NRC Publications Archive
Archives des publications du CNRC

Sequence-structure relationships in polysaccharide co-polymerase (PCP) proteins

Morona, Renato; Purins, Leanne; Tocilj, Ante; Matte, Allan; Cygler, Miroslaw

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. / La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.
For the publisher's version, please access the DOI link below./ Pour consulter la version de l'éditeur, utilisez le lien DOI ci-dessous.

Publisher's version / Version de l'éditeur:
https://doi.org/10.1016/j.tibs.2008.11.001
Trends in Biomedical Science, 34, 2, pp. 78-84, 2009-12-06

NRC Publications Record / Notice d'Archives des publications de CNRC:
https://nrc-publications.canada.ca/eng/view/object/?id=7fdb7828-5ab2-4b98-9ddb-88e414b45b80
https://publications-cnrc.canada.ca/fra/voir/objet/?id=7fdb7828-5ab2-4b98-9ddb-88e414b45b80

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at
https://nrc-publications.canada.ca/eng/copyright
READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site
https://publications-cnrc.canada.ca/fra/droits
LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.

National Research Council Canada
Conseil national de recherches Canada
Canada

Review

Cell PRESS

Sequence-structure relationships in polysaccharide co-polymerase (PCP) proteins

Renato Morona[1], Leanne Purins[1], Ante Tocilj[2], Allan Matte[3] and Miroslaw Cygler[2,3]

[1] Australian Bacterial Pathogenesis Program, School of Molecular and Biomedical Science, Discipline of Microbiology and Immunology, University of Adelaide, North Terrace, Adelaide, SA 5005, Australia
[2] Department of Biochemistry, McGill University, 3655 Promenade Sir William Osler, Montréal, Québec H4P 2R2, Canada
[3] Biotechnology Research Institute, National Research Council, 6100 Royalmount Avenue, Montréal, Québec H4P 2R2, Canada

Polysaccharides are ubiquitously distributed on the cell surface of bacteria. These polymers are involved in many processes, including immune avoidance and bacteria–host interactions, which are especially important for pathogenic organisms. In many instances, the lengths of these polysaccharides are not random, but rather distribute around some mean value, termed the modal length. A large family of proteins, called polysaccharide co-polymerases (PCPs), found in both Gram-negative and Gram-positive species regulate polysaccharide modal length. Recent crystal structures of Wzz proteins from *Escherichia coli* and *Salmonella typhimurium* provide the first atomic-resolution information for one family of PCPs, the PCP1 group. These crystal structures have important implications for the structures of other PCP families.

Wzy-dependent polysaccharide biosynthesis

The cell surface polysaccharides produced by bacteria affect the interactions with their environment and, in pathogenic bacteria, contribute to immune avoidance and host–pathogen interactions. The biosynthesis of many lipopolysaccharide (LPS) O-antigens (Oags), enterobacterial common antigen (ECA) and some capsule polysaccharides (CPSs) in Gram-negative and Gram-positive bacteria occurs by a mechanism known as the Wzy polymerase-dependent pathway [1–4]. The process relies on a putative, multifunctional protein complex involving homologues of the inner membrane proteins Wzx, Wzy and Wzz [5]. Undecaprenol pyrophosphate-linked sugar repeat units are synthesized in the cytoplasm and flipped to the periplasmic (or external) side of the inner membrane by the flippase Wzx. The repeat units are assembled in the periplasm into long chains by a poorly understood mechanism involving the polymerase Wzy. The polysaccharide chain grows at the reducing end and, once complete, it can be attached to an anchor, such as lipid A-core in the case of LPS Oag biosynthesis, by the ligase WaaL. The synthesized polysaccharides are not uniform in length but, rather, their lengths show a distribution that is a characteristic of the particular bacterial species. The polysaccharide is then transported to the outer leaflet of the outer membrane in a manner that is dependent on the type of polysaccharide being made. The length distribution is associated with the presence of a Wzz protein and shows a maximum of a certain number of repeat units (modal distribution) as determined by the specific Wzz protein. The mechanism by which Wzz proteins impact the polymerization process is unknown. In the absence of Wzz, polysaccharide chain lengths show a stochastic distribution, decreasing in quantity as the chain grows longer, characteristic of a competition between elongation and transfer to an acceptor.

Regulation of surface polysaccharide length is an important adaptation that leads to optimal survival and virulence of pathogens [6–8]. During LPS Oag biosynthesis, Wzz proteins control Oag modal length distribution through two proposed mechanisms: as a molecular timer of Oag polymerization [9] or in an organizational manner to determine the crucial ratio of the polymerase Wzy to the Oag ligase WaaL [10]. Much of the work aimed at understanding the mechanism behind Wzz-mediated chain length regulation has come from mutagenesis and genetic studies, which revealed that chain modal length regulation is a property of the entire protein [10–17]. In agreement with secondary structure predictions, genetic and biochemical analyses have shown that Wzz is largely α-helical in structure and is likely to function as an oligomer [9,12,18,19]. These approaches have been limited in giving insight into the molecular mechanism involved in polysaccharide chain modal length control. The recent structure determination of several Wzz proteins has provided a structural basis for the interpretation of these observations. In light of these exciting new data, the purpose of this communication is to review the sequence-structure relationships of PCP1 proteins in more detail and to draw some comparisons with other members of the PCP family. We also propose a model to explain how Wzz proteins function in polysaccharide biosynthesis.

PCP proteins

A large number of proteins, in addition to Wzz, can influence the length and/or export of polysaccharides. These have been named PCPs (for polysaccharide co-polymerases) and are further distinguished by their common membrane topology: nearly all harbor transmembrane

Corresponding author: Morona, R. (renato.morona@adelaide.edu.au).

