Sequence Coevolution Reveals an Intermediate Step in BamA Assisted Insertion of Outer Membrane Protein FadL

by

Jaden Anderson

Submitted to the graduate degree program in the Department of Molecular Biosciences and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Masters in Biophysics and Biochemistry

University of Kansas

Chair: Joanna S.G. Slusky

Member 2: Jamie R. Walters

Member 3: Krzysztof Kuczera

Date Defended: May 5th, 2021

The thesis committee for Jaden Anderson certifies that this is the approved version of the following thesis:

Sequence Coevolution Reveals an Intermediate Step in BamA Assisted Insertion of Outer Membrane Protein FadL

Chair: Joanna S.G. Slusky

Date Approved: May 5th, 2021

Abstract

A unique characteristic of outer membrane proteins (OMPs) is that they almost exclusively exist as β-barrels. In *E. coli*, the β-Barrel Assembly Machine (BAM) complex catalyzes insertion of OMPs from the periplasm into the outer membrane. Experimental evidence has shown that the outer membrane embedded β-barrel portion of the BAM complex BamA is essential for insertion function (1,2). However, the precise mechanism of insertion is still unknown. To better understand the interactions between BamA and its client OMPs, we explore a computational approach analyzing sequence coevolution in joint-Multiple Sequence Alignments (joint-MSAs) to determine inter-protein residue contacts between BamA and client OMPs. Our computational results using RaptroX revealed an anti-parallel contact between strand five of BamA and strand 14 of FadL. Experimentally characterizing this contact revealed that mutation does not affect the *in vitro* BAM-unassisted insertion of FadL into large unilamellar vesicles. However, single mutations on FadL that are hypothesized to contact the interior region of strand five of BamA affect the *in vitro* BAM-assisted insertion of FadL.

Table of Contents

Introduction 1
Results 5
Joint Multiple Sequence Alignment (joint-MSAs) and Sequences/Length (ω) 5
BamA FadL Contact 7
In vitro BAM-unassisted insertion of client OMP FadL 10
In vitro BAM-Assisted Insertion 11
Discussion 13
Methods 14
Joint Multiple Sequence Alignment Generation 14
Contact Map and Mutation Selection 15
Cloning of FadL and Mutations 15
Large Scale Protein Expression and Preparation for Inclusion Bodies 15
Preparation of Large Unilamellar Vesicles (LUVs) 16
Folding and SDS-PAGE 16
Analyzing SDS-PAGE Gels and Densitometry 17
References 18

Introduction

Gram-negative bacteria are bound by two separate lipid bilayers: the inner membrane and the outer membrane. Bacterial proteins in the inner membrane adopt an α-helical fold, while bacterial outer membrane proteins (OMPs) exist almost exclusively in the anti-parallel β-barrel conformation (Ralf Koebnik, 2000). In a β-barrel, the residues are oriented such that the inside facing residues point in the direction of the hydrophilic lumen, and the outside facing residues point in the direction of the hydrophobic membrane (Dhar, Feehan, & Slusky, 2021), and due to the predominate fold being a β-barrel, OMPs are structurally similar to each other. Though structurally similar, they vary widely in size from 8-36 stranded OMPs, and their sequence diversifies more than that of water-soluble proteins (Olivella, Gonzalez, Pardo, & Deupi, 2013) (Sojo, Dessimoz, Pomiankowski, & Lane, 2016). This variety in size and diversity in sequence allows OMPs to adopt several critical functions needed for bacterial survival such as assembly complexes (Cho et al., 2014), porins (Burns & Smith, 1987), import (Spector et al., 2010), efflux (Du et al., 2014), and adhesion (Krachler, Ham, & Orth, 2011).

OMPs are known to spontaneously fold and insert *in vitro* into various vesicles of different acyl chain lengths (Burgess, Dao, Stanley, & Fleming, 2008). However, OMPs do not insert into the membranes of cells spontaneously. An assembly complex is required.

The assembly complex in gram-negative bacteria is the β-barrel assembly machinery (BAM) complex which accelerates OMP folding and membrane insertion which is an essential function to bacterial survival (Noinaj et al., 2013). In *Escherichia coli*, the BAM complex is comprised of five proteins: an embedded 16-stranded OMP BamA that has five polypeptide transport-associated (POTRA) domains that extend into the periplasm, and four lipoproteins BamB-E (Seokhee Kim, 2007). To access these assembly complexes, OMPs must first be synthesized into the cytosol. From there, OMPs are post- or co-translationally directed to the Sec machinery and transported from the cytoplasm to the periplasm (Ruiz-Perez et al., 2009). Once in the periplasm, unfolded OMPs interact with a variety of chaperones,

like SurA, Skp, and DegP, to assist in transport to assembly complexes and prevent protein aggregation and misfolding (Sklar, Wu, Kahne, & Silhavy, 2007) (Ruiz-Perez et al., 2009). The essential components of this complex are BamA, which is a part of the Omp85 superfamily of OMPs that function as assembly factors (Gentle, Burri, & Lithgow, 2005), and BamD (Malinverni et al., 2006) (Lee et al., 2018). Though not required for client OMP insertion, BamD has been shown to regulate the protein-protein interactions between BamA and BamC, BamE, and client OMPs (Lee et al., 2018). BamA has a weak hydrogen bonding pattern between the first and last strand of the β-barrel that exhibits an open or closed conformation known as the lateral gate (Noinaj et al., 2013). It has been demonstrated that the arrest of this lateral gate is detrimental to OMP folding and insertion (Noinaj et al., 2013) (Doerner & Sousa, 2017). The presence of this lateral gate has led to research that has proposed steps to OMP insertion.

One of the initial steps is that the exposed first and last strand of BamA interacts with the client OMP starting at the C-terminal strand (Gruss et al., 2013) (Robert et al., 2006). This allows the client OMP to create a hybrid-intermediate with BamA (Gruss et al., 2013), suggesting that there is direct contact between BamA and client OMPs.

Experimental evidence using crosslinking has demonstrated that client OMPs do form hybrid-intermediates with BamA (Doyle & Bernstein, 2019) (Ieva, Tian, Peterson, & Bernstein, 2011) (Lee et al., 2018) (Lee et al., 2019). An N-terminally truncated form of the auto transporter EspP makes hybrid-intermediates that has asymmetric interactions with the BamA lateral gate (Doyle & Bernstein, 2019). In this intermediate, the C-terminal strand of EspP contacts the first strand of BamA in a rigid fashion, and the N-terminal strand of EspP was shown to have more dynamic and asymmetric interactions with strands fifteen and sixteen of BamA (Doyle & Bernstein, 2019). This suggests that the steps involving the lateral gate can be quite dynamic.

LptD, a twenty-six stranded OMP, has also been shown to interact with BamA (Lee et al., 2019). C-terminal strand of LptD was shown to form hybrid-intermediates with the lateral gate of BamA and the interior strands of BamA located away from the lateral gate (Lee et al., 2019). The interior strands two and ten of BamA were also shown to be important for contact with LamB, an eighteen stranded OMP, and OmpF, a sixteen stranded OMP, suggesting that interior contacts to BamA could be common and important step for OMP insertion (Lee et al., 2019).

