006) The cell
1068   surface expression of group 2 capsular polysaccharides in *Escherichia coli*: the role of KpsD,
1069   RhsA and a multi-protein complex at the pole of the cell. *Mol. Microbiol.* **59**: 907-922.

1070 Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, Gustavo A., Sonnhammer, E.L.L.,
1071   Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., Finn, R.D., and Bateman, A. (2021)
1072   Pfam: the protein families database in 2021. *Nucleic Acids Res.* **49**: D412-D419.

1073 Mori, Y., Maeda, M., Takegawa, K., and Kimura, Y. (2012) PhpA, a tyrosine phosphatase of
1074   *Myxococcus xanthus*, is involved in the production of exopolysaccharide. *Microbiology* **158**:
1075   2546-2555.

1076 Muñoz-Dorado, J., Marcos-Torres, F.J., García-Bravo, E., Moraleda-Muñoz, A., and Pérez, J. (2016)
1077   Myxobacteria: moving, killing, feeding, and surviving together. *Front. Microbiol.* **7**.

1078 Muñoz-Dorado, J., Moraleda-Muñoz, A., Marcos-Torres, F.J., Contreras-Moreno, F.J., Martin-
1079   Cuadrado, A.B., Schrader, J.M., Higgs, P.I., and Pérez, J. (2019) Transcriptome dynamics of
1080   the *Myxococcus xanthus* multicellular developmental program. *eLife* **8**: e50374.

1081 Nesper, J., Hill, C.M.D., Paiment, A., Harauz, G., Beis, K., Naismith, J.H., and Whitfield, C. (2003)
1082   Translocation of group 1 capsular polysaccharide in *Escherichia coli* serotype K30: structural
1083   and functional analysis of the outer membrane lipoprotein Wza*. *J. Biol. Chem.* **278**: 49763-
1084   49772.

1085 Nickerson, N.N., Mainprize, I.L., Hampton, L., Jones, M.L., Naismith, J.H., and Whitfield, C. (2014)
1086   Trapped translocation intermediates establish the route for export of capsular polysaccharides
1087   across *Escherichia coli* outer membranes. *Proc. Natl. Acad. Sci. USA* **111**: 8203-8208.

1088 Pavlova, A., Hwang, H., Lundquist, K., Balusek, C., and Gumbart, J.C. (2016) Living on the edge:
1089   simulations of bacterial outer-membrane proteins. *Biochim. Biophys. Acta* **1858**: 1753-1759.

1090 Peleg, A., Shifrin, Y., Ilan, O., Nadler-Yona, C., Nov, S., Koby, S., Baruch, K., Altuvia, S., Elgrably-
1091   Weiss, M., Abe, C.M., Knutton, S., Saper, M.A., and Rosenshine, I. (2005) Identification of
1092   an *Escherichia coli* operon required for formation of the O-antigen capsule. *J. Bacteriol.* **187**:
1093   5259-5266.

1094 Pérez-Burgos, M., García-Romero, I., Jung, J., Schander, E., Valvano, M.A., and Søgaard-Andersen,
1095   L. (2020) Characterization of the exopolysaccharide biosynthesis pathway in *Myxococcus*
1096   *xanthus*. *J. Bacteriol.* **202**: e00335-00320.

1097 Pérez-Burgos, M., and Søgaard-Andersen, L. (2020) Biosynthesis and function of cell-surface
1098   polysaccharides in the social bacterium *Myxococcus xanthus*. *Biol. Chem.* **401**: 1375-1387.

1099 Potter, S.C., Luciani, A., Eddy, S.R., Park, Y., Lopez, R., and Finn, R.D. (2018) HMMER web
1100   server: 2018 update. *Nucl. Acids Res.* **46**: W200-W204.

1101 Reeves, P.R., Hobbs, M., Valvano, M.A., Skurnik, M., Whitfield, C., Coplin, D., Kido, N., Klena, J.,
1102   Maskell, D., Raetz, C.R.H., and Rick, P.D. (1996) Bacterial polysaccharide synthesis and
1103   gene nomenclature. *Trends Microbiol.* **4**: 495-503.

Rouse, S.L., Hawthorne, W.J., Berry, J.-L., Chorev, D.S., Ionescu, S.A., Lambert, S., Stylianou, F., Ewert, W., Mackie, U., Morgan, R.M.L., Otzen, D., Herbst, F.-A., Nielsen, P.H., Dueholm, M., Bayley, H., Robinson, C.V., Hare, S., and Matthews, S. (2017) A new class of hybrid secretion system is employed in *Pseudomonas* amyloid biogenesis. *Nat. Commun.* **8**: 263.

Rutten, L., Geurtsen, J., Lambert, W., Smolenaers, J.J.M., Bonvin, A.M., de Haan, A., van der Ley, P., Egmond, M.R., Gros, P., and Tommassen, J. (2006) Crystal structure and catalytic mechanism of the LPS 3-*O*-deacylase PagL from *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. U. S. A.* **103**: 7071-7076.

Saïdi, F., Jolivet, N.Y., Lemon, D.J., Nakamura, A., Belgrave, A.M., Garza, A.G., Veyrier, F.J., and Islam, S.T. (2021) Bacterial glycocalyx integrity drives multicellular swarm biofilm dynamism. *Mol. Microbiol.* **116**: DOI: 10.1111/mmi.14803.

Sande, C., Bouwman, C., Kell, E., Nickerson, N.N., Kapadia, S.B., and Whitfield, C. (2019) Structural and functional variation in outer membrane polysaccharide export (OPX) proteins from the two major capsule assembly pathways present in *Escherichia coli*. *J. Bacteriol.* **201**: e00213-00219.

Sande, C., and Whitfield, C. (2021) Capsules and extracellular polysaccharides in *Escherichia coli* and *Salmonella*. *EcoSal Plus* **9**: eESP-0033-2020.

Sathiyamoorthy, K., Mills, E., Franzmann, T.M., Rosenshine, I., and Saper, M.A. (2011) The crystal structure of *Escherichia coli* Group 4 capsule protein GfcC reveals a domain organization resembling that of Wza. *Biochemistry* **50**: 5465-5476.

Seef, S., Herrou, J., de Boissier, P., My, L., Brasseur, G., Robert, D., Jain, R., Mercier, R., Cascales, E., Habermann, B.H., and Mignot, T. (2021) A Tad-like apparatus is required for contact-dependent prey killing in predatory social bacteria. *eLife* **10**: e72409.

Sharma, G., Yao, A.I., Smaldone, G.T., Liang, J., Long, M., Facciotti, M.T., and Singer, M. (2021) Global gene expression analysis of the *Myxococcus xanthus* developmental time course. *Genomics* **113**: 120-134.

Silver, R.P., Aaronson, W., and Vann, W.F. (1987) Translocation of capsular polysaccharides in pathogenic strains of *Escherichia coli* requires a 60-kilodalton periplasmic protein. *J. Bacteriol.* **169**: 5489-5495.

