

## MIT Open Access Articles

*Structural molecular modeling of bacterial integral membrane protein enzymes and their AlphaFold2 predicted water-soluble QTY variants*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** Sajeev-Sheeja, A., Zhang, S. Structural molecular modeling of bacterial integral membrane protein enzymes and their AlphaFold2 predicted water-soluble QTY variants. *J Proteins Proteom* (2024).

**As Published:** <https://doi.org/10.1007/s42485-024-00170-8>

**Publisher:** Springer Nature Singapore

**Persistent URL:** <https://hdl.handle.net/1721.1/157745>

**Version:** Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

**Terms of use:** Creative Commons Attribution





# Structural molecular modeling of bacterial integral membrane protein enzymes and their AlphaFold2 predicted water-soluble QTY variants

Akash Sajeev-Sheeja<sup>1</sup> · Shuguang Zhang<sup>2</sup>

Received: 17 June 2024 / Revised: 31 August 2024 / Accepted: 24 October 2024  
© The Author(s) 2024

## Abstract

**Context** Beta-barrel enzymes are an important area of study in the field of structural biology. These proteins serve crucial roles, acting as porins, transporters, enzymes, virulence factors, and receptors. Recent research has unveiled a novel role for beta-barrel enzymes in the bacterial integral membrane as sentinels. They remain inactive when the integral membrane is intact but activate to carry out enzymatic catalysis in response to host immune responses and antibiotics that breach this barrier. Understanding their structure and function is pivotal in grasping their sentinel role in the bacterial integral membrane. Here we present our structural molecular modeling analyses on four bacterial integral membrane beta-barrel enzymes: (a) OMPLA, (b) OmpT, (c) PagP from *E. coli*, and (d) PagL from *Pseudomonas aeruginosa*. We superposed the structures of native beta-barrel integral membrane enzymes with their AlphaFold2-predicted QTY variant structures that showed remarkable similarity despite the replacement of at least 22.95% amino acids in transmembrane regions, the superposed structures displayed notable structural similarity, indicated by RMSD values ranging from 0.181 Å to 0.286 Å. We also analyze the hydrophobicity patches and the enhanced hydrophilic surfaces. Our research provide insights into the structural similarity of hydrophobic and hydrophilic beta-barrel enzymes, validating the utility of the QTY code for investigating beta-barrel membrane enzymes. Our results not only demonstrate that the QTY code serves as a straightforward tool for designing water-soluble membrane proteins across various biological contexts, but it may also stimulate experiments to validate our molecular modeling studies.

**Methods** All the QTY variant beta-barrel enzyme structure prediction was performed using the AlphaFold2 program (<https://github.com/sokrypton/ColabFold>) following the provided instructions. Computations were carried out on 11th Gen Intel Core i5-11300H processor with 16 GB RAM and Iris Xe Graphics, 512 GB NVMe SSD. The structures are publicly available on the AlphaFold2 database (<https://alphafold.ebi.ac.uk>) at the European Bioinformatics Institute (EBI). A custom Python script was used to extract the relevant information from the UniProt database. To predict the structures of the QTY variants, AlphaFold2 was utilized. The native sequences for these enzymes were retrieved from UniProt <https://www.uniprot.org>, and AlphaFold2 structural predictions were performed using the open-source implementation at <https://github.com/sokrypton/ColabFold>. The predicted variant structures were then superposed with the native structures using PyMOL <https://pymol.org/2/> for structural analysis and comparison. This work leverages public databases PDB, UniProt and open-source software AlphaFold2 and PyMOL to computationally model and analyze QTY variant integral membrane beta-barrel enzyme structures.

---

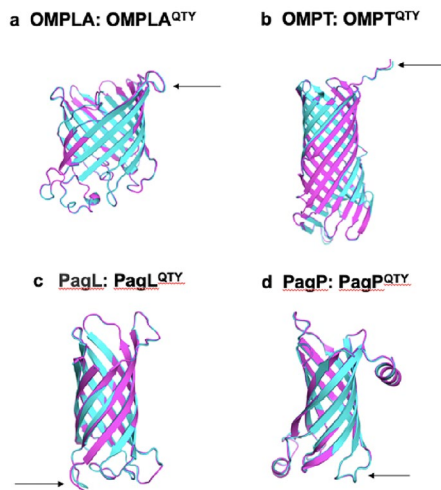
✉ Shuguang Zhang  
Shuguang@MIT.EDU

Akash Sajeev-Sheeja  
akashadarshss@gmail.com

<sup>1</sup> Department of Chemistry, Indian Institute of Science Education and Research, Tirupati, Srinivasapuram, Yerpedu Mandal, Tirupati Dist, Andhra Pradesh 517619, India

<sup>2</sup> Laboratory of Molecular Architecture, Media Lab, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

## Graphical abstract



**Superposed images of four bacterial native integral membrane beta-barrel enzymes and their AlphaFold2-predicted water-soluble QTY variants.**

The crystal structures of the native integral membrane beta-barrel enzymes (colored magenta) and the predicted structures of their water-soluble QTY variants generated by AlphaFold2 (colored cyan) are depicted.

**Keywords** Hydrophobic to hydrophilic conversion · Membrane enzymes · Protein design · QTY code · Water-insoluble to water-soluble

## Introduction

Outer membrane proteins (OMPs) represent an important class of proteins located in the outer membrane of gram-negative bacteria, mitochondria, and chloroplasts that carry out vital biological functions (Zhang et al. 2014; Xu et al. 2023). The integral membrane possesses an asymmetric phospholipid bilayer, with phospholipids occupying the inner leaflet and lipopolysaccharide (LPS) glycolipids constituting the integral leaflet (Sun et al. 2022; Funahara and Nikaido 1980). The proteins found in the integral membrane fall into two categories: lipoproteins and beta-barrel proteins. Lipoproteins possess lipid portions that are linked to a cysteine residue at the amino end of the protein (Sankaran and Wu 1994), these proteins are not thought to be transmembrane (TM) proteins (Silhavy et al. 2010). Nearly all integral transmembrane proteins form  $\beta$ -barrel structures (Silhavy et al. 2010; Hermansen et al. 2022). These integral membrane proteins (IMPs) are found in the integral membranes of Gram-negative bacteria and exhibit antiparallel beta-barrel structures. Despite their prevalence, very few of these proteins function as enzymes (Bishop 2008). In addition, they have been adapted to perform a variety of essential cellular functions such as porins, transporters, enzymes, virulence factors, and receptors (Fairman 2011).

Recent research has demonstrated that beta-barrel enzymes also serve as sentinels, remaining dormant when the integral membrane permeability barrier is intact, but being activated by host immune defenses and antibiotics

that disrupt this barrier. The integral membrane is the first target of external assault but it receives the least support from cellular utilities, making it susceptible to irreversible damage. It is essential to comprehend the structure and function of beta-barrel enzymes and their role as sentinels in the bacterial integral membrane (Bishop 2008). Beta-barrel enzymes are crucial in how pathogens induce disease. Understanding how beta-barrel enzymes function could aid in developing new treatments for bacterial infections (Fairman 2011).

