Rational design of a fusion partner for membrane protein expression in *E. coli*

Jianying Luo, Julie Choulet, and James C. Samuelson*

New England Biolabs, Inc. Gene Expression Division, 240 County Road, Ipswich, Massachusetts 01938

Received 30 January 2009; Revised 24 April 2009; Accepted 8 June 2009
DOI: 10.1002/pro.189
Published online 15 June 2009 proteinscience.org

Abstract: We have designed a novel protein fusion partner (P8CBD) to utilize the co-translational SRP pathway in order to target heterologous proteins to the *E. coli* inner membrane. SRP-dependence was demonstrated by analyzing the membrane translocation of P8CBD-PhoA fusion proteins in wt and SRP-*ffh77* mutant cells. We also demonstrate that the P8CBD N-terminal fusion partner promotes over-expression of a *Thermotoga maritima* polytopic membrane protein by replacement of the native signal anchor sequence. Furthermore, the yeast mitochondrial inner membrane protein Oxa1p was expressed as a P8CBD fusion and shown to function within the *E. coli* inner membrane. In this example, the mitochondrial targeting peptide was replaced by P8CBD. Several practical features were incorporated into the P8CBD expression system to aid in protein detection, purification, and optional *in vitro* processing by enterokinase. The basis of membrane protein over-expression toxicity is discussed and solutions to this problem are presented. We anticipate that this optimized expression system will aid in the isolation and study of various recombinant forms of membrane-associated protein.

Keywords: SRP; signal peptide; fusion partner; P8; membrane protein expression

Introduction

The study of α-helical membrane proteins (MPs) presents unique challenges due to the hydrophobic nature of their transmembrane segments (TM). *E.coli* over-expression of MPs typically results in saturation of the membrane protein biogenesis pathway and cytoplasmic accumulation of inactive aggregated protein.[1] Although 20–30% of prokaryotic and eukaryotic ORFs encode putative α-helical MPs, the collection of such protein structures in the Protein Data Bank is just 1.5% of the total number of entries.[2,3] This situation is primarily due to the inability to over-produce MPs in functional form and secondarily due to the difficulty in obtaining diffraction quality crystals. Our efforts are focused on the challenge of over-producing α-helical MPs in *E. coli* by optimizing targeting to the inner membrane. Most native *E.coli* inner membrane proteins are thought to integrate in a co-translational manner, via the Signal Recognition Particle (SRP) and the SRP receptor FtsY.[4–6] An early step in this pathway is SRP recognition of a sufficiently hydrophobic N-terminal signal sequence upon emergence from the ribosome tunnel.[7] In contrast, most exported proteins possess less hydrophobic N-terminal signals, which evade SRP recognition thereby allowing translation to be completed before targeting of the protein to the SecYEG translocase. Most soluble exported proteins are thought to remain in a partially folded, translocation-competent form by interaction with chaperones such as SecB. One notable exception is the SRP-dependent export of DsbA to the periplasm.[8]

SRP-dependent membrane translocation is important to prevent release of nascent membrane proteins into the cytoplasm where the hydrophobic segments promote aggregation. Also, tightly folded protein domains may jam the translocation machinery if they are presented to SecYEG protein conducting channel

Additional Supporting Information may be found in the online version of this article.

JCS is an owner of New England Biolabs Stock.

Grant sponsor: NIH-SBIR; Grant number: 1R43 GM 083413-01.

*Correspondence to: James C. Samuelson, New England Biolabs, Inc., Gene Expression Division, 240 County Road, Ipswich, Massachusetts 01938. E-mail: samuelson@neb.com

after translation has been completed. Thus, engineering a protein to utilize the SRP pathway may directly result in reduced cytoplasmic toxicity and increased throughput through the SecYEG translocase. Bowers *et al.*[9] recently reported one such example of protein engineering. For many years, the LamB-LacZ42-1 fusion protein has been employed as a model to study the secretion process. This fusion protein depends on the SecB chaperone for post-translational targeting to the SecYEG translocase where it jams the secretion pathway, presumably by rapid, stable folding of LacZ. However, Bowers *et al.* demonstrated that Sec pathway jamming may be alleviated simply by increasing the hydrophobicity of the LamB-LacZ signal peptide to promote SRP recognition. The finding that SRP-dependence may reduce over-expression toxicity is especially relevant when attempting to produce heterologous MPs in *E. coli*. In particular, in the case of some eukaryotic organellar MPs, a proper signal for *E.coli* inner membrane targeting may not be present.

The hydrophobic character of a protein is certainly not the only requirement for efficient membrane insertion in *E. coli*. The net charge distribution of acidic and basic residues flanking TM segments is clearly important for the determination of membrane topology.[10] Remarkably, addition of a single arginine or lysine has been shown to block the membrane translocation of a model protein in *E. coli*.[11] Realization of the higher frequency of basic residues within the cytoplasm segments of bacterial membrane proteins in the late 1980s led von Heijne to propose the "positive-inside rule"[12] to guide researchers in predicting α-helical membrane protein topology. In subsequent years, many studies have demonstrated that a net negative charge in extra-cytoplasmic sequences flanking either a TM segment or a signal peptide promotes membrane translocation.[13–15] This phenomenon is related to the electro-chemical membrane potential across the inner membrane of *E. coli*. Altering the amino acid charge distribution of the model MP "Pf3 coat" results in a complete reversal of membrane topology.[16] This study also demonstrated that the membrane potential-dependent translocation of Pf3 coat protein requires the presence of acidic residues in the translocated N-terminal segment. The major coat protein of M13 bacteriophage (M13 Procoat, pVIII or P8 for simplicity) has also been used as a model protein to study membrane potential-dependent translocation. Several studies have concluded that the P8 protein possesses an optimal amino acid charge distribution for efficient inner membrane insertion without the aid of the SecYEG translocase.[11,17]

After contemplating the wealth of knowledge concerning membrane protein biogenesis in *E. coli*, we have designed the P8CBD fusion partner to aid in the over-expression of heterologous membrane proteins. The membrane protein fusion partner P8CBD possesses the following desirable characteristics: (1) SRP-dependence, (2) optimal charge distribution of the N-terminal 73 amino acids which dictates the topology of the fused protein of interest and (3) a net negative charge at the end of TM2 (at the fusion junction) to promote membrane translocation of the fused protein of interest. Other practical features have been incorporated to enable P8CBD fusion protein detection, purification, and optional *in vitro* cleavage by enterokinase (P8CBDek).