The C-terminal region where the β-signal of OMPs is localized is known to be conserved (Hoang et al., 2011). The β-signal has been shown to be an important factor for OMP signaling and recognition (Robert et al., 2006), and experimental results have shown that BamA interacts with the C-terminal strands of client OMPs (Robert et al., 2006) (Doyle & Bernstein, 2019) (Lee et al., 2019). To probe the possibility of an evolutionary relationship between BamA and client OMPs, we utilize at computational approach known as sequence coevolution. This method has been useful in revealing protein-protein interactions between proteins with and without a known structure (Goh & Cohen, 2002).

There exists a variety of sequence coevolution algorithms that utilize mutual information, direct coupling analyses , and machine learning (Simonetti, Teppa, Chernomoretz, Nielsen, & Marino Buslje, 2013) (Kamisetty, 2013) (Wang, Li, Yu, & Xu, 2017). RaptorX is a sequence coevolution algorithm that ranked number 1 in CASP12 (Schaarschmidt, Monastyrskyy, Kryshtafovych, & Bonvin, 2018) and CASP13 (Senior et al., 2019) in contact prediction and ranks highly in the fully automated, live benchmark CAMEO (Peng & Xu, 2011). This algorithm excels in structure and contact prediction for protein sequences without close homologs in the Protein Data Bank for soluble and membrane proteins alike (Wang et al., 2017), and utilizes a combinatorial approach with mutual information, direct machine learning.

RaptorX generates intra-protein contact maps using the evolutionary information present in multiple sequence alignments (MSAs). In this study, we created joint multiple sequence alignments

(joint-MSAs) for use with RaptorX to determine inter-protein contacts, specifically between the BamA β-barrel and client OMPs. The BamA-client OMP maps revealed an anti-parallel β-strand contact between strand five of BamA and the C-teriminal strand fourteen of FadL. FadL is an essential fourteen stranded OMP that imports long-chain fatty acids (Ginsburgh, Black, & Nunn, 1984). We then experimentally investigated this hypothesized contact between BamA and FadL by mutating FadL positions in the contact and determining the mutational effect on BamA-unassisted and BamA-assisted insertion. We found that single mutations on the fourteenth strand of FadL do not affect BamA-unassisted folding. However, the mutations affect BamA-assisted folding when mutations are on the lipid-facing surface of FadL that is hypothesized to contact the hydrophilic-facing lumen of BamA.

Results

Joint Multiple Sequence Alignment (joint-MSAs) and Sequences/Length (ω)

Protein	Strands	BamA Residues	Client Residues	Total Parent Length	Total Sequences	ω
BamA TolC	12	385	428	813	2508	3.08
BamA BtuB	22	385	590	975	2304	2.36
BamA FhuA	22	385	706	1091	2259	2.07
BamA FadL	14	385	427	812	1366	1.68
BamA FhaC	16	385	554	939	1539	1.64
BamA LptD	26	385	577	962	1424	1.48
BamA OmpW	8	385	197	582	788	1.35
BamA OmpA	8	385	171	556	707	1.27
BamA OprH	8	385	179	564	715	1.27
BamA EstA	12	385	632	1017	1235	1.21
BamA PorB	16	385	327	712	694	0.97
BamA Omp32	16	385	332	717	686	0.96
BamA OmpF	16	385	340	725	684	0.94
BamA OmpC	16	385	346	731	649	0.89
BamA OprO	16	385	429	814	695	0.85
BamA OmpLA	12	385	275	660	492	0.75
BamA YadA	12	385	105	490	298	0.61
BamA FimD	24	385	558	943	373	0.40
BamA CarO	8	385	233	618	232	0.38
BamA VDAC	19	385	295	680	250	0.37
BamA CarO1	8	385	255	640	194	0.30
BamA OpdN	18	385	406	791	234	0.30
BamA Intimin	12	385	242	627	181	0.29
BamA Invasin	12	385	245	630	176	0.28
BamA Secretin	16	385	311	696	142	0.20
BamA OmpT	10	385	297	682	125	0.18
BamA AIL	8	385	157	542	77	0.14
BamA NanC	12	385	214	599	78	0.13
BamA AlgE	18	385	458	843	49	0.06
BamA OmpG	14	385	280	665	25	0.04

Table 1 Joint-MSA Characteristics: The resulting table of joint-MSAs. Table includes protein name as BamA Client Protein, number of strands for the client OMP, number of residues for BamA (4N75), number of residues for Client OMP, Total Parent Length which is the combination of both BamA sequence length and Client OMP sequence length, total sequences in the joint-

MSA, and the ω score the joint-MSA has. Above the black line indicates joint-MSAs with an ω above 1.5

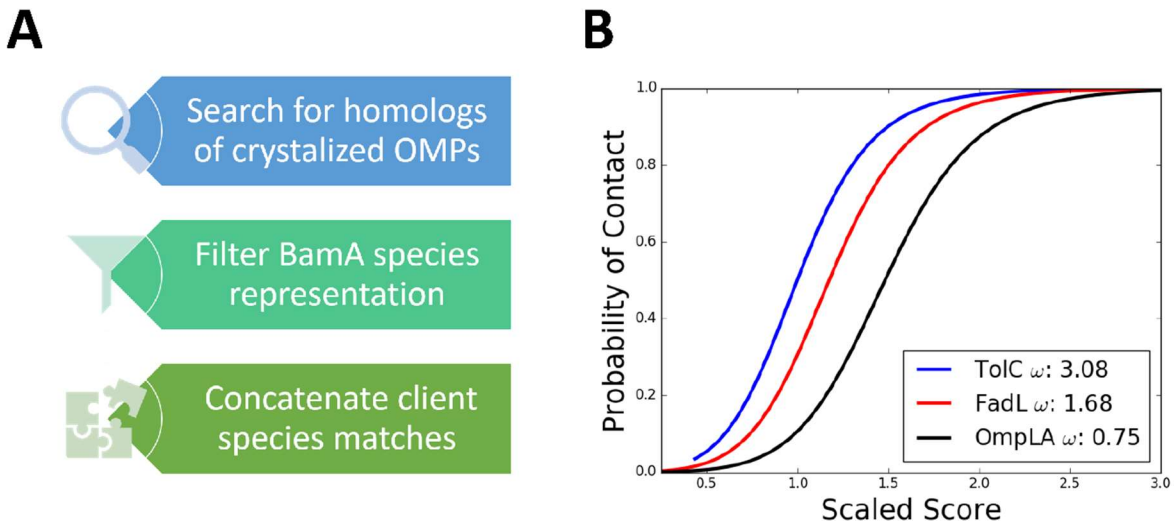


Figure 1 Joint-MSA Screening A) Outline of joint-MSA protocol. First we use HHBlits to search for homologs of crystalized OMPs. Then we use the resulting MSA to build a BamA species representation to be applied to client OMP MSAs. Finally, we concatenate BamA Client species matches into a joint-MSA. B) GREMLIN probability distribution of RaptorX contact prediction results to narrow down BamA Client OMPs for experimental characterization. Proteins with high number of ω shown, TolC in blue, FadL in red, and OmpLA in black.

