Title: Redesigning OmpA Loops Using Canonical Outer Membrane Protein Loop Structures

Authors: Meghan W. Franklin[1], Joanna Krise[2], Jacqueline J. Stevens[2], and Joanna S.G. Slusky[1,2*]

Affiliations:
[1] Center for Computational Biology, The University of Kansas, 2030 Becker Dr., Lawrence, KS 66045-7534.

[2] Department of Molecular Biosciences, The University of Kansas, 1200 Sunnyside Ave. Lawrence KS 66045-3101

*To whom correspondence should be addressed, slusky@ku.edu

ABSTRACT

Protein loops can be difficult to design and predict. There have been multiple different algorithms developed to predict the structure of loops. Outer membrane proteins are all beta barrels and these barrels have a variety of well-documented loop conformations. Here we test three different algorithms to predict the structure of outer membrane protein loops. We find the PETALS algorithm is superior for this purpose. We then experimentally test the effect of replacing the long loops of outer membrane protein OmpA with twelve shorter designed loops. Though we succeeded in creating the smallest known outer membrane barrel, we find that the designed loops do not have a strong effect on OmpA folding.

INTRODUCTION

Loops are the most difficult parts of proteins to accurately structurally predict or computationally design. Their greater range of φ and ψ angles and lower levels of homology among similar proteins make them the most difficult structure to predict (Gront, et al. 2012). Moreover, loops are often more dynamic than α-helices and β-sheets.

However, the difficulty in designing loops or predicting their structure makes them no less important. Because loops are frequently observed on the surface of proteins, loops are often important in substrate recognition, receptor binding, and enzymatic activity in globular proteins (Rose, et al. 1985).

Due to their large numbers of loops, outer membrane proteins (OMPs) can be used to study structure prediction of loops and how loops affect folding. Almost all OMPs have the same fold. They are β-barrels with each strand connected to the next strand via a loop-structure. The periplasmic loop-structures are known to be shorter and are traditionally referred to as 'turns'. The outer membrane loop-structures are longer and more varied and are referred to as 'loops'. The loops of OMPs are often the business end of the protein. Within OMPs the loops are known to control channel gating (Yildiz, et al. 2006; Zahn, et al. 2015), receptor binding (Fox, et al. 2014), and enzymatic activity (Kingma, et al. 2000; Vandeputte-Rutten, et al. 2001).

Outer membrane loops are plentiful. Each protein has half the number of loops as the number of β-strands in the barrel. For prototypical, single chain barrels, the number of β-strands per barrel vary from 8-26 (Franklin, et al. 2018). The most frequent number of amino acids per outer loop

is five. These five-residue outer membrane loops have five preferred geometries each with a preferred sequence hallmark (Franklin and Slusky 2018).

A particularly well studied OMP is OmpA. OmpA is an 8-stranded, 4-looped barrel. Its folding has been extensively characterized (Surrey and Jähnig 1992; Kleinschmidt, et al. 1999; Kleinschmidt and Tamm 1999, 2002; Kleinschmidt, et al. 2011; Danoff and Fleming 2017). OmpA can be folded in the presence of vesicles and its folding can be visualized and quantified with SDS PAGE. Here we compare three methods of loop structure prediction for three outer membrane canonical loop-types. We then use the sequence hallmarks of the outer membrane loop-types to design twelve loops and we test the effects of those loops on the rates of OMP folding.

RESULTS

Structure prediction of outer membrane protein loops

Because of the functional importance of OMP loops, we began by testing that modeling strategies are capable of recapturing the native OMP loop conformation. Our previous work identified four types of 5-residues loops. Each loop was named based on the part of the Ramachandran map in which each of the three middle positions are localized as previously defined by North et al (North, et al. 2011) (figure S1), for example AAP has the three central residues in the alpha, alpha, and proline regions of the Ramachandran map.

There are 72 five-residue loops in structurally-characterized OMPs. We tested three loop modeling systems by their ability to predict the structure of these OMP loops. Each of these loop modeling programs was readily available for a Linux system: PETALS (Wong, et al. 2017), DiSGro (Tang, et al. 2014) as implemented in m-DiSGro (Tang, et al. 2015), and the Rosetta KIC method (Mandell, et al. 2009). All three methods are template free or mostly template-free. Both PETALS and DiSGro use energy functions to evaluate loops and both use the CSJD closure method (Coutsias, et al. 2003). PETALS grows loops from the N and C termini with the loop closing in the middle of the loop. DiSGro grows loops from the N-terminus and closes at the C-terminus. PETALS is template free whereas the m-DiSGro implementation of DiSGro is nearly template free using templates only in cases of correcting course after steric clashes. PETALS does a more exhaustive search by filtering out high energy conformations and then evaluates all the remaining loop confirmations whereas DiSGro uses a more Monte Carlo based approach with less restrictive filter criteria and then sampling among the remaining confirmations to evaluate a larger set of confirmations. Rosetta's KIC method also uses kinematics to close loops (from N to C) but in this method the loop closure is done many times with fewer energy filters and then the loops are optimized from the closed positions.

PETALS allows for the implementation of different energy functions, and we used both DBB and OSCAR-o which we tested separately. mDiSGro and PETALS are advantageous for the speed at which they produce models. Rosetta KIC is ~1000 times slower but has the advantage of being integrated with the rest of Rosetta's design algorithms. The center three residues of each loop were modeled individually using the coordinates of i and i+4 positions as starting points in the context of the whole structure although all three programs are capable of building multiple loops simultaneously. The modeled structures were ranked according to the scores assigned by each program, and the top 25 models from each program were further analyzed. We rated performance by two metrics - similarity to atom placement as measured by backbone RMSD from native structure (Figure 1 A-C, scatter plots) and similarity to native angles as measured by maintaining native loop type (recapture rates, Fig 1 A-C, bars).

We tested nine PLL-loops (Figure 1A), 20 PLD-loops (Figure 1B), and nine AAL-loops (Figure 1C). For all three types, both of PETALS' scoring functions – OSCAR-o and DBB - result in nearly identical performance as measured by RMSD (Fig 1, red and orange dots). Both functions also have 100% recapture of the AAL-type loops but result in different loop types other than native for the models that are of the PLL-type and PLD-type. The OSCAR-o function in PETALS recaptures the PLL loop-type better, while the DBB function recaptures the PLD type more effectively. Examples of models generated by PETALS with low RMSD that did and did not recapture the native loop-types are shown in Fig 1D. mDiSGro has the next best performance based on RMSD (Fig 1, yellow dots). However, these models either recaptured the native loop-types or are not close to any loop-type (black bars). Finally, Rosetta KIC has the

