

Deep learning-driven insights into super protein complexes for outer membrane protein biogenesis in bacteria

Mu Gao^{1*}, Davi Nakajima An², Jeffrey Skolnick^{1*}

¹Center for the Study of Systems Biology, School of Biological Sciences, Georgia Institute of Technology, Atlanta, United States; ²School of Computer Science, Georgia Institute of Technology, Atlanta, United States

Abstract To reach their final destinations, outer membrane proteins (OMPs) of gram-negative bacteria undertake an eventful journey beginning in the cytosol. Multiple molecular machines, chaperones, proteases, and other enzymes facilitate the translocation and assembly of OMPs. These helpers usually associate, often transiently, forming large protein assemblies. They are not well understood due to experimental challenges in capturing and characterizing protein-protein interactions (PPIs), especially transient ones. Using AF2Complex, we introduce a high-throughput, deep learning pipeline to identify PPIs within the *Escherichia coli* cell envelope and apply it to several proteins from an OMP biogenesis pathway. Among the top confident hits obtained from screening ~1500 envelope proteins, we find not only expected interactions but also unexpected ones with profound implications. Subsequently, we predict atomic structures for these protein complexes. These structures, typically of high confidence, explain experimental observations and lead to mechanistic hypotheses for how a chaperone assists a nascent, precursor OMP emerging from a translocon, how another chaperone prevents it from aggregating and docks to a β -barrel assembly port, and how a protease performs quality control. This work presents a general strategy for investigating biological pathways by using structural insights gained from deep learning-based predictions.

***For correspondence:**
mu.gao@gatech.edu (MG);
skolnick@gatech.edu (JS)

Competing interest: The authors declare that no competing interests exist.

Funding: See page 14

Preprinted: 25 August 2022

Received: 22 August 2022

Accepted: 28 November 2022

Published: 28 December 2022

Reviewing Editor: José D Faraldo-Gómez, National Institutes of Health, United States

© Copyright Gao et al. This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

Editor's evaluation

The authors show that an artificial-intelligence method can be used to predict the three-dimensional structure of protein-protein complexes formed between cellular factors that promote the assembly of bacterial outer membrane proteins. The structures are compelling because they explain previously published biochemical data and provide novel insights into the function of these factors.

Introduction

A structural component unique to gram-negative bacteria is the outer membrane (OM), composed of lipopolysaccharides and phospholipids in an asymmetric bilayer with embedded lipoproteins and transmembrane β -barrel proteins (*Silhavy et al., 2010*). The latter group, OM proteins (OMPs), play vital functional roles, e.g., exchanging small molecules with the environment through their transmembrane β -barrel porins. OMPs are synthesized by cytosolic ribosomes, translocated across the inner membrane (IM), and finally delivered to the OM via the OMP biogenesis pathway (*Rollauer et al., 2015*).

The translocation and folding of OMPs involve many proteins, including essential and auxiliary ones that form multiple complexes in cooperation (*De Geyter et al., 2016; Troman and Collinson,*

eLife digest All living cells are contained within a fatty cell membrane that allows water and only certain other molecules to pass through with ease. Bacteria only consist of a single cell, making their membrane the only interface with the surrounding environment. Gram-negative bacteria – which include *Escherichia coli*, a bacterium found in the gut of all humans – have an extra layer of protection, the ‘outer membrane’. Proteins in this membrane are called ‘outer membrane proteins’ or OMPs and allow nutrients to enter the cell. But OMPs, which are made inside the cell, need to be transported to the outer membrane and folded correctly before they can perform their role. This multistep process, which involves interactions between many different proteins, is not fully understood.

The journey of an OMP from the center of the cell where it is made to the outer membrane is complicated. First, the OMP needs to pass through the cell’s inner membrane. To do this, it must interact with ‘channel proteins’ in the inner membrane that feed the OMP into the space between the two membranes, known as the bacterial envelope. This step requires the OMP to be unfolded. Once in the bacterial envelope the OMP interacts with proteins that help it fold correctly and integrate into the outer membrane.

The interactions between proteins in the bacterial envelope are short-lived, making them difficult to study using lab-based experiments. An alternative approach is predicting a protein’s structure from its amino acid sequence which is a difficult computational problem to solve. However, in 2020 developers behind the AlphaFold2, a deep learning program, were able to use a set of equations organized in a ‘neural network’ that can ‘learn’ from a library of known protein structures to predict unknown structures with high accuracy. Gao et al. used AF2Complex, a tool based AlphaFold2, tailored to predicting interactions between proteins, to investigate what interactions OMPs could be involved with on their way to the outer membrane.

With the help of a supercomputer at the Oakridge National Laboratory, Gao et al. screened nearly 1,500 *E. coli* proteins within the bacterial envelope to see how they might interact with OMPs. The screen identified previously unknown interactions between proteins that suggest that the formation of the bacterial outer membrane and the integration of proteins into it involve protein complexes and molecular mechanisms that have not yet been characterized. Additionally, the screen also identified interactions that had been previously described, confirming that the deep learning approach can correctly capture real interactions.

Overall, Gao et al.’s work inspires new hypotheses about the mechanisms through which OMPs are transported to the outer membrane, although further work will be needed to confirm the roles of protein interactions predicted by the computational model experimentally. Furthermore, the ability to design experiments based on computational predictions is exciting. If confirmed, the new protein interactions could help scientists better understand OMP transport, which is essential for bacterial biology. In the future, this could lead to the discovery of new targets for antibiotic drugs.

2021). Two core complexes are the SecYEG translocon (*Rapoport et al., 2017; Oswald et al., 2021*) and the β -barrel assembly machine (BAM; *Tomasek and Kahne, 2021; Noinaj et al., 2017*). SecYEG, a hetero-trimer composed of the secretion channel SecY and two additional subunits SecE and SecG, is anchored to the IM and is responsible for moving most cell envelope proteins across the IM (*Rapoport et al., 2017*). Periplasmic and OM proteins enter the channel in SecY via the SecA-dependent translocation pathway (*Oswald et al., 2021*). They possess signal peptides recognized by SecA, which inserts a protein substrate into SecY and powers it through the channel using energy from ATP hydrolysis (*Collinson, 2019*). The signal peptide, located at the N-terminus of a substrate, contains a hydrophobic segment that folds into a transmembrane α -helix once it exits the lateral gate of SecY. To release the substrate from the IM, a type I signal peptidase (SPase I) cleaves the signal peptide (*Paetzel, 2014*). At this point, a periplasmic protein has reached its destination, but an OMP, escorted by the chaperones SurA or Skp, continues its journey toward BAM harbored in the OM (*Sklar et al., 2007*). Five subunits constitute BAM, of which BamA plays the major role in folding a β -barrel and releasing the matured product (*Tomasek and Kahne, 2021; Noinaj et al., 2017*).

There are many open questions concerning the OMP biogenesis pathway. In *Escherichia coli*, the SecYEG translocon recruits additional proteins, such as SecA (*Li et al., 2016*), SecDF (*Tsukazaki,*

2018), YidC (Kumazaki et al., 2014), and PpiD (Antonoaea et al., 2008; Sachelaru et al., 2014), to form a variety of supercomplexes in different scenarios. We do not know the identities of all members of these supercomplexes, let alone their atomic structures. SurA plays a major role in chaperoning a nascent OMP (Sklar et al., 2007; Behrens-Kneip, 2010). How does it handle and deliver a substrate to BAM? Likewise, BAM recruits additional helpers, e.g., BepA (Narita et al., 2013), but how do they work together?

To answer these questions experimentally is challenging (Babu et al., 2018; Maddalo et al., 2011; Alvira et al., 2020; Carlson et al., 2019). Recently, deep learning approaches have made tremendous progress in predicting the structures of protein complexes (Gao et al., 2022b; Humphreys et al., 2021; Evans et al., 2021; Bryant et al., 2022). Here, we use one such method to address the above questions. The centerpiece of our approach is AF2Complex (Gao et al., 2022b), built on AlphaFold2 (AF2) (Evans et al., 2021; Jumper et al., 2021). Using AF2Complex, we combine virtual screening for protein-protein interactions (PPIs) and supercomplex modeling and apply this strategy to several important proteins in the OMP biogenesis pathway.

Results

Virtual screening for PPIs in the *E. coli* envelopome

A workflow employing AF2Complex was implemented to search for interacting proteins within the *E. coli* cell envelope. We refer to these proteins collectively as the 'envelopome,' which consists of ~1450 proteins, or ~35% of the complete *E. coli* proteome. Given a query envelope protein sequence, screening is conducted against every envelope protein, including the query itself to test if it forms a homo-oligomer. For each pair of input proteins, AF2Complex predicts their structures simultaneously as a putative dimer and evaluates the interface score (iScore) of the resulting structural models to decide if they interact. The value of iScore ranges between 0 and 1. Interacting protein pairs typically return a positive iScore, whereas non-interacting proteins usually return 0 or low iScores. Typically, 20 final structures are produced in multiple independent runs using 10 different AF2 deep learning models. The highest iScore among all predicted structures of a target is used for ranking the entire envelopome.

According to previous benchmarks on a set of ~7000 putatively non-interacting protein pairs from *E. coli*, minimum iScore thresholds of 0.40, 0.50, and 0.70 yield false positive rates of 1.2, 0.4, and <0.01%, respectively (Gao et al., 2022b). With respect to the capability of identifying and modeling true PPIs, on a set of 440 heterodimeric complexes whose experimental structures have been recently determined and were not used for AF2 model training, AF2Complex recalls 81, 74, and 34% of these benchmark targets at the same three iScore thresholds, respectively, and yields medium- or high-quality complex structures for 84, 87, and 93% in the top ranked models of these positively identified targets (Gao et al., 2022b). Hence, we consider predictions of medium, high, and very high confidence, progressively, at iScore cutoffs of 0.4, 0.5, and 0.7. Given a query protein, the top iScores of all scanned envelope proteins are then sorted to identify potential interacting partners with the query protein. Note that our computational predictions are about physical interactions between a pair of proteins subjected to screening, not about their biological roles even if they are predicted to interact physically. Moreover, the predicted physical interactions may not be relevant in a cellular environment due to various factors not considered in modeling, e.g., competition from other proteins with stronger binding affinities, post-translational modifications, etc. Thus, it is possible that many PPIs predicted by this pipeline do not necessarily have biological relevance. Nevertheless, since cognate PPIs required by their functions are more likely to be detected than randomly selected proteins, biologically interesting PPIs are enriched at the top of the screening results ranked by iScore. Thus, the screening procedure may provide valuable even critical clues for subsequent investigation. In this study, assisted by existing experimental evidence, we select from high confidence computational predictions those most likely to have significant biological implications and then predict the structures of larger complexes if more than two proteins are involved according to our predictions or based on literature information. The interactions that we ignored are either of unknown biological significance, physically interacting but biologically irrelevant, or simply false positives.