0968-0004/$ – see front matter © 2008 Elsevier Ltd. All rights reserved. doi:10.1016/j.tibs.2008.11.001 Available online 6 December 2008

helices near the N (TM1) and the C (TM2) termini, a periplasmic α-helical domain containing a predicted coiled-coil region and limited conserved sequence motifs including a Pro–Gly-rich segment overlapping TM2 [20]. Residues within this latter segment have been shown to be essential for function. These proteins have been subdivided into three groups (PCP1, PCP2 and PCP3) based on the chemical nature of the polysaccharide, their association with a Wzy-dependent or ATP-binding cassette (ABC) transporter-dependent pathway and, finally, by the presence or absence of an additional cytoplasmic domain. The PCP1 and PCP2 proteins participate in the Wzy-dependent pathway of polysaccharide biosynthesis, whereas the PCP3 proteins are involved in CPS biosynthesis via an ABC-2 type transporter pathway (for a review, see Ref. [4]). Additionally, the PCP1 group is divided into PCP1a and PCP1b, distinguishing the WzzB proteins (PCP1a) that operate on Oag from the WzzE proteins (PCP1b) that regulate the length of ECA [1]. The most striking effects on polysaccharide length have been visualized for the PCP1a proteins in Oag biosynthesis because Oag ladders can be resolved by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) such that the number of individual sugar repeat units can be counted, thus, enabling subtle differences in the degree of polymerization to be distinguished [7]. By contrast, the PCP2 and PCP3 proteins influence large molecular weight capsular polysaccharides and exopolysaccharides that currently cannot be resolved to the same level of clarity. PCP2 proteins are further divided into the PCP2a subgroup of proteins from Gram-negative bacteria and the PCP2b subgroup of proteins from Gram-positive bacteria [20].

Despite many efforts, limited structural data have been available for PCPs, slowing our efforts to rationalize the genetic data and to define a precise role for these proteins in polysaccharide biosynthesis. Wzc, a PCP2 protein required for high molecular weight capsule production in *Escherichia coli*, was the first PCP protein to be structurally characterized, with a 14 Å resolution reconstruction obtained by electron microscopy [21]. This result constituted a major advance in the field and, although the assignment of sequence to specific parts of the structure was not possible at this resolution, the image reconstruction showed the overall shape of the molecule and confirmed that it forms oligomers. Additionally, the recently published structure of Wzc in complex with the outer membrane protein Wza determined by electron microscopy shows, for the first time, the participation of a PCP protein in a periplasmic-spanning bridge [22].

In an exciting breakthrough, the first crystal structures of multiple PCP1 family members have been reported [23]. The presence of TM regions has been a major hindrance in obtaining high-resolution structural data for PCP proteins and has necessitated the use of detergents to solubilize these proteins from the inner membrane during purification, and for their crystallization. Although the periplasmic domains are expected to be the dominant functional regions of PCPs, a functional activity for the C-terminal cytoplasmic region of a PCP2b protein has been demonstrated [24] and, therefore, these segments might also be important for the activity of other PCPs. The periplasmic domains of several PCP1s crystallized readily, resulting in structures for the periplasmic regions of FepE (*E. coli* O157:H7), $WzzB_{ST}$ (*S. typhimurium*) and WzzE (*E. coli* O157:H7) and a low-resolution crystal structure for $WzzB_{EC}$ (*E. coli*). In addition, electron microscopy of negatively stained, purified, full-length WzzE showed that the dimensions of the reconstructed molecule correlated extremely well with the oligomers observed in the crystal structure of the soluble periplasmic domain alone, confirming that these oligomers represent an accurate picture of the full-length protein. These PCP1 proteins regulate either Oag chain length (FepE and $WzzB_{ST}$) or ECA chain length (WzzE).

Crystal structures of PCP1 proteins

The structures of the periplasmic domains of the three different PCP1 protomers reveal that, despite their limited sequence identity (21–27%), they all adopt a similar fold that can be divided structurally into two domains, an α-β base domain and a protruding α helical hairpin domain [23] (Figure 1a). The α-β base domain comprises the N-terminal part of the periplasmic region and is located at the bottom of the protomer close to the membrane. A single β-strand at the C terminus, just before the second TM helix, contributes to the β-sheet within the base domain. A group of conserved, hydrophobic residues interact to stabilize this region with contacts between α6 and α2 [23]. In the FepE and WzzE oligomers, helices α4 and α5 in the base domain are surface-exposed and protrude outwards, creating what looks like the rim of a bell (Figure 1a,b). The exposed loop between helices α4 and α5 is highly charged in FepE (seven residues) and is less charged, yet still polar in nature, in WzzE. The beginning of α6 is also partially surface-accessible at the base of the oligomer (Figure 1b). $WzzB_{ST}$ shows a more compact base domain relative to FepE and WzzE, having a shorter, less protruding connection between helix α2 and strand β2. This occurs because $WzzB_{ST}$ lacks helices α4 and α5 and possesses a single helix in place of α2 and α3 [23] (Figure 1a). However, this shortened connection in $WzzB_{ST}$ is also highly charged, similarly to that in FepE (Figure 1a).

The base domain is expected to participate in interactions with other Oag or ECA biosynthesis proteins located in the membrane, including Wzy and Wzx [23]. Currently, genetic evidence but no direct physical evidence supports the existence of such interactions [5]. It is possible that these exposed regions in the base domain have a role in putative interactions with these other proteins. The role of the charged, polar region between α4–α5 in FepE and WzzE or just before β2 in $WzzB_{ST}$ is unknown, but could be important in mediating such interactions. Alternatively, the close proximity of charged residues to the inner membrane enables these residues to interact with lipid head groups, anchoring the oligomer at the membrane surface [23]. Interestingly, the conserved Pro residue located at the end of β4 was observed in a *cis* configuration, which might be important for properly orienting the periplasmic domain with respect to the membrane.