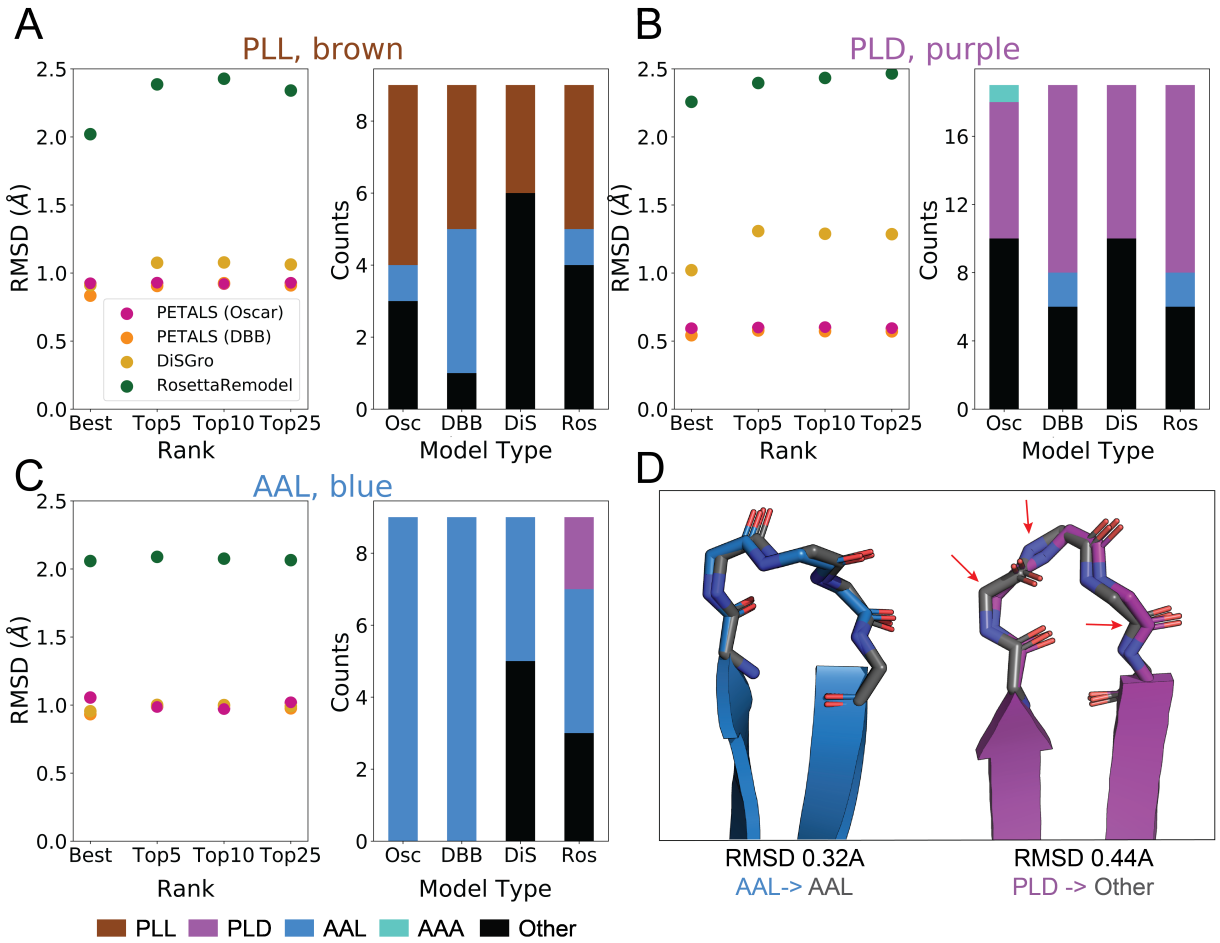


Figure 1. Testing loop modeling accuracy of three different modeling software. A-C) Average RMSD (left graphs) and loop-type recapture (right graphs) of the three loop-types from which loop sequences were chosen. The RMSD points represents the average backbone RMSD across the best designs, the top 5 designs, top 10 designs, or top 25 designs from left to right respectively. The bars represent whether the best model for each structure maintains all central angles (i+1 through i+3) within 30° of the loop-type cluster centroid; the expected color for all bars in a graph is the heading color. Black is outside every cluster; cyan is the AAA loop-type. The model types are PETALS (Oscar), PETALS (DBB), DiSGro, and Rosetta KIC from left to right. **A)** PLL, **B)** PLD, **C)** AAL. **D)** Lower RMSD models that do (left) or don't (right) maintain loop type. The native loops are colored according to their native loop-type and the model is shown in grey. Both examples maintain a low RMSD, but the left example maintained the same loop-type assignment while the right example did not maintain the same loop-type assignment.

highest RMSD (Fig 1, green dots) and lowest recapture rates, often returning model of other types instead of the native type.

Because PETALS with the OSCAR-o energy function had the best performance, we employed it for the more computationally more intensive process of determining the structure of designed loops rather than the recapitulation of native loops. The central difference between these two processes is the determination of the terminal fixed (no change in φ and ψ angles) positions. For the determination of the structure of designed loops, all five loop residues (i to i+4) have to be rebuilt (allowing for new determination of φ and ψ angles) rather than only finding angles for the three central residues of the loop (Fig 2A). Compared to building loops between residue i+1 and i+3 (Fig 2B solid bars) as described above, when building loops from i to i+4 the recapture

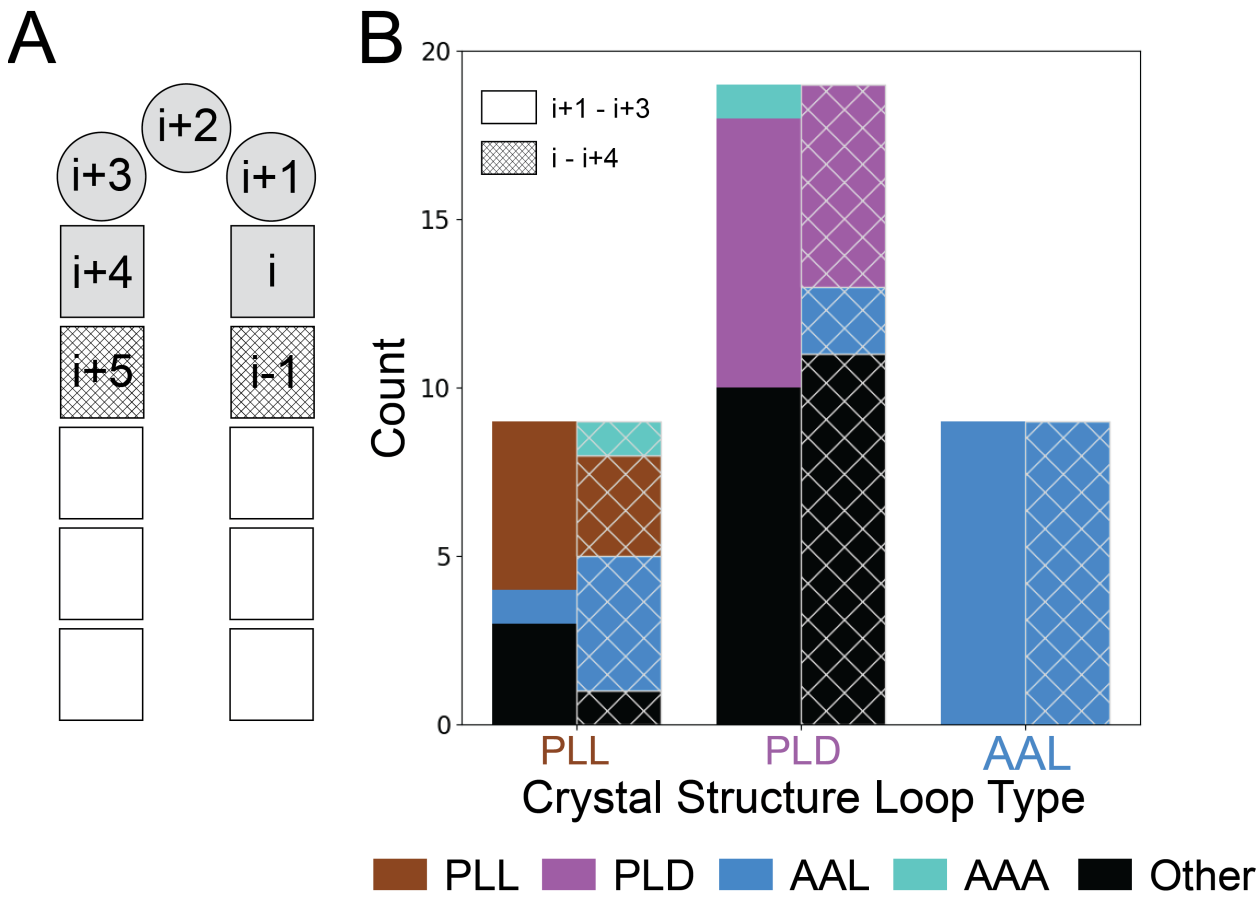


Figure 2. PETALS (OSCAR-o) loop structure predictions for structurally characterized 5-residue outer membrane protein loops A) Schematic of a loop. Residues defined as strand are boxes; residues defined as loop are circles. The residues we define as a loop are in grey and are numbered by convention from i to i+4; because outer membrane protein strands run from right to left if the loops are pointing upwards the C-terminus is on the right. The hatched bars in B have the terminal fixed residues at the i-1 and i+5 positions, while the solid bars use the i and i+4 as the terminal fixed residues. **B)** The recapture counts of each loop from the PLL, PLD and AAL loop-types using the best model from PETALS (OSCAR-o). The solid bars represent the models built with defined φ and ψ angles for the i and i+4 positions but no knowledge of the angles of the central three positions while the hatched bars represent models built with no knowledge of any of the five loop residues.