Smaldone, G.T., Jin, Y., Whitfield, D.L., Mu, A.Y., Wong, E.C., Wuertz, S., and Singer, M. (2014) Growth of *Myxococcus xanthus* in continuous-flow-cell bioreactors as a method for studying development. *Appl. Environ. Microbiol.* **80**: 2461-2467.

Söding, J., Biegert, A., and Lupas, A.N. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucl. Acids Res.* **33**: W244-W248.

Solan, R., Pereira, J., Lupas, A.N., Kolodny, R., and Ben-Tal, N. (2021) Gram-negative outer-membrane proteins with multiple β-barrel domains. *Proc. Natl. Acad. Sci. U. S. A.* **118**: e2104059118.

Sutherland, I.W., and Thomson, S. (1975) Comparison of polysaccharides produced by *Myxococcus* strains. *J. Gen. Microbiol.* **89**: 124-132.

Tan, J., Rouse, S.L., Li, D., Pye, V.E., Vogeley, L., Brinth, A.R., El Arnaout, T., Whitney, J.C., Howell, P.L., Sansom, M.S.P., and Caffrey, M. (2014) A conformational landscape for

1145 alginate secretion across the outer membrane of *Pseudomonas aeruginosa*. *Acta Crystallogr.*
1146 *D Biol. Crystallogr.* **70**: 2054-2068.

1147 Teufel, F., Almagro Armenteros, J.J., Johansen, A.R., Gíslason, M.H., Pihl, S.I., Tsirigos, K.D.,
1148 Winther, O., Brunak, S., von Heijne, G., and Nielsen, H. (2022) SignalP 6.0 predicts all five
1149 types of signal peptides using protein language models. *Nat. Biotechnol.*

1150 Toh, E., Kurtz, H.D., and Brun, Y.V. (2008) Characterization of the *Caulobacter crescentus* holdfast
1151 polysaccharide biosynthesis pathway reveals significant redundancy in the initiating
1152 glycosyltransferase and polymerase steps. *J. Bacteriol.* **190**: 7219-7231.

1153 Wang, S., Sun, S., Li, Z., Zhang, R., and Xu, J. (2017) Accurate *de novo* prediction of protein contact
1154 map by ultra-deep learning model. *PLOS Comput. Biol.* **13**: e1005324.

1155 Wang, Y., Andole Pannuri, A., Ni, D., Zhou, H., Cao, X., Lu, X., Romeo, T., and Huang, Y. (2016)
1156 Structural basis for translocation of a biofilm-supporting exopolysaccharide across the
1157 bacterial outer membrane*. *J. Biol. Chem.* **291**: 10046-10057.

1158 Wartel, M., Ducret, A., Thutupalli, S., Czerwinski, F., Le Gall, A.-V., Mauriello, E.M.F., Bergam, P.,
1159 Brun, Y.V., Shaevitz, J., and Mignot, T. (2013) A versatile class of cell surface directional
1160 motors gives rise to gliding motility and sporulation in *Myxococcus xanthus*. *PLoS Biol.* **11**:
1161 e1001728.

1162 Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009) Jalview
1163 Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**:
1164 1189-1191.

1165 Whitfield, C., and Valvano, M.A., (1993) Biosynthesis and expression of cell-surface
1166 polysaccharides in Gram-negative bacteria. In: Advances in Microbial Physiology. A.H. Rose
1167 (ed). Academic Press, pp. 135-246.

1168 Whitfield, C., Wear, S.S., and Sande, C. (2020) Assembly of bacterial capsular polysaccharides and
1169 exopolysaccharides. *Annu. Rev. Microbiol.* **74**: 521-543.

1170 Whitney, J.C., and Howell, P.L. (2013) Synthase-dependent exopolysaccharide secretion in Gram-
1171 negative bacteria. *Trends Microbiol.* **21**: 63-72.

1172 Willis, L.M., and Whitfield, C. (2013) Structure, biosynthesis, and function of bacterial capsular
1173 polysaccharides synthesized by ABC transporter-dependent pathways. *Carbohydr. Res.* **378**:
1174 35-44.

1175 Wirth, C., Condemine, G., Boiteux, C., Bernèche, S., Schirmer, T., and Peneff, C.M. (2009) NanC
1176 crystal structure, a model for outer-membrane channels of the acidic sugar-specific KdgM
1177 porin family. *J. Mol. Biol.* **394**: 718-731.

1178 Wunder, D.E., Aaronson, W., Hayes, S.F., Bliss, J.M., and Silver, R.P. (1994) Nucleotide sequence
1179 and mutational analysis of the gene encoding KpsD, a periplasmic protein involved in
1180 transport of polysialic acid in *Escherichia coli* K1. *J. Bacteriol.* **176**: 4025-4033.

1181 Yang, Y., Liu, J., Clarke, B.R., Seidel, L., Bolla, J.R., Ward, P.N., Zhang, P., Robinson, C.V.,
1182 Whitfield, C., and Naismith, J.H. (2021) The molecular basis of regulation of bacterial
1183 capsule assembly by Wzc. *Nat. Commun.* **12**: 4349.

1184 Zeng, H., Wang, S., Zhou, T., Zhao, F., Li, X., Wu, Q., and Xu, J. (2018) ComplexContact: a web
1185 server for inter-protein contact prediction using deep learning. *Nucleic Acids Res.* **46**: W432-
1186 W437.

1187   Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J.,
1188   Lupas, A.N., and Alva, V. (2018) A completely reimplemented MPI Bioinformatics Toolkit
1189   with a new HHpred server at its core. *J. Mol. Biol.* **430**: 2237-2243.

1190

1191

1192   **TABLE 1. β-barrels identified to be syntenic with OPX genes in the REP dataset.**

β-Barrel Query Template	Total β-Barrel Homologues Detected Near OPX Genes	β-Barrel Homologues Detected Near Class 1 OPX Genes	β-Barrel Homologues Detected Near Class 2A OPX Genes	β-Barrel Homologues Detected Near Class 2B OPX Genes	β-Barrel Homologues Detected Near Class 3 OPX Genes
MXAN_7418	MXAN_7418: 6	–	–	–	6
MXAN_3226	MXAN_3226: 9	–	–	–	9
MXAN_1916	MXAN_1916: 12	–	–	–	12
PgaA$_{βb}$	PgaA: 8	3	–	1	4
	PgaA$_{βb}$: 2	2	–	–	–
GfcD$_{Cterβb}$	GfcD: 66	44	4	18	–
	GfcD$_{Cterβb}$: 2	–	1	1	–
YjbH$_{Cterβb}$	YjbH: 79	55	5	19	–
	YjbH$_{Cterβb}$: 3	1	1	1	–
BcsC$_{βb}$	BcsC: 5	1	–	–	4
	BcsC$_{βb}$: –	–	–	–	–
AlgE	AlgE: 14	6	–	2	6
Wzi$_{βb}$	Wzi: 133	44	9	48	32
	Wzi$_{βb}$: 28	5	1	14	8