The second structural domain consists of the long α-helical hairpin that protrudes into the periplasm. The α-helical hairpin comprises a ~19 turn helix, α6, which

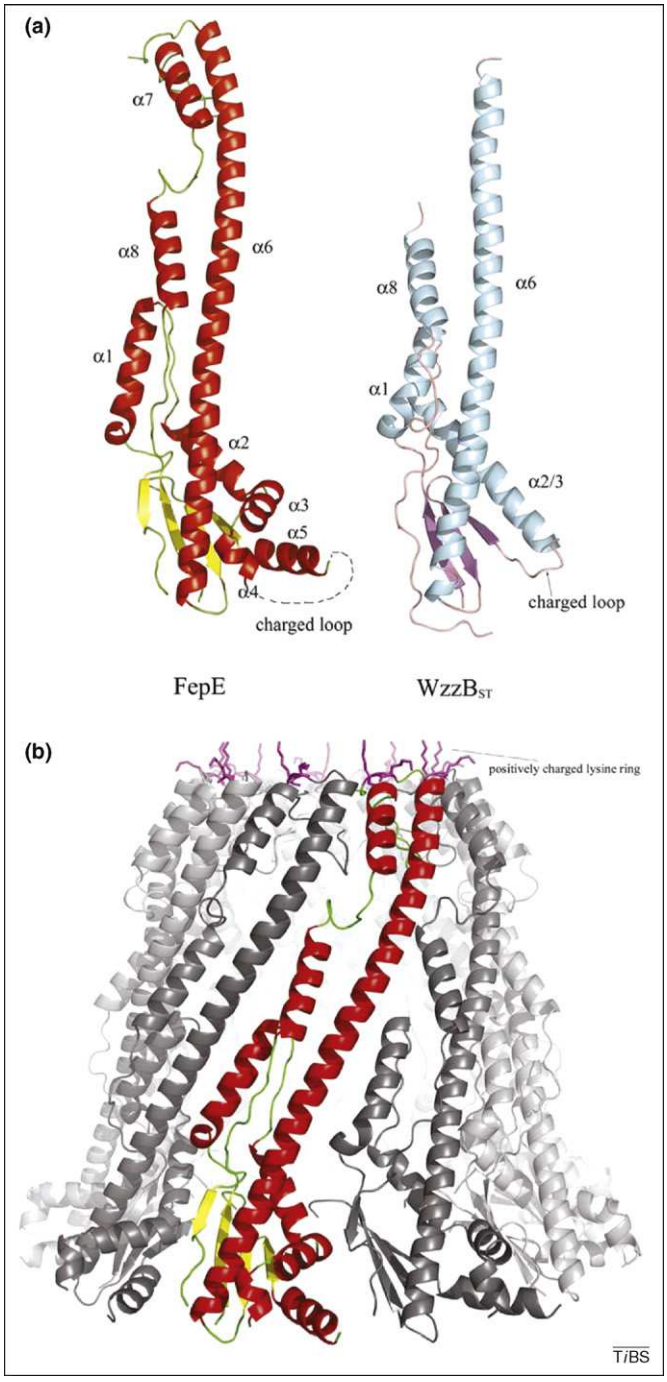

Figure 1. Different views of the atomic structures of two PCP1 proteins. The 3D structure determination has revealed that the periplasmic regions of the PCP1 proteins FepE and WzzB$_{ST}$ exist as nonameric and pentameric rings, respectively. The TM1 and TM2 regions at the N and C termini that anchor these PCPs into the bacterial cytoplasmic membrane are absent from the structures. **(a)** Protomers of FepE and WzzB$_{ST}$ side by side illustrating structural differences within the α-β domain. The FepE protomer is colored red (α helices), yellow (β strands) and green (loops), whereas the WzzB protomer is colored cyan (α helices), magenta (β strands) and orange (loops). The structure of the charged loop in FepE is not resolved and is indicated by a dashed line. **(b)** Structure of the FepE α-β oligomer with one protomer shown in color [as in part (a)] and the remaining ones in gray. The lysine ring at the top of the protein is shown in magenta.

extends ~100 Å away from the membrane and is completed by two additional helices, α7 and α8, which fold back down towards the membrane (Figure 1a). These α-helices together adopt a slight right-handed twist (Figure 1b). A group of conserved, hydrophobic residues at either end of the membrane distal loop connecting the α-helical regions interact to stabilize the hairpin in an anti-parallel coiled-coil structure involving α6 and α7 [23]. The distal loop and the following helix are disordered in the structures of WzzE and WzzB$_{ST}$ but, in FepE, the majority of the distal loop was structurally resolved and shown to lie within the top of the bell in the oligomer, resembling an internal plug. Interestingly, the well-ordered part of the distal loop lies flush with the top of the bell except for two lysine sidechains in FepE, Lys253 and Lys254, which protrude upwards creating a ring of positive charge at the top of the bell (Figure 1b). This feature of protruding charged residues at the top of the oligomer might also occur in WzzB$_{EC}$ and WzzE, which have either a lysine (Lys214) or an arginine (Arg242), respectively, in the same position as the second lysine in FepE. WzzB$_{ST}$ has a polar, conserved glutamine (Gln214) in this position; however, a lysine residue (Lys211) is found nearby. The structure of the C-terminal region of the periplasmic domain is highly conserved between all three crystal structures and each shows a propensity to reach at least 90 Å into the periplasm [23]. We speculate that the top of the oligomer has a functional role and might participate in interactions with a partner in the outer membrane, much like the periplasmic contacts between Wzc and Wza [22].