rate of the native loop-type is decreased for both the PLL-type and PLD-type loops, while both methods result in 100% recapture of the AAL-type (Fig 2B, hatched bars). Notably, the AAA loop-type which we did not test for appeared among the models for the i:i+4 native PLL loop-type, resulting in increased loop diversity in the i:1+4 loop models over the i+1:i+3 models. The i:i+4 models of the native PLD loop-type were less native-like, shown by the increase in the black bar representing untyped turns.

Designing Sequences for OmpA

We chose to redesign the loops of OmpA because of how thoroughly OmpA folding has been characterized (Surrey and Jähnig 1992; Kleinschmidt, et al. 1999; Kleinschmidt and Tamm 1999, 2002; Kleinschmidt, et al. 2011; Danoff and Fleming 2017). The loops of OmpA are natively long, 17-21 residues in length and show great flexibility in the NMR structure (Fig 3A). Similar to most loops in our study of the loop types, the terminal positions of the long loops of OmpA fall into the beta-region of the Ramachandran map because these residues are all strand by definition (Figure 3B). These initial and terminal φ and ψ angles are characteristic of two of the five types for five-residue loops. Loops 1, 3 and 4 have PLL-like terminal residues angles and loop 2 has AAL-like terminal residue angles (Figure 3B).

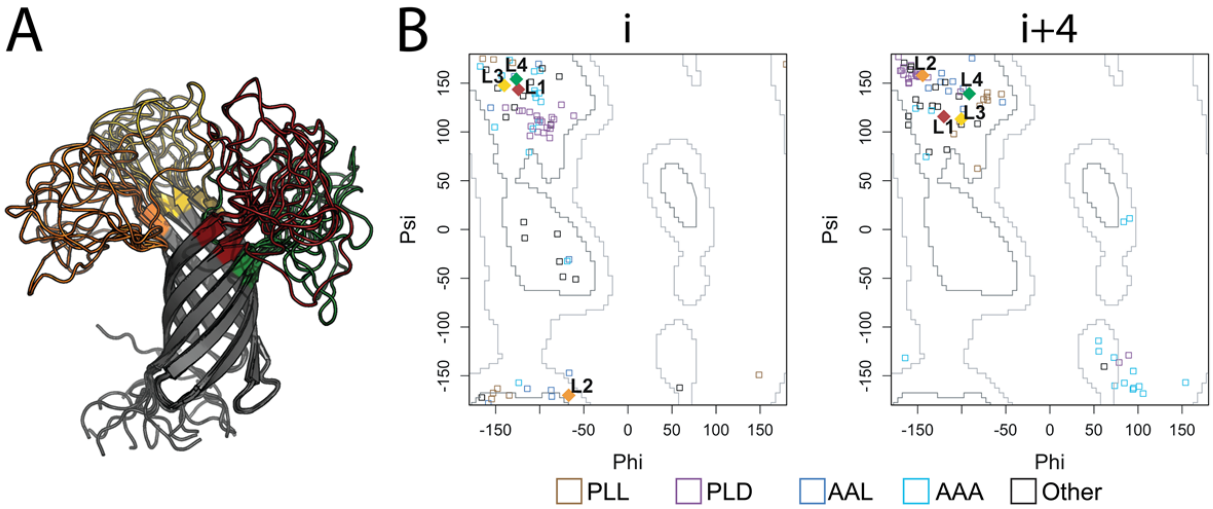
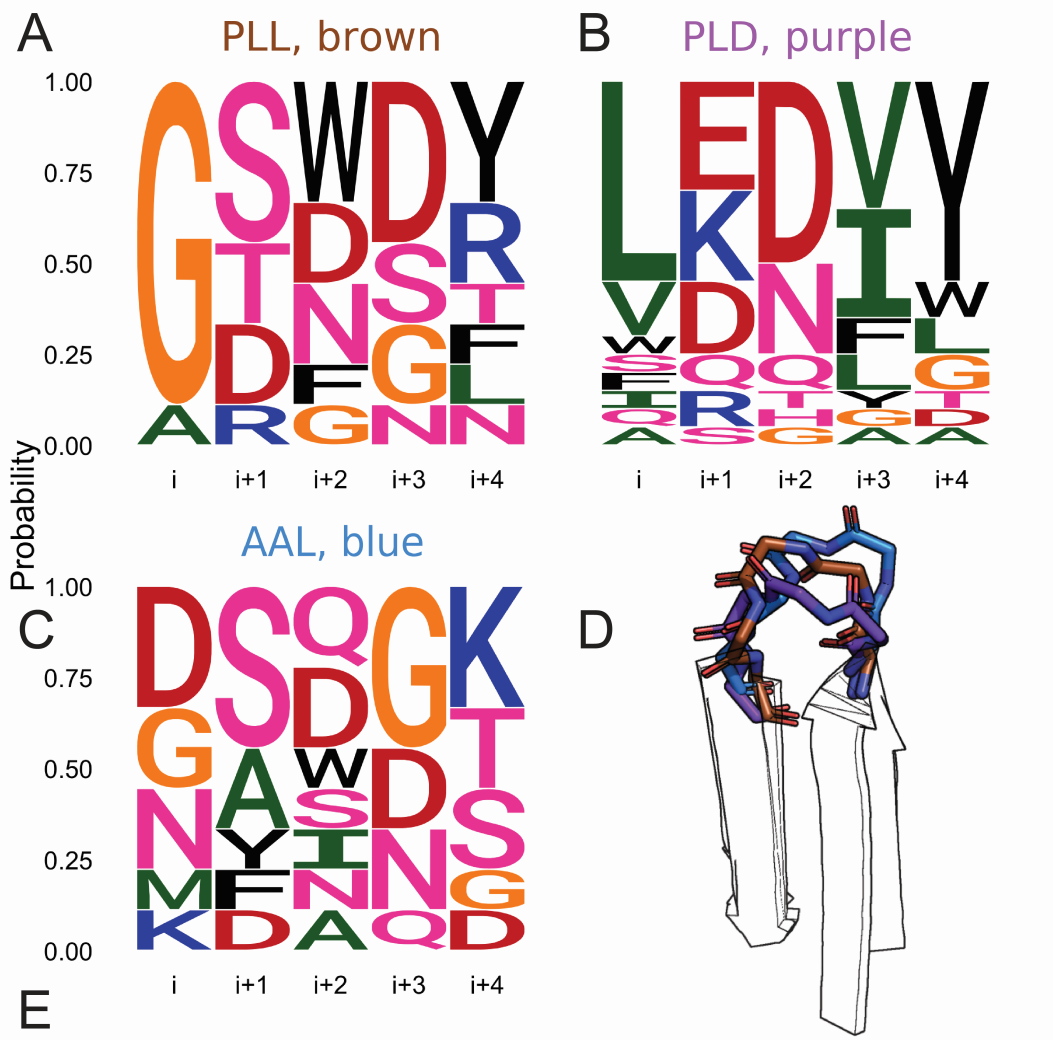


Figure 3. OmpA and its loop types A) The loops of the NMR structure of OmpA (PDB ID: 2ge4). The four loops are colored red, orange, yellow and green from C to N-terminus. **B)** The *φ and ψ* angles of the i and i+4 positions of the crystallized 5-residue loops of OMBBs. The small squares are colored as for Fig A's loop-types. The native OmpA i and i+4 angles are shown in diamonds and colored as in A.