We previously applied the QTY code (glutamine, threonine, tyrosine) (Zhang et al. 2018; Zhang and Egli 2022) for designs of detergent-free beta-barrel, alpha-helix transmembrane (TM) protein chemokine receptors, cytokine receptors and antibodies for various uses using conventional computing programs (Zhang et al. 2018; Sajeev-Sheeja et al. 2023; Qing et al. 2019; Hao et al. 2020; Li et al. 2023). QTY code is a bidirectional simple molecular code which systematically replacing water-insoluble amino acids (L, V, I and F) with water-soluble amino acids (Q, T and Y) in the transmembrane region (Zhang et al. 2018; Sajeev-Sheeja et al. 2023; Meng et al. 2023). The QTY code is based on two key similarities between amino acids: (1) some amino acids have very similar structures despite stark differences in hydrophobicity (e.g., Leu vs Gln/Asn, Ile/Val vs Thr, Phe vs Tyr), and (2) these structurally similar amino acids tend to have similar secondary structure propensities. The analyses using the highly accurate protein structure prediction tool AlphaFold2 showed that the structures of QTY variants closely overlay the

native protein structures. This demonstrates that the QTY variants retain the key structural features of the original native proteins (Sajeev-Sheela et al. 2023; Li et al. 2023; Skuhersky et al. 2021; Tao et al. 2022; Smorodina et al. 2022; Smorodina et al. 2022).

We here report structural bioinformatic studies of four bacterial beta-barrel integral membrane enzymes, they include a) OMPLA (a phospholipase) (Snijder et al. 1999), b) OmpT (a protease) (Vandeputte-Rutten et al. 2001), c) PagP (a phospholipid-lipid A palmitoyltransferase) (Ahn et al. 2004), and d) PagL (a lipid A 3-O-deacylase) (Rutten et al. 2006a). Among these, OMPLA, OmpT, and PagP are the only three  $\beta$ -barrel enzymes known to exist in the integral membrane of *E. coli* and have been characterized at the molecular structural level. Additionally, structural details have also been elucidated for PagL from *Pseudomonas aeruginosa* (Bishop 2008).

The outer membrane phospholipase A (OMPLA) was the first characterized integral membrane  $\beta$ -barrel enzyme in *E. coli* (Scandella and Kornberg 1971). OMPLA consists of 12 antiparallel  $\beta$ -strands that fold into a transmembrane  $\beta$ -barrel. Normally, this enzyme is inactive but dimerizes when phospholipids migrate to the integral leaflet, creating two active sites where  $\text{Ca}^{2+}$  ions can bind and activate OMPLA to restore integral membrane asymmetry by removing phospholipids (Bishop 2008; Dekker et al. 1997; Dekker 2000). Its activation helps cells respond to assaults on the integral membrane (Michel and Stárka 1979; Audet et al. 1974). Some pathogens lack OMPLA activity, while in others it promotes host interactions by inducing cell surface changes (Bishop 2008; Grant et al. 1997; Bukholm et al. 1997; Tannaes et al. 2005).

OmpT is an integral membrane protease in *E. coli* that specifically cleaves between paired basic amino acids (Sugimura and Nishihara 1988; Fiss et al. 1979). The enzyme forms a 10 antiparallel  $\beta$ -barrel strands, prominently rises above the integral surface of the bilayer (Vandeputte-Rutten et al. 2001). It helps provide resistance against antimicrobial peptides by degrading them when they penetrate the cell surface (Stumpe et al. 1998; Guina et al. 2000). OmpT activity helps overcome host defenses (Thomassin et al. 2012). Though OmpT homologs fulfill distinct functions, their conserved active site architectures classify them as omptins (Kukkonen and Korhonen 2004). Thus OmpT activity enables bacteria to evade immune system components and promote infection.

The PagP enzyme transfers palmitate from phospholipids to lipid A in the integral membrane of Gram-negative bacteria (Bishop et al. 2000; Bishop 2005). This palmitoylation provides resistance to cationic antimicrobial peptides and attenuates signaling through the host TLR4 pathway, helping bacteria evade the immune response (Bishop 2005; Guo et al. 1998; Kawasaki et al. 2004). The eight-stranded

antiparallel  $\beta$ -barrel of PagP contains a short alpha helix at its N-terminus and is tilted approximately 25 degrees within the plane of the membrane (Evanics et al. 2006). PagP homologs are important for disease causation in respiratory pathogens like *Legionella* and *Bordetella* by providing resistance to antibody-mediated lysis (Preston et al. 2003). Overall, this enzyme palmitoylation activity allows evasion of host defenses in some pathogens, but may conflict with virulence strategies in others.

PagL is an integral membrane lipid A 3-O-deacylase activated when L-Ara4N is absent from lipid A (Kawasaki et al. 2005, 2007). Its eight-stranded antiparallel tilted  $\beta$ -barrel structure has the catalytic triad facing outward to encounter lipid A (Rutten et al. 2006b). Unlike constitutive PagP expression, PagL is only active under PhoP/PhoQ regulation (Trent et al. 2001). This enzyme removes an acyl chain from lipid A, conferring polymyxin resistance and reducing endotoxicity (Kawasaki et al. 2004, 2007). PagL homologs are more common than PagP (Preston et al. 2003; Geurtsen et al. 2006). Inactivation of PagL during *P. aeruginosa* adaptation in cystic fibrosis is observed (Ernst et al. 2006). Dimerization may inhibit PagL when L-Ara4N is present. Thus it alters lipid A to promote antibiotic evasion and immune system evasion (Bishop 2008; Rutten et al. 2006a).

In this structural bioinformatic study, we used AlphaFold2, publicly released in 2021, to predict protein structures. This deep learning tool can accurately predict 3D protein structures, offering insights into function and aiding drug development (Jumper et al. 2021). In July 2022, DeepMind released a database of over 214 million AlphaFold2-predicted structures, covering almost all known proteins.

Recently, we asked if QTY could also solubilize these beta-barrel integral membrane enzymes, enabling their biological study without detergents.

Here, we report structural molecular modeling studies of the molecular structures of four experimentally determined beta-barrel integral membrane enzymes and their AlphaFold2-predicted water-soluble QTY variants. The native structures and their QTY variants of these beta barrel enzymes share remarkable structural similarities and superpose very well with residue mean-square distances (RMSD) between 0.181 Å and 0.286 Å despite the replacement of at least 22.95% transmembrane hydrophobic amino acids, L, V, I, F with Q, T, Y.

We also show that native and QTY variants have different hydrophobicity surface patches, and the QTY variants have more hydrophilic surfaces. Our study not only provides important insights into the differences between hydrophobic and designed hydrophilic beta-barrel, but also it validates the QTY code for studying beta-barrel membrane enzymes and perhaps other hydrophobic aggregated proteins. Our findings in this study demonstrate that the QTY code as a simple tool for studying beta-barrel integral membrane enzymes.