The N-terminal end of the periplasmic segment has a well-conserved central three-stranded β-sheet, whereas the α-helices vary in number and length. These differences are in the periphery of the domain and do not impact the overall fold, because the FepE, WzzE and WzzB$_{ST}$ protomers exhibit a high degree of similarity in their α-β base domain [23]. However, differences in the α-β domains impact on oligomerization because it has been observed that the N-terminal α-β domain has an important role in the inter-subunit interactions observed in the crystal structure, especially for WzzB$_{ST,}$ where there are almost no interactions between the tops of the protomers [23].

Structural predictions for other PCP subfamilies

To explore wider structural similarities within the PCP family [20], we have used secondary structure predictions (Figure 2) to analyze representatives of each subfamily and compare them with the atomic resolution structures of PCP1 proteins (Figure 1). *E. coli* K-12 Wzc and *Sinorhizobium meliloti* ExoP were chosen as representatives of the PCP2a subfamily. We have also analyzed CpsC from *Streptococcus pneumoniae* as a member of the PCP2b subfamily and KpsE from *E. coli* as the representative of the PCP3 subfamily.

Independent secondary structure predictions for FepE, WzzE and WzzB$_{ST}$ closely match the actual crystal structures (data not shown). Likewise, the predicted secondary structures of several other PCP1 proteins, including Wzz$_{pHS2}$ [17,25] (62% identity to FepE) and WzzB$_{SF}$ (*S. flexneri wzz* [10]; 72% identity to WzzB$_{ST}$), correlate well with the FepE and WzzB$_{ST}$ crystal structures (Figure 2). These results increased our confidence that the predictions for other PCP members would enable us to infer meaningful comparisons of their secondary structural features to members of the PCP1 subfamily and, in particular, to

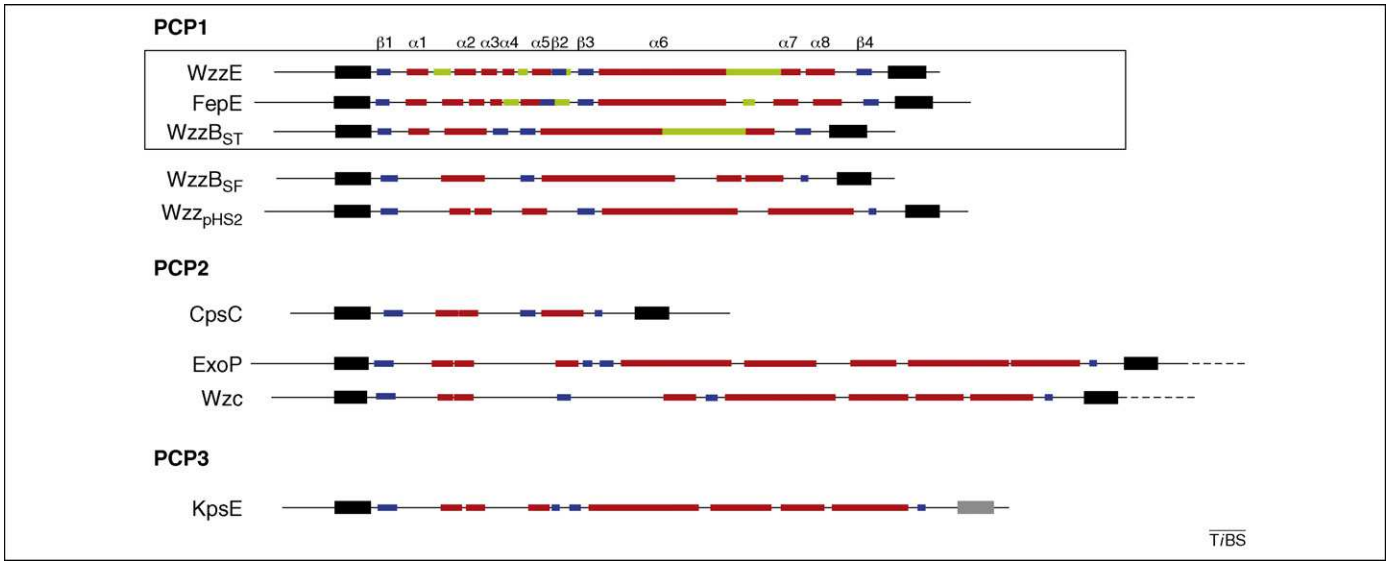

Figure 2. Secondary structure of PCP proteins in comparison to known PCP1 structures. Schematic comparison of the predicted secondary structures of selected PCP proteins with the actual secondary structures of the PCP1 proteins FepE, WzzE and WzzB$_{ST}$. Complete ORFs (open reading frames) were submitted to the Predict Protein server (http://www.predictprotein.org/) and the output from PHDhtm (Predict transmembrane) and PROF (Predict secondary structure) used to predict TM regions and secondary structure features, respectively. Manual adjustment was used where needed to align secondary structure features that were predicted with <80% probability. The structures of WzzE, FepE and WzzB$_{ST}$ (boxed) are known; the depicted secondary structures are based on these crystal structures [23]. Secondary structure color code: length of protein, black line; TM, black rectangle; TM-associated region in KpsE, grey rectangle; β strands, blue; α helices, red; regions not resolved in crystal structures, green. All proteins are shown in their entirety except for ExoP and Wzc, where only the N-terminal periplasmic domain is shown and the C-terminal tyrosine kinase domain is indicated by a dashed line. National Centre for Biotechnology Information (NCBI) accession numbers and lengths of each protein are as follows: Wzc, P76387, 450 aa (N-terminal periplasmic domain only); ExoP AAA160042, 489 aa (N-terminal periplasmic domain only); CpsC AF094575, 231 aa; WzzE, P0AG01 349 aa; WzzB$_{ST}$ QO4866, 327 aa; FepE, NP_286314, 377 aa; WzzB$_{SF}$ P37792, 325 aa; Wzz$_{pHS2}$, EU220028, 369 aa; KpsE, P62586, 382 aa.

discover regions that are structurally conserved across the entire PCP family.