We built 5-residue loops to substitute for the four naturally long loops of OmpA. Each loop position was substituted with 3 different sequences, one sequence with residues typical of an AAL loop, one sequence with residues typical of a PLL loop and one sequence with residues typical of a PLD loop. We used the position-specific amino acid preferences of each loop-type to select amino acids for the redesigned loops of OmpA (Figure 4 A-C) as previously described (Franklin and Slusky 2018). 12 sequences in total were computationally selected from the three target loop-types to replace the four loops of OmpA (Figure 4E).



Loop ID	Residues	Native Sequence	PLL (Brown)	PLD (Purple)	AAL (Blue)
L1	38-56	SQYHDTGFINNNGPTHENQ	GSFDT	WKNAA	DSNGK
L2	80-97	YDFLGRMPYKGSVENGAYKAQ	GSFSR	VDDVY	KANDT
L3	125-140	RADTKSNVYGKNHDTG	GRWDY	LKDIG	DSAGT
L4	164-183	QFTNNIGDAHTIGTRPDNG	ATWSR	VSHGY	NSDGK

Figure 4. Sequences designed for OmpA loops. A-C) Sequence logo plots showing the sequence diversity of each loop-type. The height of each letter represents the incidence of that amino acid at that position. The designed sequences are shown in the table on the bottom. **A)** PLL, **B)** PLD, **C)** AAL, **D)** Backbone trace of a representative loop for each of the three loop-types, shown in the assigned loop-type color (PLL, brown; PLD, purple; AAL, blue).E) Identity

PETALS (OSCAR-o) was used to model the fold of each designed sequence. Because of its good performance in modeling the loops of crystal structures, we anticipated that we would generally get models that fell into the intended loop-type. However, just three out of four AAL-type sequences that were so well recaptured in the crystal structures fell into the AAL-type (Figure 5). The PLD sequences didn't fall into any loop-type regardless of the number of models considered, a PLD conformation was never observed. Of the PLL loop sequences, a single loop fell into the PLL loop-type. Seven of twelve loop models did not fall into any type. When the top

10 or 25 models were considered, we saw an increase in models that fell into the AAL loop-type and never observed PLD conformation.

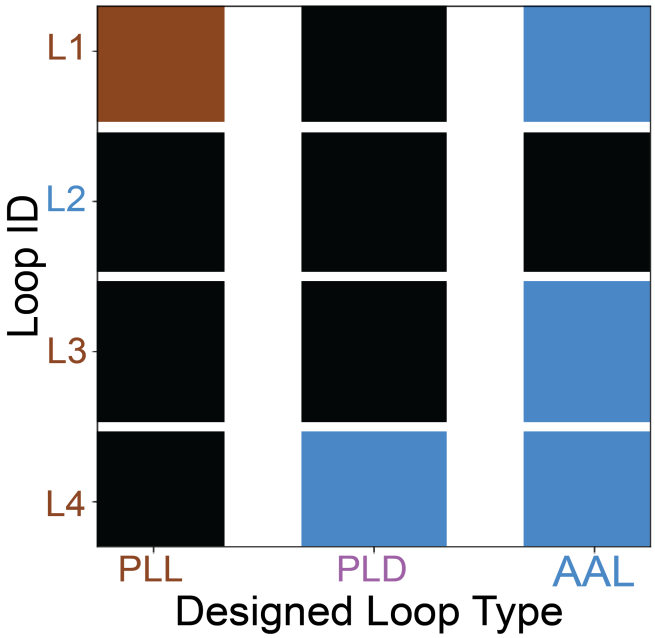


Figure 5. Loops designed by sequence are not consistently predicted to have canonical loop structures. The square indicates the loop-type to which the best model for each of the 12 designed loops was assigned. The x-axis indicates what color the squares of that column should have been. Brown is PLL, blue is AAL, black is a loop with non-canonical φ and ψ angles.

Folding Assay

DNA constructs of OmpA with designed loops were synthesized. Specifically, 15 altered OmpA proteins were made, 12 were OmpA with one loop substituted (figure 4E) and 3 were OmpA with all four loops substituted. The three constructs in which all four loops were substituted were substituted with loops of the same designed type (named by loop type, PLL-all, PLD-all, and AAL-all). The 15 proteins with designed loops and native OmpA were purified from inclusion bodies as described in the methods section.

The folding of OMPs is assessable by using their property of heat modifiability. OMPs are so stably folded that a folded OMP shows a different migration on SDS-PAGE gel than when it is unfolded by boiling in SDS buffer (Rosenbusch 1974; Beher, et al. 1980).

Folding for these proteins can occur when they are presented with a synthetic membrane to fold into. To assess folding kinetics, each protein was rapidly diluted 1:10 into folding buffer containing large unilamellar vesicles (LUVs). The protein LUV mix was incubated at 22°C while being shaken for up to six hours. Samples were taken at various time points, starting at sixty seconds. Folding was quenched with loading buffer. Samples were loaded and separated on a gradient SDS PAGE gel. Densitometry of the folded and unfolded bands were analyzed (Figure S2)

Folding was measured as by dividing the density of the folded band by the sum of the folded and unfolded bands. Folding experiments were repeated in triplicate. For each experiment, the folding kinetics were fit to both a single and double exponential equation and the best fit was chosen according to the lowest R-squared value (see supplemental equations 1 and 2). Figure 6 shows the folding of all 15 proteins over time with the fit curves. We found that ten of 15 proteins (including wild type OmpA) fit best with a two-state kinetic equation, which includes a value for the observed k_{fast} and k_{slow}. The folding efficiency (fraction of protein folded at the final

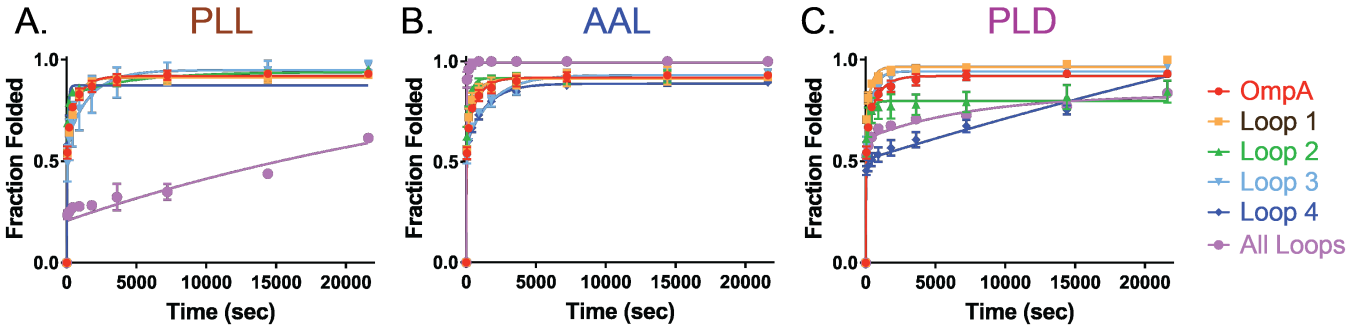


Figure 6. The fraction of OmpA folded over time at 22°C, determined by densitometry analysis of SDS-PAGE bands. Experiments were performed in triplicate with the average value shown and error bars corresponding to standard deviation; solid lines are the fitted exponential rate equation for each protein. Wild type OmpA is shown in red, mutation of the first OmpA loop to the designed sequence loop is shown in orange, mutation of second loop is shown in green, third loop is sky blue, fourth loop is royal blue, and all four loops mutated at once is shown in purple. OmpA loops have been substituted with designed **A)** PLL, **B)** PLD, and **C)** AAL loops.

time point) is shown in Figure 7A while the folding rate constants are shown in Figure 7B and C. Wild type OmpA folding kinetics is consistent with previous measurements (Burgess, et al. 2008).