1193

1194 **TABLE 2. *Myxococcus xanthus* strains used in this study.**

Strain	Genotype/Description	Source or Reference
SI1	DZ2 (wild type)	(Campos & Zusman, 1975)
TM389	Δ*pilA* (Δ*mxan_5783*)	(Ducret *et al.*, 2012)
TM469	Δ*wzaX* (i.e. Δ*mxan_7417/epsY*)	(Ducret *et al.*, 2012)
SI93	Δ*wzpX* (i.e. Δ*mxan_7418/epsX*)	This study

1195

1196   **Supplementary Figure S1. Structural homology between Wza$_{Ec}$ and KpsD$_{Ec}$. (A)** Fold-

1197   recognition analysis of KpsD$_{Ec}$ (via HHpred) revealing C-terminal structural homology to GfcC (PDB:

1198   3P42) as well as Wza$_{Ec}$ (PDB: 2J58). **(B)** Profile-based alignment of Wza$_{Ec}$ and KpsD$_{Ec}$ C-terminal

1199   sequences from Panel A. Wza$_{Ec}$ α-helix (*dark green cylinders*) and β-strand (*dark blue arrows*)

1200   structure is depicted as per the 2J58 PDB entry. KpsD$_{Ec}$ predicted α-helix (*light green cylinders*) and

1201   β-strand (*light blue arrows*) secondary structure is indicated as per PSIPRED analysis. Aligned

1202   residues have been coloured according to Jalview conservation score (out of 10). *Maroon*, 10; *red*,

1203   9; *orange*, 8; *yellow*, 7; *pink*, 6. Scores of 5 or less have been omitted to improve clarity of the

1204   figure. The end of KpsD$_{Ec}$ structural homology with the stand-alone GfcC protein has been indicated

1205   as per a previous report (Sande *et al.*, 2019). **(C)** Tertiary structure model of the KpsD$_{Ec}$ C-terminus

1206   based on structural alignment with Wza$_{Ec}$ as indicated in Panel B. N- and C-termini of the displayed

1207   peptide have been indicated.

SUPPLEMENTARY FIGURE S1

A)

B)


```
Wza_Ec  : TEFQLQPYDIVYVTTAPLVRWNRVISQLVPTISGVHDMTETVRYI : 374
KpsD_Ec : DVDSLKPGDEIMVLPKYESKNIEVTRGISTILYQLAVGAKVILSL : 558
```

End of KpsD_Ec
structural
homology
to GfcC
(PDB: 3P42)

C)
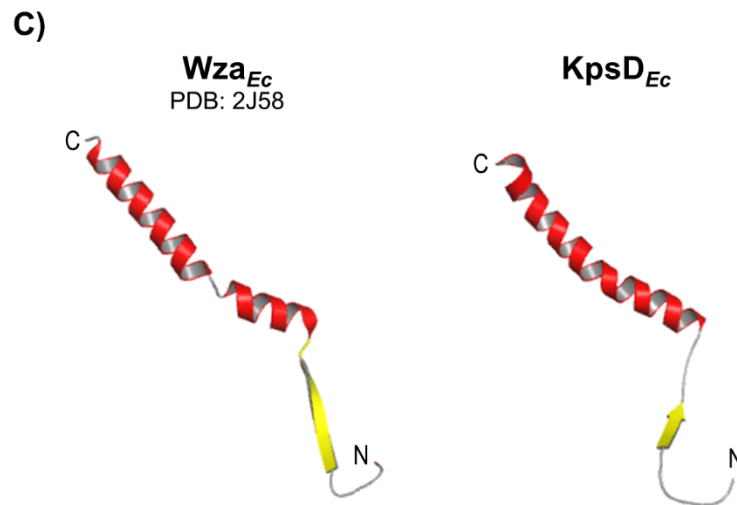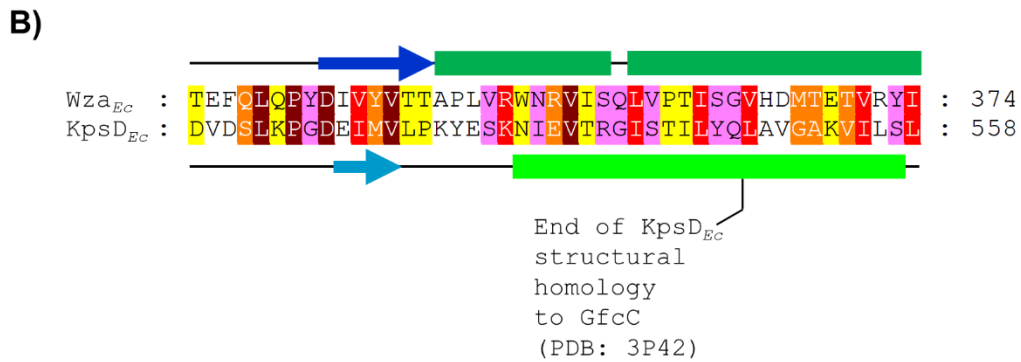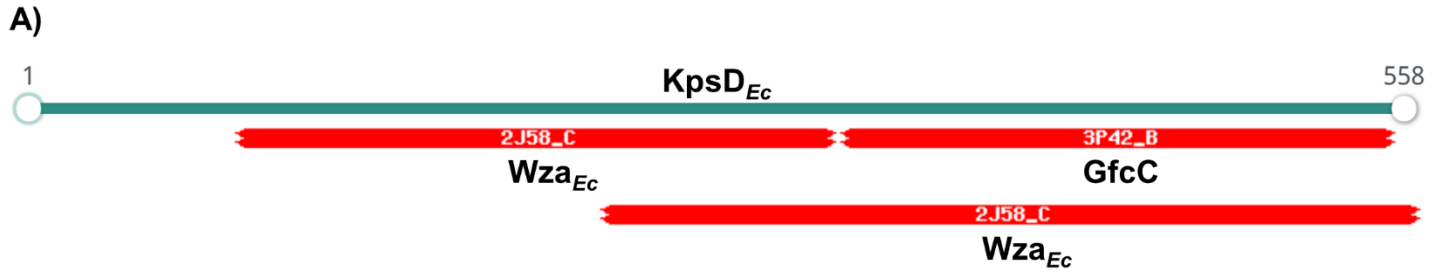
1209 **Supplementary Figure S2. Structural analysis of MXAN_7418 (WzpX). (A)** Evolutionarily-

1210 coupled amino acids within the MXAN_7418 primary structure (determined via RaptorX). **(B)** Fold-

1211 recognition analysis of MXAN_7418 (via HHpred) revealing N-terminal structural homology with two

1212 β-strands from FhuA (PDB: 4CU4) (Mathavan *et al.*, 2014), with the remainder of the protein

1213 displaying structural homology to PgaA$_{βb}$ (PDB: 4Y25) (Wang *et al.*, 2016). FhuA and PgaA$_{βb}$ β-

1214 strand (*dark blue arrows*) structure is depicted as per the respective PDB entries. MXAN_7418

1215 predicted β-strand (*light blue arrows*) secondary structure is indicated as per PSIPRED analysis.