Results

Design of the P8CBD fusion partner

The P8CBD N-terminal fusion partner was designed for optimal membrane targeting. The P8CBD sequence begins with the wild-type M13 phage major coat protein, hereby referred to as P8. Interestingly, wild-type P8 does not depend on SRP for membrane targeting, but does require the YidC insertase for proper membrane assembly.[18,19] The N-terminal processed signal peptide of P8 is only moderately hydrophobic (1.421 on the Goldman-Engelman-Steitz scale[20]). However, the P8 TM segment (amino acids 44–64) is sufficiently hydrophobic for SRP recognition (1.801 on the GES scale). Translation of the entire 73 amino acid P8 protein may be completed before the hydrophobic TM segment emerges from the ribosome tunnel, so the lack of SRP/FtsY recognition may simply be due to the short length of wild-type P8. Engelman and Steitz[21] proposed that the P8 phage protein has evolved to efficiently insert into membranes by forming a helical hairpin structure by association of its two amphipathic helices. Although wild-type P8 does not require SRP, it is conceivable that P8 fusion proteins may be recognized by SRP due to the hydrophobic nature of residues 44–64. In fact, de Gier *et al.*[18] demonstrated SRP-dependence of a protein where P8 was fused at the C-terminus to a soluble domain of 104 amino acids. Therefore, we anticipated that P8-membrane protein fusions would also be recognized by SRP.

Figure 1 shows the membrane topology of P8CBD. The P8 TM segment is drawn as a helix (the processed signal peptide is not shown). The P8 ORF was extended by fusing DNA encoding 58 amino acids of *E. coli* Signal peptidase to provide a second TM segment (TM2) in order to extend the protein fusion junction into the periplasmic space. TM2 from Signal peptidase was chosen since this peptide is known to efficiently establish the desired orientation within the inner membrane. Furthermore, TM2 from Signal peptidase acts an efficient inner membrane export signal in its native context and in the context of the P8CBD fusion partner. The chitin binding domain (CBD) was engineered into the cytoplasmic loop to serve as an optional affinity tag or detection epitope. Insertion of CBD or other affinity domains at this position does not affect the membrane targeting properties (data not shown). Finally, the Enterokinase (ek) protease

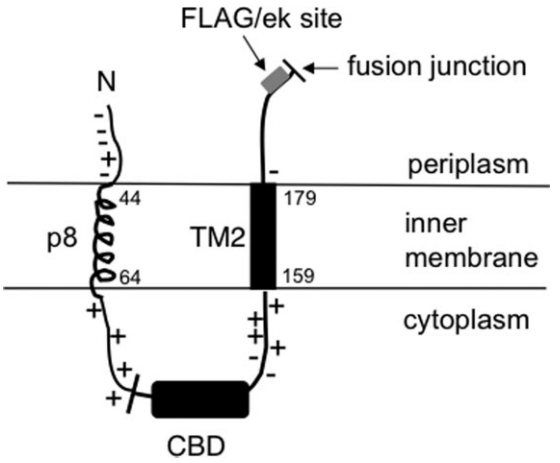

Figure 1. The P8CBDek N-terminal protein fusion partner. The P8 transmembrane segment is drawn as a helix. CBD is the chitin binding domain from *Bacillus circulans*. The TM2 is transmembrane segment 2 from *E. coli* Signal peptidase. The FLAG epitope and corresponding enterokinase (ek) protease site are positioned at the protein fusion junction. The membrane topology conforms to the "positive-inside" rule as shown. FLAG epitope (DYKDDDDK) and CBD charges are not shown for clarity.

cleavage site (DDDDK↓) and corresponding FLAG epitope (DYKDDDDK) were incorporated at the fusion junction.

SRP-dependent membrane targeting by P8CBD

The P8CBD fusion partner was designed to utilize the Signal Recognition Particle (SRP) membrane targeting pathway. SRP-dependence was demonstrated by comparing the membrane translocation of *E. coli* alkaline phosphatase (PhoA) fused to the P8CBD fusion partner in wild-type SRP and mutant SRP-*ffh77* cells. In this cell-based assay, PhoA is only functional when translocated across the inner membrane to the periplasm (after being targeted to the SecYEG translocase by either SecB or SRP). The *ffh77* strain was constructed to carry the chromosomal point mutation described by Tian and Beckwith,[22] which results in expression of an A37P variant of Ffh (SRP *F*ifty-*f*our *h*omolog). PhoA fusion proteins were induced with 100 μM IPTG and expressed for 60 min at 30°C. Each PhoA activity result is the average of two individual cultures derived from independent colonies. In the wild-type SRP strain MC1061, exceptionally high phosphatase activity was detected upon expression of each P8CBD-PhoA fusion protein [Fig. 2(A)]. In contrast, phosphatase activity levels were significantly less in *ffh77* cells, ranging from 31–37% of the level measured in MC1061 cells. Importantly, Figure 2(B) shows that the P8CBD-PhoA fusion protein expression level was similar in each strain: wt MC1061 versus *ffh77*. Accordingly, we conclude that the lower levels of PhoA activity in *ffh77* cells relate to a deficiency in SRP-mediated membrane targeting and that P8CBD fusion proteins utilize the SRP pathway. Figure 2(B) also shows that P8CBD fusions are not processed by Signal peptidase to the same extent in *ffh77* cells (Lane 8 vs. Lane 7 and Lane 10 vs. Lane 9), providing additional direct evidence for less efficient inner membrane targeting in cells where the function of the SRP pathway is compromised (doublet in Lanes 8 and 10 indicates precursor and mature forms of P8CBD-PhoA and P8CBDek-PhoA, respectively). Supporting Information Figures 2(A) and 2(B) show protease mapping results where the upper precursor band is protected from digestion thus confirming cytoplasmic

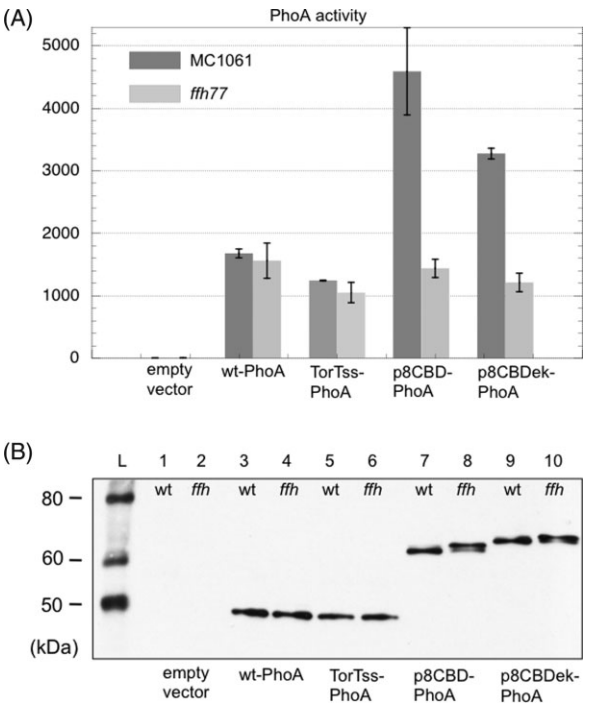

Figure 2. (A) PhoA activity is a measure of membrane targeting and membrane translocation of PhoA in wt MC1061 cells versus mutant *ffh77* cells. Columns represent the average of duplicate PhoA assays derived from individual cultures. Standard deviation is shown as error bars. MC1061 (*phoA*+) and *ffh77* cells (*phoA*+) carrying an empty vector produced negligible phosphatase activity. wt-PhoA is expression of the wt PhoA protein from parent vector pJF119. TorTss-PhoA is expression of PhoA fused to the TorT signal sequence. P8CBD-PhoA and P8CBDek-PhoA expression results in export of PhoA into the periplasm more efficiently in MC1061 (*ffh*+) cells. (B) Culture samples from (A) were analyzed by SDS-PAGE/Western blot using anti-PhoA monoclonal antibody. Lanes 1–10 exhibit the PhoA proteins assayed in equivalent lanes in panel A. Importantly, P8CBD-PhoA and P8CBDek-PhoA samples were diluted 20-fold (relative to wt-PhoA and TorTss-PhoA samples) before loading. wt-PhoA and TorTss-PhoA proteins are targeted efficiently in *ffh77* cells as shown by a single protein species (mature form). P8CBD-PhoA and P8CBDek-PhoA fusions are not targeted efficiently in *ffh77* cells (doublet in Lanes 8 and 10 indicates P8 single peptide is not processed by Signal peptidase). Lane (L) is a biotinylated ladder detected by HRP-conjugated anti-biotin antibody (Cell Signaling Technology, Beverly MA).