As a first test, we applied our PPI pipeline to the chaperones PpiD and YfgM because strong experimental evidence indicates that these two proteins belong to a super SecYEG translocon (Antonoaea

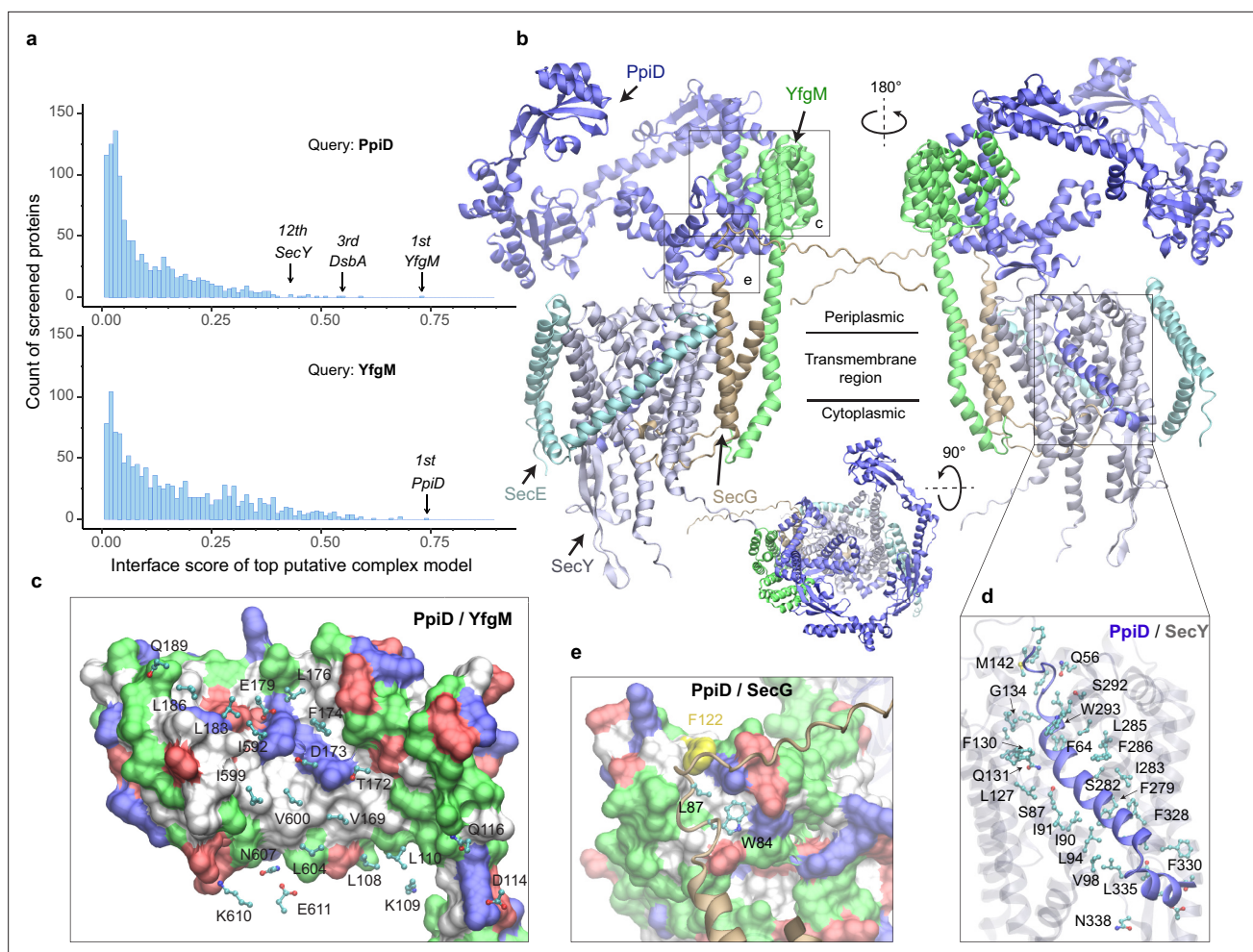


Figure 1. *E. coli* super-translocon SecYEG/PpiD/YfgM. **(a)** Computational screening for protein-protein interaction partners of PpiD and YfgM within the *E. coli* envelopome, respectively. A histogram displays the distribution of the top interface scores (iScores) of all envelope proteins screened with each query. Black arrows mark the top hits that were further studied, along with their names and overall ranks. **(b)** The top AF2Complex model of a supercomplex made of PpiD (blue), YfgM (green), SecY (silver), SecE (cyan), and SecG (tan) in three different views. Proteins are shown in a cartoon representation. Viewpoint transition, from either left to right or top to bottom, is indicated by a rotation axis (dashed line) and the rotation angle in degrees (circled arrow). **(c, d, and e)** Predicted PPI sites. The corresponding locations in **b** are indicated by black boxes. For clarity, the viewpoints and representations are adjusted. In the surface representations **c** and **e**, the color code is hydrophobic (white), polar (green), positive (blue), and negative (red), except for Phe122_{PpiD} (yellow) in **e**. The same color code for the surface representation is employed below unless noted otherwise. PPI residues are shown in a ball-and-stick representation for PpiD in **c**, SecY in **d**, and SecG in **e**; the color scheme of atoms is carbon (cyan), oxygen (red), nitrogen (blue), and sulfur (yellow). The same scheme of atoms is adopted throughout this work.

The online version of this article includes the following figure supplement(s) for figure 1:

Figure supplement 1. Comparison of a computed structure of SecY (silver) and two experimental structures (magenta).

et al., 2008; Sachelaru et al., 2014; Maddalo et al., 2011; Dartigalongue and Raina, 1998; Fürst et al., 2018; Jauss et al., 2019; Matern et al., 2010; Götzke et al., 2014), yet their structures remain elusive. **Figure 1a** shows the screening results of PpiD and YfgM against the envelopome. Reassuringly, these two proteins stand out at the very top of the predicted PPI rankings among the *E. coli* envelopome with iScores of ~0.73 in each case (**Figure 1a, Supplementary file 1**). Furthermore, PpiD hits SecY at the 12th rank (iScore = 0.43), corroborating previous experimental studies that SecY is an interacting partner of PpiD (*Antonoaea et al., 2008; Sachelaru et al., 2014; Fürst et al., 2018*).

SecYEG associates with the chaperone PpiD and assistant YfgM

Subsequently, we modeled SecYEG together with PpiD and YfgM and acquired another very high confidence structure (iScore = 0.73, mean predicted local distance difference test [pLDDT] = 79,

see Methods; **Figure 1b–e**). Between this supercomplex and the dimer structures obtained from screening, the PPI interfaces between PpiD and YfgM are very similar according to iAlign (**Gao and Skolnick, 2010**), which yields a root-mean-square deviation (RMSD) of 0.6 Å for the C_α atoms of the interface residues. Thus, we focus on the supercomplex structure below.

Individually, PpiD and YfgM in our complex model display the same architecture as their respective monomer models calculated by applying AF2 (**Varadi et al., 2022**). PpiD by itself has four domains in an open-arm shape, fitting its proposed chaperone role (**Antonoaea et al., 2008; Matern et al., 2010**), plus an N-terminal transmembrane α-helix (**Figure 1b**). The membrane anchoring helix is linked to a domain consisting of two discontinuous segments (residues 40–189, 581–623) from the terminal regions. Within this domain, two α-helices (residues 167–189, 581–609) are held by the tetratricopeptide repeat (TPR) domain of YfgM, which inserts itself into the IM via a single hydrophobic α-helix. Structurally, the PpiD/YfgM interface resembles a pair of ‘chopsticks’ (two PpiD helices) grasped by a ‘palm’ (the TPR domain). Their PPI interactions are quite extensive, including ~100 inter-protein residue-residue contacts involving both hydrophobic and polar partners (**Figure 1b and c**), which may explain the strong binding affinity observed in experiments (**Maddalo et al., 2011**).

In addition to the channel that connects the cytosol to the periplasm, SecY features a lateral gate embedded in the IM. It opens to release transmembrane segments, including the signal peptide of a substrate such as OmpA (**Li et al., 2016**). In vivo photo cross-linking studies found numerous sites at the lateral gate probed by PpiD (**Sachelaru et al., 2014; Jauss et al., 2019**). Such cross-linking suggests direct physical contacts between the two proteins because the covalent linkers formed upon UV light radiation are very short, lying within 4 Å. Because PpiD has only a single transmembrane α-helix, a logical speculation is that the α-helix interacts with the lateral gate of SecY. Indeed, our structure reveals that PpiD guards the lateral gate of SecY with its α-helix (**Figure 1b and d**), which occupies the same position as a signal peptide leaving the SecY gate (**Figure 1—figure supplement 1**). In superposition with a SecY crystal structure co-crystallized with an OmpA signal peptide, the C_α RMSD of SecY is 2.4 Å between the X-ray and computed structures. Thus, SecY in our model is open to accommodate the α-helix of PpiD, in comparison to an electron microscopy (EM) structure of *E. coli* SecY in which the gate is closed (**Figure 1—figure supplement 1**). Our model further elucidates why SecY residues located in transmembrane helices (TM2/3/7/8) can be cross-linked to PpiD, because they are located at the PpiD/SecY interface. For example, strong cross-linking signals have been found for five SecY residues (**Sachelaru et al., 2014; Jauss et al., 2019**), Ile91_{SecY}, Leu94_{SecY},

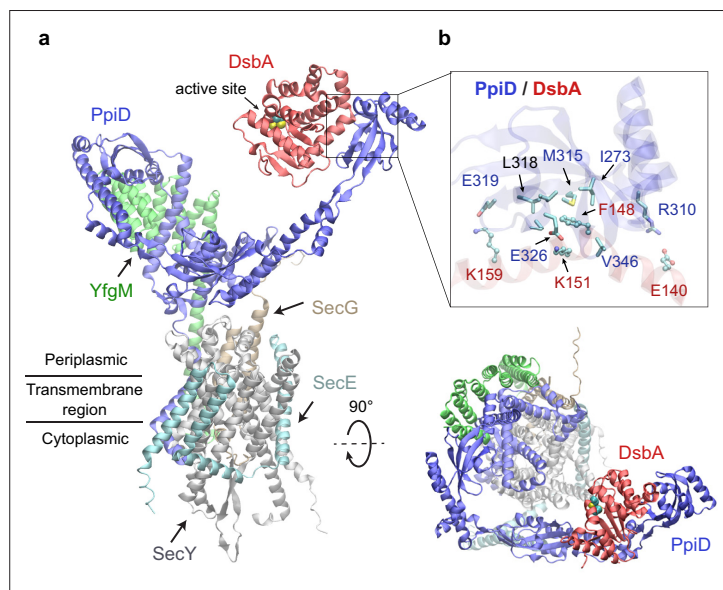


Figure 2. Structural model of the SecYEG/PpiD/YfgM/DsbA supercomplex. **(a)** Two views of the predicted structure. DsbA is shown in red, while the other proteins are colored the same as in **Figure 1**. Two cysteines, Cys49 and Cys52, essential to the enzymatic function of DsbA, are shown as spheres. **(b)** Protein-protein interaction sites between PpiD and DsbA. For clarity, tertiary structures are transparent. Key interacting residues are shown in the licorice representation for PpiD and in the ball-and-stick representation for DsbA.