1216 Aligned residues have been coloured according to Jalview conservation score (out of 10). *Maroon*,

1217 10; *red*, 9; *orange*, 8; *yellow*, 7; *pink*, 6. Scores of 5 or less have been omitted to improve clarity of

1218 the figure.

SUPPLEMENTARY FIGURE S2

A)

MXAN_7418

B)

FhuA aa484   FhuA aa518   PgaA aa517

```
FhuA(484-518): GRYDWADQESLNRVAGTTDKRDDKQFTWRGGVNYLLAELRIAGSTGID-----AEGPDSG : 536
   MXAN_7418 : ARVDSRLRSNEDPAPEAPRVAGDTDVTPLLGLEIRDGNTALQVEYAPRISLREVTARART :  83
```

```
   PgaAβb : KHDVDLTTIVYSPPLKDNWRGFAGFGYADGQF----------------------------- : 568
MXAN_7418 : EIQHMGRLA-GLWRPERGLSLRLSEELVLGSVNLLTTDLGSLPGDTPGTPDGPLRPPGNG : 142
```

```
   PgaAβb : SEGK---------GIVR-DWLAGVEWRSRN-IWLEAEYAERV---FNHEH--------KP : 606
MXAN_7418 : GPLQPLPQADTVYFLSSATTLTADTGWLGRRWNLTGSGGFSISGGLDGPAREAVPMQYGP : 202
```

```
   PgaAβb : GARLSGWYDFNDNWRIGSQLERLSHRVPLRAMKNGVTGNSAQAYVRWYQNERRKYGVSWA : 666
MXAN_7418 : RADVSLSHALSPLSAITTSVAFNYARFSTG---ANNTIVTLREAWGHRVDRRTSLEAGAG : 259
```

```
   PgaAβb : FTDFS------------------DSNQRHEVSLEGQERIWSS-PYLIVDFLPSLYYEQN : 706
MXAN_7418 : VSVVHTIPVPPGEGEPEPGPLDAPRTELLPNLNVGVSHRIPSRTADFNGRADLRLVPFTD : 319
```

```
   PgaAβb : TEHDTPYYNPIKTFDIVPAFEASHLLWRSYENSWEQIFSAGVGASWQKHYGTDVVTQLGY : 766
MXAN_7418 : RLTAR------VYPRADLTLTGTWAL----GPRVRVSAIGGGAAA-VGGSRGDSILSGGF : 368
```

```
   PgaAβb : GQRISWNDVIDAGATLRWEKRPYDGDREH--NLYVEFDMTFRF---- : 807
MXAN_7418 : TGSWILTRWVSVDTDLRGTWSRSPELPAARLTWAATLGLSVRQTGIL : 415
```
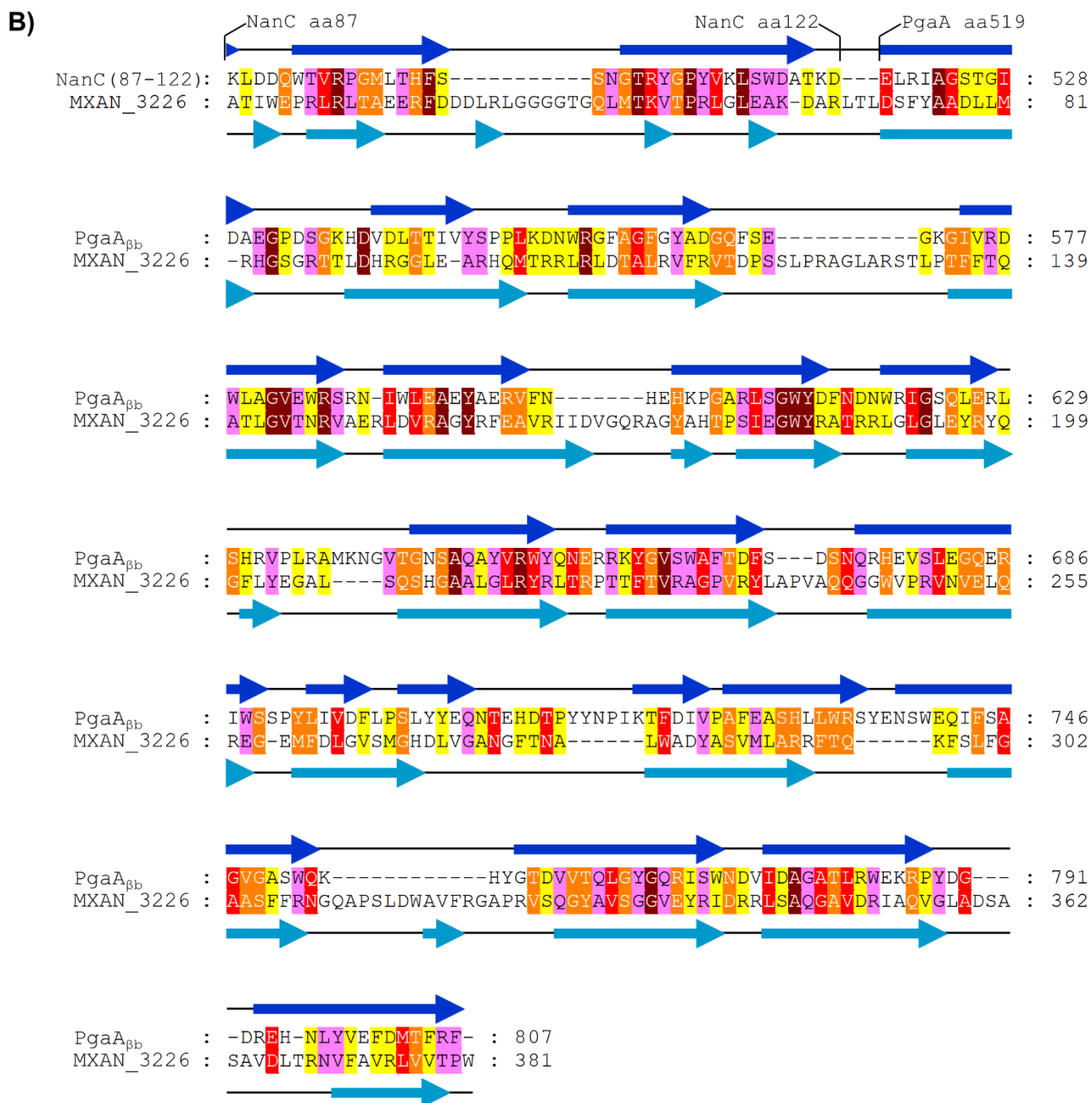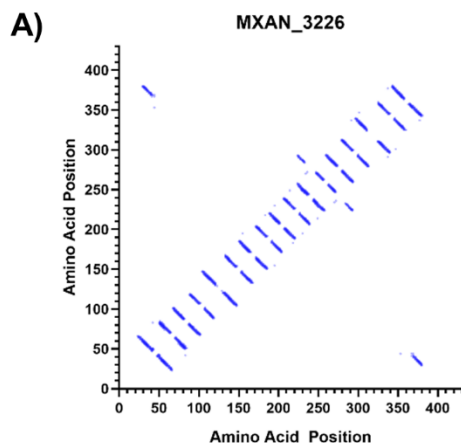
1220   **Supplementary Figure S3. Structural analysis of MXAN_3226 (WzpS). (A)** Evolutionarily-