localization of a significant fraction of the PhoA domain in *ffh77* cells. PhoA was also expressed with its wild-type signal peptide and with an exceptionally hydrophobic signal sequence from the *E. coli* protein TorT. As expected, wild-type PhoA export was not significantly affected in *ffh77* cells as the native protein is known to be exported by the post-translational SecB pathway. When expressed with the TorT signal, PhoA export also was not affected by the *ffh77* host mutation. This result was unexpected as the TorT signal was shown to promote SRP-dependent export of thioredoxin.[23] Finally, a practical consideration is that expression and membrane translocation of the PhoA domain is significantly higher when fused to P8CBD as compared to the two *E. coli* signal peptides.

Functional expression of two heterologous membrane proteins

Recombinant over-expression of membrane proteins is complicated by the fact that most membrane proteins cannot be easily assayed for activity. However, one approach to test protein function and proper membrane assembly is to use a host suitable for complementation. The *E. coli* inner membrane protein YidC is essential for cell viability.[19,24,25] In the YidC-depletion strain JS7131, the *araBAD* promoter controls YidC expression from the chromosome. In JS7131, YidC expression is induced by arabinose and the level of YidC protein in the cell may be depleted by growth in the presence of glucose. Thus, JS7131 does not grow on glucose plates unless transformed with a plasmid induced to express a complementing protein. Complementation of YidC-depletion has been shown previously: The YidC homologue from chloroplasts (Alb3) and the yeast mitochondrial protein Oxa1p were both shown to function in *E. coli* when each was expressed as a chimeric fusion to the N-terminal region of YidC.[25,26] In this study, we have demonstrated complementation of YidC depletion by expression of *Thermotoga maritima* YidC or Oxa1p as fusions to the P8CBD membrane targeting fusion partner. PSI-BLAST[27] was used to search for a homologue of the *E. coli* YidC 548 amino acid protein (accession number NP_418161) within the ORFs of the *T. maritima* MSB8 genome sequence.[28] A 445 amino acid *Tma* protein (accession number NP_229260) was identified as the homologue of *E. coli* YidC. The two YidC proteins are 48% similar and 35% identical throughout transmembrane segments 2–5. The N-terminal amino acids of YidC family proteins are less conserved. Figure 3(A) shows the membrane topology of wild-type *Tma*-YidC as predicted by the Phobius web server[29] while Figure 3(B) shows the predicted topology of a fusion protein where the wild-type *Tma*-YidC signal peptide was replaced by P8CBD. The P8CBD-*Tma*YidC fusion construct was created by replacement of the first 29 codons of the *Tma-yidC* ORF by the P8CBD coding sequence (189

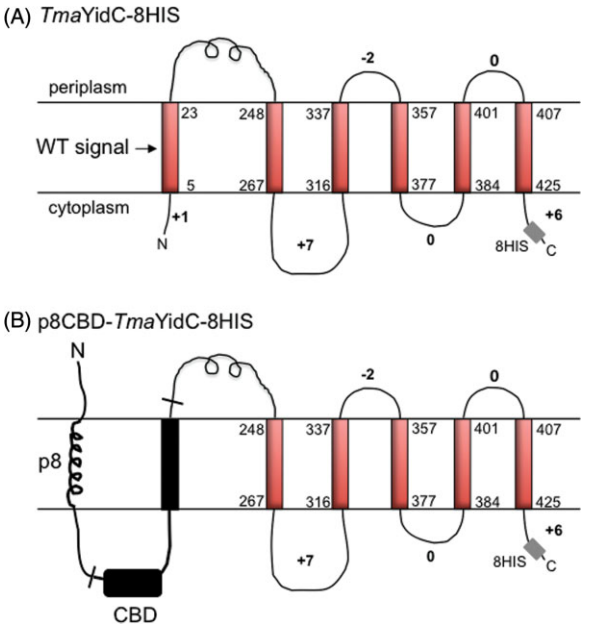

Figure 3. (A) The Phobius web server was used to predict the transmembrane topology of *Tma*-YidC (Genbank accession number NP_229260). (B) The predicted membrane topology of a P8CBD-*Tma*YidC fusion protein. Both proteins were expressed with a C-terminal eight histidine tag. Numbers shown next to the periplasmic and cytoplasmic loops indicate the net charge of the amino acid side chains within these loops. Coordinates of the TM segments are indicated by amino acid number of the wild-type protein. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

amino acids ending in SGS). When plasmids encoding either protein were transformed into JS7131, colony formation was dependent on the presence of 10 μM IPTG to induce the respective recombinant YidC protein from a *Ptac* promoter (data not shown). In this experiment, glucose was added to the media to turn-off expression of the chromosomal *ParaBAD-yidC* gene. These results show that the hypothetical *yidC* gene from *Thermotoga maritima* encodes a membrane protein able to functionally complement the loss of *E. coli* YidC. However, more importantly complementation by P8CBD-*Tma*YidC8HIS demonstrates that the P8CBD fusion partner performs as designed without interfering with the function of the fused protein of interest.

Figures 4(A,B) show the results of over-expressing the two forms of *Tma*-YidC8HIS in strain MC1061. SDS-PAGE and Western blot analyses reveal that wt *Tma*-YidC8HIS expression is limited and does not respond to 100 μM IPTG addition. However, expression of the P8CBD-*Tma*YidC8HIS fusion protein is detectable by SDS-PAGE and does respond significantly to 100 μM IPTG addition. Figure 4(C,D) show expression of P8CBD-*Tma*YidC8HIS in the preferred protein expression strain NEB Express. NEB Express (a BL21 derivative) shows inducible expression and no