RaptorX contact map analysis was applied to thirty-one OMP joint-MSAs (Table 1). Number of sequences per sequence length is a measure of relative sequence information present in a MSA and we shall define as ω. This metric is a ratio between the number of sequences in the MSA to the length of the parent sequence (Kamisetty, 2013). This ratio can be used as a cutoff for information content. GREMLIN, a sequence coevolution algorithm, demonstrates that a ω score greater than 1.5 is valuable for analyzing contact prediction (Kamisetty, 2013). Proteins with a higher ω score have a greater confidence in contact prediction (Figure 1). Scaled score is a metric of how far away from the average score a resulting contact probability is (Figure 1). The probability distribution

Equation 1

$$Prob = P(contact \mid Scaled\ Score, \omega)$$

depends on ω and contact prediction results for any BamA Client joint-MSA (Figure 1B) (Kamisetty, 2013). The resulting distribution from Equation 1 allows us to visualize the effect of ω on probability of contact. Using this in tandem with the resulting contact map helps assess the quality of information the contact map provides. The joint-MSA protocol (Figure 1A) described in methods produced several BamA client joint-MSAs for RaptorX analysis with an ω above 1.5 (Table 1). BamA-Client contact maps were produced by RaptorX to characterize coevolutionary signals between these proteins and then processed to predict contact probabilities.

BamA FadL Contact

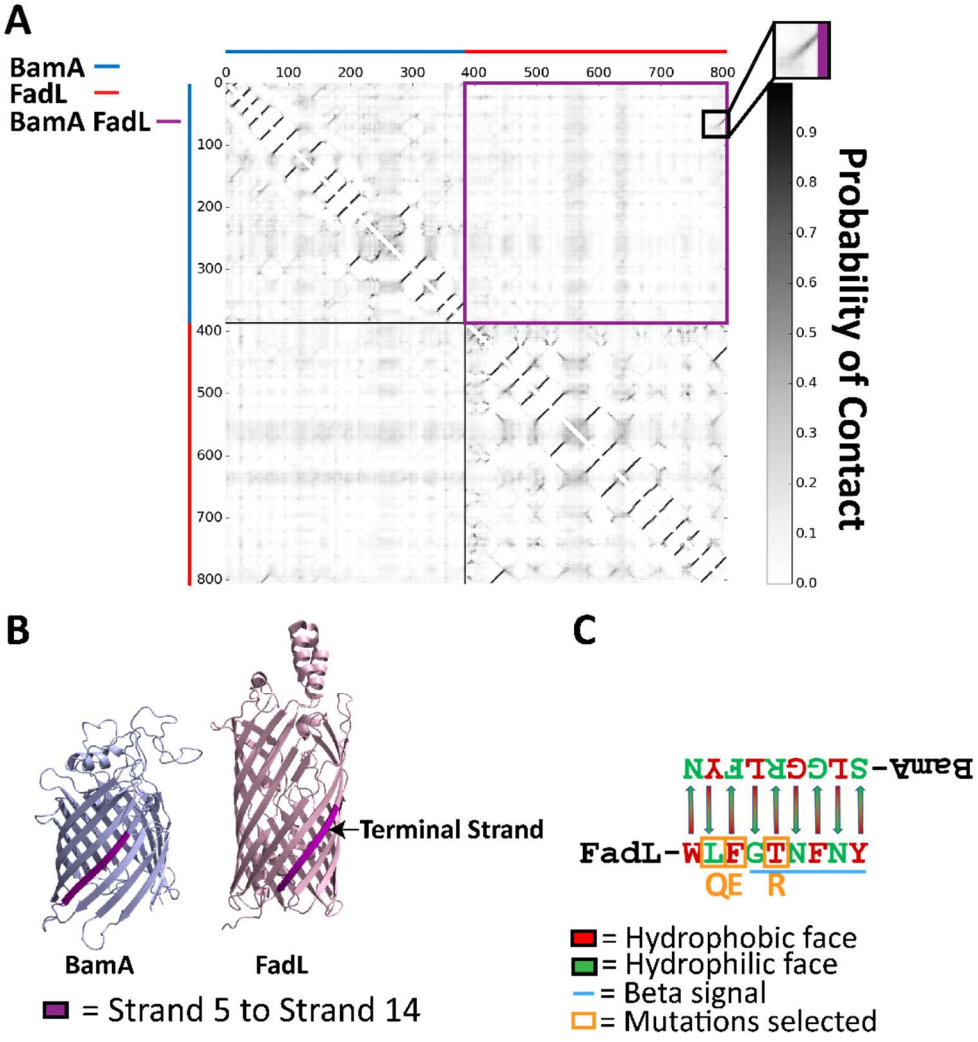


Figure 2 RaptorX Contact Prediction A) The resulting contact map prediction from the RaptorX contact prediction. BamA intra-protein contacts are bounded by blue space (Quadrant 4 of contact

map). FadL intra-protein contacts are bounded by red space (Quadrant 2 of contact map). BamA FadL inter-protein contacts are bounded by purple (Quadrant 1 & 3). There is an anti-parallel beta contact between BamA strand five and FadL strand fourteen (black box). **B)** Inter-protein contact mapped onto the structure of BamA (PDBid: 4N75, blue) and FadL (PDBid: 1T16, red). Purple strands are the predicted contact between BamA and FadL. **C)** Sequence of contact between BamA and FadL. Green colored amino acids are residues pointing into the hydrophilic lumen inside of the folded β-barrel. Red colored amino acids are residues pointing into the hydrophobic membrane exterior to the folded beta barrel. Underlined in blue is the β-signal of FadL which is present on the terminal strands of β-barrels. Orange boxes label selected residues for mutation. The identity that the position was mutated to is noted in orange below the native identity.

The RaptorX contact map (Figure 2A) between BamA and client FadL contained an anti-parallel β-strand contact between the fifth strand of BamA and the C-terminal strand of FadL (Figure 2A, square box and Figure 2B). The sequence for the nine contacts between BamA and FadL (Figure 2C) includes part of the β-signal. The β-signal of OMPs have been shown to be important for interaction with BamA (Hoang et al., 2011). Sequences for the contact are orientated such that the hydrophobic face of FadL contacts the hydrophilic face of BamA and vice versa. This is the opposite pattern seen in native β-barrels (SUPPLEMENT).