Previously we observed that QTY-variant receptors exhibited tunable ligand affinities and enhanced solubility, allowing for functional characterization without the use of detergents (Qing et al. 2019). Additionally, we found that the G protein-coupled receptor CXCR4, when designed using the QTY code, became more hydrophilic while retaining its cell signaling activity (Tegler et al. 2020). Specifically, our findings indicated that the QTY modification allowed CXCR4 to maintain its ability to bind to its ligand CXCL12 and initiate cellular signaling upon transfection into HEK293 cells. Based on our structural analysis and previous research findings with the QTY code, we believe that this modification will have significant potential impacts on protein activity.

Recently, we also demonstrated that a designed bacterial membrane protein histidine kinase using the QTY code, it became water-soluble; the histidine kinase not only retained its intact structure, but it also retained its four biological functions, exhibiting expected biophysical properties and highly preserved native molecular function, including the activities of (1) autokinase, (2) phosphotransferase, (3) phosphatase, and (4) pH and potassium signaling (Li et al. 2024).

Our structural bioinformatic study again demonstrates that the QTY code can successfully design water-soluble beta-barrel membrane enzymes, providing a simple tool to study this class of membrane enzymes without detergents. Our findings validate applying the QTY code to design water-soluble variants of beta-barrel integral membrane enzymes for functional characterization and biotechnological applications.

## Methods

### Protein sequence alignments and other characteristics

The proteins analyzed were identified on UniProt (<https://www.uniprot.org>), where protein ID, entry name, description, and FASTA sequences were retrieved. The native protein sequences of beta-barrel integral membrane enzymes and their corresponding QTY-variant sequences were aligned utilizing previously described methods. The ExPasy website (ExPasy-Compute pI/Mw tool) was used to compute the molecular weight (MW) and isoelectric point (pI) of the protein sequences.

### AlphaFold2 prediction

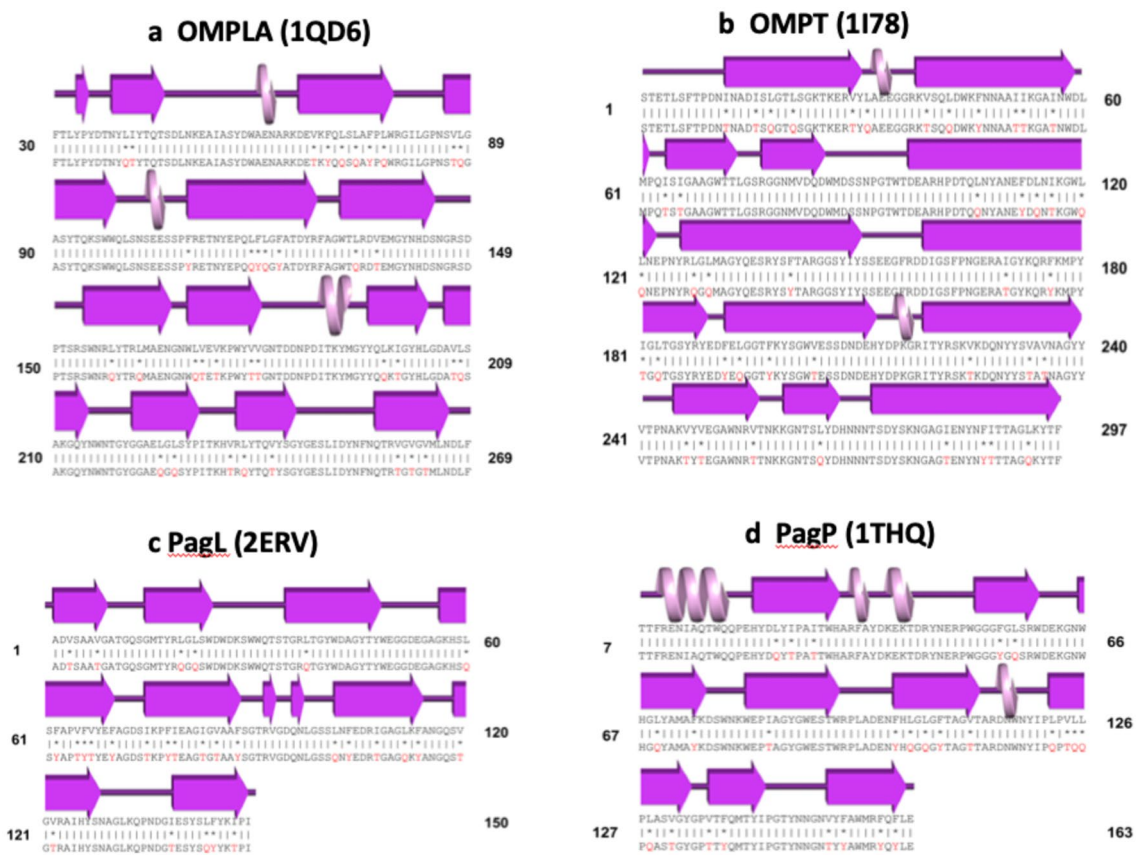
All the QTY variant beta-barrel enzyme structure prediction was performed using the AlphaFold2 program (<https://github.com/sokrypton/ColabFold>) following the provided instructions. Computations were carried out on 11th Gen Intel Core i5-11300H processor with 16 GB RAM and Iris Xe Graphics, 512 GB NVMe SSD. The structures are publicly available on the AlphaFold2 database (<https://alphafold.ebi.ac.uk>) at the European Bioinformatics Institute (EBI). A custom Python script was used to extract the relevant information from the UniProt database.

**Table 1** Characteristics of four bacterial outer membrane beta-barrel enzymes in their native form and water-soluble QTY code variants

Name	RMSD	pI	MW (kDa)	TM variations (%)	Overall variations (%)
OMPLA P0A921	–	5.01	27.69	–	–
OMPLA <sup>QTY</sup>	0.189 Å	5.01	28.01	25.53 (36/141)	15 (36/240)
OmpT P09169	–	5.38	33.46	–	–
OmpT <sup>QTY</sup>	0.240 Å	5.38	33.65	22.95 (42/183)	14.14 (42/297)
PagL Q9HVD1	–	5.01	16.09	–	–
PagL <sup>QTY</sup>	0.181 Å	5.01	16.26	30.85 (29/94)	19.33 (29/150)
PagP P37001	–	5.75	18.47	–	–
PagP <sup>QTY</sup>	0.286 Å	5.75	18.71	29.76% (25/84)	15.92 (25/157)

Note: The term “TM variations” refers to the percentage of amino acid residues that were changed within the transmembrane domains of these enzymes. The numbers in parentheses represent the percentage of altered residues in the transmembrane regions, which is calculated by dividing the number of substituted residues in the TM domains by the total number of residues that make up those domains. “Overall variations” indicates the number of amino acids that were replaced with the QTY code compared to the total count of amino acids in the full enzyme





**Fig. 1** Sequence alignments of four native integral membrane beta-barrel enzymes and their corresponding water-soluble QTY variants. Vertical lines (|) and asterisks (\*) denote identical and different amino acid residues, respectively. Note the glutamine (Q), threonine (T), and tyrosine (Y) substitutions (highlighted in red). Beta-pleated sheets (colored magenta) are displayed above the protein sequences.