1221   coupled amino acids within the MXAN_3226 primary structure (determined via RaptorX). **(B)** Fold-

1222   recognition analysis of MXAN_3226 (via HHpred) revealing N-terminal structural homology with two

1223   β-strands from NanC (PDB: 2WJR) (Wirth *et al.*, 2009), with the remainder of the protein displaying

1224   structural homology to PgaA$_{βb}$ (PDB: 4Y25) (Wang *et al.*, 2016). NanC and PgaA$_{βb}$ β-strand (*dark*

1225   *blue arrows*) structure is depicted as per the respective PDB entries. MXAN_3226 predicted β-

1226   strand (*light blue arrows*) secondary structure is indicated as per PSIPRED analysis. Aligned

1227   residues have been coloured according to Jalview conservation score (out of 10). *Maroon*, 10; *red*,

1228   9; *orange*, 8; *yellow*, 7; *pink*, 6. Scores of 5 or less have been omitted to improve clarity of the
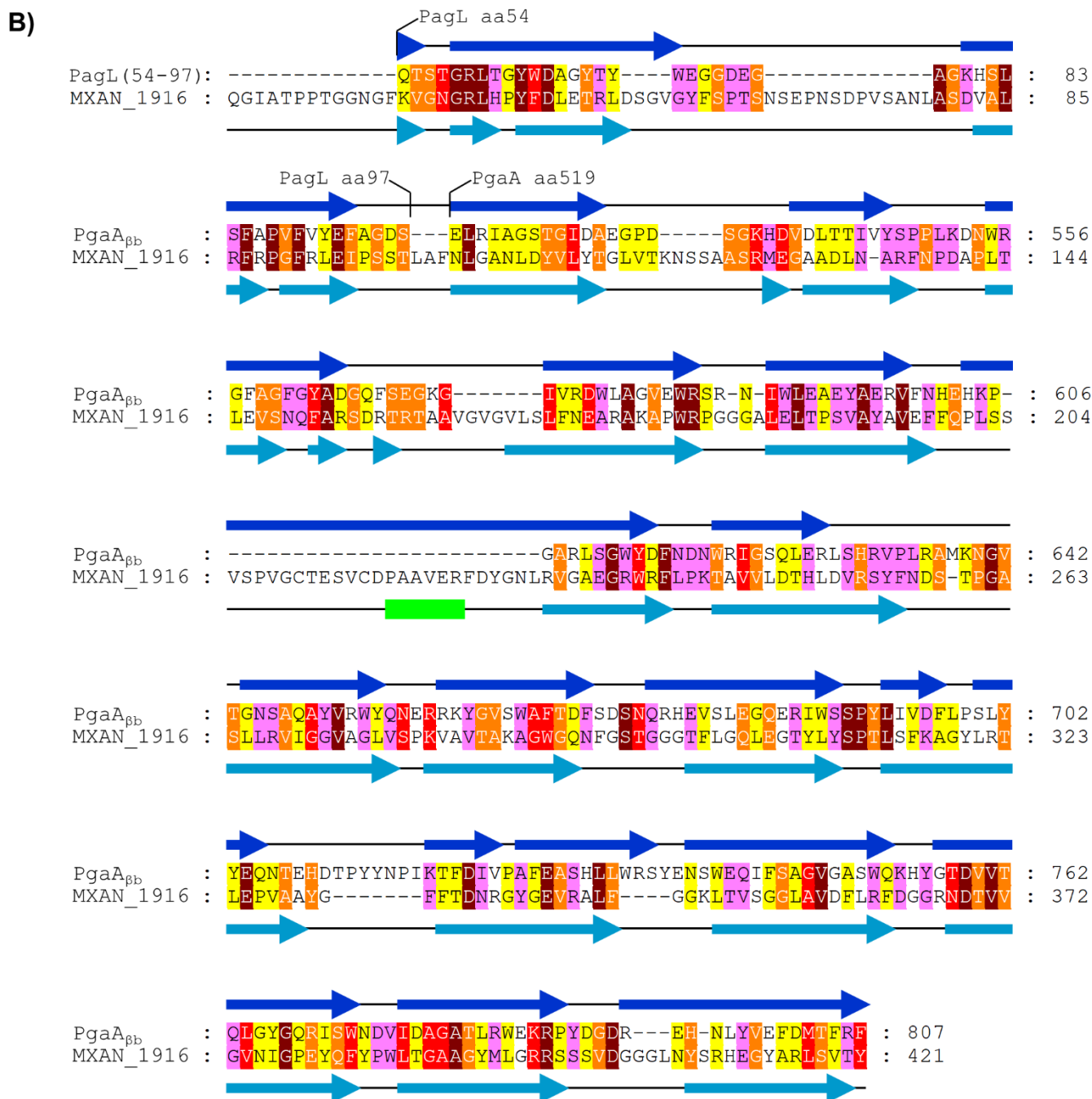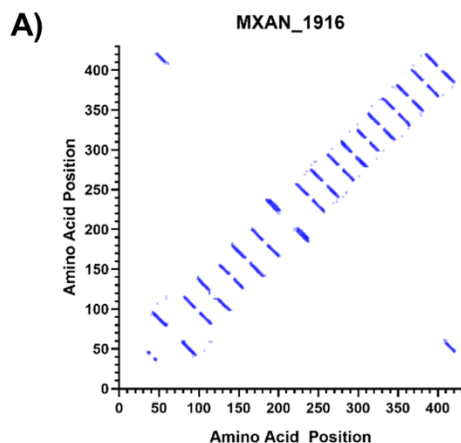
1229   figure.