Leu127_{SecY}, Phe130_{SecY}, and Phe286_{SecY}, and all make direct physical contacts (<4.5 Å) with PpiD in our structural model (**Figure 1d**). Moreover, we identified direct contacts between a hydrophobic pocket of PpiD and SecG, explaining the experimentally observed crosslinking of Phe122_{PpiD} to SecG (**Fürst et al., 2018; Figure 1e**). The interaction assigns SecG a functional role in coordinating with PpiD. Overall, the predicted structure rationalizes the results of several experimental studies.

The chaperone PpiD interacts with the disulfide isomerase DsbA

Among the top hits for PPI partners of PpiD, a fascinating discovery is DsbA, ranked third with a high confidence iScore of 0.55. DsbA donates its disulfide bond to a substrate in need; thus, it is critical to the folding of a nascent protein as it leaves the SecY channel (**Kadokura and Beckwith, 2009; Goemans et al., 2014**). How does DsbA coordinate with the SecY translocon? The predicted interaction immediately provides an answer, i.e., by associating with the chaperone PpiD that is a part of a super translocon complex (SecYEG/PpiD/YfgM/DsbA). In a predicted structure of this supercomplex (iScore = 0.71, mean pLDDT = 78, **Figure 2a**), PpiD uses its 'hand' to grip DsbA, as observed in the top model of the pair because their interactions do not interfere with other members of the translocon. The hand of PpiD is a parvulin-like domain, but it is devoid of the peptidyl-prolyl isomerase (PPIase) activity canonical to parvulin, due to the mutation of a critical histidine (**Weininger et al., 2010**). As our structure reveals, the catalytically inactive pocket of the PPIase domain is hydrophobic and buries the aromatic sidechain of Phe148_{DsbA}, which becomes a hot-spot residue surrounded by six hydrophobic residues of PpiD upon complexation (**Figure 2b**). In addition, three pairs of salt-bridges are formed around the hydrophobic contacts, contributing to specific recognition. Notably, the catalytic cysteines of DsbA, Cys49_{DsbA} and Cys52_{DsbA}, are on the opposite side to the protein-protein interface (**Figure 2a**); thus, they are open to engage a substrate as it emerges from the SecY channel. Based on this model, we hypothesize that DsbA resides on PpiD transiently to improve its chance of encountering a substrate.

SPase I LepB accesses an OMP substrate received by PpiD/YfgM

Polypeptide translocation triggers the dissociation of PpiD from SecY (**Sachelaru et al., 2014**). The modeled complex suggests that PpiD dissociation is realized by pushing a substrate out of the SecY gate, e.g., the signal peptide, which then repels the PpiD helix bound to the gate. For an OMP, a key next step is the removal of the signal peptide by the peptidase LepB in *E. coli* (**Paetzel, 2014**). How does it operate? The dissociation of PpiD from SecYEG vacates a space for LepB to approach a substrate. We generated a structural model of LepB in complex with a precursor OmpA (proOmpA) chain (residue 1–87) grasped by PpiD (iScore = 0.64, mean pLDDT = 80; **Figure 3**). The OmpA cleavage site, Ala21_{OmpA}, directly contacts the LepB active site triad, Ser89_{LepB}, Ser91_{LepB}, and Lys146_{LepB} at ~ 4 Å

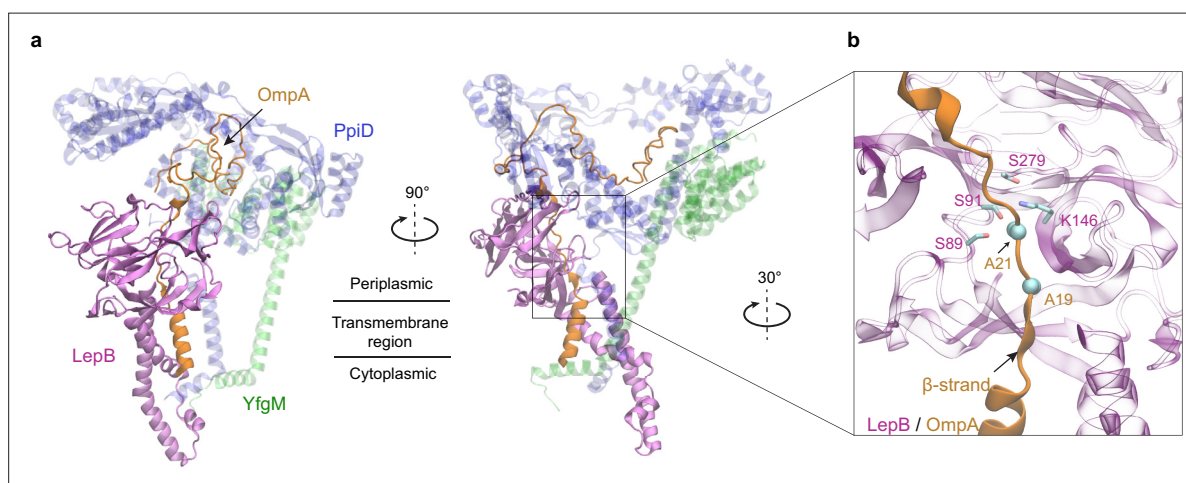


Figure 3. Predicted structure of the PpiD/YfgM/LepB/OmpA supercomplex. **(a)** Two views in the cartoon representation are shown. Colors: PpiD (blue), YfgM (green), LepB (magenta), and OmpA (residue 1–87, yellow). For clarity, representations of PpiD and LepB are transparent. **(b)** Close-up view of the OmpA signal peptide in the active site of LepB. Essential catalytic residues, Ser89, Ser91, Lys146, and Ser279 of LepB are shown in a licorice representation, and the cleavage site Ala21 and Ala19 of OmpA is shown as spheres.

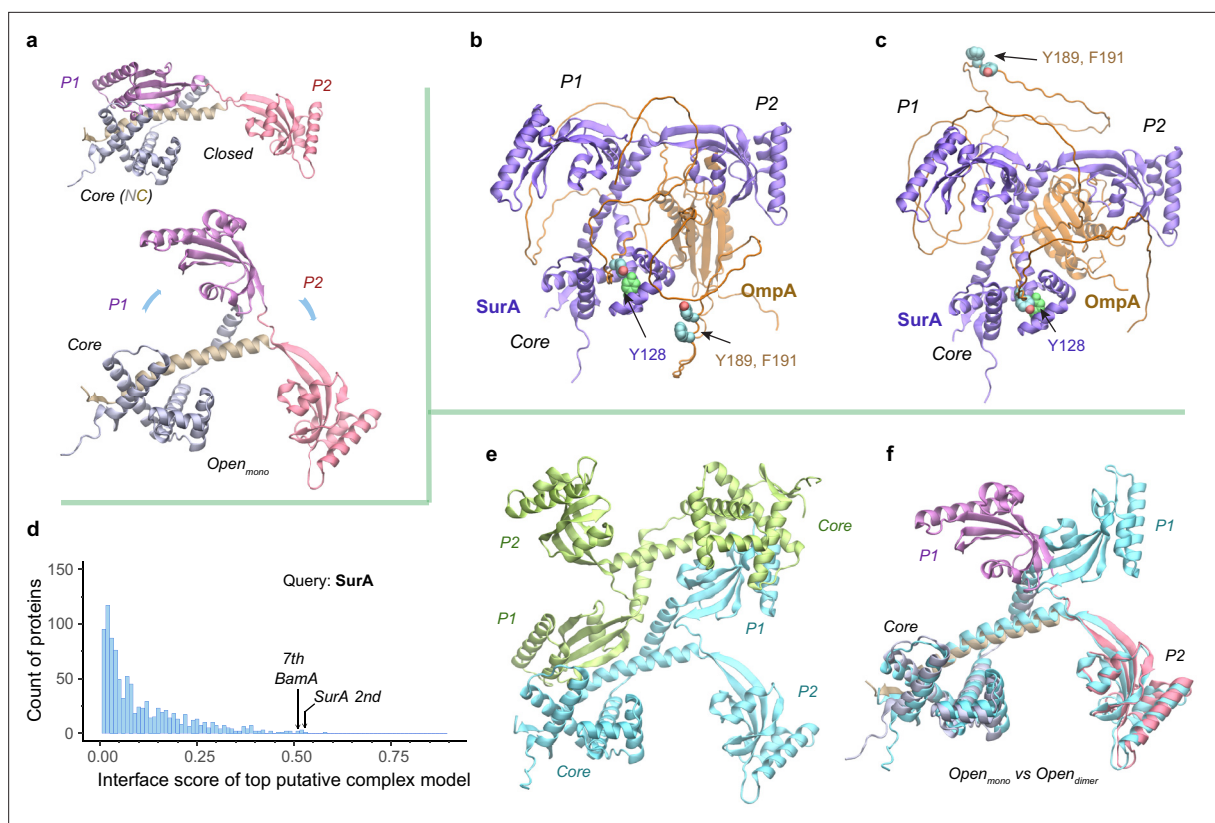


Figure 4. Structural models of SurA in the absence and presence of an OmpA substrate. **(a)** Open and closed conformations of monomeric SurA, consisting of the core domain (N-terminal region in gray and C-terminal in tan), P1 (purple), P2 (red). **(b and c)** Two structures of SurA in the presence of an OmpA substrate from two separate modeling runs. In both, SurA is open as in **a**. The β -barrel domain of OmpA is completely unfolded and generally does not maintain the same residue-residue contacts with SurA, except that an OmpA aromatic residue consistently makes π - π interactions with Tyr128_{SurA} located in the crevice of the SurA core domain. Two β -signal residues, Y189 and F191 of OmpA, are also shown as spheres. The folded periplasmic domain of OmpA is bound to the P2 domain of SurA. **(d)** Envelopome protein-protein interaction screening of SurA identifies itself and BamA among the top hits. **(e)** Two views of the top predicted structure of a SurA dimer (green and cyan). **(f)** Superimposition of two open conformations from the monomeric and dimeric SurA. Subscripts indicate the stoichiometry. The color schemes correspond to those used in **a** and **e**. Only a single SurA from the dimeric model is shown in the superposition.