The RaptorX sequence coevolution results show both sides of the β-strands interacting with each other (Figure 2C). In this case, what becomes the inward-facing side of the FadL strand is predicted to contact the outward-facing side of the BamA strand and the outward-facing side of the FadL strand is predicted to contact the inward-facing contact of BamA (Figure 2C). It is unlikely that both sides of the strands interact with one another at the same time. Sequence coevolution results often place statistical dependencies across distal pairs of predicted interacting residues (Burger & van Nimwegen, 2010). To illustrate this point, hypothetical contact pairs A-B, B-C, and D-E exist in some space and have a high statistical dependency (Burger & van Nimwegen, 2010). Since A-B and B-C are strongly linked through these calculations, algorithms will often put strong dependence on A-C (Burger & van Nimwegen, 2010). This inferred interaction between A-C muddies the water, and puts less dependency on D-E due to

proximity (Burger & van Nimwegen, 2010). This often makes it difficult to ascertain direct interacting from indirect interacting residues (Burger & van Nimwegen, 2010). Thus, we experimentally characterize the hypothesized contact between BamA and FadL to determine if FadL contacts the interior hydrophilic face or the exterior hydrophobic face of BamA.

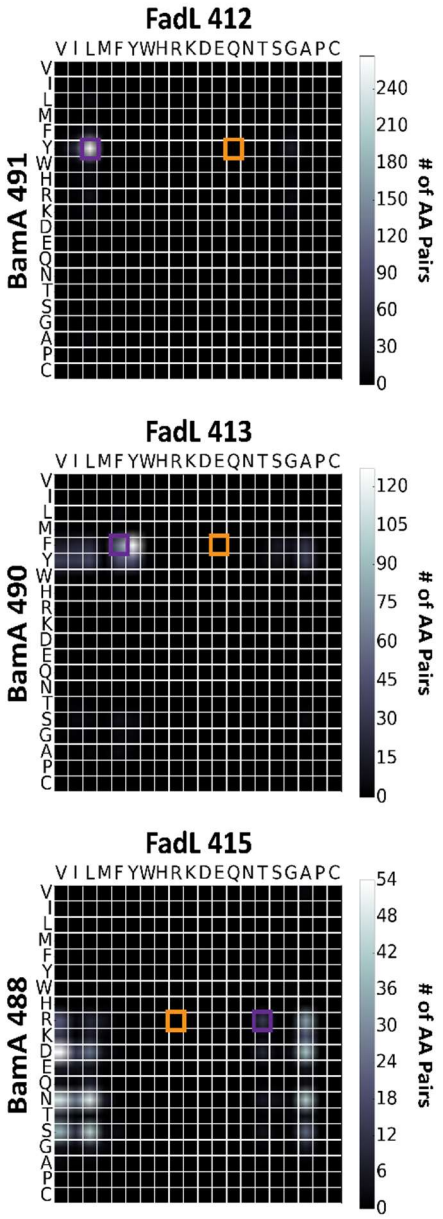


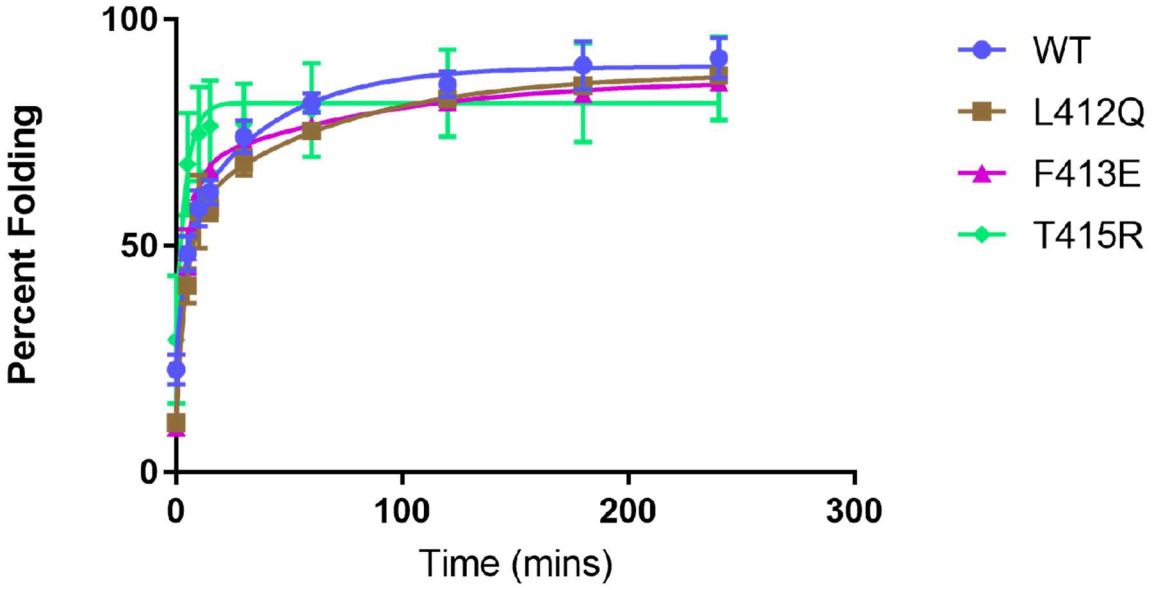
Figure 3 Mutation Selection Heatmaps of positions BamA 491 FadL412, BamA 490 FadL 413, and BamA 488 FadL. Purple squares denote the native *E. coli* x *E. coli* contact between BamA and FadL at that position. Positions that are dispreferred by FadL evolution were chosen to mutate (orange boxes). L412Q, F413E, and T415R were chosen because FadL dispreferred large, charged residues in these positions (# of AA pairs = 0).

To ascertain if the interaction could be disrupted and if the interaction is between the outward face of FadL strand fourteen and the inward face of BamA strand five or the inward face of FadL strand fourteen and that outward face of BamA strand five, we selected three positions to mutate. Nine contacts between BamA and FadL were characterized using the joint-MSA to plot heatmaps of the amino acid pairs that exist between BamA and FadL at the contact positions in the joint-MSA (SUPPLEMENT). Figure 3 shows the three heatmaps of chosen positions for mutation. To choose what amino acid to replace the native residues, we analyzed the inter-protein contacts present in the joint-MSA. We chose combinations that are dispreferred by evolution, FadL L412Q, F413E, and T415R. These combinations may prevent contact between BamA

and FadL. Thus, mutations on FadL that were not prevalent in combination with the native BamA amino acid identity were chosen (Figure 3). If these positions do contact each other during insertion, positions that are dispreferred by evolution may prevent contact between these proteins and thereby hinder insertion. If the positions do not contact, then a position dispreferred by evolution may not affect insertion. Two of the FadL mutants (F413E and T415R) are predicted by the contact map to interact with the lumen of BamA, and one FadL mutant (L412Q) is predicted to interact with the exterior of BamA.