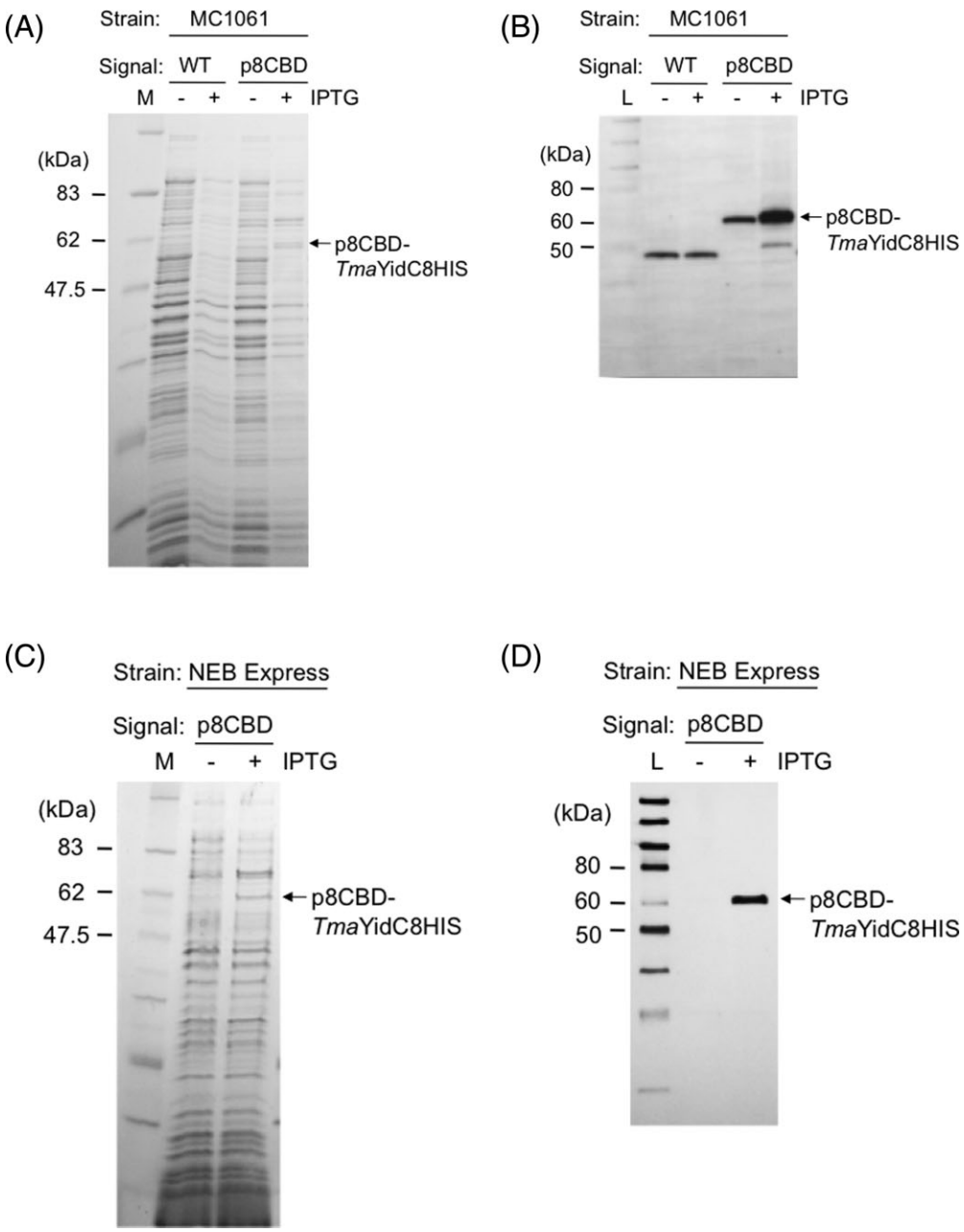

Figure 4. (A) SDS-PAGE analysis of *Tma*YidC8HIS over-expression in MC1061 cells either mediated by the wt signal peptide or the P8CBD fusion partner. Lane (M) is a pre-stained protein marker (NEB #P7708) (+) IPTG indicates induction with 100 μ*M* IPTG. (B) Western detection (anti-His tag antibody) of the same samples analyzed in panel A. Lane (L) is a biotinylated ladder detected by HRP-conjugated anti-biotin antibody. (C) SDS-PAGE analysis of P8CBD-*Tma*YidC8HIS over-expression in NEB Express cells. (+) IPTG indicates induction with 100 μ*M* IPTG. (D) Western detection (anti-His tag antibody) of the same samples analyzed in panel C. The calculated molecular weight of of wt *Tma*YidC is 51 kDa and the calculated molecular weight of mature P8CBD-*Tma*YidC is 67 kDa. Both proteins run anomalously in SDS gels due to their hydrophobic nature.

degradation of P8CBD-*Tma*YidC8HIS when detected by anti-His tag immunoblotting.

We also designed an Oxa1p fusion protein in order to test function and over-expression in *E. coli*. In yeast, Oxa1p is required for the biogenesis of certain respiratory chain complexes,[30] a function that has been shown to be conserved among Oxa1p/YidC family members.[24] In *Saccharomyces cerevisiae*, Oxa1p is encoded by a nuclear gene (*OXA1*) and the protein is targeted to the mitochondrial inner membrane by a mitochondrial matrix targeting peptide (amino acids 1-42). The P8CBD-Oxa1p fusion construct was designed so that the Oxa1p ORF begins at codon 43. Therefore, the matrix targeting peptide is replaced by P8CBD. In addition, the Oxa1p ORF was extended at the 3′ end to encode a C-terminal 8-HIS tag. Figure 5(A) shows a genetic representation of the initial *OXA1* construct tested for complementation. We