The online version of this article includes the following figure supplement(s) for figure 4:

Figure supplement 1. Structural models of the OmpA polypeptide in the absence of SurA.

away from the alanine. In our model, the proOmpA chain contributes a β -strand to form a parallel β -sheet with LepB, as has been speculated (Paetzl, 2014). Overall, the structure provides a model for how LepB accesses a substrate received by PpiD/YfgM to cleave a signal peptide.

The chaperone SurA opens to load an OMP substrate

After passage through a SecYEG translocon, a nascent OMP polypeptide is relayed to SurA, which conveys the substrate toward its next stop, BAM. A crystal structure of SurA exhibits three domains: a core domain (split into N- and C-terminal subdomains) and two PPLase domains, P1 and P2 (Bitto and McKay, 2002). P2 retains PPLase activity but P1 does not (Behrens et al., 2001). In this crystal structure, P1 further packs with the core domain forming a cradle-like structure (Bitto and McKay, 2002), referred to as the closed conformation of SurA. By contrast, another crystal structure of the SurA homodimer lacking P2s displays an open conformation, where two P1s are uncoupled from the core domains to hold a dodecapeptide at the dimeric interface between P1s (Xu et al., 2007). To explore conformations of SurA, we predicted structures for a single, full-length SurA and obtain two major conformations (Figure 4a). One conformation closely resembles the cradle-like closed conformation with a backbone C α RMSD of 2.4 Å from the crystal structure. The other mimics the open conformation but incorporates P2 absent in the X-ray structure. Overall, the open conformation resembles

a ‘three-prong hook,’ where both P1 and P2 swing away from their resting position in the closed conformation.

SurA functions mainly as a monomer but may act as a dimer for a large client (*Li et al., 2018; Calabrese et al., 2020*). How does SurA chaperone a substrate such as OmpA? OmpA consists of a β -barrel (N-terminal) (*Pautsch and Schulz, 1998*) and a periplasmic (C-terminal) domain (*Ishida et al., 2014*). To model the OmpA polypeptide, we reduced the size of multiple sequence alignments provided to AF2Complex and removed all its structural templates. Consequently, we generated a monomeric OmpA structure with a non-native, collapsed N-terminal domain and a native-like periplasmic domain (*Figure 4—figure supplement 1*). In the presence of SurA, the periplasmic domain maintains the same fold, but remarkably, the non-native β -barrel region completely unravels and wraps around SurA (*Figure 4b–c*). This is consistent with small angle neutron scattering data that suggests an expanded and unfolded OmpA substrate (*Marx et al., 2020*). The SurA/OmpA dimer is not stable as different wrapping configurations were observed in two separate structure models, leading to very low iScores of ~ 0.05 , reflecting uncertainty due to the possibility of many conformations. Nevertheless, the SurA/OmpA models appear physical and provide a hypothetical basis for how the chaperone SurA could prevent a polypeptide chain from aggregating and present an unfolded polypeptide to BAM for its final assembly. Intriguingly, in both predicted complex structures, the OmpA polypeptide passes through a SurA crevice at its core domain, where Tyr126_{SurA} consistently attracts a substrate aromatic residue with π - π interactions. This may be one mechanism by which SurA recognizes OMPs, which are enriched with aromatic residues (*Xu et al., 2007; Hennecke et al., 2005*). Moreover, the disordered β -barrel region has few interactions with SurA could explain how the system avoids high-energy expenditure for β -barrel release. Of note, two key ‘ β -signal’ residues (*Tomassen, 2010*), Y189 and F191 of OmpA, do not make direct physical contact with SurA in both models.

To identify other interacting partners, we searched the *E. coli* envelopome for PPIs with SurA. Interestingly, SurA recognizes itself at a high confidence iScore of 0.53 (*Figure 4d*). The top model of a SurA homodimer exhibits twofold rotational symmetry, in which P1 domains are swapped, leading to an open conformation quite different from the above one (*Figure 4e–f*). This conformation, like a

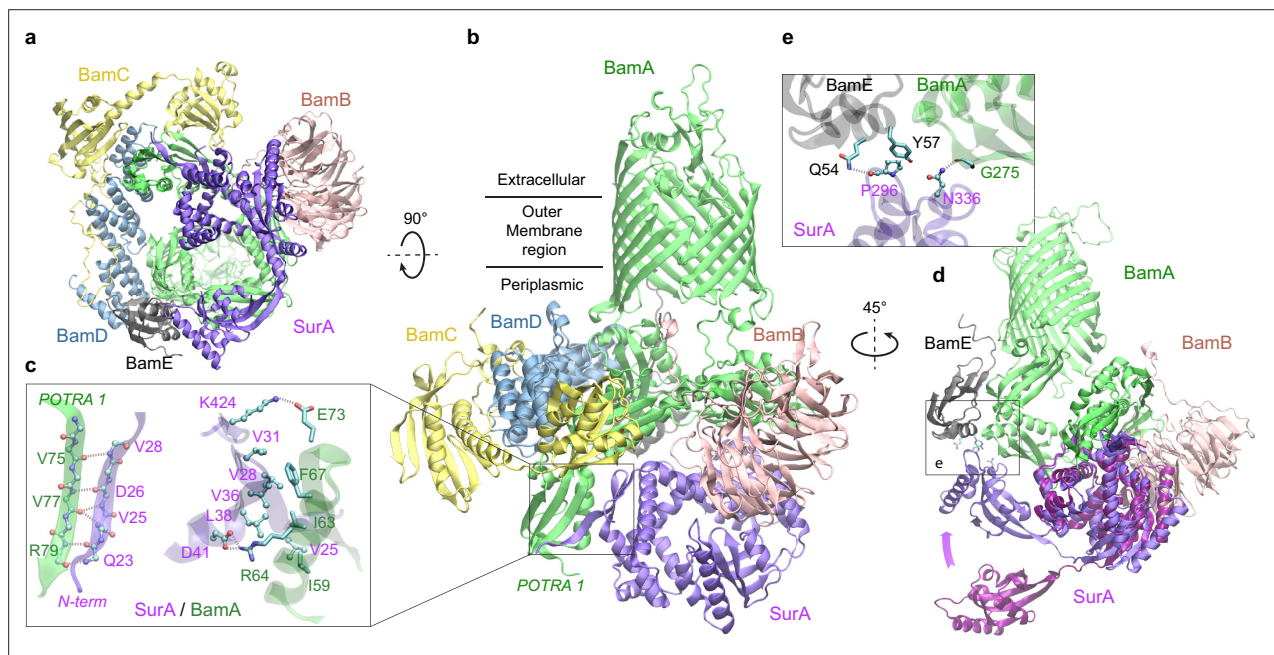


Figure 5. Structural model of SurA docked to β -barrel assembly machine (BAM). **(a and b)** Two views of the top ranked supercomplex model in the cartoon representation. The BAM constituents are BamA (green), BamB (pink), BamC (yellow), BamD (blue), and BamE (black). The N-terminal POTRA1 domain of BamA provides the main interaction sites for SurA (violet). **(c)** Close-up view of the interaction sites at POTRA1 and the core domain of SurA. Interacting residues are shown in the licorice (SurA) and ball-and-stick (BamA) representations. **(d)** Crystal structure of SurA (magenta, PDB 1M5Y) superimposed onto the computed structure of the supercomplex. The magenta arrow indicates the change of location in P2 between two structures. BamCD are omitted for clarity. **(e)** Protein-protein interaction between P2 of SurA and POTRA4 of BamA and BamE.

four-prong hook, provides SurA with another configuration to handle clients, potentially large ones, consistent with experimental conclusions (Li *et al.*, 2018; Calabrese *et al.*, 2020).

SurA and BamA specifically recognize each other

Excitingly, envelopome screening also confidently detected that SurA interacts with BamA, the anchoring subunit of BAM at the final stop of OMPs. To understand how SurA interacts with BAM, we subsequently used AF2Complex to probe potential interactions of SurA with all five BAM constituents (BamABCDE) and acquired a high confidence complex structure (iScore = 0.75, mean pLDDT = 84; Figure 5a–b). Because BAM has been extensively studied structurally (Tomasek and Kahne, 2021; Wu *et al.*, 2020), we focus on describing its interaction with SurA, though the predicted BAM structure closely mimics a known crystal structure of the complex determined at 2.9 Å resolution (PDB 5D00; Gu *et al.*, 2016). The alignment of the two complex structures yielding a very high TM-score of 0.94. Overall, SurA mainly interacts with BamA, with similar interactions observed in both the top supercomplex and the BamA/SurA dimer structures, in the N-terminal domains of both proteins (Figure 5b–c). Gln23_{SurA} to Val28_{SurA}, largely missing in the crystal structures of SurA (Bitto and McKay, 2002; Xu *et al.*, 2007), now form a β -sheet with the β 2 strand of POTRA1, the N-terminal domain of BamA. In addition, hydrophobic contacts and two salt bridges are present in the same region. The structure explains the result that Asp26_{SurA} is cross-linked to BamA (Wang *et al.*, 2016) because Asp26_{SurA} is located at the center of the SurA/BamA interface. Furthermore, SurA adopts a closed conformation in our model, but the P2 domain rotates $\sim 45^\circ$ from its crystal position (Bitto and McKay, 2002) to engage additional contacts with BamA and BamE (Figure 5d–e). For example, Asn336_{SurA} forms a hydrogen bond with the backbone oxygen atom of Gly275_{BamA}; in the same cross-linking study

