Applied and Environmental Microbiology, Sept. 2010, p. 5852–5859 Vol. 76, No. 17
0099-2240/10/$12.00 doi:10.1128/AEM.00963-10
Copyright © 2010, American Society for Microbiology. All Rights Reserved.

Simple Genetic Selection Protocol for Isolation of Overexpressed Genes That Enhance Accumulation of Membrane-Integrated Human G Protein-Coupled Receptors in *Escherichia coli*▽

Georgios Skretas[1,4,5] and George Georgiou[1,2,3,4]*

Department of Chemical Engineering,[1] Department of Biomedical Engineering,[2] Section of Microbiology and Molecular Genetics,[3] and Institute of Cellular and Molecular Biology,[4] The University of Texas at Austin, Austin, Texas 78712, and Institute of Biological Research and Biotechnology, National Hellenic Research Foundation, 48 Vassileos Constantinou Ave., Athens 11635, Greece[5]

Received 20 April 2010/Accepted 5 July 2010

The efficient production of membrane proteins in bacteria remains a major challenge. In this work, we sought to identify overexpressed genes that enhance the yields of recombinant membrane proteins in *Escherichia coli*. We developed a genetic selection system for bacterial membrane protein production, consisting of membrane protein fusions with the enzyme β-lactamase and facile selection of high-production strains on ampicillin-containing media. This system was used to screen the ASKA library, an ordered library of plasmids encoding all the known *E. coli* open reading frames (ORFs), and several clones with the ability to accumulate enhanced amounts of recombinant membrane proteins were selected. Notably, coexpression of *ybaB*, a gene encoding a putative DNA-binding protein of unknown function, was found to enhance the accumulation of a variety of membrane-integrated human G protein-coupled receptors and other integral membrane proteins in *E. coli* by up to 10-fold. The results of this study highlight the power of genetic approaches for identifying factors that impact membrane protein biogenesis and for generating engineered microbial hosts for membrane protein production.

Between 20 and 30% of the genes of every living organism are dedicated to the synthesis of membrane proteins (41). Membrane proteins mediate crucial cellular processes, such as signal transduction, maintenance of structure, energy production, transport of chemicals, etc. (26). From a biotechnological standpoint, membrane proteins are valuable targets for pharmaceutical discovery: it is estimated that approximately 50% of all drugs in use or under development act by modulating the function of this class of proteins (18). Biochemical, biophysical, and structural characterizations of membrane proteins could greatly accelerate the discovery of this type of therapeutics. Such studies, however, are strongly impeded by the difficulty in production and isolation of sufficient amounts of protein for characterization (26, 41).

Apart from a few exceptions, membrane proteins are found in their native cells and tissues in very low abundance and therefore cannot be isolated easily in their native state (26, 41). Their production relies, thus, on recombinant overexpression in mammalian cell cultures, insect cells, or microbial hosts, such as yeasts or bacteria (16). Among the available expression hosts, *Escherichia coli* has proven very useful for the production and biochemical and structural characterization of a number of membrane proteins (41). However, despite this success, the production of membrane proteins in simple microbial hosts remains notoriously problematic, especially for proteins of eukaryotic origin (41).

Membrane protein expression in bacteria usually results in low yields of membrane-integrated and correctly folded polypeptide; instead, the polypeptide either is degraded or accumulates in cytoplasmic inclusion bodies, which are difficult to denature and refold (26). Furthermore, the expression of membrane proteins has profound effects on cell growth and physiology (40), further limiting production yields. An example of the problems associated with membrane protein expression in bacteria is that, out of more than 100 putative membrane proteins from *Mycobacterium tuberculosis* that were studied, less than 10% could accumulate in the membrane fraction, while almost 50% did not even yield sufficient protein for detection by Western blotting (22).

Due to our incomplete understanding of the pathways and physiological processes that govern membrane protein biogenesis, optimization of membrane protein production is still carried out in an empirical fashion. Optimal conditions that maximize membrane protein production are usually identified by testing different basic protein production parameters, such as expression constructs (vectors, promoters) and strains, growth media, incubation times, and temperatures, etc. These techniques, however, usually result in limited success, while they treat the issue of membrane protein expression as a "black box".

Recently, protein engineering approaches were deployed to improve the production of membrane proteins in bacteria. For example, directed protein evolution was used to isolate membrane protein variants that could be produced in markedly

* Corresponding author. Mailing address: Department of Chemical Engineering, University of Texas at Austin, 1 University Station C0400, Austin, TX 78712-0231. Phone: (512) 471-6975. Fax: (512) 471-7963. E-mail: gg@che.utexas.edu.

▽ Published ahead of print on 16 July 2010.

Downloaded from http://aem.asm.org/ on July 24, 2015 by Univ of Texas Austin

enhanced yields (30, 35). However, engineering a protein for expression can have unintended consequences for its structure and/or function.