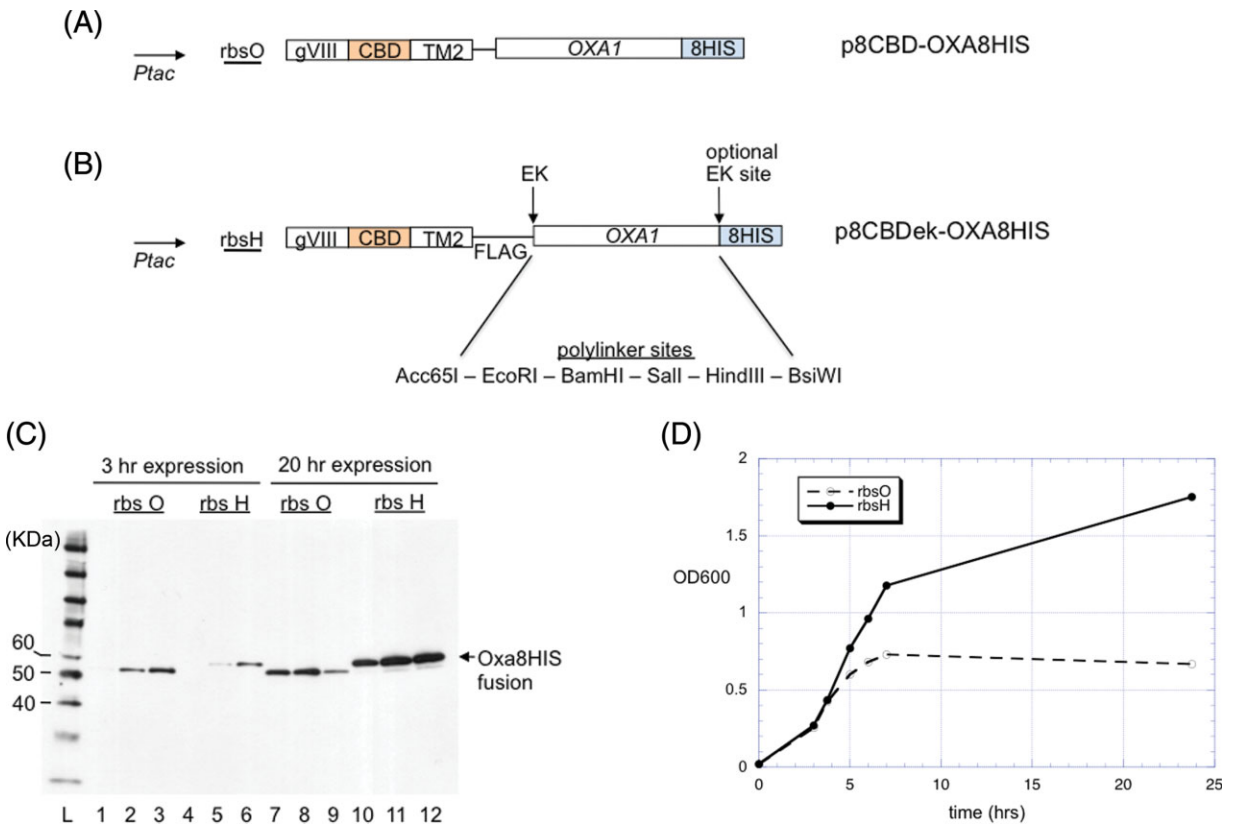

Figure 5. (A) a genetic representation of the initial *OXA1* construct tested for complementation. rbsO is the original wt ribosomal binding site of M13 phage gene VIII (gVIII) that expresses the P8 major coat protein. (B) a genetic representation of the second *OXA1* construct tested for complementation. rbsH is a weaker translational initiation signal than rbsO. The *OXA1* gene was cloned into the Acc65I and BamHI sites of the improved p8CBDek polylinker. (C) Results of over-expressing the Oxa8HIS fusion protein in NEB Express carrying the rbsO and rbsH constructs shown in A and B. Oxa8HIS fusion protein was detected by Western blot using anti-His tag antibody. Lane (L) is a biotinylated ladder detected by HRP-conjugated anti-biotin antibody. Each set of three lanes (1–3, 4–6, 7–9, 10–12) represent induction with 0, 40, or 400 μ*M* IPTG, respectively. Note that the weaker rbsO construct generates a significantly greater level of P8CBDek-Oxa8HIS fusion protein after 20 h of induction. (D) Alteration of rbs strength relieves cell growth arrest observed upon induction of a model membrane protein. Displayed is the optical density of two cultures grown in shake flasks and induced to express P8CBD-Oxa8HIS with 400 μ*M* IPTG at 4 h post-inoculation until 24 h post-inoculation (Lanes 9 and 12 in part C display the relative level of protein expressed in the two selected cultures). Expression from the rbsH construct results in a 2.5-fold higher cell density after a 20 h induction period. In addition, protein expression per cell is significantly greater when using rbsH versus rbsO (compare Lane 12 vs Lane 9 in part C). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

observed that the P8CBD-Oxa8HIS protein complemented the depletion of YidC when the fusion protein was induced with 10 μ*M* IPTG (data not shown). However, agar plates containing 100 μ*M* IPTG did not produce colonies of JS7131 indicating that this level of Oxa1p induction may be lethal to *E. coli*. Heterologous membrane protein over-expression toxicity is well-documented. One primary cause of such toxicity is that the SecYEG translocase becomes overwhelmed with the over-expressed protein and Sec-dependent endogenous proteins accumulate within the cytoplasm.[1] A simple remedy to alleviate such toxicity is to moderate the expression of the heterologous membrane protein to a level where host cell growth is maintained. We accomplished this situation by mutating the ribosomal binding site (rbs) of the *OXA1* construct to moderate the level of translation. Work within our laboratory previously confirmed that rbsH from supplemental Figure 2(b) of Gardner *et al.*[31] is a less efficient translation signal than the original wild-type P8 ribosomal binding site (labeled as rbsO). Figure 5(B) shows an outline of the rbsH-*OXA1* construct that was secondarily tested for complementation. As expected, complementation of YidC depletion using the rbsH-*OXA1* construct required the addition of higher levels of IPTG (100 μ*M*) to produce colonies. The practical benefits of employing a weaker rbs (in combination with a *tac* promoter) for membrane protein expression are demonstrated in Figure 5(C,D). Over-expression of the Oxa8HIS fusion protein is improved significantly when comparing the final yield of protein after 20 h of induction with either 40 or 400 μ*M* IPTG (Fig. 5(B) Lanes 11 and 12 vs. Lanes 8 and 9, respectively). Strikingly, even with saturating

levels of IPTG (400 μM), cells expressing the rbsH construct continued to grow in liquid culture for at least 20 h [Fig. 5(D)] and produced the target protein. The yield of protein per cell was in fact greater using the weaker rbsH. All expression experiments were conducted using Terrific Broth and induction was carried out at 20°C. These conditions consistently produced the best results when expressing membrane proteins fused to P8CBD.

The p8CBDek expression vector contains all necessary features

When creating the final *OXA1*-rbsH construct, a multipurpose polylinker was incorporated [Fig. 5(B)]. Furthermore, nucleotide sequence coding for the enterokinase recognition site (DDDDK) and corresponding FLAG epitope (DYKDDDDK) was inserted in-frame at the protein fusion junction. Figure 6 displays the results of a pilot experiment where the P8CBDek-Oxa8HIS fusion protein was purified by Ni-NTA chromatography and subsequently cleaved by Enterokinase. Interestingly, a higher order form of the fusion protein is detected that is stable during SDS-PAGE analysis. Gel migration indicates that the Oxa1p fusion is present as a tetramer, consistent with a report of the native protein homo-oligomeric state when purified from *Neurospora crassa* mitochondria.[32] The polylinker of the p8CBDek vector also codes for an optional ek site following the target protein. This allows for removal of a C-terminal affinity tag (if desired) during proteolytic processing of the fusion protein. Several detection epitopes are present within the P8CBDek fusion partner: First, the mature N-terminus of P8 may be detected by monoclonal antibody B62-FE2.[33] Second, the CBD tag may be detected by polyclonal or monoclonal antibody products (NEB). The FLAG epitope may be detected by the Anti-FLAG M2 antibody (Sigma) and in our studies, C-terminal 8HIS-tagged proteins were detected by monoclonal antibody.