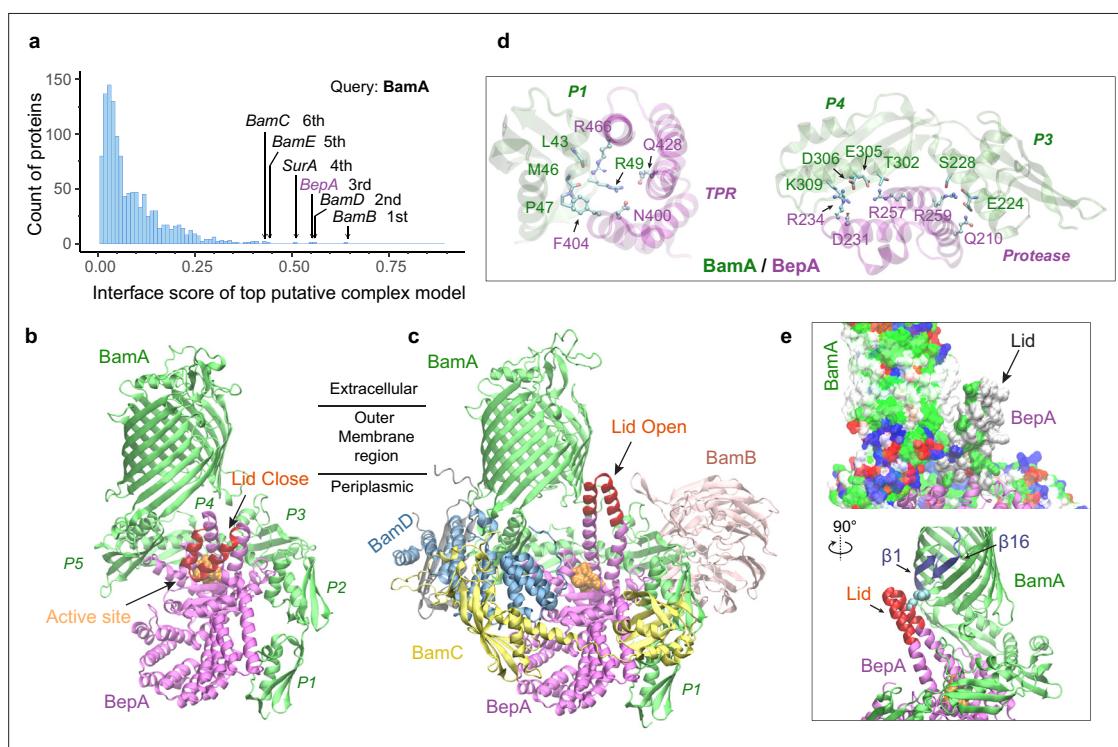


Figure 6. Structural models of β -barrel assembly machine (BAM) and BepA. **(a)** Computational protein-protein interaction screening identifies BepA as a top hit to BamA. **(b)** Top structural model of the heterodimeric complex of BamA (green) and BepA (purple) in cartoon representation. The lid of BepA is colored red. The active sites with the protease domain of BepA are shown in a surface representation (orange). The five POTRA domains of BamA are labeled P1–P5. **(c)** Predicted structure of the BAM/BepA supercomplex. The lid of BepA extends to an open conformation. The image was created from the same viewpoint as **b**. **(d)** Specific residue-residue contacts between BamA and the tetratricopeptide repeat (TPR) and protease domains of BepA. **(e)** Close-up views of the lid and the BamA β -barrel in the surface (top) and cartoon representations (bottom). A hydrophobic contact between Ala180_{BepA} and Leu780_{BamA} is shown as spheres, and the lateral gate of BamA is between the β 1 and β 16 strands (dark blue).

The online version of this article includes the following figure supplement(s) for figure 6:

Figure supplement 1. Predicted structures of BepA compared to two experimental structures.

(Wang et al., 2016), Asn336_{SurA} is implied to interact with BamA. These results invite a hypothesis that SurA and BamA initiate payload transfer via specific docking at their N-terminal domains.

Metalloprotease BepA flips a lid in complex with BAM

SurA led us to BamA, for which we conducted another envelopome PPI screening. The top six hits include all expected ones, the other four BAM factors and SurA, but unexpectedly, BepA ranked third with high confidence (Figure 6a–b). BepA was not anticipated because we were not aware of its relevance. However, a literature search quickly revealed that BepA is highly relevant because it cleans up stalled OMP folding at BAM (Narita et al., 2013; Soltes et al., 2017), and its interactions with BAM have been documented (Narita et al., 2013; Daimon et al., 2017), though no structure of the complex is available. Consequently, we modeled BepA and the full BAM complex altogether, obtaining a high confidence model (iScore = 0.68, mean pLDDT = 85, Figure 6c). Overall, the BepA/BamA interfaces are quite similar (interface C_α RMSD 1.2 Å) between the dimer and the supercomplex structure. Extensive contacts are present between all five periplasmic domains (POTRA1–5) of BamA and the two domains (protease and TPR) of BepA. They can be clustered into two main groups scattered between the POTRA1 and TPR and between POTRA3-4 and the protease domains (Figure 6d). Strong photo cross-linking signals to BamA have been observed previously for several residues of TPR including Phe404_{BepA} (Narita et al., 2013; Daimon et al., 2017). According to our model, Phe404_{BepA} belongs to the first interaction cluster with five BamA residues, particularly Pro47_{BamA}. Another residue cross-linked to BamA is Gln428_{BepA}, which establishes a specific hydrogen bond with Arg49_{BamA} in our model.

Crystal structures of BepA alone have been solved, but a segment (residue 154–192) of the metalloprotease domain is absent in these structures, presumably due to its flexibility (Daimon et al., 2017; Bryant et al., 2020; Shahrizal et al., 2019; Figure 6—figure supplement 1). The segment, termed an ‘active-site lid,’ was speculated to be important to the protease function of BepA because its movement could either expose (open state) or cover (closed state) the catalytic site, the HEXXH motif (Bryant et al., 2020). Intriguingly, we only see the closed conformation in 40 models of the BamA/BepA heterodimer (Figure 6b), but both the open and closed states in 8 supercomplex models. In the open state, two α-helices comprising the lid are in a straight-up configuration (Figure 6c), whose upper half (residue 157–186) rotates ~105° to cover the enzymatic site in the closed configuration (Figure 6b). Moreover, the flexible lid, with a large hydrophobic surface similar to the β-barrel of BamA, is in a position as if it is inserted into the inner leaflet of the OM and is in the proximity of the lateral gate of BamA (Figure 6e). As such, BepA could control its protease activity via probing a substrate β-barrel stemming out of BamA. Hence, it is hypothesized that a normal budding β-barrel blocks the lid opening, whereas a stalled OMP within BamA permits lid opening and subsequent substrate cleavage. It is also possible that the lid of BepA somehow detects an abnormal β-barrel popping out of BamA, as suggested in an experimental study that shows BepA degrades a LptD mutant stalled at a late stage of β-barrel folding (Soltes et al., 2017).

Discussion

Here, we have demonstrated a deep learning strategy that combines virtual PPI screening over the *E. coli* envelopome and supercomplex structure modeling. By applying it to several key proteins in the OMP biogenesis pathway, we have identified their functional partners within the top 1% ranking of ~1450 proteins screened for PPIs per query. Thanks to high confidence structures underlying the top predictions, one can understand many experimental phenomena, particularly in vivo site-directed photo cross-linking data. For example, cross-linked products found from the SecYEG or BAM supercomplexes may be explained by direct physical interactions revealed in our predicted structures. Moreover, previously speculated conformations are captured for SurA and BepA. Most importantly, these revealing atomic structures suggest mechanistic hypotheses for various steps of the OMP biogenesis pathway as summarized in Figure 7, where we present their diagrams along with some predicted supercomplex structures.

One unexpected discovery is the DsbA/PpiD interaction. It was known that DsbA crucially transfers its disulfide bond to a nascent polypeptide translocated by SecYEG (Kadokura and Beckwith, 2009; Goemans et al., 2014), but how does the translocon interact with DsbA? The predicted supercomplex

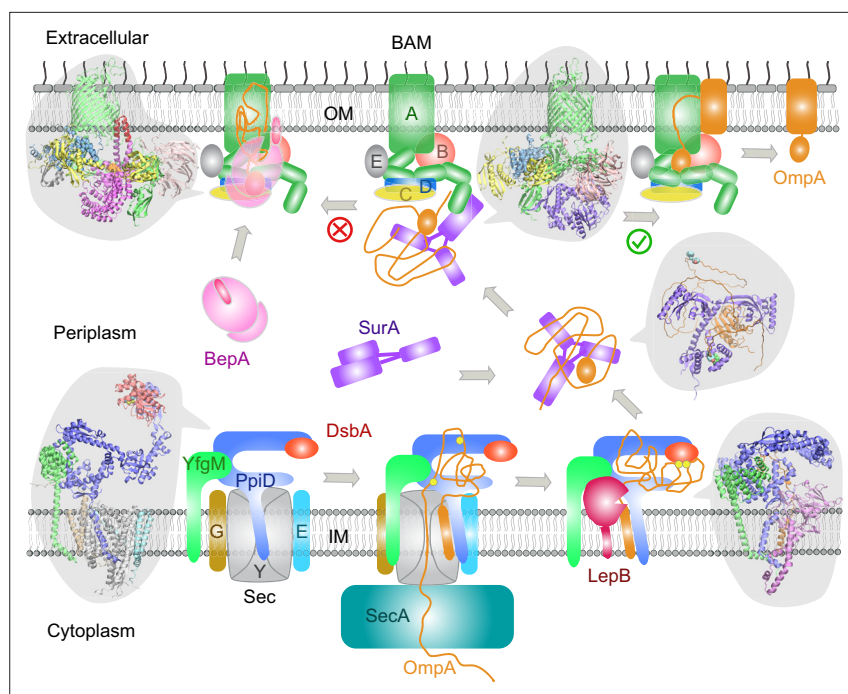


Figure 7. Proposed mechanisms involved in the outer membrane protein (OMP) biogenesis pathway in *E. coli*. Complex structures resulting from this study accompany relevant cartoon diagrams. Powered by SecA, a precursor OmpA polypeptide (orange line) first passes through the SecYEG translocon in complex with PpiD and YfgM. PpiD, held in place by YfgM, senses the translocating substrate via its N-terminal α -helix bound to the lateral gate of SecY and temporarily dissociates from the translocon upon receiving the substrate OmpA. Protein disulfide isomerase DsbA is recruited by PpiD and promotes formation of a disulfide bond between two cysteine residues (yellow spheres) of OmpA. Meanwhile, peptidase LepB fills the vacancy left by SecYEG and cleaves the transmembrane signal peptide from OmpA, which is then handed over to chaperone SurA. At this point, the periplasmic domain of OmpA is folded, but the unfolded β -barrel region wraps around SurA, which carries OmpA to BAM. Lastly, SurA docks to POTRA1, the N-terminal domain of BamA, where the β -barrel domain of OmpA is folded and released from the lateral gate of BamA. If this folding and assembly process are stalled for some reason, metalloprotease BepA senses the failure with its flexible lid and cleans up by cleaving a stuck substrate. For clarity, the peptidoglycan layer in the periplasm is not shown, and the schematic drawings are not to scale.

of SecYEG/PpiD/YfgM/DsbA provides a compelling answer for three reasons: first, the transient residence of DsbA on PpiD at the translocon dramatically improves its efficiency versus a random search in a crowded periplasm. Second, it predicts a function for the PPIase domain of PpiD that mysteriously lacks the PPIase function. Third, since PpiD is anchored to the IM, it can keep DsbA close to DsbB, an IM protein that recycles DsbA. Notably, two other *E. coli* chaperones, SurA and FkpA, possess two PPIase domains each, and AF2Complex predicts that DsbA interacts with FkpA (iScore = 0.47 versus 0.55 for PpiD) but not with SurA (iScore = 0.03).