## Superposed structures

The molecular structures presented in this work were obtained from the PDB <https://www.rcsb.org> and include the following beta-barrel enzymes: OMPLA (PDB: 1QD6), OmpT (PDB: 1I78), PagL (PDB: 2ERV), and PagP (PDB: 1THQ). To predict the structures of the QTY variants, AlphaFold2 was utilized. The native sequences for these enzymes were retrieved from UniProt <https://www.uniprot.org>, and AlphaFold2 structural predictions were performed using the open-source implementation at <https://github.com/sokrypton/ColabFold>. The predicted variant structures were then superposed with the native structures using PyMOL <https://pymol.org/2/> for structural analysis and comparison. This work leverages public databases PDB, UniProt and open-source software AlphaFold2 and PyMOL to computationally model and analyze QTY variant integral membrane beta-barrel enzyme structures.

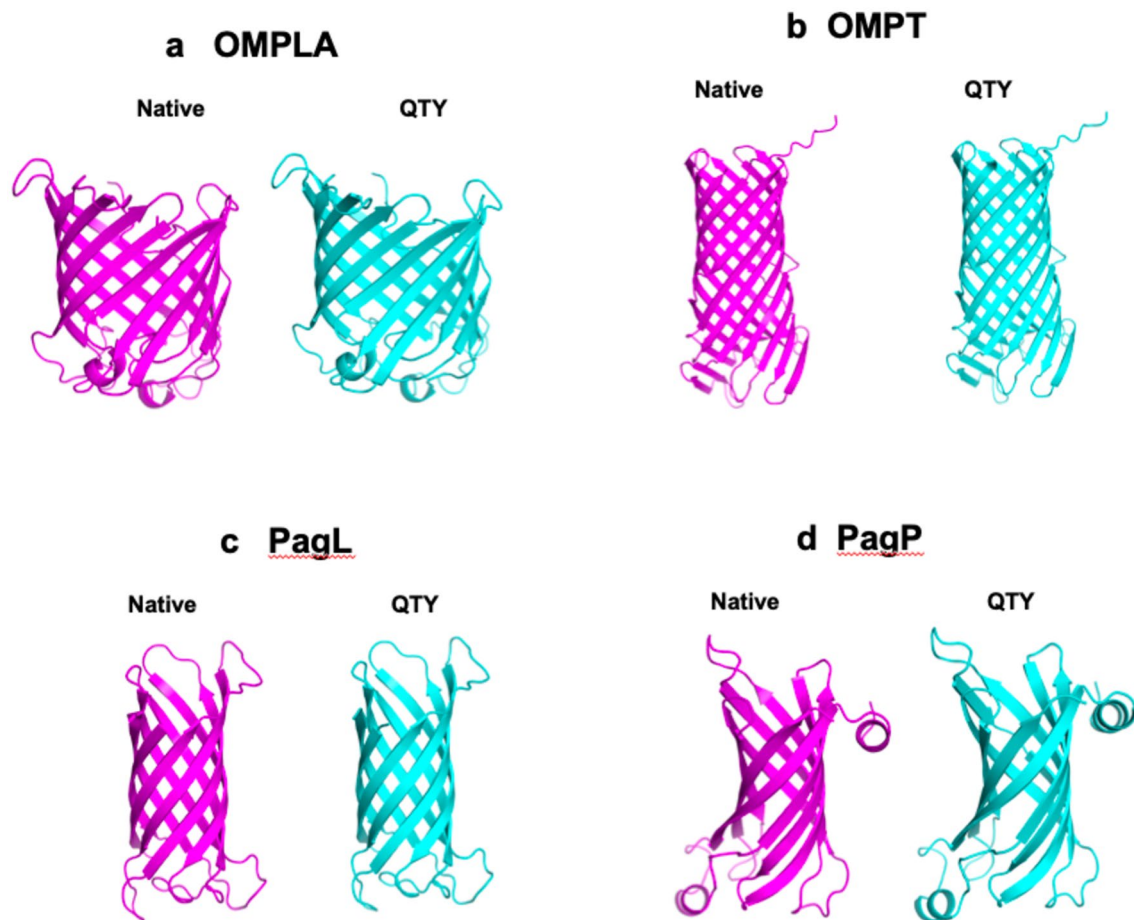
Though alpha-helical structures are present in the membrane beta-barrel enzymes, they are omitted from the QTY design modifications. The alignments shown are: **a** OMPLA vs OMPLA<sup>QTY</sup>, **b** OmpT vs OmpT<sup>QTY</sup>, **c** PagL vs PagL<sup>QTY</sup> and **d** PagP vs PagP<sup>QTY</sup>. Refer to the Supplementary Figure S1 for enlarged details

## Structure visualization

Two main software tools were used for visualizing and analyzing the predicted molecular structures. PyMOL (<https://pymol.org/2/>) and UCSF ChimeraX 1.4 (<https://www.rbvi.ucsf.edu/chimera/>). PyMOL enabled structural superposition and comparison of the native and QTY variant models. ChimeraX was utilized to compute and visualize hydrophobicity surfaces for the beta-barrel enzyme models, providing insights into the hydrophobic patches. This combined approach using PyMOL and ChimeraX allowed detailed structural analysis of the modeled mutant enzymes in relation to their native counterparts.

## Data availability

The QTY variants of these four beta-barrel enzymes including: a) OMPLA<sup>QTY</sup>, b) OmpT<sup>QTY</sup>, c) PagL<sup>QTY</sup>, and



**Fig. 2** Structures of the four bacterial native integral membrane beta-barrel enzymes and their AlphaFold2-predicted water-soluble QTY variants. The crystal structures of the native integral membrane beta-barrel enzymes are colored magenta, while the predicted structures of

the water-soluble QTY variant enzymes generated by AlphaFold2 are colored cyan. These enzymes are shown in the identical orientations. **a** OMPLA vs OMPLA<sup>QTY</sup>, **b** OmpT vs OmpT<sup>QTY</sup>, **c** PagL vs PagL<sup>QTY</sup>, and **d** PagP vs PagP<sup>QTY</sup>

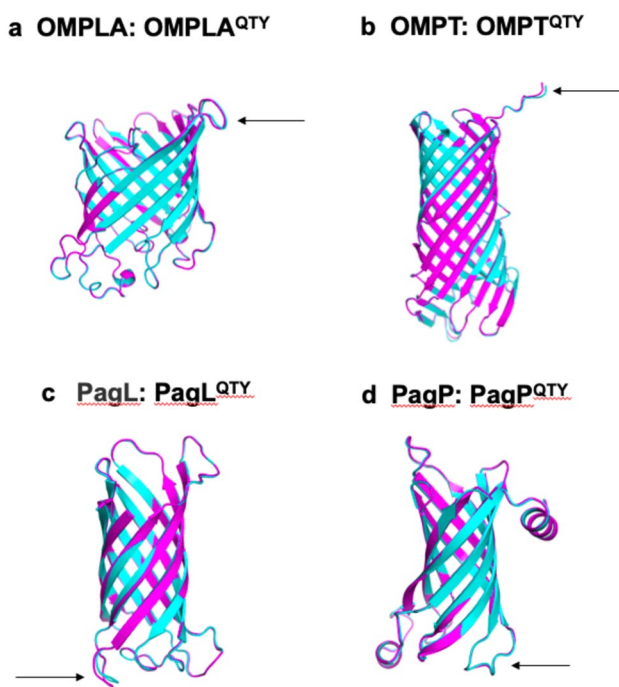
d) PagP<sup>QTY</sup> are curated at <https://github.com/Akash2000A/OMPs>