The structure of the related PCP2a protein, Wzc from *E. coli* K30, has been determined at low resolution (~14 Å) by cryo- electron microscopy (CEM) [21], enabling us to make some comparisons with the Wzz structures. Wzc is an inner membrane protein that is required in *E. coli* for the biosynthesis of high molecular weight group 1 polysaccharides in a process coupled to export via contacts with the OMP Wza [4]. Single particle averaging of Wzc, phosphorylated on the cytoplasmic domain, from CEM images indicate that it forms a tetramer, with contacts between the periplasmic domains forming a nearly sealed chamber above the inner membrane [4]. Unlike Wzz proteins, Wzc has a cytoplasmic C-terminal domain with tyrosine kinase activity and an ATP-binding motif [26,27]. A structurally important difference between Wzc and Wzz proteins is that Wzz protrudes ~100 Å into the periplasm [23], whereas Wzc seems to protrude, at most, 45 Å into the periplasm [21,22] despite its longer periplasmic domain of 376 residues, compared with the 277 residues of the FepE periplasmic domain. The size of the Wzc periplasmic domain is sufficient, however, to form a periplasm-spanning complex with Wza, a protein that reaches down into the periplasm ~100 Å from the outer membrane [22,28]. Secondary structure predictions for the Wzc periplasmic domain indicate the presence of β-strands at similar topological locations as in Wzz proteins: β1 at the N-terminal end of the periplasmic region distal to TM1, a β2–β3 hairpin followed by four long helices and, finally, β4 at the C terminus of this region, proximal to TM2. However, between β2 and β3, several additional β-strands are predicted (albeit with low certainty). It is, therefore, possible that Wzc has a more extended β-sheet within the core of the base domain. The long helices that correspond topologically to the extended helical domain in Wzz proteins are predicted to be <60 residues long in Wzc, in comparison with the ~70 residue-long helix α6 in FepE. Despite very low sequence identity between the Wzc and Wzz proteins, it seems that Wzc folds in a somewhat similar manner to that of Wzz, with an elongated α-helical domain and a compact base domain, possibly having a more extended β-sheet within its core. If this is the case, we would expect that the α-helices in the Wzc periplasmic domain are highly inclined towards the membrane because the length of the longest helix, ~75 Å, is much greater than the 45 Å distance that Wzc projects into the periplasm, based on CEM analyses [21,22].

ExoP, which is involved in the production of high molecular weight exopolysaccharides, is classified as a PCP2a protein [29,30]. This protein contains a C-terminal domain with tyrosine kinase activity, showing clear sequence similarity to the analogous domain in Wzc. In both cases, secondary structure predictions indicate the presence of at least four β-strands interspersed with α-helices. The first and last β-strands are near the N and C termini of the periplasmic segment, with the overall topology indicating a fold similar to the base domain of PCP1 proteins. After strand β3, there are four long helices, with the last one being ~90 residues long. The helical domain might, therefore, differ somewhat from that found in PCP1 proteins. The four helices could form an up-and-down four helix bundle, with the longest helix measuring ~125 Å. Another difference with the PCP1 group is the very short predicted connections between these helices compared with the long, flexible loops in PCP1 proteins.

CpsC, a member of the PCP2b subgroup, is associated with CPS biosynthesis in *Streptococcus pneumoniae* [31,32]. In contrast to Wzc, the CpsC domain located between the two TMs is much shorter, a feature that can be attributed to its lack of a predicted, extended α-helical domain at the C-terminal end of the molecule (Figure 2). Predictions for CpsC clearly reveal four β-strands, with several short helices between strands β1 and β2. Instead of several long α-helices forming a hairpin after strand β3, CpsC contains one short α-helix followed by strand β4. Therefore, the arrangement of secondary structure elements indicates that CpsC has a fold similar to the base domain of PCP1 proteins, with a rather short α-helical hairpin instead of the long α-helical domain of PCP1s. CpsC is found in Gram-positive organisms that do not face the challenge of exporting polysaccharides across a second membrane, perhaps accounting for its lack of a domain protruding far from the membrane, as previously noted [20,33].

PCP2a and PCP2b proteins either harbor or enhance tyrosine kinase activity, respectively. Because PCP2-mediated phosphorylation is known to be involved in other cellular functions, including polymyxin resistance and phosphorylation of phage integrase single-stranded DNA-binding (SSB) proteins and sigma factors (i.e. functions different from polysaccharide synthesis), the tyrosine kinases of the two classes of PCPs have been classified within the bacterial protein tyrosine (BY)-kinase family [34].

Recently, structural information has become available for two BY-kinases, the cytoplasmic domain of the Wzc homologue Etk, from *E. coli* [35] and *Staphylococcus aureus* CapA–CapB [36]. CapA, homologous to CpsC, belongs to the PCP2b family and is associated, rather than fused, with the BY-kinase CapB (homologous to *S. pneumoniae* CpsD). CapA shares 33% sequence identity with the kinase domain of *E. coli* Wzc across its entire cytoplasmic domain (residues 452–721) and 27% sequence identity to the Wzc middle region (residues 340–470) that includes the second TM helix. In determining the 3D structure of CapB, the 29 C-terminal residues of CapA were fused to the CapB N terminus, generating a construct that corresponds to the entire cytoplasmic domain of *E. coli* Wzc (residues 452–721). In the crystal, the Etk kinase domains pack as linear chains; however, the unphosphorylated form of the CapAB fusion protein assembles into octameric rings with a large, 50 Å hole in the center, with the kinase active site located on the external side of the ring [36]. The C-terminal segment from CapA is part of the substrate-binding site and so complex formation between CapA and CapB is required for CapB enzymatic activity. The significant sequence identity between CapB and the Wzc cytoplasmic domain indicates that the octameric form of unphosphorylated CapB and the tetrameric form of phosphorylated Wzc [21,22] might represent two different functional states of PCP2 molecules. Like CapB, the Wzc periplasmic domain could also initially associate into octamers, which would correspond well to the octameric structure of Wza [37] and, upon auto-phosphorylation, partially dissociate to the tetrameric form [21,22]. The self-association of the CapAB fusion protein into a large ring structure corroborates well with the observed ring-like structures of PCP1 proteins [23].