Interestingly, the AF2 neural networks can model a partially unfolded polypeptide accompanied by chaperones, even though AF2 was trained on folded proteins. We attained structural models that appear to be at least partially physical. One example is the model of PpiD/YfgM/LepB/proOmpA, in which proOmpA is posed for cleavage by peptidase LepB. The cleavage alanine of OmpA is ~ 4 Å away from the catalytic triad of LepB. Considering that only apo or inhibitor bound structures of LepB are available, this computed model implies that AF2 has learned physical representations. More intriguing examples are the models of SurA/OmpA, where the periplasmic domain of OmpA is folded, but the β -barrel domain is completely unfolded and loosely wrapped around SurA. The predicted structures echo the NMR structures of an unfolded polypeptide bound to SecB, a cytosolic chaperone involved in the early stage of the SecA-dependent translocation pathway (Huang et al., 2016). Despite the low confidence due to weak interactions, the predicted structures delineate a picture for how SurA prevents OmpA from aggregating. Moreover, since it transports OmpA with a relatively small number of intermolecular contacts, the free energy required to dissociate OmpA from SurA is

small. Notwithstanding these considerations, we caution that artifacts likely exist in these predicted structural models.

These results reinforce the notion that deep learning is a promising way to explore the conformational ensemble of proteins and to uncover the molecular mechanisms of biosystems (Gao *et al.*, 2022b). The combination of advanced AF2 deep learning models, an effective PPI ranking metric, and a workflow optimized for large-scale screening generates illuminating structures of protein complexes. In the presented examples, we focused on those with obvious biological relevance according to the literature. There are of course other confident PPI predictions that were not described here. While some are promiscuous interactions or physically possible but biologically irrelevant, or simply false, there are likely functional interactions yet to be explored. More generally, these results are an example of a deep learning-based strategy that helps to elucidate mechanistic aspects of complex biochemical pathways.

Methods

E. coli cell envelopome

The proteome for *E. coli* strain K12 MG1655, consisting of 4400 protein sequences, was retrieved from UniProt (Wu *et al.*, 2006) in March 2022 (Proteome ID [UP000000625](https://www.uniprot.org/proteomes/UP000000625)). Then, the subcellular location of each protein was parsed to collect all known and predicted cell envelope proteins. To be conservative, we included all proteins whose primary subcellular location is not the cytosol, yielding 1466 proteins defined as the envelopome. The set contains all proteins located in the IM, periplasm, OM, and extracellular surface, and some cytosolic proteins located primarily at the periphery of the periplasmic IM, e.g., SecA. For AF2Complex modeling, we removed 11 sequences that contain at least one non-standard amino acid.

AF2Complex

An updated version of AF2Complex was built upon AlphaFold version 2.2.0 (Evans *et al.*, 2021). AF2Complex supports three different set of deep learning neural network models provided with AF2, i.e., the original models for monomer prediction (AF version 2.0.1 Jumper *et al.*, 2021) and two set of models for multimer prediction (version 2.1.0 and version 2.2.0; Evans *et al.*, 2021). Numerous changes were made; we list three major ones here. First, the iScore metric was introduced to rank the confidence of a predicted complex model (Gao *et al.*, 2022b). The iScore metric was derived from the predicted alignment errors that give an estimated distance for interface residue j from its position in the experimental structure, as viewed from a local frame of residue interface residue i (Gao *et al.*, 2022b; Jumper *et al.*, 2021). To better estimate confidence, the contribution of each interface residue to the iScore is calculated using local frames not located within the same protein chain, i.e., residue i and j belonging to different chains. Second, the data pipeline for generating input features was split from the neural network inference. This allows rapid assembly of input features for individual proteins of a putative complex from pre-generated input features of the full proteome. Third, five options for multiple sequence alignment (MSA) pairing are provided: no pairing, all paired, cyclic, linear, and arbitrary pairing. The last three options are experimental, and we only employed the first two MSA pairing modes in this work. All structural models, either of a monomer or a multimer, were predicted by AF2Complex using either the 'monomer_ptm' deep learning models without any MSA pairing or the 'multimer_v2' models with all MSA pairing; the latter only used for structure prediction of complexes. The input features of all monomers were derived using standard sequence libraries (The Uniprot Consortium, 2019; Mitchell *et al.*, 2020; Tnyasuvunakool *et al.*, 2021) and a version of the Protein Data Bank (Berman *et al.*, 2000) released in November 2021. The confidence of an output structure is evaluated using three different metrics: the iScore for protein interface (Gao *et al.*, 2022b), the predicted TM-score for global structure of a monomer (Zhang and Skolnick, 2004), and the pLDDT score for local domain structure (Jumper *et al.*, 2021; Mariani *et al.*, 2013). According to AF2, a mean pLDDT score higher than 70 indicates high confidence in a predicted structure when evaluated on individual domain(s).

Envelopome PPI screening

We adapted a workflow for proteome-scale monomeric structure prediction with AF2 implemented on the Summit supercomputer at Oak Ridge National Laboratory (Gao et al., 2022a). The workflow was applied to four query proteins: PpiD, YfgM, SurA, and BamA. Given a query sequence and an envelopome protein sequence, AF2Complex assembles the pre-generated monomeric input features and then feeds the composite features to the AF2 neural network models for inference. 10 different neural network models are used as mentioned above, and each deep learning model is invoked twice with two different random seeds and up to eight recycles. This procedure typically gives 20 final structural models per target for ranking, and the top ranked model by iScore is retained. Although the 'multimer_v2' set of models greatly reduces unphysical clashes in predicted structures compared to structures predicted by the original set of multimer deep learning models, there is still a small chance of generating severe clashes, which could yield an artifactually high iScore. We filter out these structures by applying a minimum interface clash indicator of 0.4 (Gao et al., 2022b). We did not apply model relaxation to these predicted structures because most are not actual protein complexes; as such, it is not worth expending the extra computing time. To avoid memory overflow, we limited the input MSA depth to 5000 for each monomer and allowed a maximum of four structural templates per monomer. A maximum of 1600 amino acids was imposed on the total size of each complex target. This limitation slightly reduced the total number proteins in the envelopome for screening, typically to ~1450 for the four proteins studied. Because about 30% of envelopome proteins have a signal peptide that is absent in their mature chains, only mature chains were used for screening. This is readily realized with AF2Complex because it provides an option to crop arbitrary segments of a monomer input feature during feature assembly. The residue ranges of the mature chains were obtained from the UniProt knowledgebase.

Modeling the OmpA polypeptide

It was necessary to model OmpA as a substrate of PpiD or SurA. OmpA was chosen mainly because it is a model OMP for studying the OMP biogenesis pathway (Reusch, 2012). Many experimental data exist for validation. To minimize potential 'memory' effects due to large MSAs or structural templates, we reduced the number of sequences in the MSAs and removed all structural templates in the input features of OmpA. We tested MSA depths of 1, 10, 20, 50, and 100 to predict a structural model of proOmpA in complex with SecYEG. The goal is to generate a structural model that mimics the crystal structure of SecYEG translocating a proOmpA polypeptide (Li et al., 2016). Our tests found that an MSA depth below 50 can yield a model structure of an unfolded OmpA peptide through the channel of SecY, when used with either 'model_1_ptm' or 'model_3_ptm' of all AF2 neural network models. In this study, we used an MSA depth of 20 and the 2 AF2 models to predict structural models involving OmpA.

Computational resources

E. coli envelopome PPI screening was performed on the Summit supercomputer, typically using 923 nodes for several hours of wall clock time. Each Summit node hosts six Nvidia 16 GB V100 GPUs. The structure predictions of various supercomplexes were conducted locally using about 10 workstations each with four Nvidia RTX6000 GPUs, each with 24 GB of GPU memory.

Analysis

The program VMD (Humphrey et al., 1996) was used to inspect predicted structural models and create all molecular images. APoc was used to align monomeric protein structures (Gao and Skolnick, 2013), and iAlign was used to perform protein-protein interface comparison (Gao and Skolnick, 2010).

Acknowledgements

We thank Jerry M Parks for stimulating discussions and critical reading of the manuscript. This work was supported in part by the DOE Office of Science, Office of Biological and Environmental Research (DOE DE-SC0021303) and the Division of General Medical Sciences of the National Institutes Health (NIH R35GM118039). The research used resources supported in part by the Advanced Scientific

Computing Research (ASCR) Leadership Computing Challenge (ALCC) program, and by the Partnership for an Advanced Computing Environment (PACE) at the Georgia Institute of Technology.

Additional information

Funding

Funder	Grant reference number	Author
U.S. Department of Energy	DE-SC0021303	Jeffrey Skolnick
National Institute of General Medical Sciences	R35GM118039	Jeffrey Skolnick
Advanced Scientific Computing Research		Mu Gao Jeffrey Skolnick
Advanced Computing Environment (PACE)		Jeffrey Skolnick

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Mu Gao, Conceptualization, Resources, Data curation, Software, Formal analysis, Supervision, Funding acquisition, Investigation, Visualization, Methodology, Writing - original draft, Writing - review and editing; Davi Nakajima An, Software, Methodology, Writing - review and editing; Jeffrey Skolnick, Conceptualization, Resources, Supervision, Funding acquisition, Project administration, Writing - review and editing

Author ORCIDs

Mu Gao  <http://orcid.org/0000-0002-0378-3704>

Jeffrey Skolnick  <http://orcid.org/0000-0002-1877-4958>

Decision letter and Author response

Decision letter <https://doi.org/10.7554/eLife.82885.sa1>

Author response <https://doi.org/10.7554/eLife.82885.sa2>

Additional files

Supplementary files

- Supplementary file 1. Top hits from the protein-protein interaction (PPI) screening over the *E. coli* envelopome with AF2Complex for query proteins: PpiD, YfgM, SurA, and BamA.
- MDAR checklist

Data availability

The input features for the full *E. coli* proteome used for envelopome screening by AF2Complex, and the computational models presented in this study are available at Zenodo (<https://doi.org/10.5281/zenodo.6846915>). The source code of AF2Complex is freely available at <https://github.com/FreshAirTonight/af2complex>.