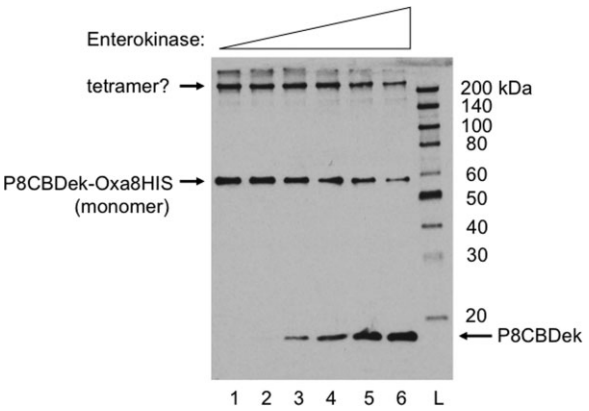

Figure 6. Pilot experiment demonstrating cleavage of P8CBDek-Oxa8HIS fusion protein by various amounts of Enterokinase. The ratio of Enterokinase to protein are as follows: Lane 1, 0.001%; Lane 2, 0.005%; Lane 3, 0.01%; Lane 4, 0.02%; Lane 5, 0.05%; Lane 6, 0.1%. Lane L is a biotinylated protein ladder. P8CBDek and P8CBDek-Oxa8HIS were detected by anti-P8 monoclonal antibody.

Discussion

The correlation of SRP-dependence to signal sequence hydrophobicity has been documented in several studies. Huber demonstrated that most *E. coli* signal peptides with hydrophobicity scores greater than that of DsbA mediated SRP-dependent export of the rapidly folding thioredoxin protein, but there were clear exceptions. For example, the very hydrophobic TreA signal did not direct export of thioredoxin. The motivation for using an SRP-dependent signal peptide is generally to promote the co-translational export or membrane integration of the target protein to increase the probability that the target protein is presented to the SecYEG translocase in an export-competent, unfolded state. Our evaluation of TorTss-PhoA export indicates that the simple solution of fusing a hydrophobic signal peptide to a protein of interest does not dictate SRP-dependent inner membrane translocation. However, each variant of P8CBD displayed SRP-dependence when fused to the PhoA domain. When P8CBD was fused to two different polytopic membrane proteins (*Tma*YidC and Oxa1p) SRP-dependence was not directly analyzed. Both proteins possess large periplasmic domains in addition to multiple TM segments so membrane integration after release into the cytoplasm is extremely unlikely. Importantly, both *Tma*YidC and Oxa1p were shown to be functional when moderately over-expressed as fusions to the P8CBDek membrane targeting fusion partner. In the case of *Tma*YidC, P8CBD simply replaced the native uncleaved N-terminal signal anchor. For Oxa1p expression, P8CBDek replaced the mitochondrial matrix targeting peptide. We anticipate that other heterologous membrane proteins may be over-expressed in *E. coli* using the P8CBD strategy by replacement of the native N-terminal signal or targeting sequence. These examples are especially encouraging for the over-expression of other α-helical membrane proteins with an N-out topology (e.g. GPCRs).

The P8CBD membrane-targeting fusion partner was rationally designed with practical needs in mind. Firstly, multiple epitopes allow for immuno-detection and verification of full-length expression of the protein of interest. The N-terminus of mature P8 was readily detected in our studies by monoclonal antibody B62-FE2 (see Fig. 6). The cytoplasmic CBD domain may also be detected by poly- or monoclonal antibody reagents (NEB). Furthermore, FLAG immuno-detection of the enterokinase (ek) recognition site is possible for P8CBDek fusion proteins. Poly-histidine sequences may be incorporated to aid protein purification and immuno-detection. We expect that C-terminal tag addition will be preferred so that only proteins expressed in full-length are isolated by metal-chelate

chromatography. N-terminal His-tagging is not a viable approach as the N-terminus of P8 is subject to Signal peptidase processing. However, a second option for poly-histidine tagging is to replace the CBD tag sequence in the cytoplasmic loop. Polyhistidine sequences in extra-cytoplasmic regions are not recommended as membrane translocation may be affected by basic amino acid residues. This may explain the reduced export of PhoA when fused to P8CBDek versus P8CBD [Fig. 2(A)], as the only difference in the two constructs is the insertion of the DYKDDDDKGS amino acid sequence at the fusion junction. Even with the observed reduction in PhoA export, the P8CBDek fusion partner performed very well in a direct comparison to membrane targeting mediated by two different *E. coli* signal peptides. In particular, the P8CBDek fusion partner outperformed the wild-type PhoA signal in terms of producing periplasmic functional PhoA activity. Current work in the lab involves testing a P8CBD fusion partner containing a Signal peptidase (sp) site at the end of TM2 (vector p8CBDsp) so that soluble proteins may be constitutively delivered into the periplasmic space.

Protein expression level is especially sensitive to the DNA sequence at the 5′ end of the ORF. For example, Qing *et al.*[34] reported that translation initiation in *E. coli* is enhanced by A/T-rich sequences downstream of the start codon. Of note is that the P8 ORF (gene VIII from M13 phage) happens to begin with the coding sequence 5′-ATGAAAAAG... Efficient translation of P8CBD apparently provides a running start for the expression of membrane protein fusions and may minimize potential bias against translation of heterologous proteins where the genes have not been codon-optimized for *E. coli*. The *Thermatoga maritima yidC* gene contains ten AGA and two AGG arginine codons that are rarely used in *E. coli*. Specifically, an AGA codon at position four of the native signal might be limiting the expression of the wild-type *Tma* protein in *E. coli* as rare codons at the 5′ end of an ORF have been shown to be more influential than rare codons in other segments of an ORF. The finding that a weaker ribosomal binding site results in a greater level of over-expressed membrane protein might be puzzling. However, one must simply imagine the SecYEG translocase as a bottleneck to understand why tuning the expression level of a membrane protein is so critical.

Our results in over-expressing *Tma*-YidC as a P8CBD fusion have prompted the notion of pursuing structural studies since this thermostable member of the YidC family may be more amenable to x-ray crystallography. Other future studies will include testing the over-expression of human membrane receptors and quantifying the level of protein capable of ligand binding. Alternative protease sites may be tested at the fusion junction for their effect on membrane translocation and for susceptibility to *in vitro* cleavage. Finally, we expect that the P8CBD fusion technology will help to advance many studies of membrane protein structure and function.

Materials and Methods

Bacterial strains and plasmids

This MC1061-*ffh77* strain was modeled after the *ffh77* strain isolated by Tian and Beckwith.[22] The *ffh77* allele refers to a single base change resulting in an alanine to proline substitution at position 37. MC1061-*ffh77* was derived from MC1061[35] using the allele exchange method described by Hamilton *et al.*[36] NEB 10-beta competent *E. coli* (New England Biolabs, Ipswich MA) was employed for cloning and propagation of plasmid DNA. The genotype of NEB 10-beta is: *fhuA mcrA ϕ80D(lacZ)M15 DlacX74 endA1 recA1 araD139 D(ara,leu)7697 galU galK rpsL nupG D(mrr-hsdRMS-mcrBC*). The genotype of NEB Express is: *fhuA2 [lon] ompT gal sulA11 R(mcr-73::miniTn10–TetS)2 [dcm] R(zgb-210::Tn10–TetS) endA1 Δ(mcrC-mrr)114::IS10*. JS7131[19] was used for YidC-complementation assays. JS7131 growth requires Rich media or agar containing 10 g/L sodium chloride and 0.2% L-arabinose. Complementation assays were performed using rich agar plates containing 0.1% glucose, 10 g/L sodium chloride and either 0, 10, or 100 μ*M* IPTG. Colonies appearing on complementation plates were re-streaked on glucose/IPTG plates versus glucose plates to confirm the IPTG requirement.