The third group within the PCP family, the PCP3 proteins, is represented here by KpsE. The PCP3 family is associated with capsule biosynthesis using the ABC-2 type transporter biosynthesis pathway. KpsE is part of a protein complex responsible for group 2 capsule biosynthesis in *E. coli* [4,38]. Two TM helices are predicted within this 382 residue-long protein that contains a ~310 residue-long periplasmic region. The second predicted TM contains a lysine residue in its C-terminal half and was shown to be associated with the periplasmic face of the inner membrane rather than traversing the membrane [39]. Secondary structure predictions for KpsE reveal that it also shows similarities to the crystal structures of the PCP1 group, having four β-strands with similar topology to the PCP1 base domain. Four long α-helices follow strand β3, probably forming a domain similar to the α-helical hairpin of PCP1 proteins.

Overall, the striking conservation of secondary structure topology of the PCP proteins indicates that, despite their different lengths, all PCPs adopt similar protomer folds, supporting our previous hypothesis that the structures of FepE, WzzE and WzzB are representative of other PCPs [23]. In particular, the four (or more)-stranded β-sheet forming the central feature of the base domain is predicted in all of these proteins. With the exception of CpsC, all other sequences show the presence of an extended region containing long α-helical hairpins in the C-terminal half of the periplasmic segment that, like in PCP1s, are located between strands β3 and β4. In the PCP1 crystal structures, the β-strands from each end of the periplasmic domain have a key role in defining the tertiary structure by contributing to a single β-sheet, bringing the N and C termini and, presumably, the TM segments in close proximity to each other.

Current model for Wzy-dependent Oag biosynthesis

The recent structural data obtained for Wzz proteins have prompted the proposal of a new model for Wzy-dependent Oag biosynthesis [23]. In this model, a Wzz oligomer serves as a scaffold for Wzy molecules to facilitate Oag polymerization by a mechanism involving the transfer of the growing chain from one Wzy molecule to the next (Figure 3). Different modal lengths would result from the different Wzz protein oligomer sizes, which in turn would determine the number of neighboring Wzy molecules available for continuing polymerization. One of the many unanswered questions pertinent to the model is whether Wzz interacts directly with Oag to regulate chain length. *In vitro* studies have shown an interaction between Oag with Wzz, as determined by changes in Wzz circular dichroism spectra interpreted as conformational changes in Wzz in the presence of Oag [17]. The structure, however, does not provide an obvious oligosaccharide-binding site for this interaction to occur and, furthermore, PCP1a proteins lack specificity for Oag structure [11,40], thereby making the existence of specific binding interactions uncertain. The long length of the periplasmic α-helical hairpin makes it tempting to speculate that, in addition to its putative role in coordinating Wzy molecules,

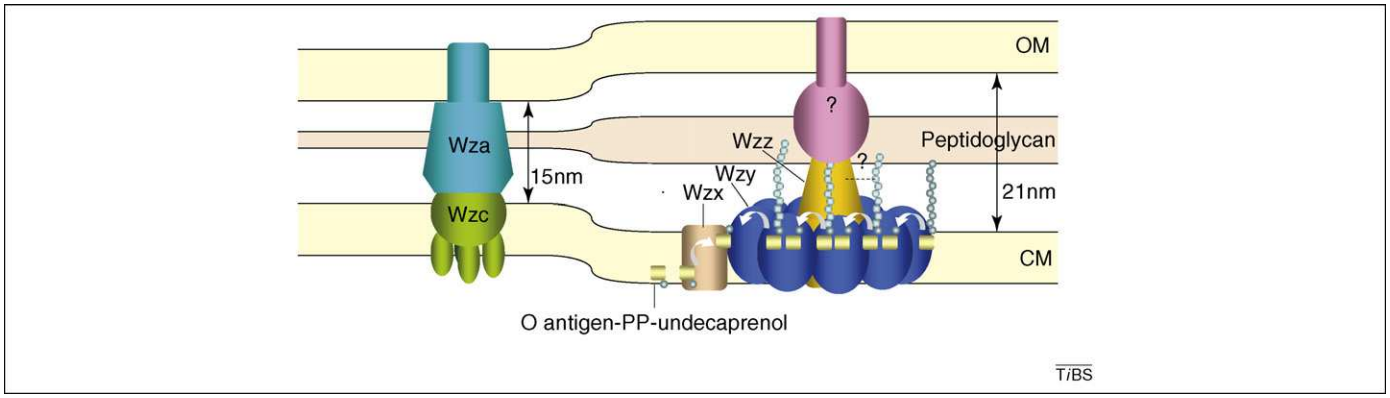

Figure 3. Model for Wzz function in O antigen biosynthesis. A proposed model for the participation of Wzz oligomers in Oag biosynthesis, in comparison to the Wzc–Wza complex, is illustrated. On the right hand side, a Wzz oligomer is shown in a scale representation of a Gram negative cell envelope. The dimensions are based on CEM of *E. coli* K-12 [41]. On the left hand side, a model of Wza–Wzc is shown for comparison to Wzz, revealing the smaller periplasmic space that would result from contact between Wza and Wzc [22]. Individual sugar subunits (pale blue circles), linked to the lipid carrier undecaprenol pyrophosphate (pale yellow squares), are flipped to the periplasmic side of the cell membrane (CM) by Wzx (beige). The Wzz oligomer (orange) is shown at the centre of a putative complex with Wzy (blue) whereby individual Wzy proteins might be organized into a ring by the Wzz oligomer. The subunits are putatively assembled by Wzy in a processive manner with the newly lengthened chain passed to a waiting Wzy molecule that can add the growing chain to a new undecaprenol pyrophosphate-linked subunit, after which it is passed to the next Wzy molecule and so on. A growing polymer chain is shown in transparent mode. A putative OM protein (or complex; pink) is shown interacting with Wzz. A possible interaction between Oag and Wzz is indicated by a double-headed arrow and question mark.