A promising alternative approach for membrane protein expression is the engineering of the bacterial host itself. In early studies, Miroux and Walker isolated mutants of *E. coli* BL21(DE3) that suppressed the cytotoxicity induced by the expression of membrane proteins from the strong T7 promoter (29). Not surprisingly, the reduced cytotoxicity resulted from mutations that reduced the transcriptional efficiency of the T7 RNA polymerase (42).

In other studies, coexpression of the genes encoding the chaperone/cochaperone DnaK/DnaJ was shown to result in a significant increase in the bacterial production of the magnesium transporter CorA (3), whereas coexpression of the gene encoding the membrane-bound protease FtsH resulted in a dramatic increase in the production of the membrane-incorporated and detergent-soluble form of certain human G protein-coupled receptors (GPCRs) (27). *E. coli* mutants carrying chromosomal lesions or transposon insertions that favorably affect membrane protein accumulation have also been reported (28, 39). These results demonstrate that *E. coli* can potentially be engineered to serve as a specialized membrane protein-producing cell factory, in a fashion similar to what has been achieved for the production of natural products and small molecules (7).

In this work, we employed a simple genetic selection system that links cell growth to the expression of properly folded heterologous membrane proteins. This system is based on fusions of the N terminus of members of the N_{out} class of membrane proteins (in which the N terminus is outside the cytoplasm) to β-lactamase. We used this selection to screen the ASKA library, a complete library of all individual *E. coli* genes expressed from a strong inducible promoter (21), and identified overexpressed genes that enhance membrane protein production. Several such genes were isolated in this manner. Among these, the gene *ybaB*, encoding a small protein of unknown function, was shown to enhance greatly the expression of multiple membrane proteins, including that of human GPCRs.

MATERIALS AND METHODS

Strains and plasmids. The *E. coli* strain MC4100A (MC4100 *ara*$^+$) (34) was utilized for all experiments. The plasmid pBADBla was generated by amplification of the gene encoding the mature sequence of TEM-1 β-lactamase (Bla) (lacking its native Sec-dependent signaling sequence) from the vector pTrc99a (GE Healthcare) with the PCR primers ssDsbABlaSacI (AAAAAA GAGCTCAGGAGGAAACGATGAAAAAGATTTGGCTGGCGCTGGCT GGTTTAGTTTTAGCGTTTAGCGCATCGGCGGCGCAGTATGAAGAT CTGCAGCACCCAGAAACGCTG) and BlaHindIII (GCCTCACTGATTA AGCATTGGTCTAGATAAGCTGCGAAGCTTAAAAAA). These primers append to the 5′ end of the *bla* sequence a SacI restriction site, an optimized Shine-Dalgarno sequence, the signal recognition particle (SRP)-dependent signaling sequence of DsbA, a PstI restriction site, and, on the 3′ end of *bla*, a stop codon flanked by the two restriction sites XbaI and HindIII. This PCR product was digested with SacI and HindIII and ligated into a similarly digested pBAD-33 vector (17). pBADBla-BR2 was constructed by PCR amplification of the human bradykinin receptor 2 (BR2)-encoding gene from pASKBR2, which contains a codon-optimized BR2 gene for expression in *E. coli* (27), and insertion into the XbaI and HindIII sites of pBADBla. pBADBla-YidC was constructed by PCR amplification of *yidC* from the *E. coli* chromosome and insertion into the XbaI and HindIII sites of pBADBla. pBADBla-YegH was constructed by PCR amplification of *yegH* from the *E. coli* chromosome, digestion with XbaI-SphI, and three-way ligation with the SacI-XbaI-digested *bla* gene described above into the SacI and SphI sites of pBAD-33. pBADSmRBla-BR2 was generated by interrupting the chloramphenicol acetyltransferase gene of pBADBla-BR2 at its MscI site with the *aadA* (Sm^r) gene amplified from plasmid pCDF-1b (Novagen). pBADBR2-GFP, pBADCB1-GFP, pBADNKR1-GFP, and pBADBR2 were constructed by subcloning BR2-GFP, human central cannabinoid receptor 1 (CB1)-GFP, human neurokinin (substance P) receptor 1 (NKR1)-GFP, and BR2 from pASKBR2-GFP, pASKCB1-GFP, pASKNKR1-GFP, and pASKBR2 (27), respectively, into pBAD30 (17) by utilizing the available XbaI-HindIII restriction sites. pBADYidC-GFP, pBADCstA-GFP, and pBADSCD-GFP were constructed by PCR amplification of the membrane protein-encoding genes and insertion into the XbaI and PstI sites of pBADBR2-GFP after removal of the BR2-encoding gene.