## Results and discussions

### Protein sequence alignments and other characteristics

The QTY code was designed to substitute four hydrophobic amino acids including leucine (L), isoleucine (I), valine (V), and phenylalanine (F) with three neutrally polar amino acids glutamine (Q), threonine (T), and tyrosine (Y). In the transmembrane segments, the hydrophobic amino acids were replaced with Q, T, and Y residues. Table 1 shows the overall variations seen in the membrane enzymes after using the QTY code on the native membrane proteins. Remarkably, these variations had little effect on the protein overall structure (Fig. 2, Fig. 3).

We protein-sequence-aligned the native integral membrane beta-barrel enzymes with known crystal structures to their corresponding QTY variants (Fig. 1). Despite substantial QTY replacement of hydrophobic residues (22.95–30.85%) in the transmembrane domains of the integral membrane  $\beta$ -barrel enzymes, the isoelectric point (pI) and molecular weight (MW) remained comparable (Table. 1). This is attributable to the uncharged, polar nature of the Q, T, and Y amino acids, which introduce hydrophilic side chains. The  $-NH_2$  side chains of Q (glutamine) can form 4 hydrogen bonds with water, with two donors from  $-NH_2$  and two acceptors through the oxygen of  $-C=O$ . The  $-OH$  groups of T (threonine) and Y (tyrosine) can form three hydrogen bonds with water molecules, with one donor from H (hydrogen) and two acceptors from O (oxygen). As such, the hydrophobicity of the transmembrane  $\beta$ -barrels is markedly reduced. For instance, the transmembrane  $\beta$ -sheets in the protein sequences of PagL and OmpT show differences



**Fig. 3** Superposed images of four bacterial native integral membrane beta barrel enzymes and their AlphaFold2-predicted water-soluble QTY variants. The crystal structures of the native integral membrane beta-barrel enzymes (colored magenta) and the predicted structures of their water-soluble QTY variants generated by AlphaFold2 (colored cyan) are depicted, with arrows indicating deviations in unstructured loop regions. For the superimposed structures, the (RMSD) is given in Å (°). These superposed structures are shown: **a** OMPLA vs OMPLA<sup>QTY</sup> (0.189 Å), **b** OmpT vs OmpT<sup>QTY</sup> (0.240 Å), **c** PagL vs PagL<sup>QTY</sup> (0.181 Å), **d** PagP vs PagP<sup>QTY</sup> (0.286 Å). The superposed native and water-soluble QTY variant structures are very similar

of 30.85% and over 22%, respectively, compared to their water-soluble QTY variants (Table 1).

The selected targets have a range of isoelectric focusing points (pIs), all being acidic. Notably, the pIs are identical between the native and QTY variants. All four beta barrel enzymes have mild acidic pIs, OMPLA (pI 5.01), OmpT (pI 5.38), PagL (pI 5.01), and PagP (pI 5.75), despite the substantial number of QTY amino acid substitutions. The amino acids glutamine, threonine, and tyrosine (Q, T, Y) are uncharged at neutral pH. Consequently, the isoelectric point (pI) of the protein is largely unaffected by substitutions involving these residues. This observation is particularly relevant since changes in pI can promote nonspecific protein interactions, underscoring the need to carefully consider the impact of specific amino acid substitutions on protein function and behavior.

X-ray crystallography determines protein structures by measuring electron density maps from X-ray diffraction patterns (Dessau and Modis 2011). These maps reveal electron densities that are compared to known amino acid

structures to deduce the protein's amino acid sequence and 3D structure. The proposed amino acid placements are iteratively fitted into the experimental electron density maps to derive the final atomic-resolution structure (Yao and Moseley 2020). As evident in the electron density maps (Zhang and Egli 2022), leucine (L) with glutamine (Q), isoleucine (I) and valine (V) with threonine (T), and phenylalanine (F) with tyrosine (Y) are among the amino acid pairs with the greatest structural similarity. Readers can refer to our previous studies for more details on the QTY code and electron density maps. By substituting the CH<sub>3</sub>- groups on leucine and valine with -OH groups on glutamine (Q) and threonine (T), and by adding an -OH group to phenylalanine (F) to form tyrosine (Y), the transmembrane  $\beta$ -sheets of these four beta-barrel enzymes underwent QTY substitutions ranging from 22.95 to 30.85%. These substitutions slightly increased the molecular weight of each protein (Table 1). Visual inspection (Fig. 2) and RMSD values (Table 1) revealed minute structural changes in the transmembrane  $\beta$ -sheets after QTY substitution. The  $\beta$ -barrel structures appear highly similar, as evident in the superimposed structures shown in Fig. 3.