SUPPLEMENTARY FIGURE S3

A)

MXAN_3226

B)

NanC aa87   NanC aa122   PgaA aa519

```
NanC(87-122): KLDDQWTVRPGMLTHFS-----------SNGTRYGPYVKLSWDATKD---ELRIAGSTGI : 528
  MXAN_3226 : ATIWEPRLRLTAEERFDDDLRLGGGGTGQLMTKVTPRLGLEAK-DARLTLDSFYAADLLM :  81
```

```
    PgaAββ : DAEGPDSGKHDVDLTTIVYSPPLKDNWRGFAGFGYADGQFSE-----------GKGIVRD : 577
  MXAN_3226 : -RHGSGRTTLDHRGGLE-ARHQMTRRLRLDTALRVFRVTDPSSLPRAGLARSTLPTFFTQ : 139
```

```
    PgaAββ : WLAGVEWRSRN-IWLEAEYAERVFN-------HEHKPGARLSGWYDFNDNWRIGSQLERL : 629
  MXAN_3226 : ATLGVTNRVAERLDVRAGYRFEAVRIIDVGQRAGYAHTPSIEGWYRATRRLGLGLEYRYQ : 199
```

```
    PgaAββ : SHRVPLRAMKNGVTGNSAQAYVRWYQNERRKYGVSWAFTDES---DSNQRHEVSLEGQER : 686
  MXAN_3226 : GFLYEGAL----SQSHGAALGLRYRLTRPTTFTVRAGPVRYLAPVAQQGGWVPRVNVELQ : 255
```

```
    PgaAββ : IWSSPYLIVDFLPSLYYEQNTEHDTPYYNPIKTFDIVPAFEASHLLWRSYENSWEQIFSA : 746
  MXAN_3226 : REG-EMFDLGVSMGHDLVGANGFTNA------LWADYASVMLARRFTQ------KFSLFG : 302
```

```
    PgaAββ : GVGASWQK------------HYGTDVVTQLGYGQRISWNDVIDAGATIRWEKRPYDG--- : 791
  MXAN_3226 : AASFFRNGQAPSLDWAVFRGAPRVSQGYAVSGGVEYRIDRRLSAQGAVDRIAQVGLADSA : 362
```

```
    PgaAββ : -DREH-NLYVEFDMTFRF- : 807
  MXAN_3226 : SAVDLTRNVFAVRLVVTPW : 381
```
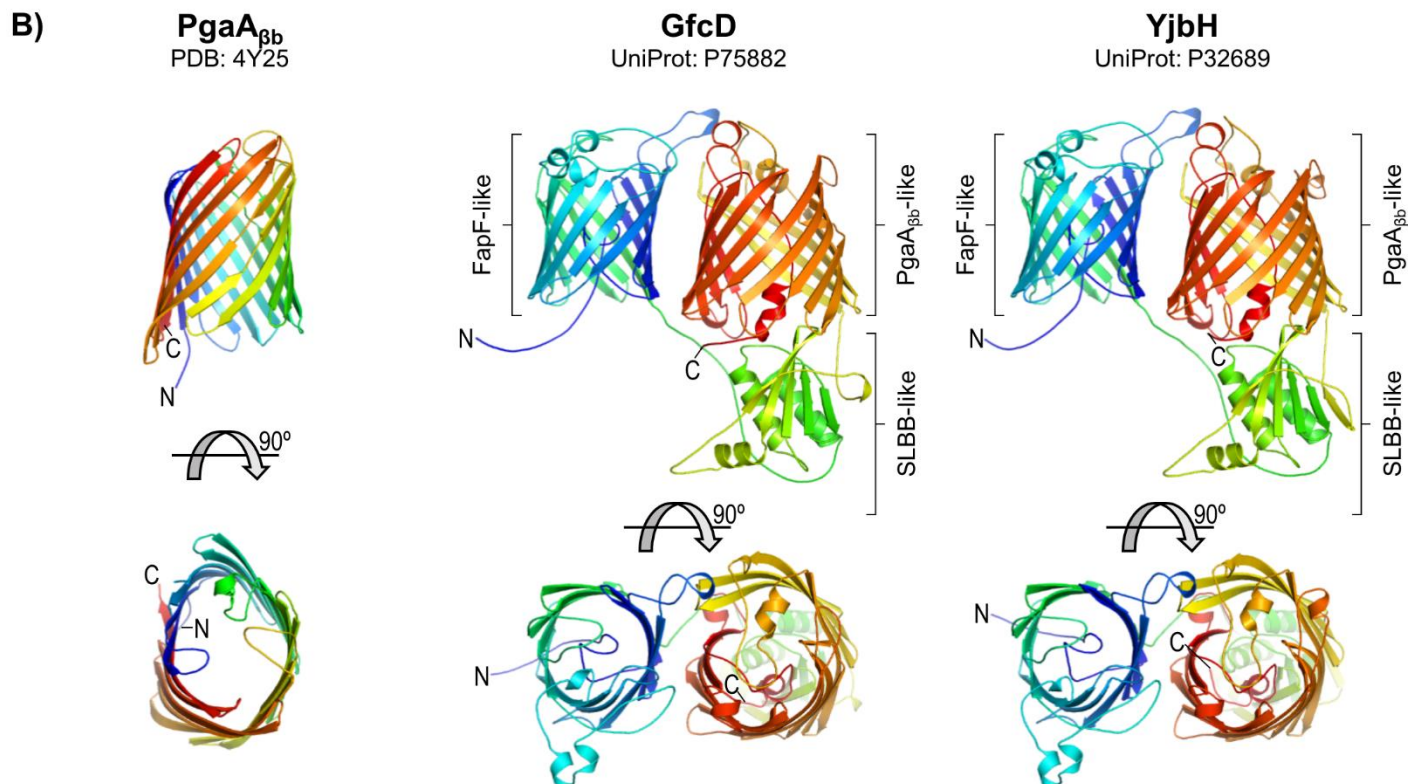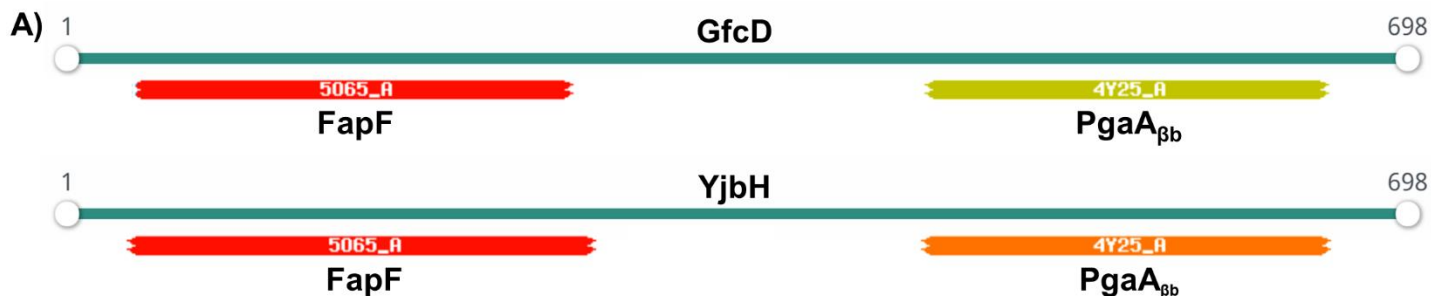
1231   **Supplementary Figure S4. Structural analysis of MXAN_1916 (WzpB). (A)** Evolutionarily-

1232   coupled amino acids within the MXAN_1916 primary structure (determined via RaptorX). **(B)** Fold-

1233   recognition analysis of MXAN_1916 (via HHpred) revealing N-terminal structural homology with two

1234   β-strands from PagL (PDB: 2ERV) (Rutten *et al.*, 2006), with the remainder of the protein displaying

1235   structural homology to PgaA$_{\beta b}$ (PDB: 4Y25) (Wang *et al.*, 2016). PagL and PgaA$_{\beta b}$ β-strand (*dark*

1236   *blue arrows*) structure is depicted as per the respective PDB entries. MXAN_1916 predicted α-helix

1237   (*light green cylinders*) and β-strand (*light blue arrows*) secondary structure is indicated as per

1238   PSIPRED analysis. Aligned residues have been coloured according to Jalview conservation score

1239   (out of 10). *Maroon*, 10; *red*, 9; *orange*, 8; *yellow*, 7; *pink*, 6. Scores of 5 or less have been omitted

1240   to improve clarity of the figure.