Similar to wild type OmpA, most loops fold were able to fold to ~90% completion, however, PLL-all loops substituted only folded to ~61%. Similarly, most k_{fast} values were consistent with wild type OmpA with the exception of the slower PLL-all loops, PLL loop 4 and PLD loop 4. From this it appears that the rates of folding were most affected by substitutions for the fourth loop though folding could still be completed with these changes.

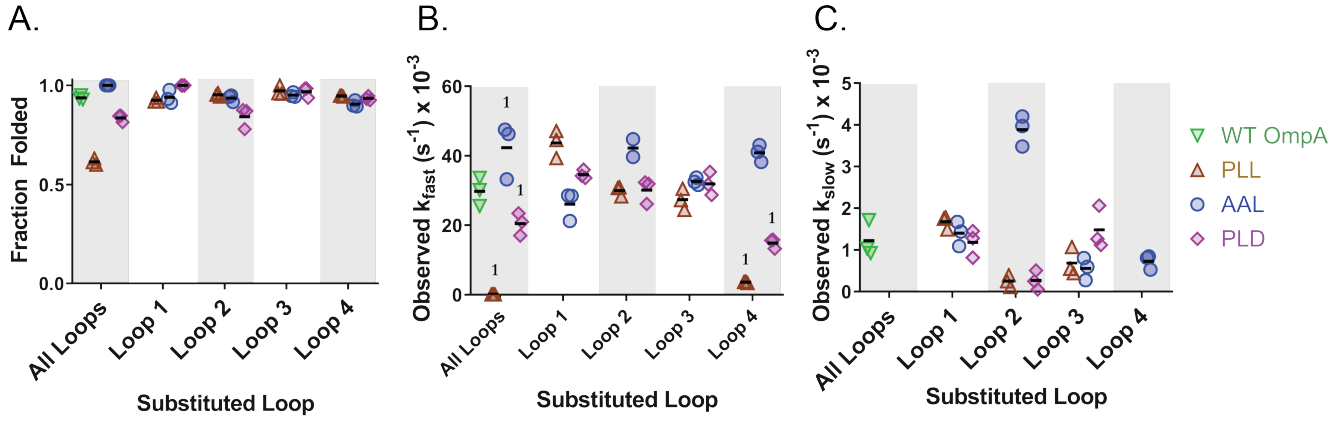


Figure 7. Analysis of kinetic data by substituted loop number. Each point represents the value obtained from each triplicate analysis, the average value is shown as a horizontal line. **A)** The folding efficiency of substituted loops at the end-point (6 hours). Wild type OmpA is shown in green, PLL loops are brown, PLD loops are purple, and AAL loops are blue. Observed **B)** k_{fast} and **C)** k_{slow} rate constants of OmpA with substituted loops are shown, when applicable. For proteins where the folding kinetics fit best with a single exponential curve, the observed rate constant is shown in B with a "1" above the data point. Alternating gray and white lanes are provided to guide the eye to the relevant loop.

DISCUSSION

Here we assess three loop structure prediction algorithms. We then design 12 loops, predict their structure with the best of the three algorithms, and test the effect of the 12 loops on folding. We find differences in how well the structure prediction algorithms predict OMP loop structures. We also find that many designed loops allow for native-like or almost native-like folding and that very few loop designs significantly hinder folding.

Smallest OmpA-like barrel, smallest barrel

In a previous study (Koebnik 1999), Koebnik shortened what were at the time believed to be the loops of OmpA (Vogel and Jähnig 1986) to three-residue loops creating a 130-residue barrel. In that study, both the individual loops and the all loops substituted together were found to fold *in vivo*. Since then the NMR structure of OmpA (Arora, et al. 2001) was solved showing significantly longer loops and shorter strands than had been thought, suggesting that while Koebnik had meant to create three residue long loops he had inadvertently created loops between five and ten residues long. More recent evidence indicates that long loops such as these may be structured in the native outer membrane LPS though not in vesicles or micelles (Schubeis, et al. 2020).

The barrels of our three constructs with all four loops replaced are only 117 amino acids. Though these barrels were expressed along with the 154-residue C-terminal periplasmic domain of OmpA that folds independently (Ishida, et al. 2014), the barrels themselves are 13 residues smaller than the previously known smallest barrel. Our barrels are also smaller than the smallest known native barrel, PagP which has a 139-residue barrel. Though our barrels are small, the majority of them still have undiminished folding efficiency. Such small barrels may be useful as minimalist folding models of outer membrane proteins.

Assessing loop prediction algorithms

Of the three secondary structure types, loops are by far the most diverse and difficult to predict. We tested loop prediction algorithms for the particular case of 5-residue OMP loops which are the most frequent loop type in OMPs (Franklin and Slusky 2018).

We judged three algorithms for rebuilding the 5-residues of the native loops in crystal structures of types AAL, PLL, and PLD. These algorithms were scored based on RMSD and by recapture of loop type by backbone angles. We found PETALS to be best on both metrics and to be especially good at recapturing AAL loops. PETALs may be better at OMP loop prediction because of its method of growing the loop from both the N and C terminus. AAL loops may be the easiest to recapture of the loop types used because it is the only one of the three loop types that has been found as a cluster within soluble protein loop types as well. Given that all of these structure prediction algorithms use knowledge-based potentials, all of the algorithms are likely to be biased towards building loops in globular proteins. This bias is likely due to the relative scarcity of membrane protein crystal structures and the documented differences in soluble loop angle preferences from those of OMPs (Franklin and Slusky 2018). Implementation of more membrane protein structures in the databases for loop construction algorithm would likely significantly improve their ability to effectively predict OMP loop conformations.

Because $PETALS_{OSCAR}$ scored the best in the recapturing native loops, we then used the PETALS algorithm with the OSCAR energy function to predict the structure of native loops when allowing the end positions (i and i+4) to vary in φ and ψ angles as well. These positions need to be rebuilt when changing their amino acid identities. Therefore this test is closer to the real problem of loop reconstruction. We find that although prediction of AAL loops remain consistent

when rebuilding under this more rigorous challenge, both PLD loops and PLL loops are more difficult to accurately predict by φ and ψ angles. Given that rebuilding the end positions lessens the successful prediction of previously solved loops we conclude that there is some memory of the real loop type retained in the predictions where the i and i+4 positions are fixed.

Loop design

When we used the sequence hallmarks of the loop types to design loops with conformations of the AAL, PLL, and PLD loop types, PETALS only predicted that we designed the intended structure one third of the time, with our most frequently correct design being the AAL loops. Because we did not crystalographically resolve any of our proteins we do not know if this is a failure of the design or of structure prediction. It is notable that the only AAL loop we designed that was not predicted to be AAL is loop 2. This is surprising because the angles of the i and i+4 positions of loop 2 in native OmpA have angles that are most consistent with the AAL loop type. This may demonstrate that our method of loop structure prediction is as significantly biased by 'memory' of the original scaffold. However, the differences in success between the i:i+4 and i+1:i+3 predictions (figure 2B) shows there is still some memory imparted by the loop termini.