References

- Alvira S, Watkins DW, Troman L, Allen WJ, Lorrigan JS, Degliesposti G, Cohen EJ, Beeby M, Daum B, Gold VA, Skehel JM, Collinson I. 2020. Inter-Membrane association of the Sec and Bam translocons for bacterial outer-membrane biogenesis. *eLife* **9**:e60669. DOI: <https://doi.org/10.7554/eLife.60669>, PMID: 33146611
- Antonovaa R, Fürst M, Nishiyama K-I, Müller M. 2008. The periplasmic chaperone PpiD interacts with secretory proteins exiting from the SecYEG translocon. *Biochemistry* **47**:5649–5656. DOI: <https://doi.org/10.1021/bi800233w>, PMID: 18439025
- Babu M, Bundalovic-Torma C, Calmettes C, Phanse S, Zhang Q, Jiang Y, Minic Z, Kim S, Mehla J, Gagarinova A, Rodionova I, Kumar A, Guo H, Kagan O, Pogoutse O, Aoki H, Deineko V, Caufield JH, Holtzapfel E, Zhang Z,

- et al. 2018. Global landscape of cell envelope protein complexes in *Escherichia coli*. *Nature Biotechnology* **36**:103–112. DOI: <https://doi.org/10.1038/nbt.4024>, PMID: 29176613
- Behrens S, Maier R, de Cock H, Schmid FX, Gross CA. 2001. The sura periplasmic ppiase lacking its parvulin domains functions in vivo and has chaperone activity. *The EMBO Journal* **20**:285–294. DOI: <https://doi.org/10.1093/emboj/20.1.285>, PMID: 11226178
- Behrens-Kneip S. 2010. The role of SurA factor in outer membrane protein transport and virulence. *International Journal of Medical Microbiology* **300**:421–428. DOI: <https://doi.org/10.1016/j.ijmm.2010.04.012>, PMID: 20447864
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000. The protein data bank. *Nucleic Acids Research* **28**:235–242. DOI: <https://doi.org/10.1093/nar/28.1.235>, PMID: 10592235
- Bitto E, McKay DB. 2002. Crystallographic structure of SurA, a molecular chaperone that facilitates folding of outer membrane porins. *Structure* **10**:1489–1498. DOI: [https://doi.org/10.1016/S0969-2126\(02\)00877-8](https://doi.org/10.1016/S0969-2126(02)00877-8), PMID: 12429090
- Bryant JA, Cadby IT, Chong Z-S, Boelter G, Sevastyanovich YR, Morris FC, Cunningham AF, Kritikos G, Meek RW, Banzhaf M, Chng S-S, Lovering AL, Henderson IR. 2020. Structure-Function characterization of the conserved regulatory mechanism of the *Escherichia coli* M48 metalloprotease bepa. *Journal of Bacteriology* **203**:e00434-20. DOI: <https://doi.org/10.1128/JB.00434-20>, PMID: 33106348
- Bryant P, Pozzati G, Elofsson A. 2022. Improved prediction of protein-protein interactions using alphafold2. *Nature Communications* **13**:1265. DOI: <https://doi.org/10.1038/s41467-022-28865-w>, PMID: 35273146
- Calabrese AN, Schiffrin B, Watson M, Karamanos TK, Walko M, Humes JR, Horne JE, White P, Wilson AJ, Kalli AC, Tuma R, Ashcroft AE, Brockwell DJ, Radford SE. 2020. Inter-domain dynamics in the chaperone sura and multi-site binding to its outer membrane protein clients. *Nature Communications* **11**:2155. DOI: <https://doi.org/10.1038/s41467-020-15702-1>, PMID: 32358557
- Carlson ML, Stacey RG, Young JW, Wason IS, Zhao Z, Rattray DG, Scott N, Kerr CH, Babu M, Foster LJ, Duong Van Hoa F. 2019. Profiling the *Escherichia coli* membrane protein interactome captured in peptidic libraries. *eLife* **8**:e46615. DOI: <https://doi.org/10.7554/eLife.46615>, PMID: 31364989
- Collinson I. 2019. The dynamic ATP-driven mechanism of bacterial protein translocation and the critical role of phospholipids. *Frontiers in Microbiology* **10**:1217. DOI: <https://doi.org/10.3389/fmicb.2019.01217>, PMID: 31275252
- Daimon Y, Iwama-Masui C, Tanaka Y, Shiota T, Suzuki T, Miyazaki R, Sakurada H, Lithgow T, Dohmae N, Mori H, Tsukazaki T, Narita S-I, Akiyama Y. 2017. The TPR domain of bepa is required for productive interaction with substrate proteins and the β -barrel assembly machinery complex. *Molecular Microbiology* **106**:760–776. DOI: <https://doi.org/10.1111/mmi.13844>, PMID: 28960545
- Dartigalongue C, Raina S. 1998. A new heat-shock gene, PpiD, encodes a peptidyl-prolyl isomerase required for folding of outer membrane proteins in *Escherichia coli*. *The EMBO Journal* **17**:3968–3980. DOI: <https://doi.org/10.1093/emboj/17.14.3968>, PMID: 9670013
- De Geyter J, Tsigotaki A, Orfanoudaki G, Zorzini V, Economou A, Karamanou S. 2016. Protein folding in the cell envelope of *Escherichia coli*. *Nature Microbiology* **1**:16107. DOI: <https://doi.org/10.1038/nmicrobiol.2016.107>, PMID: 27573113
- Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, Židek A, Bates R, Blackwell S, Yim J, Ronneberger O, Bodenstern S, Zielinski M, Bridgland A, Potapenko A, Cowie A, Tunyasuvunakool K, Jain R, Clancy E, Kohli P, et al. 2021. Protein Complex Prediction with AlphaFold-Multimer. *bioRxiv*. DOI: <https://doi.org/10.1101/2021.10.04.463034>
- Fürst M, Zhou Y, Merfort J, Müller M. 2018. Involvement of PpiD in Sec-dependent protein translocation. *Biochimica et Biophysica Acta. Molecular Cell Research* **1865**:273–280. DOI: <https://doi.org/10.1016/j.bbamcr.2017.10.012>, PMID: 29097228
- Gao M, Skolnick J. 2010. IAlign: a method for the structural comparison of protein-protein interfaces. *Bioinformatics* **26**:2259–2265. DOI: <https://doi.org/10.1093/bioinformatics/btq404>, PMID: 20624782
- Gao M, Skolnick J. 2013. Apoc: large-scale identification of similar protein pockets. *Bioinformatics* **29**:597–604. DOI: <https://doi.org/10.1093/bioinformatics/btt024>, PMID: 23335017
- Gao M, Coletti M, Davidson RB, Prout R, Abraham S, Hernandez B, Sedova A. 2022a. Proteome-scale Deployment of Protein Structure Prediction Workflows on the Summit Supercomputer. 2022 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW). Lyon, France. DOI: <https://doi.org/10.1109/IPDPSW55747.2022.00045>
- Gao M, Nakajima An D, Parks JM, Skolnick J. 2022b. AF2Complex predicts direct physical interactions in multimeric proteins with deep learning. *Nature Communications* **13**:1744. DOI: <https://doi.org/10.1038/s41467-022-29394-2>, PMID: 35365655
- Goemans C, Denoncin K, Collet JF. 2014. Folding mechanisms of periplasmic proteins. *Biochimica et Biophysica Acta* **1843**:1517–1528. DOI: <https://doi.org/10.1016/j.bbamcr.2013.10.014>, PMID: 24239929
- Götzke H, Palombo I, Muheim C, Perrody E, Genevaux P, Kudva R, Müller M, Daley DO. 2014. YfgM is an ancillary subunit of the SecYEG translocon in *Escherichia coli*. *The Journal of Biological Chemistry* **289**:19089–19097. DOI: <https://doi.org/10.1074/jbc.M113.541672>, PMID: 24855643
- Gu Y, Li H, Dong H, Zeng Y, Zhang Z, Paterson NG, Stansfeld PJ, Wang Z, Zhang Y, Wang W, Dong C. 2016. Structural basis of outer membrane protein insertion by the Bam complex. *Nature* **531**:64–69. DOI: <https://doi.org/10.1038/nature17199>, PMID: 26901871
- Hennecke G, Nolte J, Volkmer-Engert R, Schneider-Mergener J, Behrens S. 2005. The periplasmic chaperone SurA exploits two features characteristic of integral outer membrane proteins for selective substrate