SUPPLEMENTARY FIGURE S4

A)

MXAN_1916

B)


```
                                    PagL aa54

PagL(54-97): --------------QTSTGRLTGYWDAGYTY----WEGGDEG-------------AGKHSL :  83
MXAN_1916  : QGIATPPTGGNGFKVGNGRLHPYFDLETRLDSGVGYFSPTSNSEPNSDPVSANLASDVAL :  85


              PagL aa97        PgaA aa519

PgaA_βb    : SFAPVFVYEFAGDS---ELRIAGSTGIDAEGPD-----SGKHDVDLTTIVYSPPLKDNWR : 556
MXAN_1916  : RFRPGFRLEIPSSTLAFNLGANLDYVLYTGLVTKNSSAASRMEGAADLN-ARFNPDAPLT : 144


PgaA_βb    : GFAGFGYADGQFSEGKG-------IVRDWLAGVEWRSR-N-IWLEAEYAERVFNHEHKP- : 606
MXAN_1916  : LEVSNQFARSDRTRTAAVGVGVLSLFNEARAKAPWRPGGGALELTPSVAYAVEFFQPLSS : 204


PgaA_βb    : ------------------------GARLSGWYDFNDNWRIGSQLERLSHRVPLRAMKNGV : 642
MXAN_1916  : VSPVGCTESVCDPAAVERFDYGNLRVGAEGRWRFLPKTAVVLDTHLDVRSYFNDS-TPGA : 263


PgaA_βb    : TGNSAQAYVRWYQNERRKYGVSWAFTDFSDSNQRHEVSLEGQERIWSSPYLIVDFLPSLY : 702
MXAN_1916  : SLLRVIGGVAGLVSPKVAVTAKAGWGQNFGSTGGGTFLGQLEGTYLYSPTLSFKAGYLRT : 323


PgaA_βb    : YEQNTEHDTPYYNPIKTFDIVPAFEASHLLWRSYENSWEQIFSAGVGASWQKHYGTDVVT : 762
MXAN_1916  : LEPVAAYG-------FFTDNRGYGEVRALF----GGKLTVSGGLAVDFLRFDGGRNDTVV : 372


PgaA_βb    : QLGYGQRISWNDVIDAGATLRWEKRPYDGDR---EH-NLYVEFDMTFRF : 807
MXAN_1916  : GVNIGPEYQFYPWLTGAAGYMLGRRSSSVDGGGLNYSRHEGYARLSVTY : 421
```

1242 **Supplementary Figure S5. Structural homology between Pgaβb, GfcD, and YjbH. (A)** Fold-

1243 recognition analysis of GfcD and YjbH (via HHpred) revealing N-terminal structural homology of

1244 each to the amyloid secretion β-barrel FapF (PDB: 5O65) (Rouse *et al.*, 2017) and C-terminal

1245 structural homology of each to PgaA$_{βb}$ (PDB: 4Y25) (Wang *et al.*, 2016). **(B)** AlphaFold2-generated

1246 tertiary structure models for GfcD and YjbH, displayed alongside the Pga$_{βb}$ X-ray crystal structure for

1247 comparison. Proteins have been coloured with a spectrum, from the N-terminus (*blue*) to the C-

1248 terminus (*red*). **(C)** Multiple-sequence alignment of the PgaA$_{βb}$ (aa 511-807), GfcD$_{Cterβb}$ (aa 425-

1249 698), and YjbH$_{Cterβb}$ (aa 423-698) segments. Aligned residues have been coloured according to

1250 Jalview conservation score (out of 10). *Maroon*, 10; *red*, 9; *orange*, 8; *yellow*, 7; *pink*, 6. Scores of 5

1251 or less have been omitted to improve clarity of the figure.

SUPPLEMENTARY FIGURE S5

A)

GfcD 1 — 698
FapF (5065_A) / PgaA$_{\beta b}$ (4Y25_A)

YjbH 1 — 698
FapF (5065_A) / PgaA$_{\beta b}$ (4Y25_A)

B)

PgaA$_{\beta b}$ — PDB: 4Y25
GfcD — UniProt: P75882
YjbH — UniProt: P32689

C)

```
GfcD_Cterβb : FSYSFNPTLSQSLGGPEDFYMFQLG----LMSSARYWFTDHLLLDGGIFTNIYNNYDKFK : 480
YjbH_Cterβb : FDFHIDPVLNQSVGGPENFYMYQLG----VMGTADLWLTDHLLTTGSLFANLANNYDKFN : 478
PgaA_βb     : -AVDVHNLAELRIAGSTGIDAEGPDSGKHDVDLTTIVYSPPLKDNWRGFAGFGYADGQFS : 569

GfcD_Cterβb : SSLLPADSTLPRVRTHIRDYVRNDVYLNNLQANYFADLG--NGFYGQVYGGYLETMYAGV : 538
YjbH_Cterβb : YTNPPQDSHLPRVRTHVREYVQNDVYVNNLQANYFQHLG--NGFYGQVYGGYLETMFGGA : 536
PgaA_βb     : EGKGIVRDWLAGVEWRSRNIWLEAEYAERVFNHEHKPGARLSGWYDFNDNWRIGSQLERL : 629

GfcD_Cterβb : GSELLYRPLDACWALGVDVNYVK--QRDWD---NMMRFTDYS--------TPTGFVTAYW : 585
YjbH_Cterβb : GAEVLYRPLDSNWAFGLDANYVK--QRDWRSAKDMMKFTDYS--------VKTGHLTAYW : 586
PgaA_βb     : SHRVPLRAMKNGVTGNSAQAYVRWYQNERRKYGVSWAFTDFSDSNQRHEVSLEGQERIWS : 689

GfcD_Cterβb : NPPTLNGVLMKLSVGQYLAKDKG-----ATIDVAKRFDSGVAVGVWAAISNVSKDDYGEG : 640
YjbH_Cterβb : TPSFAQDVLVKASVGQYLAGDKG-----GTLEIAKRFDSGVVVGGYATITNVSKEEYGEG : 641
PgaA_βb     : SPYLIVDFLPSLYYEQNTEHDTPYYNPIKTFDIVPAFEASHLLWRSYENSWEQIFSAGVG : 749

GfcD_Cterβb : GFSKGFYISIPFDLMTIGPNRNRAVVSWTPLTRDGGQMLSRKYQLYPMTAFREVPVGQ : 698
YjbH_Cterβb : DFTKGVYVSVPLDLFSSGPTRSRAAIGWTPLTRDGGQQLGRKFQLYDMTSDRSVNFR- : 698
PgaA_βb     : ASWQKHYGTDVVTQLGYGQRISWNDVIDAGATLRWEKRPYDGDREHNLYVEFDMTFRF : 807
```