Loop folding

Given the variety of loops it was surprising that almost all single loop mutants folded to completion similar to wild type. The AAL loops were both most predicted to correctly fold as AAL loops and folded most consistently. None of the PLD loops were predicted to get their designed conformation and they had the most folding variability. The PLD loop that was predicted to fold as an AAL loop folded with the lowest rate of any of the single mutants. Finally, though there was a lot of heterogeneity in the prediction of the PLL, loops only the substitution of all four loops substantially affected folding.

Although recent studies have shown that OmpA folding is most accurately modeled by using up to 11 states (Danoff and Fleming 2017), we have used the more usual single and double state kinetics for ease of comparison with previous studies from multiple labs over the past decades (Surrey and Jähnig 1995; Pocanschi, et al. 2006; Burgess, et al. 2008).

Many careful folding studies have been conducted to map out the OmpA folding pathway. These studies have illustrated the order of which parts of the protein contact the membrane (Kleinschmidt, et al. 1999; Kleinschmidt and Tamm 1999); the order of which parts of OmpA contact each other (Kleinschmidt, et al. 2011); and how folded different intermediates are (Danoff and Fleming 2017). Early work suggested that the loops contact the membrane first (Kleinschmidt, et al. 1999; Kleinschmidt and Tamm 1999) and that some hairpins were formed earlier than others (Kleinschmidt, et al. 2011). We had hypothesized that varying loops with shorter sequences that were predicted to easily form hydrogen bonding would facilitate a zipping-up of the beta strands and increase folding rates of OmpA. We did not find results that support that hypothesis. Our folding studies showed that though we find particular instances where we can increase and decrease the folding rate, overall most loop replacements leave the folding mostly unchanged. We anticipate that this is consistent with the conclusions of previous work that it may be the turns and not the loops that facilitate the zipping (Danoff and Fleming 2017). However, the lack of importance of loops to the folding rate may not be universal for all OMPs as shortening of the loops of BamA lead to systematic slowing of its folding (Tomasek, et al. 2020). More work will need to be done to determine the universality of the relative importance of the loops and turns in folding.

CONCLUSION

Overall, we find that the $PETALS_{OSCAR}$ algorithm is particularly well suited to predict the structures of loops in OMPs. Moreover we find that though we removed a larger portion of the loops of OmpA than have ever been removed before, folding of this protein remains remarkably consistent.

METHODS

Loop Definitions

Our prior work described the amino acid preferences observed in the 4-, 5- and 6-residue strand connectors of outer membrane β-barrels (OMBBs). The strand connectors in OMBB crystal structures were clustered by the φ and ψ angles of the central residues to define the amino acid preferences of each position (Franklin and Slusky 2018). We use the same definitions outlined therein. Briefly, a loop is the extracellular strand connector between two adjacent strands, inclusive of the last residue of the first strand and the first residue of the second strand, while turns are periplasmic strand connectors. When describing the five-residue loops designed here they are numbered by convention, i to i+4; by definition, the i and i+4 residues are strand residues.

Predicting loop structure

For each loop structure we independently generated an ensemble of 500 models using mDiSGro (Tang, et al. 2014; Tang, et al. 2015), PETALS (Wong, et al. 2017), and KIC in Rosetta using the remodel application (Mandell, et al. 2009).In the first round for each loop, knowledge of the angles of the i and i+4 residues were maintained while the coordinates of the central three residues (non-terminal residues) were predicted. For the second round of prediction and for the designed sequences, all loop coordinates were removed and the sequences were modeled from the loop-adjacent i-1 and i+5 residue angles known and the angles for i to i+4 unknown (Fig 2). Default parameters for each program were used to generate each ensemble, except as follows: in mDiSGro, the number of conformations was 10,000 and retained conformations was 1,000. In PETALS, the number of seeds was set to 5,000. Multiple scoring functions are implemented in PETALS and reported in the results; the default for loop building is DBB but each final model was rescored using several other scoring functions, so we also used the OSCAR-o scores for comparison. For KIC in Rosetta, the flag max_kic_build_attempts was set to 250; in the required blueprint file, when using for modeling non-native loops, the i and i+5 residues were also allowed to be remodeled (Fig 2).

The modeled structures were ranked according to the scores assigned during loop building, and the top 25 models were further analyzed for RMSD and recapture rates to determine the best model. RMSD values were calculated for the backbone atoms of all amino acids in the loop structure. A loop model was determined to be “recaptured” if all the φ and ψ angles of the non-terminal amino acid positions were less than 30° from the native structure.

Designed Loops

For each of the three loop-types that represented a spectrum of side-chain:backbone hydrogen bonding (Franklin and Slusky 2018) - PLL (brown), PLD (purple), and AAL (blue), sequences were generated from the amino acid preferences for each position. Specifically, for each position a script chose an amino acid based on the weighted probability of each amino acid identity for each position.

Cloning and Expression of OmpA Modified Loop proteins

Mature OmpA without the 22 amino acid signal sequence in the pET303 vector was purchased from Genscript for cloning of the modified loop constructs. Primers were designed for the alteration of each of the 4 loops using Q5 polymerase (NEB) PCR reaction and a KLD enzyme mix (NEB). The resulting plasmids were transformed into chemically competent DH5α cells and sequences were confirmed by Sanger double stranded DNA sequencing. All confirmed plasmids were transformed into BL21(DE3) cells for expression. Transformed cells were grown in Terrific Broth (Millipore) with 0.04 % v/v glycerol at 37°C with shaking at 250 rpm to an optical density of 0.6 at 600 nm. Expression was induced by the addition of 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) and incubation at 25°C with shaking at 200 rpm for 18 hours. Cells were then harvested by centrifugation for 30 minutes at 4,700 g at 4°C. Pellets were re-suspended in 50 mM Tris pH 8.5, 0.1 mg/mL lysozyme and 1 μM phenylmethylsulfonyl fluoride (PMSF) and incubated for 30 minutes on ice before being sonicated for 5 minutes on 50% duty cycle with 10 seconds off, 10 seconds on. Magnesium chloride (5 μM) and DNAse (5 μg/ml) were then added and incubation on ice continued another 30 minutes. Brij-L23 (Alfa Aesar) was added to a final concentration of 0.1 % v.v. Inclusion bodies were centrifuged at 4700 g for 30 minutes at 4°C then washed 2 times with 10 mM Tris, 2 mM EDTA, pH 8.2. The pellet was then re-suspended in wash buffer and divided into 2 tubes before a final centrifugation (4700 g, 30 min., 4°C). Supernatant was removed and the inclusion body pellets were stored at -20°C.

Preparation of Large Unilamellar Vesicles (LUVs)
LUVs were prepared using the methods previously described (Burgess, et al. 2008). 1,2-didecanoyl-sn-glycero-3-phosphocholine (PC-di_{C10}) (Avanti Polar Lipids) was dissolved in chloroform at 5.66 mg/mL then dried to a thin film in amber glass vials under a stream of nitrogen. The vials of lipid were then lyophilized overnight to remove residual solvent and then stored at -20°C until use. The lipid was then re-constituted in 1 mL of 20 mM Borate buffer pH 10 to give a final concentration of 10 mM, then gently vortexed. LUVs were prepared by extruding the 10 mM lipids 15 times through a 0.1 μm filter using a mini-extruder (Avanti Polar Lipids) and stored at 4°C until use.