Wzz couples Oag polymerization to export via contact with an outer membrane protein, as is the case for Wzc–Wza [22].

Concluding remarks and future perspectives

The Wzz proteins provide a striking example of how a common structural template can control the synthesis or length of different polymers, while at the same time promote differences in protein quaternary structure. These PCP1 structures shed a broad light on the entire PCP family, revealing, to a large extent, a similar structural framework in other PCPs. The nature of oligomerization in other PCP families and the functional importance of oligomer formation are key questions that will require further work (Box 1). The crystallization of Wzz constitutes a major advance towards our goal of understanding Wzy-dependent polysaccharide biosynthesis; indeed, it opens the doorway for functional analyses and structure determination of rationally designed mutants. Further progress in our understanding of the mode of action of PCPs will require assignment of functional roles to different parts of the PCP protomer structure. The biosynthesis of cell surface polysaccharides that involves proteins from the PCP2 group awaits the structure determination of at least one representative member. The relationship between the phosphorylation state of the cytoplasmic tyrosine kinase domain, oligomer conformation and, in the case of PCP2a proteins, the interaction of their periplasm regions with outer membrane proteins (e.g. Wza) also need to be explored using structure determination. In addition, the spatial relationship between PCPs, polysaccharide polymerases, lipid A ligases, flippases and other biosynthesis proteins awaits resolution.

Box 1. Outstanding questions

- Do PCP1 proteins interact with an outer membrane protein as observed for PCP2a proteins?
- Do PCP1 proteins have a cytoplasmic component as seen in PCP2 and PCP3 proteins?
- What are the physical interactions between the PCP proteins and other components of the Wzy-dependent polysaccharide biosynthesis machinery?
- How do PCPs recognize the length of the synthesized polymer?
- What is the role of the PCP TM regions in oligomer formation?
- How can a Wzz PCP1 protein function in different bacteria to control the modal chain length of structurally (or serologically) distinct O antigen polysaccharides?
- How can a given PCP1 protein function with different Wzy proteins given that the latter have little sequence similarity?

Acknowledgements

Our work is supported in part by the National Research Council of Canada (to M.C. and A.M.), the Canadian Institutes of Health Research (CIHR) grant MOP-48370 (to M.C.) and by an Australian National Health and Medical Research Council Program Grant (to R.M.). This is publication number 49562 from the National Research Council of Canada.

References

1 Barr, K. *et al.* (1999) The modality of enterobacterial common antigen polysaccharide chain lengths is regulated by o349 of the *wec* gene cluster of *Escherichia coli* K-12. *J. Bacteriol.* 181, 6564–6568
2 Bentley, S.D. *et al.* (2006) Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. *PLoS Genet.* 2, e31
3 Raetz, C.R. and Whitfield, C. (2002) Lipopolysaccharide endotoxins. *Annu. Rev. Biochem.* 71, 635–700
4 Whitfield, C. (2006) Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. *Annu. Rev. Biochem.* 75, 39–68
5 Marolda, C.L. *et al.* (2006) Interplay of the Wzx translocase and the corresponding polymerase and chain length regulator proteins in the translocation and periplasmic assembly of lipopolysaccharide O antigen. *J. Bacteriol.* 188, 5124–5135
6 Hong, M. and Payne, S.M. (1997) Effect of mutations in *Shigella flexneri* chromosomal and plasmid-encoded lipopolysaccharide genes on invasion and serum resistance. *Mol. Microbiol.* 24, 779–791
7 Murray, G.L. *et al.* (2006) Altering the length of the lipopolysaccharide O antigen has an impact on the interaction of *Salmonella enterica* serovar Typhimurium with macrophages and complement. *J. Bacteriol.* 188, 2735–2739
8 Van Den Bosch, L. *et al.* (1997) Regulation of O-antigen chain length is required for *Shigella flexneri* virulence. *Mol. Microbiol.* 23, 765–775
9 Bastin, D.A. *et al.* (1993) Repeat unit polysaccharides of bacteria: a model for polymerization resembling that of ribosomes and fatty acid synthetase, with a novel mechanism for determining chain length. *Mol. Microbiol.* 7, 725–734