1253 **Supplementary Figure S6. Structural schematic of polymer translocation across the**

1254 **periplasm in *M. xanthus*.** Components from the BPS secretion pathway have been used as

1255 representative proteins for those in the EPS and MASC pathways as well. All proteins, spaces, and

1256 distances have been depicted at the same relative scale across a representative 327 Å periplasmic

1257 space in a *M. xanthus* cell. X-ray crystal structures for Wza$_{Ec}$ (chain A) and Wzc$_{Ec}$ (chains F and G)

1258 have been provided as per the PDB files 2J58 (Dong *et al.*, 2006) and 7NII (Yang *et al.*, 2021)

1259 (respectively) for size references. Structure models for WzaB (Fig. 1B) and WzpB (Fig. 5A) were

1260 already generated in this investigation. Models for the PCP protein WzcB were generated using

1261 either AlphaFold2 (resulting in an extended conformation), or MODELLER (specifically against the

1262 7NII_F template, resulting in a compact conformation). High-confidence co-evolving amino acids

1263 between WzcB and WzaB have been highlighted with *green* (86% probability) and *cyan* (81%

1264 probability) spheres.
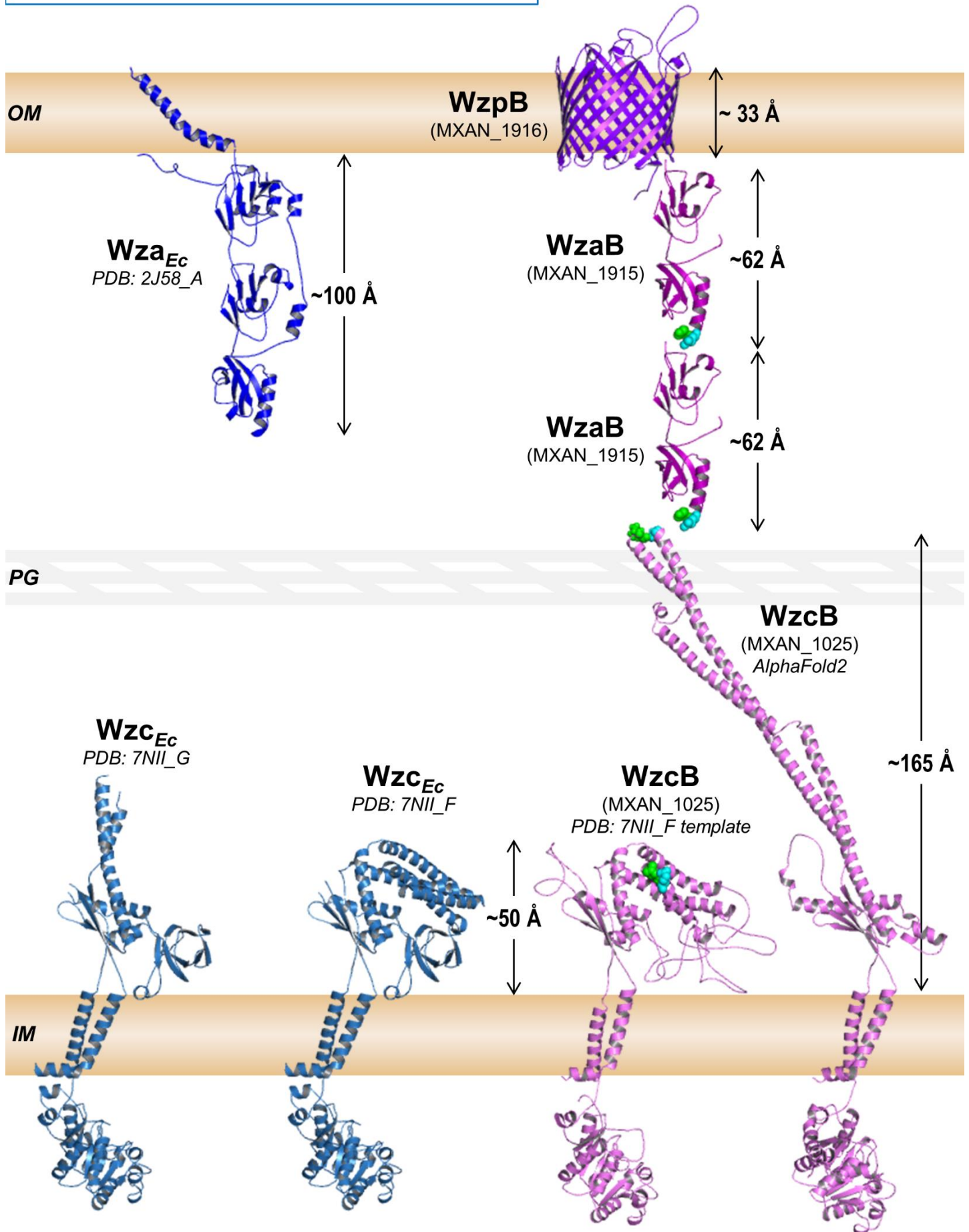
SUPPLEMENTARY FIGURE S6

SUPPLEMENTARY TABLE LEGENDS

Supplementary Table S1. MYXO dataset analysis. (A) Protein-wise OPX classification within 61 myxobacterial genomes. **(B)** Distribution of OPX-protein types within myxobacterial genomes. **(C)** Synteny analysis of OPX-protein types with β-barrel protein homologues.

Supplementary Table S2. REP dataset analysis. (A) Protein-wise OPX classification within 3662 Representative/Reference genomes. **(B)** Distribution of OPX-protein Classes within 3662 genomes. **(C)** Synteny distribution of OPX-protein types with β-barrel protein homologues. **(D)** Arrangement of syntenic OPX-protein types with β-barrel protein homologues.

Supplementary Table S3. NR dataset analysis. (A) Protein-wise OPX classification within NR-database proteins.

Supplementary Table S4. Distribution of OPX-protein types at Phylum, Class, Order, Family, and Genus level taxonomy.

Supplementary Table S5. Evolutionary couplings between proteins. Co-evolving amino acids between PCP and OPX pairs, as well as OPX and Wzp β-barrel pairs, are presented for constituents of the EPS, BPS, and MASC pathways.