Folding and SDS-PAGE
Inclusion body pellets were solubilized using 8 M urea, 10 mM borate pH 10, and 2 mM EDTA. After centrifugation at 18,000 g for 10 minutes the supernatant was filtered through a 0.45 μm syringe filter. Protein concentration was determined by measuring the absorbance at 280 nm and proteins were then diluted to 50 μM. Confirmation of the identity of the proteins was carried out using SDS-PAGE and MALDI-TOF mass spectrometry. Folding buffer was prepared to contain 3.2 mM LUVs, 1 M Urea and 2 mM EDTA in borate buffer pH 10. Folding was initiated with a one in eleven dilution of 50 μM protein into folding buffer LUVs in an amber glass vial, for a final concentration of 4.5 μM protein, 1.6 M Urea, 2.9 mM PC-diC10 LUVs, 9 mM borate pH 10, 1.8 mM EDTA. Folding samples were incubated at 22°C with shaking at 500 rpm in a thermal mixer block (Thermo Scientific) and time points were taken starting at 1 minute. At each time point, folding was quenched by taking 10 μL of sample and adding it to 10 μL of 2x Laemmli SDS gel-loading buffer without reducing agent (BioRad), vortexing and storing at 4°C until analysis. Unfolded protein in 8M Urea, 10 mM Borate pH 10 was prepared at 5 μM and loaded onto a 4-20% Mini-Protean TGX pre-cast gel (BioRad). 10 μL of each folding sample was subsequently loaded onto the same gel. After electrophoresis the gel was stained with Sypro Ruby overnight (chosen for its broad linear dynamic range). After de-staining in 10 % methanol, 7% acetic acid in water the gel was imaged using the BioRad Gel Doc.

Data Analysis and Calculations
Using densitometry the fraction folded was calculated by dividing the intensity of the folded band by the sum of the folded and unfolded bands. For each experiment, the folding kinetics were fit

to both a single and double exponential equation in Graphpad Prism 7 and the best fit was chosen according to R-squared value, see supplemental equations 1 and 2. When the single-state association was preferred, the observed k value is reported beside the observed k_{fast} value of the proteins with two-state folding. k_{fast} and k_{slow} are the two rate constants expressed in inverse seconds. The reported values and error bars (standard deviation) represent the average of three independent experiments.

ACKNOWLEDGMENTS
We gratefully acknowledge Jimmy Budiardjo, Rik Dhar, and Ryan Feehan for helpful conversations and NIH awards DP2GM128201, T32-GM008359, NSF MCB160205, The Gordon and Betty Moore Inventor Fellowship and KU-startup funds.

Supplemental

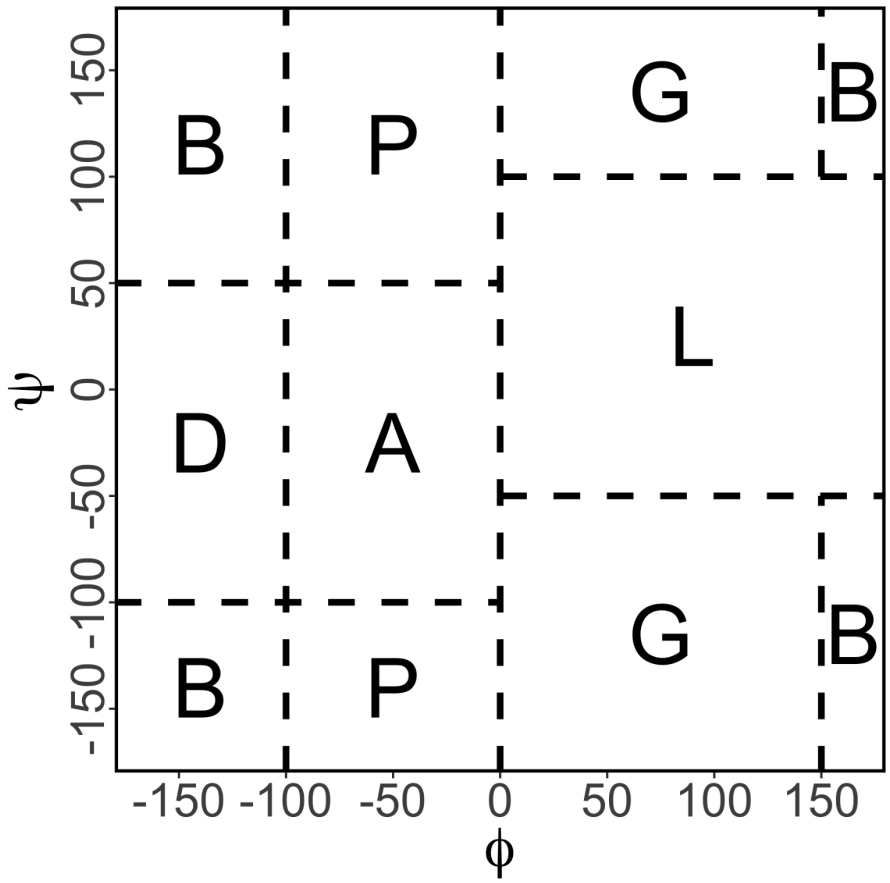


Figure S1. Divided Ramachandran Map. The Ramachandran map divided up into five parts for describing backbone angles of loops. A is for alpha; B is for beta; G is for glycine; P is for proline; L is for loop. A five residue loop with backbone angles for the i+1 position in A, the i+2 position in A and the i+3 position in L would be called an AAL loop.

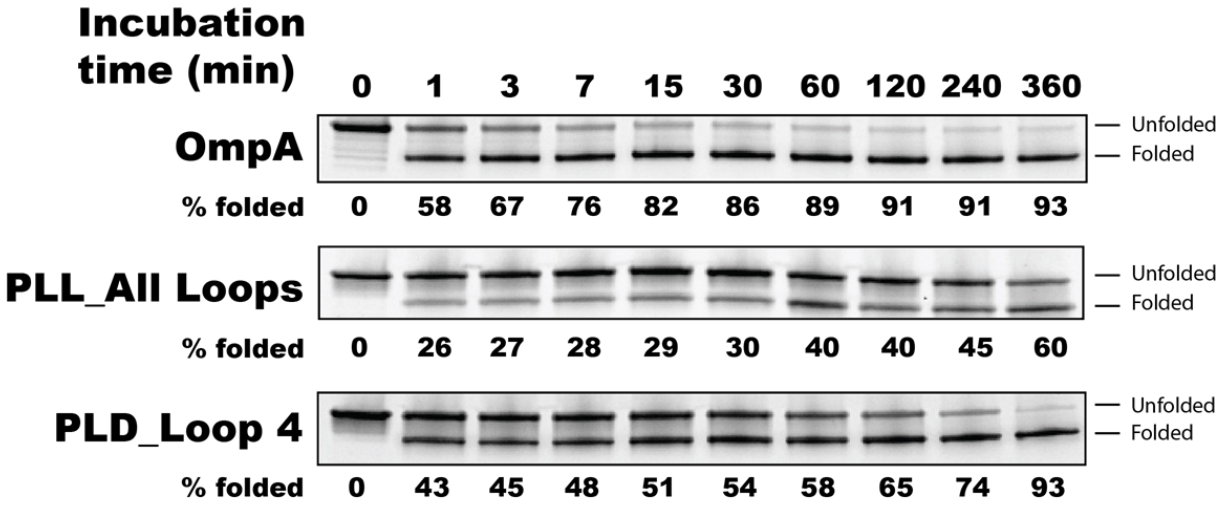


Figure S2. Representative de-stained SDS-Page gel of folding time course of substituted OmpA loops. Lane one is the unfolded protein in solubilization buffer (8 M urea) while lanes 2-10 are time points after mixing with LUVs in folding buffer. The folded protein runs faster than the unfolded protein. Data was collected in triplicate but only one gel of each is shown. Densitometry analysis was performed and the amount of protein was calculated as $Intensity_{Folded}/(Intensity_{Folded} + Intensity_{Unfolded})$, shown in Figure F.