### Structural alignment of native $\beta$ -barrel membrane enzymes and their water-soluble QTY variants

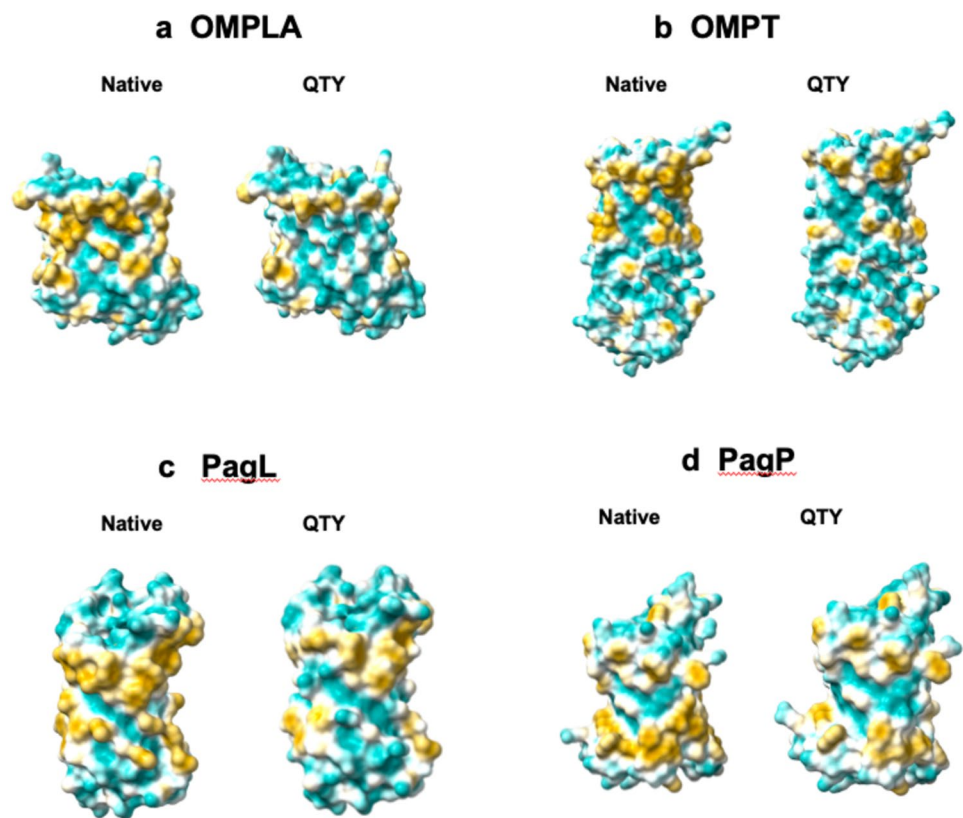
Structural alignments were performed between native integral membrane  $\beta$ -barrel enzymes solved by X-ray crystallography, and their corresponding QTY variants that were predicted by AlphaFold2. The molecular structures of the native enzymes are available for OMPLA (PDB: 1QD6), OmpT (PDB: 1I78), PagL (PDB: 2ERV), and PagP (PDB: 1THQ). We superposed the experimentally determined native structures (magenta) with the AlphaFold2 predicted QTY variant structures (cyan) (Fig. 3). The transmembrane  $\beta$ -barrel enzymes showed a high degree of structural similarity between the native and QTY-substituted structures, as evident in Fig. 3. These structural superpositions suggest that the QTY code is applicable to the  $\beta$ -sheets of these transmembrane enzymes. However, as expected, some deviations were observed in the unstructured loops since AlphaFold2 is less reliable for predicting such disordered regions. These loop regions exhibiting deviations are indicated by arrows in Fig. 3.

### Analysis of hydrophobic surface for native and QTY-substituted $\beta$ -barrel membrane enzymes

The hydrophobic characteristics of the transmembrane domain in native  $\beta$ -barrel enzymes were well-studied. The solubility and stability of these proteins can vary



**Fig. 4** Hydrophobic surfaces of four bacterial integral membrane beta barrel enzymes and their AlphaFold2 predicted water-soluble QTY variants. Surface hydrophobicity changes for integral membrane beta-barrel enzymes after mutating hydrophobic residues L,I,V,F to hydrophilic Q, T, Y. Brownish hydrophobic surfaces of native enzymes (left) become more cyan and hydrophilic in QTY variants (right) due to substitutions on transmembrane beta-sheets. Native and variant structures shown side-by-side for comparison. **a** OMPLA vs OMPLA<sup>QTY</sup>, **b** OmpT vs OmpT<sup>QTY</sup>, **c** PagL vs PagL<sup>QTY</sup>, and **d** PagP vs PagP<sup>QTY</sup>. The hydrophobic surfaces are apparently reduced



substantially depending on the specific  $\beta$ -barrel and its environment. Since integral  $\beta$ -barrel transmembrane enzymes are inherently hydrophobic, detergents are essential to solubilize and stabilize them in aqueous solutions after extraction from the native membrane. In crystal structures of native  $\beta$ -barrels membrane enzymes, the  $\beta$ -sheets are directly embedded in the hydrophobic lipid bilayer, with hydrophobic side chains including leucine (L), isoleucine (I), valine (V), and phenylalanine (F) interacting with the membrane lipids.

The hydrophobic surfaces are reduced by substituting hydrophobic amino acids, including L, I, V, and F, with the more hydrophilic residues Q, T, and Y (Zhang et al. 2018; Zhang and Egli 2022). The alterations in hydrophobicity arising from the QTY substitutions that convert the hydrophobic  $\beta$ -sheets to hydrophilic ones are shown in Fig. 4.

The rationale underlying the QTY code's pairwise substitution approach for predicting protein secondary structure from amino acid sequences is the presence of a conserved molecular motif in several protein classes. Beta-pleated sheets, based on their chemical properties, can be categorized into three discrete types. Type I beta-sheets exhibit hydrophilicity and water-solubility and are found within globular proteins such as green fluorescent protein (GFP) (Ormö et al. 1996). In contrast, Type II beta-sheets display hydrophobicity and water-insolubility and are present

in transmembrane beta-barrel proteins (Horne et al. 2020; Diederichs et al. 2020). Amphipathic Type III beta-sheets, containing both hydrophobic and hydrophilic faces, are characteristic silk-like proteins (Huang et al. 2018; Preda et al. 2013). The QTY code enhances the solubility, structural preservation, and characterization of beta-barrel integral membrane enzymes while potentially retaining their biologically functional capabilities for biotechnological applications. Despite variations in chemical properties, all forms of beta-sheets share similarities in structure, with adjacent strands aligned and stabilized through backbone hydrogen bonding (Eisenberg 2003).

### AlphaFold2 predictions

Predicting how proteins fold into 3D shapes has been a major goal in structural biology since the 1960s, but has been very difficult to achieve. Recently, a new AI tool AlphaFold2 has made breakthrough in this area. AlphaFold2 uses machine learning to predict protein structures with high accuracy. Before AlphaFold2, many protein structures were impossible to decipher, especially membrane proteins. AlphaFold2 now allows us to predict these hard-to-study protein molecular structures. This is opening up many new discoveries not only for protein function,

especially for membrane proteins but also for designing new proteins.

In this study, we used AlphaFold2 to predict structures of water-soluble QTY variants of integral membrane beta-barrel enzymes. The predicted QTY variant structures closely matched the native beta-barrel conformations. This study again validates that the QTY code can confer water-solubility while maintaining the structure and may apply to other transmembrane beta-barrel or alpha-helical membrane enzymes. This demonstrates using AlphaFold2 to gain structural insights into engineered water-soluble integral membrane enzyme QTY variants.

The growing problem of antibiotic resistance underscores the urgent need for new classes of antibiotics (Mancuso et al. 2021). Recent data from the World Health Organization highlights the global rise in antibiotic resistance, which leads to extended hospitalizations and patient deaths (Martin-Loeches et al. 2022). Without new solutions, common infections and minor injuries could once again become lethal. We desperately need to spur the development of multiple creative antibiotic options if we want to get ahead of bacterial antibiotic resistance (Zerbe et al. 2017). The water-soluble integral membrane enzymes QTY variants are useful for designing new antibiotics.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s42485-024-00170-8>.

**Author contributions** S.Z. conceived and directed the project. A.S–S prepared Figs. 1, 2, 3, 4 and Table. 1. A.S–S. and S.Z. wrote the main manuscript text. All authors reviewed the manuscript.

**Funding** Open Access funding provided by the MIT Libraries. No funding is used for this manuscript. Akash Sajeew-Sheeja was graduated in July 2024 as an undergraduate student. No grants was ever used for this research. We used the online free software including AlphaFold2 and Pymol.