In vitro BAM-unassisted insertion of client OMP FadL

A



B

	WT	L412Q	F413E	T415R
KFast	0.25±0.10s^{-1}	0.20±0.04s^{-1}	0.22±0.02s^{-1}	0.37±0.18s^{-1}
KSlow	0.03±0.01s^{-1}	0.02±0.01s^{-1}	0.01±0.01s^{-1}	0.01±0.02s^{-1}

Figure 4 BAM-Unassisted *in vitro* Insertion of FadL A) Time scale of folding percent of the *in vitro* BamA-unassisted folding of FadL. L412Q (brown), F413E (purple), and T415R (green), all overlap with the Wildtype folding values (blue). B) Table showing the kinetics rates, K_{fast} and K_{slow}, of the *in vitro*

BamA-unassisted folding of FadL. All K_{fast} rates overlap with the wildtype values. All K_{slow} values but F413E overlap with the wildtype values.

To determine if the mutations to the exterior or interior sides of the C-terminal FadL strand affects FadL folding, we folded FadL *in vitro* into large unilamellar vesicles (LUVs). We found that N-cyclohexyl-2-hydroxyl-3-aminopropanesulfonic acid (CAPS) buffer at pH 10.5 was required promote insertion into LUVs. FadL, like most OMPs, is thermodynamically stable enough to run folded on SDS PAGE (Burgess et al., 2008). Upon boiling, OMPs exhibit a shift in molecular weight from a folded to unfolded product, thus we can characterize the folding rates of FadL using densitometry (Burgess et al., 2008). This phenomenon is described as heat modifiability (Beher, Schnaitman, & Pugsley, 1980). Using this concept, we can track FadL folding over time to calculate folding percentages and kinetic rates. We find that wild type (WT) and all single mutant variants fold to a similar percent folding value (Figure 4A and supplement table 1). Kinetic rates of folding were calculated by using GraphPad Prism 7.04. The percent folded values were plotted and fitted to either a single or double exponential. Kinetic parameters were calculated by Prism when determining the best exponential association. The kinetic rates, K_{Fast} and K_{Slow}, are similar among WT and single mutant variants (Figure 4B). This indicates that the mutations do not diminish FadL folding *in vitro* in a BamA independent system.

In vitro BAM-Assisted Insertion

To determine if FadL shows BAM dependence, and whether the mutations selected (Figure 3) disrupt OMP insertion, we performed an *in vitro* BAM dependence assay (Hussain, Peterson, & Bernstein, 2020). The BAM complex purified into proteoliposomes can successful insert OMPs *in vitro* in POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) vesicles that otherwise preclude OMP insertion (Hussain et al., 2020). Using the same heat modifiability principle applied in the *in vitro* folding of FadL, we can analyze BAM dependent OMP folding. Once inserted, OMPs will be shielded from proteinase K digestion showing a folded band, and when boiled the band shifts in molecular weight

denoting the unfolded state (Hussain et al., 2020). Digestion of FadL by proteinase K would indicate a lack of BAM-dependent insertion possibly due to the missing contact between BamA strand five and FadL strand fourteen resulting in FadL digestion. FadL does not insert into proteolipsomes in the absence of the BAM complex (Figure 5A, lane 1-4). We find that FadL shows BAM dependence as it successfully inserts and is shielded from proteinase K digestion (Figure 5B, lanes 1-4). The single mutant variant of FadL, L412Q, that is hypothesized to contact the hydrophobic exterior of BamA also shows successful insertion into the proteoliposomes (Figure 5C, lanes 1-4). However, the single mutant variants of FadL, F413E and T415R, that are hypothesized to contact the hydrophilic interior show unsuccessful insertion of FadL as they get digested by proteinase K (Figure 5D & 5E, lanes 3 & 4). This shows that a single mutant in the hypothesized region that contacts the interior lumen of the BamA β-barrel is enough to disrupt the insertion process of BAM (Figure 5F).

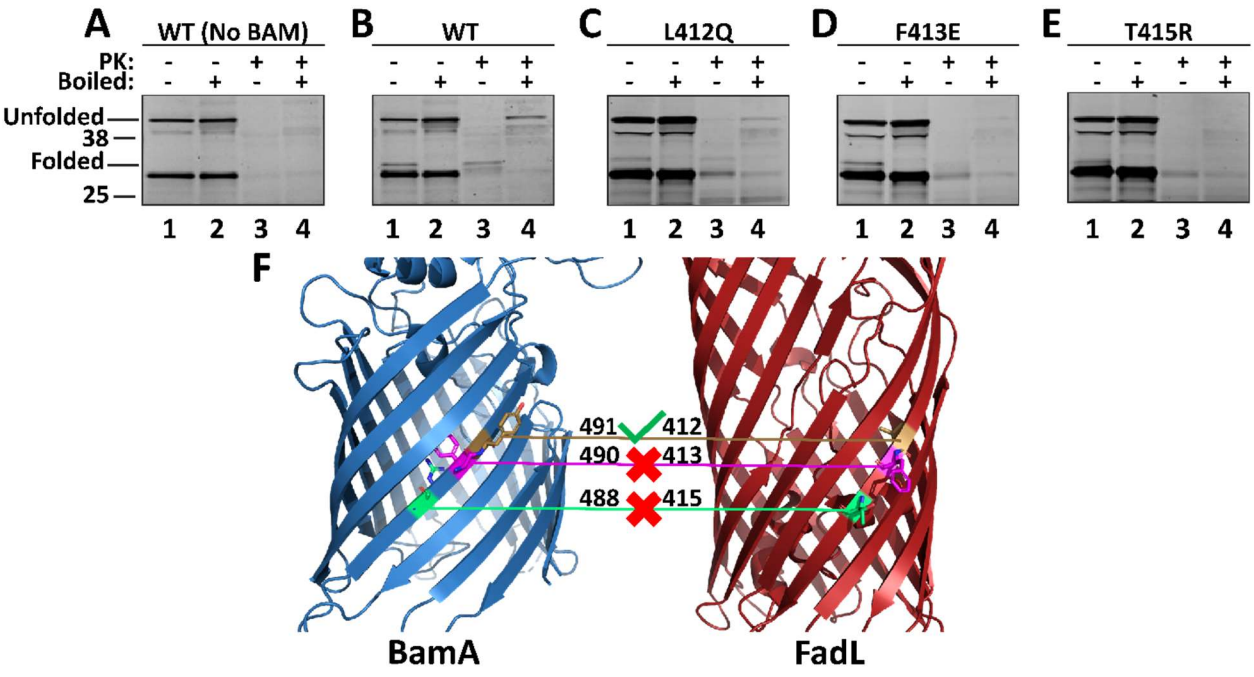


Figure 5 *in vitro* BAM-Assisted Insertion A) Wildtype FadL does not insert into proteoliposomes without BAM. Labeled for samples digested with proteinase K (PK), Boiled indicates samples boiled at 95C for 15 minutes. Lane 1: a lack of FadL folded product, lane 2: boiled sample from lane 1, lane 3: complete digestion of FadL from lane 1, lane 4: boiled sample from lane 3 denoting lack of insertion. **B)** Wildtype FadL insets into proteoliposomes with BAM. Lane 1:

folded FadL product. Lane 2: boiling of folded product from lane 1, lane 3: folded product shielded from PK treatment indicating shielding due to successful insertion of FadL, lane 4: boiled product from lane 3 denoting heat modifiability of FadL. **C)** L412Q FadL does show insertion into proteoliposomes with BAM. Lane 1: Folded product, lane 2: boiled folded product from lane 1, lane 3: proteinase K digestion of folded product from lane 1, lane 4: boiled proteinase K product from lane 3. **D)** F413E FadL does not insert into proteolipsomes without BAM. Lane 1: FadL folded product that is heat modifiable, but upon PK treatment (lane 3) it is completely digested denoting lack of insertion. **E)** T415R FadL does not insert into proteolipsomes without BAM. Lane 1: FadL folded product that is heat modifiable, lane 2: boiled product from lane 1, lane 3: complete digestion of FadL by proteinase K, lane 4: boiled product from lane 3. **E)** Hypothesized contacts between BamA (4N75) and FadL (1T16). BamA 491 FadL 412 successfully inserts into proteoliposomes (green check mark). BamA 490 FadL 413 and BamA 488 FadL 415 do not successfully insert into proteoliposomes (red X).

Discussion

The ongoing research into BAM mediated insertion continues to reveal new insights (Doyle & Bernstein, 2019) (Lee et al., 2019). Our computational results revealed a coevolutionary signal between the fifth strand of BamA and the fourteenth strand of the client OMP FadL. The experimental evidence is suggestive of a contact between outward-facing FadL residues 413 and 415 of the fourteenth strand of FadL and inward-facing residues 490 and 488 of the fifth strand of BamA. These results lead to the conclusion that BamA utilizes the interior lumen of its β-barrel to assist with membrane insertion of the client OMP FadL.

These findings are consistent with other studies. The interior strands of BamA are functionally important to the insertion of LptD, LamB, and OmpF (Lee et al., 2019). Specifically, BamA residue positions 439, 610, and 666 on strand two, ten, and loop six respectively were shown to be general binding candidates for LptD, LamB, and OmpF (Lee et al., 2019). LptD is also known to form contacts with residues positions 488 and 490 which are highlighted in our study as important contacts for BAM mediated insertion (Lee et al., 2019). This suggests the possibility that strand five on BamA is a candidate for general interaction with client OMPs.

C-terminal strands of OMPs have been shown to be conserved, and that conservation may be a result to maintain the folding and insertion pathway of OMPs (Robert et al., 2006) (Gruss et al., 2013)

(Doyle & Bernstein, 2019) (Lee et al., 2019). Our results revealed that the C-terminal strand of the client OMP FadL has a coevolutionary link with interior of BamA and is important for FadL insertion. Other client OMPs have been shown to have C-terminal strand interactions with the BAM complex (Doyle & Bernstein, 2019) (Lee et al., 2019) (Heuck, Schleiffer, & Clausen, 2011). However, our results show there's additional steps that are coevolutionary linked with the interior of BamA that is important for insertion. Future analysis using coevolution may reveal more intermediate steps in client OMP insertion pathway.

Methods

Joint Multiple Sequence Alignment Generation

Protein sequences for all crystalized outer membrane proteins were downloaded from the NCBI database. Version 2.2.26 Blastp (Camacho et al., 2009) was used to expand the sequence set for each protein using an E-value of 10^{-3}. The output from blastp was aligned using PSI-BLAST (Altschul et al., 1997) with a coverage of 80%. The resulting a3m file was filtered using an ID of 90% and coverage of 20%. Alignment generation can be found in Evolutionary PAirwise Distance-dependent potential (EPAD) in the RaptorX software package downloaded from http://raptorx.uchicago.edu/download/ (Zhao & Xu, 2012). To concatenate sequences into a joint MSA (joint-MSA), a template BamA (PDBid: 4N75) was chosen to generate a species representation. Duplicate species in the BamA set were removed to reduce the number of paralogs, and the species list was passed over the sequence files to find joint species pairs for BamA client proteins. If there is a species match between the BamA species list and a client protein, the lowest scoring E-value is taken, then those two sequences are concatenated into a joint-MSA. The total number of sequences in the joint-MSAs can be found in MRF_HMM_sp_spreadsheet.xlsx. A joint-MSA was then passed through the script seq_len.pl to generate a square alignment and filter out gapped positions that are greater than 25%. The resulting square alignment was submitted to the RaptorX Contact Prediction webserver

(http://raptorx.uchicago.edu/ContactMap/) for sequence coevolution calculation. A homology cutoff of 90% was used to filter joint-MSAs to 31 unique OMPs (table 1).

Contact Map and Mutation Selection

The RaptorX contact probabilities were fed through the script contact_map.py to generate a contact map. The BamA FadL antiparallel contact was extracted from the contact probabilities. The sequence positions in the joint-MSA was then analyzed for each BamA FadL antiparallel contact using the script heat_map.py, and heatmaps were generated to bioinformatically chose mutations. Positions L412Q, F413E, and T415R on FadL were chosen for mutation as they were not present as available contact space.

Cloning of FadL and Mutations

Wildtype FadL was cloned into pET vectors 3b and 303 for outer membrane and inclusion body expression respectively from an IDT gBlock. All necessary primers needed for that cloning can be found in the spreadsheet primers_used.xlsx. The primers used for mutation had an annealing temperature of 68C for use in a Q5 polymerase reaction.

Large Scale Protein Expression and Preparation for Inclusion Bodies

Protein was expressed using a pET 303 vector in *E. coli* BL21(DE3) cells. Cells were grown in 1L of LB at 37C and 250 rpm with 100 ug/mL carbenicillin until OD reached 0.6 in (the name of our large incubators). Once OD was reached, Isopropyl ß-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 1mM and cells continued to grow for 4 hours. Cells were centrifuged at 4000g for 30 minutes and pellets were stored at -20°C.

Pellets were resuspended in 40mL of 20mM Tris pH 8. 10 mM $MgCl_2$, 0.25µg/mL DNAse, and 0.25mg/mL Lysozyme were added. Cells were then sonicated on an ice bath at 40% amplitude with 4 seconds on 6 seconds off for 3 minutes. The inclusion bodies were centrifuged at 4000g for 30 minutes at 4°C. Supernatant was discarded, and pellet was resuspended with 40mL of 20mM Tris pH 8, 1mM EDTA, 1% TritonX-100. A second sonication on an ice bath was carried out at 40% amplitude with 4 seconds on and

6 seconds off for 1 minute. The inclusion bodies were centrifuged at 4000g for 30 minutes, and resuspended in 40mL of 20mM Tris pH 8, 1mM EDTA. If inclusion bodies are still sticky during re-suspension add up to 1% TritonX-100 and re-sonicate for another minute. A final wash was done with 40mL of 20mM Tris pH 8, 1mM EDTA and spun at 4000g for 30 minutes. Discard the supernatant and weigh the pellet to re-suspend the inclusion bodies to 100 mg/mL in 20mM Tris pH 8, 1mM EDTA. Take 5 µL of solution and add 5µL of 8M Urea and add 10 µL of 2X Lammeli's Buffer and ran on an SDS-PAGE gel at 300V for 15 minutes. Stain the gel with Insta Blue (Fisher) to check protein expression, if protein expressed the remainder of the inclusion bodies were made into 1 mL aliquots, and stored at -20°C.