The p8CBDek vector was derived from a plasmid clone referred to as Procoat-Lep in Samuelson *et al.*[37] An ApoI fragment from the *lepB* gene encoding 58 amino acids of *E. coli* Signal peptidase was ligated into the EcoRI site of the Procoat-Lep plasmid (replacing the coding sequence for the LepP2 domain). This created plasmid p8-TM2. The *Bacillus circulans* CBD gene fragment (165 bp) was PCR-amplified from pTYB11 (New England Biolabs) and inserted into the p8-TM2 ORF to create the p8CBD plasmid. Other modifications including addition of a polylinker to create the final p8CBDek vector were accomplished using the Phusion site-directed mutagenesis kit (New England Biolabs). The entire sequence of the p8CBDek (rbsH) vector beginning at the *tac* promoter is provided in Supporting Information Figure 1. The backbone of the p8CBDek vector contains the *lacIq* gene, the β-lactamase gene and the *T1T2 rrnb* transcription terminators derived from pJF119HE described in Furste *et al.*[38] The mutation of *rbsO* to *rbsH* resulted in the following nucleotide change: T*GGAAAC*TTCCT-Catg to A*AAGGA*CGGCCGGatg.

PhoA assays

Washed cells (100 μL) were added to 900 μL of assay buffer (1 m*M* Tris-HCl pH 8, 1 m*M* $MgCl_2$, 1 m*M* $ZnCl_2$). After addition of one drop of 1% SDS and two drops of methylene chloride (CH_2Cl_2), cells were vortexed and incubated 5 min at 37°C. The reaction was

initiated by the addition of 0.1 mL of 25 m*M* *p*-nitrophenyl phosphate (pNPP) and stopped by addition of 0.12 mL of Stop Solution (1 volume of 0.5 M EDTA plus 5 volumes of 2.5 *M* K_2HPO_4). The cells were pelleted by centrifugation and the optical density of the supernatant was read at 420 nm and 550 nm.

$$\text{PhoA units} = \frac{\text{OD420 nm} - (1.75 \times \text{OD550 nm})}{\text{OD600 nm} \times \text{volume of sup(mL)} \times \text{r} \times \text{n time (min)}}$$

where OD600 nm is a measure of cell density, OD420 nm is a measure of *p*-nitrophenol absorbance and OD550 nm is a measure of scatter from cell debris.

To create *phoA* genetic fusions, the *E. coli phoA* gene was amplified from MG1655 genomic DNA and digested with MfeI. The compatible ends were ligated into the EcoRI site of either p8CBD or p8CBDek. This resulted in a PhoA fusion protein beginning at mature amino acid #6 (proline) equivalent to the Tn*phoA* fusions first described by Manoil and Beckwith.[39] Wild-type *phoA* and *torTss-phoA* constructs were created so that the 5′untranslated region was exactly the same as the p8CBD-*phoA* fusions. The TorT signal sequence is: MRVLLFLLLSLFMLPAFS and the GES hydrophobicity score is 1.824.

Hydrophobicity calculations

Signal peptide and transmembrane segment hydrophobicity scores were calculated using the Goldman-Engelman-Steitz (GES) scale.[20] This calculation may be performed using the TopPred algorithm[40] available at (http://mobyle.pasteur.fr/cgi-bin/portal.py?form=toppred). We calculated the threshold GES hydrophobicity score for SRP dependence based on the work by Huber *et al.*[23] where the DsbA signal sequence was shown to mediate the SRP-dependent export of thioredoxin. The TopPred algorithm assigned a GES score of 1.521 to the DsbA signal sequence (using a default core window of 11 and a full window size of 21 amino acids). Thus, any signal sequence with a GES score >1.521 is expected to mediate SRP-dependent protein export.

Media and Growth Conditions

Terrific Broth[41] was used for protein expression experiments. Ampicillin was supplemented at 100 ug/mL. Outgrowth temperature was 30°C and membrane protein induction was carried out at 20°C. PhoA assay cultures were grown and induced and 30°C. PhoA assay cultures were prepared using a modified LB media (1% soy tryptone, 0.5% yeast extract, 1% w/v NaCl, 0.1% w/v dextrose, 0.1% magnesium chloride and NaOH was added to pH 7.2).

Reagents

Restriction enzymes and DNA modifying enzymes were supplied by New England Biolabs. Oligonucleotides were synthesized by the NEB organic synthesis department and Integrated DNA Technologies. Monoclonal antibody B62-FE2 was purchased from PROGEN Biotechnik GmbH (Heidelberg, Germany). B62-FE2 recognizes the mature N-terminal region of M13 P8 (A*EGDD*PA*K*A). Anti-FLAG M2 monoclonal antibody, PhoA monoclonal antibody and pNPP were purchased from Sigma. *T. maritima* MSB8 genomic DNA was purchased from ATCC. The yeast *OXA1* gene was amplified from *Saccharomyces cerevisiae* strain BY4734 genomic DNA (provided by Anne-Lise Fabre).

Acknowledgment

The authors thank Chris Taron, Paul Riggs, Mehmet Berkmen, Marion Sibley, Elisabeth Raleigh, and Bryce Causey of New England Biolabs for many valuable discussions. They thank Lixin An for competent cell preparation and Laurie Mazzola and Beth Ann Cantin for DNA sequencing.

References

1. Wagner S, Baars L, Ytterberg AJ, Klussmeier A, Wagner CS, Nord O, Nygren PA, van Wijk KJ, de Gier JW (2007) Consequences of membrane protein overexpression in *Escherichia coli*. Mol Cell Proteomics 6:1527–1550.
2. Tusnady GE, Dosztanyi Z, Simon I (2004) Transmembrane proteins in the Protein Data Bank: identification and classification. Bioinformatics 20: p. 2964–2972.
3. Tusnady GE, Dosztanyi Z, Simon I (2005) PDB_TM: selection and membrane localization of transmembrane proteins in the protein data bank. Nucleic Acids Res 33 (Database issue):D275–D278.
4. Luirink J, Sinning I (2004) SRP-mediated protein targeting: structure and function revisited. Biochim Biophys Acta 1694:17–35.
5. Angelini S, Deitermann S, Koch HG (2005) FtsY, the bacterial signal-recognition particle receptor, interacts functionally and physically with the SecYEG translocon. EMBO Rep 6:476–481.
6. Bahari L, Parlitz R, Eitan A, Stjepanovic G, Bochkareva ES, Sinning I, Bibi E (2007) Membrane targeting of ribosomes and their release require distinct and separable functions of FtsY. J Biol Chem 282:32168–32175.
7. Lee HC, Bernstein HD (2001) The targeting pathway of *Escherichia coli* presecretory and integral membrane proteins is specified by the hydrophobicity of the targeting signal. Proc Natl Acad Sci USA 98:3471–3476.
8. Schierle CF, Berkmen M, Huber D, Kumamoto C, Boyd D, Beckwith J. (2003) The DsbA signal sequence directs efficient, cotranslational export of passenger proteins to the *Escherichia coli* Periplasm via the signal recognition particle pathway. J Bacteriol 185:5706–5713.
9. Bowers CW, Lau F, Silhavy TJ (2003) Secretion of LamB-LacZ by the signal recognition particle pathway of *Escherichia coli*. J Bacteriol 185:5697–5705.
10. Dalbey RE (1990) Positively charged residues are important determinants of membrane protein topology. Trends Biochem Sci 15:253–257.
11. Cao G, Kuhn A, Dalbey RE (1995) The translocation of negatively charged residues across the membrane is driven by the electrochemical potential: evidence for an