10 Morona, R. *et al.* (1995) Molecular, genetic, and topological characterization of O-antigen chain length regulation in *Shigella flexneri*. *J. Bacteriol.* 177, 1059–1068
11 Burrows, L.L. *et al.* (1997) *Pseudomonas aeruginosa* B-band O-antigen chain length is modulated by Wzz (Ro1). *J. Bacteriol.* 179, 1482–1489
12 Daniels, C. and Morona, R. (1999) Analysis of *Shigella flexneri* Wzz (Rol) function by mutagenesis and cross-linking: Wzz is able to oligomerize. *Mol. Microbiol.* 34, 181–194
13 Franco, A.V. *et al.* (1998) The Wzz (Cld) protein in *Escherichia coli*: amino acid sequence variation determines O-antigen chain length specificity. *J. Bacteriol.* 180, 2670–2675
14 Klee, S.R. *et al.* (1997) Influence of different *rol* gene products on the chain length of *Shigella dysenteriae* type 1 lipopolysaccharide O antigen expressed by *Shigella flexneri* carrier strains. *J. Bacteriol.* 179, 2421–2425
15 Murray, G.L. *et al.* (2005) Inducible serum resistance in *Salmonella typhimurium* is dependent on *wzz*(*fepE*)-regulated very long O antigen chains. *Microbes Infect.* 7, 1296–1304
16 Marolda, C.L. *et al.* (2008) Functional analysis of predicted coiled coil regions in the *Escherichia coli* K-12 O antigen polysaccharide chain length determinant Wzz. *J. Bacteriol.* 190, 2128–2137
17 Purins, L. *et al.* (2008) Coiled-coil regions play a role in the function of the *Shigella flexneri* O-antigen chain length regulator WzzpHS2. *Microbiology* 154, 1104–1116
18 Guo, H. *et al.* (2006) Overexpression and characterization of Wzz of *Escherichia coli* O86:H2. *Protein Expr. Purif.* 48, 49–55
19 Tang, K.H. *et al.* (2007) Investigation of the conformational states of Wzz and the Wzz.O-antigen complex under near-physiological conditions. *Biochemistry* 46, 11744–11752
20 Morona, R. *et al.* (2000) Evaluation of Wzz/MPA1/MPA2 proteins based on the presence of coiled-coil regions. *Microbiology* 146, 1–4
21 Collins, R.F. *et al.* (2006) Periplasmic protein-protein contacts in the inner membrane protein Wzc form a tetrameric complex required for the assembly of *Escherichia coli* group 1 capsules. *J. Biol. Chem.* 281, 2144–2150
22 Collins, R.F. *et al.* (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
23 Tocilj, A. *et al.* (2008) Bacterial polysaccharide co-polymerases share a common framework for control of polymer length. *Nat. Struct. Mol. Biol.* 15, 130–138
24 Soulat, D. *et al.* (2006) *Staphylococcus aureus* operates protein-tyrosine phosphorylation through a specific mechanism. *J. Biol. Chem.* 281, 14048–14056
25 Stevenson, G. *et al.* (1995) A plasmid-borne O-antigen chain length determinant and its relationship to other chain length determinants. *FEMS Microbiol. Lett.* 125, 23–30
26 Paiment, A. *et al.* (2002) Impact of phosphorylation of specific residues in the tyrosine autokinase, Wzc, on its activity in assembly of group 1 capsules in *Escherichia coli*. *J. Bacteriol.* 184, 6437–6447
27 Wugeditsch, T. *et al.* (2001) Phosphorylation of Wzc, a tyrosine autokinase, is essential for assembly of group 1 capsular polysaccharides in *Escherichia coli*. *J. Biol. Chem.* 276, 2361–2371
28 Nesper, J. *et al.* (2003) Translocation of group 1 capsular polysaccharide in *Escherichia coli* serotype K30. Structural and functional analysis of the outer membrane lipoprotein Wza. *J. Biol. Chem.* 278, 49763–49772
29 Glucksmann, M.A. *et al.* (1993) Genes needed for the modification, polymerization, export, and processing of succinoglycan by *Rhizobium meliloti*: a model for succinoglycan biosynthesis. *J. Bacteriol.* 175, 7045–7055
30 Niemeyer, D. and Becker, A. (2001) The molecular weight distribution of succinoglycan produced by *Sinorhizobium meliloti* is influenced by specific tyrosine phosphorylation and ATPase activity of the cytoplasmic domain of the ExoP protein. *J. Bacteriol.* 183, 5163–5170
31 Morona, J.K. *et al.* (1999) Comparative genetics of capsular polysaccharide biosynthesis in *Streptococcus pneumoniae* types belonging to serogroup 19. *J. Bacteriol.* 181, 5355–5364
32 Morona, J.K. *et al.* (1999) Analysis of the 5′ portion of the type 19A capsule locus identifies two classes of cpsC, cpsD, and cpsE genes in *Streptococcus pneumoniae*. *J. Bacteriol.* 181, 3599–3605
33 Reid, A.N. and Whitfield, C. (2005) Functional analysis of conserved gene products involved in assembly of *Escherichia coli* capsules and exopolysaccharides: evidence for molecular recognition between Wza and Wzc for colanic acid biosynthesis. *J. Bacteriol.* 187, 5470–5481
34 Grangeasse, C. *et al.* (2007) Tyrosine phosphorylation: an emerging regulatory device of bacterial physiology. *Trends Biochem. Sci.* 32, 86–94
35 Lee, D.C. *et al.* (2008) Structure of *Escherichia coli* tyrosine kinase Etk reveals a novel activation mechanism. *EMBO J.* 27, 1758–1766
36 Olivares-Illana, V. *et al.* (2008) Structural basis for the regulation mechanism of the tyrosine kinase CapB from *Staphylococcus aureus*. *PLoS Biol.* 6, e143
37 Dong, C. *et al.* (2006) Wza the translocon for *E. coli* capsular polysaccharides defines a new class of membrane protein. *Nature* 444, 226–229
38 McNulty, C. *et al.* (2006) The cell surface expression of group 2 capsular polysaccharides in *Escherichia coli*: the role of KpsD, RhsA and a multi-protein complex at the pole of the cell. *Mol. Microbiol.* 59, 907–922
39 Arrecubieta, C. *et al.* (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
40 Batchelor, R.A. *et al.* (1992) Nucleotide sequences of the genes regulating O-polysaccharide antigen chain length (*rol*) from *Escherichia coli* and *Salmonella typhimurium*: protein homology and functional complementation. *J. Bacteriol.* 174, 5228–5236
41 Matias, V.R. *et al.* (2003) Cryo-transmission electron microscopy of frozen-hydrated sections of *Escherichia coli* and *Pseudomonas aeruginosa*. *J. Bacteriol.* 185, 6112–6118