Preparation of Large Unilamellar Vesicles (LUVs)

1,2-didecanoyl-sn-glycero-3-phosphocholine (PC-dic10) (Avanti Polar Lipids) were dissolved in chloroform for a final concentration of 10mg/mL in amber glass vials, and then dried to a thin film under a constant stream of nitrogen. The lipid vials were then lyophilized overnight to remove excess solvent, and then stored at -20°C until use. LUV stocks were resuspended to 3.33mM lipids in 50mM CAPS pH 10.5 1mM EDTA. The lipids were re-constituted, gently vortexed, and left to incubate at room temperature for 30 minutes. LUVs were prepared by extruding 15 times through a 0.1um filter using a mini-extruder (Avanti Polar Lipids) and incubated at 37°C for on the day of experimental use.

Folding and SDS-PAGE

Folding experiments were done in triplicate. Inclusion bodies pellets were thawed at room temperature and spun down using a tabletop centrifuge at 13,200 rpm for 25 seconds. The supernatant was discarded, and the pellets were solubilized using 8 M Urea, 50mM CAPS pH 10.5, 1mM EDTA. Concentration of the protein was determined by absorbance at 280nm and diluted to a final concentration of 50µM. A one in twenty dilution of 50 µM protein into prewarmed 37C 3.33mM LUVs was made in an amber glass vial and was incubated at 37°C with shaking at 250 rpm in a thermal mixer block (Thermo Scientific). Forward timepoints were taken at 0, 5, 10, 15, 30, 60, 120, and 240 minutes and diluted in 2X Laemmli's buffer for a final protein concentration of 1.25 µM. Samples were run on an

SDS-PAGE gel at 180 V for 35 minutes in 1x TGS with buffer line at max volume and stained overnight using Insta Blue (Fisher).

Analyzing SDS-PAGE Gels and Densitometry

All SDS-PAGE gels were imaged on a (insert scanner name here), and gel densitometry was taken using ImageJ. The .tiff image files were processed using Subtract Background feature of ImageJ. Densitometry values were taken using the AUC feature for the unfolded and folded bands. Percent folded values were calculated by lane using equation 2.

Equation 2

$$\%\ Folded = \frac{Folded}{(Folded + Unfolded)}$$

The % Folded values were plotted using GraphPad Prism 7.04 and fitted to either a one-phase or two-phase association using the comparison feature. Equation 3 is the rate equation used by GraphPad where y describes the fraction folded at time t. The fraction folded as time approaches infinity is describe with y_0. k_{fast}, and k_{slow} are the rate constants and A_{fast} and A_{slow} are the negative amplitudes of each rate constant.

Equation 3

$$y = y_0 + A_{fast}e^{-k_{fast}t} + A_{slow}e^{-k_{slow}t}$$

Kinetic parameters were calculated by Prism when determining the best phase association.

References

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res, 25*(17).

Beher, M. G., Schnaitman, C. A., & Pugsley, A. P. (1980). Major Heat-Modifiable Outer Membrane Protein in Gram-Negative Bacteria: Comparison with the OmpA Protein of *Escherichia coli*. *Bacteriology, 143*, 906-913.

Burger, L., & van Nimwegen, E. (2010). Disentangling direct from indirect co-evolution of residues in protein alignments. *PLoS Comput Biol, 6*(1), e1000633. doi:10.1371/journal.pcbi.1000633

Burgess, N. K., Dao, T. P., Stanley, A. M., & Fleming, K. G. (2008). Beta-barrel proteins that reside in the Escherichia coli outer membrane in vivo demonstrate varied folding behavior in vitro. *J Biol Chem, 283*(39), 26748-26758. doi:10.1074/jbc.M802754200

Burns, J. L., & Smith, A. L. (1987). A Major Outer-membrane Protein Functions as a Porin in Haemophilus influenzae. *General Microbiology, 133*, 1273-1277.

Camacho, C., Coulouris, G., Vahram, A., Ma, N., Papadopoulos, J., Bealer Kevin, & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*. doi:https://doi.org/10.1186/1471-2105-10-421

Cho, S. H., Szewczyk, J., Pesavento, C., Zietek, M., Banzhaf, M., Roszczenko, P., . . . Collet, J. F. (2014). Detecting envelope stress by monitoring beta-barrel assembly. *Cell, 159*(7), 1652-1664. doi:10.1016/j.cell.2014.11.045

Dhar, R., Feehan, R., & Slusky, J. S. G. (2021). Membrane barrels are taller, fatter, inside-out soluble barrels. *bioRxiv*. doi:10.1101/2021.01.30.428970

Doerner, P. A., & Sousa, M. C. (2017). Extreme Dynamics in the BamA beta-Barrel Seam. *Biochemistry, 56*(24), 3142-3149. doi:10.1021/acs.biochem.7b00281

Doyle, M. T., & Bernstein, H. D. (2019). Bacterial outer membrane proteins assemble via asymmetric interactions with the BamA beta-barrel. *Nat Commun, 10*(1), 3358. doi:10.1038/s41467-019-11230-9

Du, D., Wang, Z., James, N. R., Voss, J. E., Klimont, E., Ohene-Agyei, T., . . . Luisi, B. F. (2014). Structure of the AcrAB-TolC multidrug efflux pump. *Nature, 509*(7501), 512-515. doi:10.1038/nature13205

Gentle, I. E., Burri, L., & Lithgow, T. (2005). Molecular architecture and function of the Omp85 family of proteins. *Mol Microbiol, 58*(5), 1216-1225. doi:10.1111/j.1365-2958.2005.04906.x

Gessmann, D., Chung, Y. H., Danoff, E. J., Plummer, A. M., Sandlin, C. W., Zaccai, N. R., & Fleming, K. G. (2014). Outer membrane beta-barrel protein folding is physically controlled by periplasmic lipid head groups and BamA. *Proc Natl Acad Sci U S A, 111*(16), 5878-5883. doi:10.1073/pnas.1322473111

Ginsburgh, C. L., Black, P. N., & Nunn, W. D. (1984). Transport of Long Chain Fatty Acids in *Escherichia coli*. *Biological Chemistry, 259*(13).