The folding kinetics were fit to a one-state folding curve using a single exponential equation,

$$y = y_0 + Ae^{-kt} \quad (1)$$

or a two-state association curve using a double exponential equation,

$$y = y_0 + A_{fast}e^{-k_{fast}t} + A_{slow}e^{-k_{slow}t} \quad (2)$$

where *y* describes the fraction folded at time *t*. The fraction folded as time approaches infinity is described 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.

Other notes:

While our conditions (borate buffer instead of Tris, higher pH, lack of reducing agent, lower temperature) may not have been identical to other folding experiments for OmpA, our folding rate of k_fast = 29.7 x 10^-3 and k_slow = 1.3 x 10^-3 is similar to the published value of k_fast = 35 x 10^-3 and k_slow = 3 x 10^-3 (Burgess and Flemming, 2008).

References

Arora A, Abildgaard F, Bushweller JH, Tamm LK. 2001. Structure of outer membrane protein A transmembrane domain by NMR spectroscopy. Nature Structural Biology 8:334-338.

Beher MG, Schnaitman CA, Pugsley AP. 1980. Major heat-modifiable outer membrane protein in gram-negative bacteria: comparison with the ompA protein of Escherichia coli. Journal of Bacteriology 143:906-913.

Burgess NK, Dao TP, Stanley AM, Fleming KG. 2008. β-Barrel Proteins That Reside in the Escherichia coli Outer Membrane in Vivo Demonstrate Varied Folding Behavior in Vitro. Journal of Biological Chemistry 283:26748-26758.

Coutsias EA, Seok C, Jacobson MP, Dill KA. 2003. A kinematic view of loop closure. Journal of Compuational Chemistry 25:510-528.

Danoff EJ, Fleming KG. 2017. Novel Kinetic Intermediates Populated along the Folding Pathway of the Transmembrane β-Barrel OmpA. Biochemistry 56:47-60.

Fox DA, Larsson P, Lo RH, Kroncke BM, Kasson PM, Columbus L. 2014. Structure of the Neisserial Outer Membrane Protein Opa60: Loop Flexibility Essential to Receptor Recognition and Bacterial Engulfment. Journal of the American Chemical Society 136:9938-9946.

Franklin MW, Nepomnyachyi S, Feehan R, Ben-Tal N, Kolodny R, Slusky JSG. 2018. Evolutionary Pathways of Repeat Protein Topology in Bacterial Outer Membrane Proteins. eLife 7:e40308.

Franklin MW, Slusky JSG. 2018. Tight Turns of Outer Membrane Proteins: An Analysis of Sequence, Structure, and Hydrogen Bonding. J Mol Biol 430:3251-3265.

Gront D, Kmiecik S, Blaszczyk M, Ekonomiuk D, Koliński A. 2012. Optimization of protein models. Wiley Interdisciplinary Reviews: Computational Molecular Science 2:479-493.

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 (BBA) - Biomembranes 1838:3014-3024.

Kingma RL, Fragiathaki M, Snijder HJ, Dijkstra BW, Verheij HM, Dekker N, Egmond MR. 2000. Unusual Catalytic Triad of Escherichia coli Outer Membrane Phospholipase A. Biochemistry 39:10017-10022.

Kleinschmidt JH, Bulieris PV, Qu J, Dogterom M, den Blaauwen T. 2011. Association of Neighboring β-Strands of Outer Membrane Protein A in Lipid Bilayers Revealed by Site-Directed Fluorescence Quenching. Journal of Molecular Biology 407:316-332.

Kleinschmidt JH, den Blaauwen T, Driessen AJM, Tamm LK. 1999. Outer Membrane Protein A of Escherichia coli Inserts and Folds into Lipid Bilayers by a Concerted Mechanism. Biochemistry 38:5006-5016.

Kleinschmidt JH, Tamm LK. 2002. Secondary and Tertiary Structure Formation of the β-Barrel Membrane Protein OmpA is Synchronized and Depends on Membrane Thickness. Journal of Molecular Biology 324:319-330.

Kleinschmidt JH, Tamm LK. 1999. Time-Resolved Distance Determination by Tryptophan Fluorescence Quenching: Probing Intermediates in Membrane Protein Folding†. Biochemistry 38:4996-5005.

Koebnik R. 1999. Structural and Functional Roles of the Surface-Exposed Loops of the beta - Barrel Membrane Protein OmpA from Escherichia coli. J. Bacteriol. 181:3688-3694.

Mandell DJ, Coutsias EA, Kortemme T. 2009. Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. Nature Methods 6:551-552.

North B, Lehmann A, Dunbrack RL. 2011. A New Clustering of Antibody CDR Loop Conformations. Journal of Molecular Biology 406:228-256.

Pocanschi CL, Patel GJ, Marsh D, Kleinschmidt JH. 2006. Curvature Elasticity and Refolding of OmpA in Large Unilamellar Vesicles. Biophysical Journal 91:L75-L77.

Rose GD, Gierasch LM, Smith JA. 1985. Turns in peptides and proteins. Adv Protein Chem 37:1-109.

Rosenbusch JP. 1974. Characterization of the Major Envelope Protein from Escherichia coli : REGULAR ARRANGEMENT ON THE PEPTIDOGLYCAN AND UNUSUAL DODECYL SULFATE BINDING. Journal of Biological Chemistry 249:8019-8029.

Schubeis T, Le Marchand T, Daday C, Kopec W, Tekwani Movellan K, Stanek J, Schwarzer TS, Castiglione K, de Groot BL, Pintacuda G, et al. 2020. A β-barrel for oil transport through lipid membranes: Dynamic NMR structures of AlkL. Proceedings of the National Academy of Sciences 117:21014-21021.

Surrey T, Jähnig F. 1995. Kinetics of Folding and Membrane Insertion of a β-Barrel Membrane Protein. Journal of Biological Chemistry 270:28199-28203.

Surrey T, Jähnig F. 1992. Refolding and oriented insertion of a membrane protein into a lipid bilayer. Proceedings of the National Academy of Sciences of the United States of America 89:7457-7461.

Tang K, Wong SWK, Liu JS, Zhang J, Liang J. 2015. Conformational sampling and structure prediction of multiple interacting loops in soluble and b-barrel membrane proteins using multi-loop distance-guided chain-growth Monte Carlo method. Bioinformatics 31:2646-2652.

Tang K, Zhang J, Liang J. 2014. Fast Protein Loop Sampling and Structure Prediction Using Distance-Guided Sequential Chain-Growth Monte Carlo Method. PLOS Computational Biology 10:e1003539.

Tomasek D, Rawson S, Lee J, Wzorek JS, Harrison SC, Li Z, Kahne D. 2020. Structure of a nascent membrane protein as it folds on the BAM complex. Nature 583:473-478.

Vandeputte-Rutten L, Kramer RA, Kroon J, Dekker N, Egmond MR, Gros P. 2001. Crystal structure of the outer membrane protease OmpT from Escherichia coli suggests a novel catalytic site. EMBO J 20:5033-5039.

Vogel H, Jähnig F. 1986. Models for the structure of outer-membrane proteins of Escherichia coli derived from raman spectroscopy and prediction methods. Journal of Molecular Biology 190:191-199.

Wong SWK, Liu JS, Kou SC. 2017. Fast de novo discovery of low-energy protein loop conformations. Proteins: Structure, Function, and Bioinformatics 85:1402-1412.

Yildiz Ö, Vinothkumar KR, Goswami P, Kühlbrandt W. 2006. Structure of the monomeric outer-membrane porin OmpG in the open and closed conformation. The EMBO Journal 25:3702-3713.

Zahn M, Bhamidimarri SP, Basle A, Winterhalter M, van den Berg B. 2015. Structural Insights into Outer Membrane Permeability of Acinetobacter baumannii. Structure 24:221-231.