- recognition. *The Journal of Biological Chemistry* **280**:23540–23548. DOI: <https://doi.org/10.1074/jbc.M413742200>, PMID: 15840585
- Huang C, Rossi P, Saio T, Kalodimos CG. 2016. Structural basis for the antifolding activity of a molecular chaperone. *Nature* **537**:202–206. DOI: <https://doi.org/10.1038/nature18965>, PMID: 27501151
- Humphrey W, Dalke A, Schulten K. 1996. VMD: visual molecular dynamics. *Journal of Molecular Graphics* **14**:33–38. DOI: [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5), PMID: 8744570
- Humphreys IR, Pei J, Baek M, Krishnakumar A, Anishchenko I, Ovchinnikov S, Zhang J, Ness TJ, Banjade S, Bagde SR, Stancheva VG, Li X-H, Liu K, Zheng Z, Barrero DJ, Roy U, Kuper J, Fernández IS, Szakal B, Branzei D, et al. 2021. Computed structures of core eukaryotic protein complexes. *Science* **374**:eabm4805. DOI: <https://doi.org/10.1126/science.abm4805>, PMID: 34762488
- Ishida H, Garcia-Herrero A, Vogel HJ. 2014. The periplasmic domain of *Escherichia coli* outer membrane protein A can undergo a localized temperature dependent structural transition. *Biochimica et Biophysica Acta* **1838**:3014–3024. DOI: <https://doi.org/10.1016/j.bbamem.2014.08.008>, PMID: 25135663
- Jauss B, Petriman N-A, Drepper F, Franz L, Sachelaru I, Welte T, Steinberg R, Warscheid B, Koch H-G. 2019. Noncompetitive binding of PpiD and YidC to the SecYEG translocon expands the global view on the SecYEG interactome in *Escherichia coli* *The Journal of Biological Chemistry* **294**:19167–19183. DOI: <https://doi.org/10.1074/jbc.RA119.010686>, PMID: 31699901
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, et al. 2021. Highly accurate protein structure prediction with alphafold. *Nature* **596**:583–589. DOI: <https://doi.org/10.1038/s41586-021-03819-2>, PMID: 34265844
- Kadokura H, Beckwith J. 2009. Detecting folding intermediates of a protein as it passes through the bacterial translocation channel. *Cell* **138**:1164–1173. DOI: <https://doi.org/10.1016/j.cell.2009.07.030>, PMID: 19766568
- Kumazaki K, Chiba S, Takemoto M, Furukawa A, Nishiyama K, Sugano Y, Mori T, Dohmae N, Hirata K, Nakada-Nakura Y, Maturana AD, Tanaka Y, Mori H, Sugita Y, Arisaka F, Ito K, Ishitani R, Tsukazaki T, Nureki O. 2014. Structural basis of Sec-independent membrane protein insertion by YidC. *Nature* **509**:516–520. DOI: <https://doi.org/10.1038/nature13167>, PMID: 24739968
- Li L, Park E, Ling J, Ingram J, Ploegh H, Rapoport TA. 2016. Crystal structure of a substrate-engaged SecY protein-translocation channel. *Nature* **531**:395–399. DOI: <https://doi.org/10.1038/nature17163>, PMID: 26950603
- Li G, He C, Bu P, Bi H, Pan S, Sun R, Zhao XS. 2018. Single-Molecule detection reveals different roles of Skp and SurA as chaperones. *ACS Chemical Biology* **13**:1082–1089. DOI: <https://doi.org/10.1021/acscchembio.8b00097>, PMID: 29543429
- Maddalo G, Stenberg-Bruzell F, Götzke H, Toddo S, Björkholm P, Eriksson H, Chovanec P, Genevaux P, Lehtiö J, Ilag LL, Daley DO. 2011. Systematic analysis of native membrane protein complexes in *Escherichia coli*. *Journal of Proteome Research* **10**:1848–1859. DOI: <https://doi.org/10.1021/pr101105c>, PMID: 21210718
- Mariani V, Biasini M, Barbato A, Schwede T. 2013. LDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics* **29**:2722–2728. DOI: <https://doi.org/10.1093/bioinformatics/btt473>, PMID: 23986568
- Marx DC, Plummer AM, Faustino AM, Devlin T, Roskopf MA, Leblanc MJ, Lessen HJ, Amann BT, Fleming PJ, Krueger S, Fried SD, Fleming KG. 2020. Sura is a cryptically grooved chaperone that expands unfolded outer membrane proteins. *PNAS* **117**:28026–28035. DOI: <https://doi.org/10.1073/pnas.2008175117>, PMID: 33093201
- Matern Y, Barion B, Behrens-Kneip S. 2010. Ppid is a player in the network of periplasmic chaperones in *Escherichia coli*. *BMC Microbiology* **10**:251. DOI: <https://doi.org/10.1186/1471-2180-10-251>, PMID: 20920237
- Mitchell AL, Almeida A, Beracochea M, Boland M, Burgin J, Cochrane G, Crusoe MR, Kale V, Potter SC, Richardson LJ, Sakharova E, Scheremetjev M, Korobeynikov A, Shlemov A, Kunyavskaya O, Lapidus A, Finn RD. 2020. MGnify: the microbiome analysis resource in 2020. *Nucleic Acids Research* **48**:D570–D578. DOI: <https://doi.org/10.1093/nar/gkz1035>, PMID: 31696235
- Narita S, Masui C, Suzuki T, Dohmae N, Akiyama Y. 2013. Protease homolog bepa (yfgc) promotes assembly and degradation of β -barrel membrane proteins in *Escherichia coli*. *PNAS* **110**:E3612–E3621. DOI: <https://doi.org/10.1073/pnas.1312012110>, PMID: 24003122
- Noiraj N, Gumbart JC, Buchanan SK. 2017. The β -barrel assembly machinery in motion. *Nature Reviews. Microbiology* **15**:197–204. DOI: <https://doi.org/10.1038/nrmicro.2016.191>, PMID: 28216659
- Oswald J, Njenga R, Natriashvili A, Sarmah P, Koch HG. 2021. The dynamic secyeg translocon. *Frontiers in Molecular Biosciences* **8**:664241. DOI: <https://doi.org/10.3389/fmolb.2021.664241>, PMID: 33937339
- Paetzel M. 2014. Structure and mechanism of *Escherichia coli* type I signal peptidase. *Biochimica et Biophysica Acta* **1843**:1497–1508. DOI: <https://doi.org/10.1016/j.bbamcr.2013.12.003>, PMID: 24333859
- Pautsch A, Schulz GE. 1998. Structure of the outer membrane protein a transmembrane domain. *Nature Structural Biology* **5**:1013–1017. DOI: <https://doi.org/10.1038/2983>, PMID: 9808047
- Rapoport TA, Li L, Park E. 2017. Structural and mechanistic insights into protein translocation. *Annual Review of Cell and Developmental Biology* **33**:369–390. DOI: <https://doi.org/10.1146/annurev-cellbio-100616-060439>, PMID: 28564553
- Reusch RN. 2012. Insights into the structure and assembly of *Escherichia coli* outer membrane protein A. *The FEBS Journal* **279**:894–909. DOI: <https://doi.org/10.1111/j.1742-4658.2012.08484.x>, PMID: 22251410

- Rollauer SE**, Soreshjani MA, Noinaj N, Buchanan SK. 2015. Outer membrane protein biogenesis in gram-negative bacteria. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **370**:1679. DOI: <https://doi.org/10.1098/rstb.2015.0023>, PMID: 26370935
- Sachelaru I**, Petriman NA, Kudva R, Koch HG. 2014. Dynamic interaction of the Sec translocon with the chaperone PpiD. *The Journal of Biological Chemistry* **289**:21706–21715. DOI: <https://doi.org/10.1074/jbc.M114.577916>, PMID: 24951590
- Shahrizal M**, Daimon Y, Tanaka Y, Hayashi Y, Nakayama S, Iwaki S, Narita S-I, Kamikubo H, Akiyama Y, Tsukazaki T. 2019. Structural basis for the function of the β -barrel assembly-enhancing protease bepa. *Journal of Molecular Biology* **431**:625–635. DOI: <https://doi.org/10.1016/j.jmb.2018.11.024>, PMID: 30521812
- Silhavy TJ**, Kahne D, Walker S. 2010. The bacterial cell envelope. *Cold Spring Harbor Perspectives in Biology* **2**:a000414. DOI: <https://doi.org/10.1101/cshperspect.a000414>, PMID: 20452953
- Sklar JG**, Wu T, Kahne D, Silhavy TJ. 2007. Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in *Escherichia coli*. *Genes & Development* **21**:2473–2484. DOI: <https://doi.org/10.1101/gad.1581007>, PMID: 17908933
- Soltes GR**, Martin NR, Park E, Sutterlin HA, Silhavy TJ, O'Toole G. 2017. Distinctive roles for periplasmic proteases in the maintenance of essential outer membrane protein assembly. *Journal of Bacteriology* **199**:e00418-17. DOI: <https://doi.org/10.1128/JB.00418-17>, PMID: 28784813
- The Uniprot Consortium**. 2019. Uniprot: a worldwide hub of protein knowledge. *Nucleic Acids Research* **47**:D506–D515. DOI: <https://doi.org/10.1093/nar/gky1049>
- Tomasek D**, Kahne D. 2021. The assembly of β -barrel outer membrane proteins. *Current Opinion in Microbiology* **60**:16–23. DOI: <https://doi.org/10.1016/j.mib.2021.01.009>, PMID: 33561734
- Tomassen J**. 2010. Assembly of outer-membrane proteins in bacteria and mitochondria. *Microbiology* **156**:2587–2596. DOI: <https://doi.org/10.1099/mic.0.042689-0>, PMID: 20616105
- Troman L**, Collinson I. 2021. Pushing the envelope: the mysterious journey through the bacterial secretory machinery, and beyond. *Frontiers in Microbiology* **12**:782900. DOI: <https://doi.org/10.3389/fmicb.2021.782900>, PMID: 34917061
- Tsukazaki T**. 2018. Structure-based working model of secdf, a proton-driven bacterial protein translocation factor. *FEMS Microbiology Letters* **365**:12. DOI: <https://doi.org/10.1093/femsle/fny112>, PMID: 29718185
- Tunyasuvunakool K**, Adler J, Wu Z, Green T, Zielinski M, Židek A, Bridgland A, Cowie A, Meyer C, Laydon A, Velankar S, Kleywegt GJ, Bateman A, Evans R, Pritzel A, Figurnov M, Ronneberger O, Bates R, Kohl SAA, Potapenko A, et al. 2021. Highly accurate protein structure prediction for the human proteome. *Nature* **596**:590–596. DOI: <https://doi.org/10.1038/s41586-021-03828-1>, PMID: 34293799
- Varadi M**, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, Židek A, Green T, Tunyasuvunakool K, Petersen S, Jumper J, Clancy E, Green R, Vora A, Lutfi M, Figurnov M, et al. 2022. AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research* **50**:D439–D444. DOI: <https://doi.org/10.1093/nar/gkab1061>, PMID: 34791371
- Wang Y**, Wang R, Jin F, Liu Y, Yu J, Fu X, Chang Z. 2016. A supercomplex spanning the inner and outer membranes mediates the biogenesis of β -barrel outer membrane proteins in bacteria. *The Journal of Biological Chemistry* **291**:16720–16729. DOI: <https://doi.org/10.1074/jbc.M115.710715>, PMID: 27298319
- Weininger U**, Jakob RP, Kovermann M, Balbach J, Schmid FX. 2010. The prolyl isomerase domain of PpiD from *Escherichia coli* shows a parvulin fold but is devoid of catalytic activity. *Protein Science* **19**:6–18. DOI: <https://doi.org/10.1002/pro.277>, PMID: 19866485
- Wu CH**, Apweiler R, Bairoch A, Natale DA, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Mazumder R, O'Donovan C, Redaschi N, Suzek B. 2006. The universal protein resource (uniprot): an expanding universe of protein information. *Nucleic Acids Research* **34**:D187–D191. DOI: <https://doi.org/10.1093/nar/gkj161>, PMID: 16381842
- Wu R**, Stephenson R, Gichaba A, Noinaj N. 2020. The big bam theory: an open and closed case? *Biochimica et Biophysica Acta. Biomembranes* **1862**:183062. DOI: <https://doi.org/10.1016/j.bbmem.2019.183062>, PMID: 31520605
- Xu X**, Wang S, Hu YX, McKay DB. 2007. The periplasmic bacterial molecular chaperone SurA adapts its structure to bind peptides in different conformations to assert a sequence preference for aromatic residues. *Journal of Molecular Biology* **373**:367–381. DOI: <https://doi.org/10.1016/j.jmb.2007.07.069>, PMID: 17825319
- Zhang Y**, Skolnick J. 2004. Scoring function for automated assessment of protein structure template quality. *Proteins* **57**:702–710. DOI: <https://doi.org/10.1002/prot.20264>, PMID: 15476259