Goh, C.-S., & Cohen, F. E. (2002). Co-evolutionary Analysis Reveals Insights into Protein–Protein Interactions. *Journal of Molecular Biology, 324*(1), 177-192. doi:10.1016/s0022-2836(02)01038-0

Gruss, F., Zahringer, F., Jakob, R. P., Burmann, B. M., Hiller, S., & Maier, T. (2013). The structural basis of autotransporter translocation by TamA. *Nat Struct Mol Biol, 20*(11), 1318-1320. doi:10.1038/nsmb.2689

Heuck, A., Schleiffer, A., & Clausen, T. (2011). Augmenting beta-augmentation: structural basis of how BamB binds BamA and may support folding of outer membrane proteins. *J Mol Biol, 406*(5), 659-666. doi:10.1016/j.jmb.2011.01.002

Hoang, H. H., Nickerson, N. N., Lee, V. T., Kazimirova, A., Chami, M., Pugsley, A. P., & Lory, S. (2011). Outer membrane targeting of Pseudomonas aeruginosa proteins shows variable dependence on the components of Bam and Lol machineries. *mBio, 2*(6). doi:10.1128/mBio.00246-11

Hussain, S., Peterson, J. H., & Bernstein, H. D. (2020). Bam complex-mediated assembly of bacterial outer membrane proteins synthesized in an in vitro translation system. *Sci Rep, 10*(1), 4557. doi:10.1038/s41598-020-61431-2

Ieva, R., Tian, P., Peterson, J. H., & Bernstein, H. D. (2011). Sequential and spatially restricted interactions of assembly factors with an autotransporter beta domain. *Proc Natl Acad Sci U S A, 108*(31), E383-391. doi:10.1073/pnas.1103827108

Kamisetty, H., Ovchinnikov, Sergey, and Baker, David. (2013). Assessing the utility of coevolution-based residue-residue contact predictions in a sequence- and structure-rich era. *Proceedings of the National Academy of Sciences, 110*(46), 18734-18734. doi:10.1073/pnas.1319550110

Krachler, A. M., Ham, H., & Orth, K. (2011). Outer membrane adhesion factor multivalent adhesion molecule 7 initiates host cell binding during infection by gram-negative pathogens. *Proc Natl Acad Sci U S A, 108*(28), 11614-11619. doi:10.1073/pnas.1102360108

Lee, J., Sutterlin, H. A., Wzorek, J. S., Mandler, M. D., Hagan, C. L., Grabowicz, M., . . . Kahne, D. (2018). Substrate binding to BamD triggers a conformational change in BamA to control membrane insertion. *Proc Natl Acad Sci U S A, 115*(10), 2359-2364. doi:10.1073/pnas.1711727115

Lee, J., Tomasek, D., Santos, T. M., May, M. D., Meuskens, I., & Kahne, D. (2019). Formation of a beta-barrel membrane protein is catalyzed by the interior surface of the assembly machine protein BamA. *Elife, 8*. doi:10.7554/eLife.49787

Malinverni, J. C., Werner, J., Kim, S., Sklar, J. G., Kahne, D., Misra, R., & Silhavy, T. J. (2006). YfiO stabilizes the YaeT complex and is essential for outer membrane protein assembly in Escherichia coli. *Mol Microbiol, 61*(1), 151-164. doi:10.1111/j.1365-2958.2006.05211.x

Noinaj, N., Kuszak, A. J., Gumbart, J. C., Lukacik, P., Chang, H., Easley, N. C., . . . Buchanan, S. K. (2013). Structural insight into the biogenesis of beta-barrel membrane proteins. *Nature, 501*(7467), 385-390. doi:10.1038/nature12521

Olivella, M., Gonzalez, A., Pardo, L., & Deupi, X. (2013). Relation between sequence and structure in membrane proteins. *Bioinformatics, 29*(13), 1589-1592. doi:10.1093/bioinformatics/btt249

Peng, J., & Xu, J. (2011). RaptorX: exploiting structure information for protein alignment by statistical inference. *Proteins, 79 Suppl 10*, 161-171. doi:10.1002/prot.23175

Ralf Koebnik, K. P. L., Patrick Van Gelder. (2000). Structure and function of bacterial outer membrane proteins: barrels in a nutshell. *Molecular Microbiology, 37*(2), 239-253.

Robert, V., Volokhina, E. B., Senf, F., Bos, M. P., Van Gelder, P., & Tommassen, J. (2006). Assembly factor Omp85 recognizes its outer membrane protein substrates by a species-specific C-terminal motif. *PLoS Biol, 4*(11), e377. doi:10.1371/journal.pbio.0040377

Ruiz-Perez, F., Henderson, I. R., Leyton, D. L., Rossiter, A. E., Zhang, Y., & Nataro, J. P. (2009). Roles of periplasmic chaperone proteins in the biogenesis of serine protease autotransporters of Enterobacteriaceae. *J Bacteriol, 191*(21), 6571-6583. doi:10.1128/JB.00754-09

Schaarschmidt, J., Monastyrskyy, B., Kryshtafovych, A., & Bonvin, A. (2018). Assessment of contact predictions in CASP12: Co-evolution and deep learning coming of age. *Proteins, 86 Suppl 1*, 51-66. doi:10.1002/prot.25407

Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., . . . Hassabis, D. (2019). Protein structure prediction using multiple deep neural networks in the 13th Critical Assessment of Protein Structure Prediction (CASP13). *Proteins, 87*(12), 1141-1148. doi:10.1002/prot.25834

Seokhee Kim, J. C. M., Piotr Sliz, Thomas J. Silhavy, Stephen C. Harrison, Daniel Kahne. (2007). Structure and Function of an Essential Component of the Outer Membrane Protein Assembly Machine. *Science, 317*.

Simonetti, F. L., Teppa, E., Chernomoretz, A., Nielsen, M., & Marino Buslje, C. (2013). MISTIC: Mutual information server to infer coevolution. *Nucleic Acids Res, 41*(Web Server issue), W8-14. doi:10.1093/nar/gkt427

Sklar, J. G., Wu, T., Kahne, D., & Silhavy, T. J. (2007). Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in Escherichia coli. *Genes Dev, 21*(19), 2473-2484. doi:10.1101/gad.1581007

Sojo, V., Dessimoz, C., Pomiankowski, A., & Lane, N. (2016). Membrane Proteins Are Dramatically Less Conserved than Water-Soluble Proteins across the Tree of Life. *Mol Biol Evol, 33*(11), 2874-2884. doi:10.1093/molbev/msw164

Spector, J., Zakharov, S., Lill, Y., Sharma, O., Cramer, W. A., & Ritchie, K. (2010). Mobility of BtuB and OmpF in the Escherichia coli outer membrane: implications for dynamic formation of a translocon complex. *Biophys J, 99*(12), 3880-3886. doi:10.1016/j.bpj.2010.10.029

Wang, S., Li, Z., Yu, Y., & Xu, J. (2017). Folding Membrane Proteins by Deep Transfer Learning. *Cell Syst, 5*(3), 202-211 e203. doi:10.1016/j.cels.2017.09.001

Zhao, F., & Xu, J. (2012). A position-specific distance-dependent statistical potential for protein structure and functional study. *Structure, 20*(6), 1118-1126. doi:10.1016/j.str.2012.04.003

One-phase and Two-phase association was performed using GraphPad Prism 7.0.4 for windows, GraphPad Software.