electrophoresis-like membrane transfer mechanism. EMBO J 14:866–875.
12. von Heijne G, Gavel Y (1988) Topogenic signals in integral membrane proteins. Eur J Biochem 174:671–678.
13. Geller B, Zhu HY, Cheng S, Kuhn A, Dalbey RE (1993) Charged residues render pro-OmpA potential dependent for initiation of membrane translocation. J Biol Chem 268:9442–9447.
14. Kajava AV, Zolov SN, Kalinin AE, Nesmeyanova MA (2000) The net charge of the first 18 residues of the mature sequence affects protein translocation across the cytoplasmic membrane of gram-negative bacteria. J Bacteriol 182:2163–2169.
15. Kaderbhai MA, Davey HM, Kaderbhai NN (2004) A directed evolution strategy for optimized export of recombinant proteins reveals critical determinants for preprotein discharge. Protein Sci 13:2458–2469.
16. Kiefer D, Hu X, Dalbey R, Kuhn A (1997) Negatively charged amino acid residues play an active role in orienting the sec-independent pf3 coat protein in the escherichia inner membrane. Embo J 16:2197–2204
17. Kuhn A (1988) Alterations in the extracellular domain of M13 Procoat protein make its membrane insertion dependent on SecA and SecY. Eur J Biochem 177:267–271.
18. de Gier JW, Scotti PA, Sääf A, Valent QA, Kuhn A, Luirink J, von Heijne G (1998) Differential use of the signal recognition particle translocase targeting pathway for inner membrane protein assembly in *Escherichia coli*. Proc Natl Acad Sci USA 95:14646–14651.
19. Samuelson JC, Chen M, Jiang F, Moller I, Wiedmann M, Kuhn A, Phillips GJ, Dalbey RE (2000) YidC mediates membrane protein insertion in bacteria. Nature 406: 637–641.
20. Engelman DM, Steitz TA, Goldman A (1986) Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. Annu Rev Biophys Biophys Chem 15:321–353.
21. Engelman DM, Steitz TA (1981) The spontaneous insertion of proteins into and across membranes: the helical hairpin hypothesis. Cell 23:411–422.
22. Tian H, Beckwith J (2002) Genetic screen yields mutations in genes encoding all known components of the *Escherichia coli* signal recognition particle pathway. J Bacteriol 184:111–118.
23. Huber D, Boyd D, Xia Y, Olma MH, Gerstein M, Beckwith J (2005) Use of thioredoxin as a reporter to identify a subset of *Escherichia coli* signal sequences that promote signal recognition particle-dependent translocation. J Bacteriol 187:2983–2991.
24. van der Laan M, Urbanus ML, ten Hagen-Jongman CM, Nouwen N, Oudega B, Harms N, Driessen AJM, Luirink J (2003) A conserved function of YidC in the biogenesis of respiratory chain complexes. Proc Natl Acad Sci USA 100:5801–5806.
25. van Bloois E, Nagamori S, Koningstein G, Ullers RS, Preuss M, Oudega B, Harms N, Kaback R, Herrmann JM, Luirink J (2005) The Sec-independent function of *Escherichia coli* YidC is evolutionary-conserved and essential. J Biol Chem 280:12996–13003.
26. Jiang F, Yi L, Moore M, Chen M, Rohl T, Van Wijk KJ, De Gier JW, Henry R, Dalbey RE (2002) Chloroplast YidC homolog Albino3 can functionally complement the bacterial YidC depletion strain and promote membrane insertion of both bacterial and chloroplast thylakoid proteins. J Biol Chem 277:19281–19288.
27. Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567–580.
28. Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Nelson WC, Ketchum KA, McDonald L, Utterback TR, Malek JA, Linher KD, Garrett MM, Stewart AM, Cotton MD, Pratt MS, Phillips CA, Richardson D, Heidelberg J, Sutton GG, Fleischmann RD, Eisen JA, White O, Salzberg SL, Smith HO, Venter JC, Fraser CM (1999) Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. Nature 399:323–329.
29. Käll L, Krogh A, Sonnhammer EL (2007) Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server Nucleic Acids Res 35: W429–W432.
30. Kermorgant M, Bonnefoy N, Dujardin G (1997) Oxa1p, which is required fro cytochrome c oxidase and ATP synthase complex formation, is embedded in the mitochondrial inner membrane. Curr Genet 31:302–307.
31. Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in *Escherichia coli*. Nature 403: 339–342.
32. Nargang FE, Preuss M, Neupert W, Herrmann JM (2002) The Oxa1 protein forms a homo-oligomeric complex and is an essential part of the mitochondrial export translocase in *Neurospora crassa*. J Biol Chem 277:12846–12853.
33. Kneissel S, Queitsch I, Petersen G, Behrsing O, Micheel B, Dubel S (1999) Epitope structures recognized by antibodies against the major coat protein (g8p) of filamentous bacteriophage fd (Inoviridae). J Mol Biol 288:21–28.
34. Qing G, Xia B, Inouye M (2003) Enhancement of translation initiation by A/T-rich sequences downstream of the initiation codon in *Escherichia coli*. J Mol Microbiol Biotechnol 6:133–144.
35. Casabadan MJ, Cohen SN (1980) Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. J Mol Biol 138:179–207.
36. Hamilton CM, Aldea M, Washburn BK, Babitzke P, Kushner SR (1989) New method for generating deletions and gene replacements in *Escherichia coli*. J Bacteriol 171: 4617–4622.
37. Samuelson JC, Jiang F, Yi L, Chen M, de Gier J-W, Kuhn A, Dalbey RE (2001) Function of YidC for the insertion of M13 Procoat protein in *E. coli*: translocation of mutants that show differences in their membrane potential dependence and Sec requirement. J Biol Chem 276:34847–34852.
38. Furste JP, Pansegrau W, Frank R, Blocker H, Scholz P, Bagdasarian M, Lanka E (1986) Molecular cloning of the plasmid RP4 primase region in a multi-host-range *tacP* expression vector Gene 48:119–131.
39. Manoil C, Beckwith J (1985) TnphoA: A transposon probe for protein export signals Proc Natl Acad Sci USA 82:8129–8133.
40. Claros MG, von Heijne G (1994) TopPred II: an improved software for membrane protein structure predictions. Comput Appl Biosci 10:685–686.
41. Tartof KD, Hobbs CA (1988) New cloning vectors and techniques for easy and rapid restriction mapping. Gene 30:169–182.