**Data availability** No datasets were generated or analysed during the current study.

## Declarations

**Conflict of interest** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Ahn VE, Lo EI, Engel CK, Chen L, Hwang PM, Kay LE et al (2004) A hydrocarbon ruler measures palmitate in the enzymatic acylation of endotoxin. *EMBO J* 23:2931–2941
- Audet A, Nantel G, Proulx P (1974) Phospholipase A activity in growing *Escherichia coli* cells. *Biochim Biophys Acta* 348:334–343
- Bishop RE (2005) The lipid A palmitoyltransferase PagP: molecular mechanisms and role in bacterial pathogenesis. *Mol Microbiol* 57:900–912
- Bishop RE (2008) Structural biology of membrane-intrinsic  $\beta$ -barrel enzymes: Sentinels of the bacterial outer membrane. *Biochim Biophys Acta - Biomembranes* 1778:1881–1896
- Bishop RE, Gibbons HS, Guina T, Trent MS, Miller SI, Raetz CRH (2000) Transfer of palmitate from phospholipids to lipid A in outer membranes of gram-negative bacteria. *EMBO J* 19:5071–5080
- Bukholm G, Tannaes T, Nedenskov P, Esbensen Y, Grav HJ, Hovig T et al (1997) Colony variation of *Helicobacter pylori*: pathogenic potential is correlated to cell wall lipid composition. *Scand J Gastroenterol* 32:445–454
- Dekker N (2000) Outer-membrane phospholipase A: known structure, unknown biological function. *Mol Microbiol* 35:711–717
- Dekker N, Tommassen J, Lustig A, Rosenbusch JP, Verheij HM (1997) Dimerization regulates the enzymatic activity of *Escherichia coli* outer membrane phospholipase A. *J Biol Chem* 272:3179–3184
- Dessau MA, Modis Y (2011) Protein crystallization for X-ray crystallography. *J vis Exp*. <https://doi.org/10.3791/2285>
- Diederichs KA, Ni X, Rollauer SE, Botos I, Tan X, King MS et al (2020) Structural insight into mitochondrial  $\beta$ -barrel outer membrane protein biogenesis. *Nat Commun* 11:3290
- Eisenberg D (2003) The discovery of the alpha-helix and beta-sheet, the principal structural features of proteins. *Proc Natl Acad Sci U S A* 100:11207–11210
- Ernst RK, Adams KN, Moskowitz SM, Kraig GM, Kawasaki K, Stead CM et al (2006) The *Pseudomonas aeruginosa* lipid A deacylase: selection for expression and loss within the cystic fibrosis airway. *J Bacteriol* 188:191–201
- Evanics F, Hwang PM, Cheng Y, Kay LE, Prosser RS (2006) Topology of an outer-membrane enzyme: Measuring oxygen and water contacts in solution NMR studies of PagP. *J Am Chem Soc* 128:8256–8264
- Fairman JW (2011) Crystal structures of the outer membrane domain of intimin and invasin from enterohemorrhagic *E. coli* and enteropathogenic *Y. pseudotuberculosis*. *Structure* 19:1233–1243
- Fiss EH, Hollifield WC, Neilands JB (1979) Absence of ferric enterobactin receptor modification activity in mutants of *Escherichia coli* K12 lacking protein a. *Biochem Biophys Res Commun* 91:29–34
- Funahara Y, Nikaido H (1980) Asymmetric localization of lipopolysaccharides on the outer membrane of *Salmonella typhimurium*. *J Bacteriol* 141:1463–1465
- Geurtsen J, Steeghs L, Hamstra H-J, Ten Hove J, de Haan A, Kuipers B et al (2006) Expression of the lipopolysaccharide-modifying enzymes PagP and PagL modulates the endotoxic activity of *Bordetella pertussis*. *Infect Immun* 74:5574–5585
- Grant KA, Beldandia IU, Dekker N, Richardson PT, Park SF (1997) Molecular characterization of *pldA*, the structural gene for a phospholipase A from *Campylobacter coli*, and its contribution to cell-associated hemolysis. *Infect Immun* 65:1172–1180
- Guina T, Yi EC, Wang H, Hackett M, Miller SI (2000) A PhoP-regulated outer membrane protease of *Salmonella enterica* serovar typhimurium promotes resistance to alpha-helical antimicrobial peptides. *J Bacteriol* 182:4077–4086

- Guo L, Lim KB, Poduje CM, Daniel M, Gunn JS, Hackett M et al (1998) Lipid A acylation and bacterial resistance against vertebrate antimicrobial peptides. *Cell* 95:189–198
- Hao S, Jin D, Zhang S, Qing R (2020) QTY code-designed water-soluble fc-fusion cytokine receptors bind to their respective ligands. *QRB Discov* 1:e4
- Hermansen S, Linke D, Leo JC (2022) Transmembrane  $\beta$ -barrel proteins of bacteria: from structure to function. *Membrane proteins*. Elsevier, pp 113–161
- Horne JE, Brockwell DJ, Radford SE (2020) Role of the lipid bilayer in outer membrane protein folding in Gram-negative bacteria. *J Biol Chem* 295(30):10340–10367
- Huang W, Ling S, Li C, Omenetto FG, Kaplan DL (2018) Silk worm silk-based materials and devices generated using bio-nanotechnology. *Chem Soc Rev*. <https://doi.org/10.1039/c8cs00187a>
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O et al (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596:583–589
- Kawasaki K, Ernst RK, Miller SI (2004) 3-O-deacylation of lipid A by PagL, a PhoP/PhoQ-regulated deacylase of salmonella typhimurium, modulates signaling through toll-like receptor 4. *J Biol Chem* 279:20044–20048
- Kawasaki K, Ernst RK, Miller SI (2005) Inhibition of Salmonella enterica serovar Typhimurium lipopolysaccharide deacylation by aminoarabinose membrane modification. *J Bacteriol* 187:2448–2457
- Kawasaki K, China K, Nishijima M (2007) Release of the lipopolysaccharide deacylase PagL from latency compensates for a lack of lipopolysaccharide aminoarabinose modification-dependent resistance to the antimicrobial peptide polymyxin B in *Salmonella enterica*. *J Bacteriol* 189:4911–4919
- Kukkonen M, Korhonen TK (2004) The ompT family of enterobacterial surface proteases/adhesins: from housekeeping in *Escherichia coli* to systemic spread of *Yersinia pestis*. *Int J Med Microbiol* 294:7–14
- Li M, Wang Y, Tao F, Xu P, Zhang S (2023) QTY code designed antibodies for aggregation prevention: a structural bioinformatic and computational study. *Proteins*. <https://doi.org/10.1002/prot.26603>
- Li M, Tang H, Qing R, Wang Y, Xu P, Zhang S, Tao F (2024) Design of a water-soluble transmembrane receptor kinase with intact molecular function by QTY code. *Nat Commun* 15(1):4293
- Mancuso G, Midiri A, Gerace E, Biondo C (2021) Bacterial antibiotic resistance: the most critical pathogens. *Pathogens* 10:1310
- Martin-Loeches I, Garduno A, Pova P, Nseir S (2022) Choosing antibiotic therapy for severe community-acquired pneumonia. *Curr Opin Infect Dis* 35:133–139
- Meng R, Hao S, Sun C, Hou Z, Hou Y, Wang L et al (2023) Reverse-QTY code design of active human serum albumin self-assembled amphiphilic nanoparticles for effective anti-tumor drug doxorubicin release in mice. *Proc Natl Acad Sci U S A* 120(21):e2220173120
- Michel GP, Stárka J (1979) Phospholipase A activity with integrated phospholipid vesicles in intact cells of an envelope mutant of *Escherichia coli*. *FEBS Lett* 108:261–265
- Ormö M, Cubitt AB, Kallio K, Gross LA, Tsien RY, Remington SJ (1996) Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science* 273:1392–1395
- Preda RC, Leisk G, Omenetto F, Kaplan DL (2013) Bioengineered silk proteins to control cell and tissue functions. *Methods Mol Biol* 996:19–41
- Preston A, Maxim E, Toland E, Pishko EJ, Harvill ET, Caroff M et al (2003) *Bordetella bronchiseptica* PagP is a Bvg-regulated lipid A palmitoyl transferase that is required for persistent colonization of the mouse respiratory tract. *Mol Microbiol* 48:725–736
- Qing R, Han Q, Skuhersky M, Chung H, Badr M, Schubert T et al (2019) QTY code designed thermostable and water-soluble chimeric chemokine receptors with tunable ligand affinity. *Proc Natl Acad Sci U S A* 116:25668–25676
- Rutten L, Geurtsen J, Lambert W, Smolenaers JJ, Bonvin AM, De Haan A et al (2006a) 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
- Rutten L, Geurtsen J, Lambert W, Smolenaers JJM, Bonvin AM, de Haan A et al (2006b) 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
- Sajeev-Sheela A, Smorodina E, Zhang S (2023) Structural bioinformatics studies of bacterial outer membrane beta-barrel transporters and their AlphaFold2 predicted water-soluble QTY variants. *PLoS ONE* 18:e0290360
- Sankaran K, Wu HC (1994) Lipid modification of bacterial prolipoprotein. Transfer of diacylglycerol moiety from phosphatidylglycerol. *J Biol Chem* 269:19701–19706
- Scandella CJ, Kornberg A (1971) Membrane-bound phospholipase A1 purified from *Escherichia coli*. *Biochemistry* 10:4447–4456
- Silhavy TJ, Kahne D, Walker S (2010) The bacterial cell envelope. *Cold Spring Harb Perspect Biol* 2:a000414–a000414
- Skuhersky MA, Tao F, Qing R, Smorodina E, Zhang JD (2021) Comparing native crystal structures and AlphaFold2 predicted water-soluble G protein-coupled receptor QTY variants. *Life (Basel)*. 11(12):1285
- Smorodina E, Diankin I, Tao F, Qing R, Yang S, Zhang S (2022) Structural informatic study of determined and AlphaFold2 predicted molecular structures of 13 human solute carrier transporters and their water-soluble QTY variants. *Sci Rep* 12:20103
- Smorodina E, Tao F, Qing R, Yang JD, Zhang S (2022) Comparing 2 crystal structures and 12 AlphaFold2-predicted human membrane glucose transporters and their water-soluble glutamine, threonine and tyrosine variants. *QRB Discovery* 3:e5
- Snijder HJ, Ubarretxena-Belandia I, Blaauw M, Kalk KH, Verheij HM, Egmond MR et al (1999) Structural evidence for dimerization-regulated activation of an integral membrane phospholipase. *Nature* 401:717–721
- Stumpe S, Schmid R, Stephens DL, Georgiou G, Bakker EP (1998) Identification of OmpT as the protease that hydrolyzes the antimicrobial peptide protamine before it enters growing cells of *Escherichia coli*. *J Bacteriol* 180:4002–4006
- Sugimura K, Nishihara T (1988) Purification, characterization, and primary structure of *Escherichia coli* protease VII with specificity for paired basic residues: identity of protease VII and OmpT. *J Bacteriol* 170:5625–5632
- Sun J, Rutherford ST, Silhavy TJ, Huang KC (2022) Physical properties of the bacterial outer membrane. *Nat Rev Microbiol* 20:236–248
- Tannaes T, Bukholm IK, Bukholm G (2005) High relative content of lysophospholipids of *Helicobacter pylori* mediates increased risk for ulcer disease. *FEMS Immunol Med Microbiol* 44:17–23
- Tao F, Tang H, Zhang S, Li M, Xu P (2022) Enabling QTY server for designing water-soluble alpha-helical transmembrane proteins. *Mbio* 13(1):e03604
- Tegler L, Corin K, Pick H, Brookes J, Skuhersky M, Vogel H et al (2020) The G protein coupled receptor CXCR4 designed by the QTY code becomes more hydrophilic and retains cell signaling activity. *Sci Rep* 10:21371. <https://doi.org/10.1038/s41598-020-77659-x>
- Thomassin J-L, Brannon JR, Gibbs BF, Gruenheid S, Le Moual H (2012) OmpT outer membrane proteases of enterohemorrhagic and enteropathogenic *Escherichia coli* contribute differently to the degradation of human LL-37. *Infect Immun* 80:483–492
- Trent MS, Pabich W, Raetz CR, Miller SI (2001) A PhoP/PhoQ-induced Lipase (PagL) that catalyzes 3-O-deacylation of lipid

- A precursors in membranes of *Salmonella typhimurium*. *J Biol Chem* 276:9083–9092
- Vandeputte-Rutten L, Kramer RA, Kroon J, Dekker N, Egmond MR, Gros P (2001) Crystal structure of the outer membrane protease OmpT from *Escherichia coli* suggests a novel catalytic site. *EMBO J* 20:5033–5039
- Xu Q, Guo M, Yu F (2023) B-barrel assembly machinery (BAM) complex as novel antibacterial drug target. *Molecules* 28:3758
- Yao S, Moseley HNB (2020) A chemical interpretation of protein electron density maps in the worldwide protein data bank. *PLoS ONE* 15:e0236894
- Zerbe K, Moehle K, Robinson JA (2017) Protein Epitope mimetics: From new antibiotics to supramolecular synthetic vaccines. *Acc Chem Res* 50:1323–1331
- Zhang S, Egli M (2022) Hiding in plain sight: three chemically distinct  $\alpha$ -helix types. *Q Rev Biophys* 55:e7
- Zhang J, Qian J, Zhang X, Zou Q (2014) Outer membrane inflammatory protein A, a new virulence factor involved in the pathogenesis of *Helicobacter pylori*. *Mol Biol Rep* 41:7807–7814
- Zhang S, Tao F, Qing R, Tang H, Skuhersky M, Corin K et al (2018) QTY code enables design of detergent-free chemokine receptors that retain ligand-binding activities. *Proc Natl Acad Sci U S A* 115:E8652–E8659