Conditions of the genetic selection system. For initial testing of the genetic selection system, 5-μl specimens of *E. coli* MC4100A cells transformed with the appropriate pBADBla-membrane protein expression vector were spotted onto Luria-Bertani (LB) agar plates containing 40 μg/ml chloramphenicol and 0.1% L-arabinose, in the absence or presence of different ampicillin concentrations. For the ASKA library selection, cells were doubly transformed by electroporation with pBADSpRBla-BR2 and the ASKA library, plated onto LB agar plates containing 50 μg/ml spectinomycin, 30 μg/ml chloramphenicol, 80 μg/ml ampicillin, and 0.1% L-arabinose, either with or without 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG), and incubated at room temperature for approximately 2 days.

Protein overexpression and isolation of the total membrane fraction. *E. coli* MC4100A cells freshly transformed with the appropriate expression vector were used for all protein production experiments. Single bacterial colonies were used to inoculate liquid LB cultures containing the appropriate combination of antibiotics (100 μg/ml ampicillin, 100 μg/ml spectinomycin, or 40 μg/ml chloramphenicol). These saturated cultures were used with a 1:100 dilution to inoculate fresh LB cultures containing 0.1 mM IPTG, which were grown at 37°C to an optical density at 600 nm (OD_{600}) of 0.5 to 0.7 with shaking. The temperature was then decreased to 25°C, and after a temperature equilibration period of 5 to 20 min, membrane protein expression was induced by the addition of 0.1% L-arabinose for approximately 5 h (unless otherwise specified). Total membrane fractions were prepared from cells harvested from 500 ml LB cultures, resuspended in 10 ml of cold lysis buffer (10 mM imidazole, 300 mM NaCl, 50 mM NaH_2PO_4, 15% glycerol, 5 mM dithiothreitol), and lysed by double passage through a French press. Cell lysates were then centrifuged twice at 8,000 rpm for 20 min using a Beckman-Coulter JA-10 rotor, and the supernatant was collected (soluble fraction) and subjected to ultracentrifugation in a Beckman 50 Ti rotor at 50,000 rpm for 1.5 h at 4°C. Total membranes corresponding to the pellet of the ultracentrifugation step were resuspended in 10 ml of cold lysis buffer and homogenized.

Flow cytometry. The fluorescence of cells expressing membrane protein-green fluorescent protein (GFP) fusions (530/30 nm) was monitored using a Becton-Dickinson FACSAria flow cytometer and analyzed with FACSDiva software. Cells were gated based on size on a side-scatter–versus–forward-scatter plot. Approximately 10^4 cells were analyzed per sample.

In-gel fluorescence and Western blot analysis. In-gel fluorescence and Western blot analysis were performed as described previously (39).

RESULTS

Development of a genetic selection system for membrane protein production in *E. coli*. In order to isolate *E. coli* variants with the ability to produce increased amounts of recombinant membrane proteins rapidly, we utilized fusions of proteins to β-lactamase, a bacterial enzyme which must be exported to the periplasm in order to confer significant resistance to β-lactam antibiotics, such as ampicillin. We constructed chimeric genes that encode fusions of the TEM-1 β-lactamase (Bla) with the N termini of different integral membrane proteins exhibiting various topologies: the *E. coli* membrane integrase YidC (six transmembrane helices, N_{in}-C_{in} topology [i.e., both the N and C termini are localized in the cytoplasm]), the *E. coli* putative transporter YegH (seven transmembrane helices, N_{out}-C_{in} topology), and the human G protein-coupled receptor (GPCR) bradykinin receptor 2 (BR2) (seven transmembrane helices,

Downloaded from http://aem.asm.org/ on July 24, 2015 by Univ of Texas Austin

TABLE 1. Membrane proteins studied in this work

Membrane protein	Organism	Function	No. of transmembrane helices	Topology	Molecular mass (kDa)
YidC	*E. coli*	Membrane protein integrase	6	N_{in}-C_{in}	61.5
YegH	*E. coli*	Putative transporter	7	N_{out}-C_{in}	59.5
CstA	*E. coli*	Peptide transporter	18	N_{in}-C_{in}	75.1
BR2	*Homo sapiens*	Bradykinin receptor 2 (GPCR)	7	N_{out}-C_{in}	44.5
CB1	*H. sapiens*	Central cannabinoid receptor (GPCR)	7	N_{out}-C_{in}	52.9
NKR1	*H. sapiens*	Neurokinin (substance P) receptor 1 (GPCR)	7	N_{out}-C_{in}	46.3
SCD	*H. sapiens*	Stearoyl-CoA desaturase	4	N_{in}-C_{in}	41.1

N_{out}-C_{in} topology) (Table 1). Since most native as well as the majority of the heterologously produced membrane proteins in *E. coli* are targeted to the inner membrane via the signal recognition particle (SRP) pathway (32), we replaced the SRP-independent signaling sequence of β-lactamase (1) with that of the SRP-dependent exported protein DsbA (ssDsbA) (Fig. 1A, top; Table 2). ssDsbA-tagged proteins are exported to the bacterial periplasm through the SRP pathway (37). Bacterial cells producing membrane proteins with their N termini facing the periplasm and fused to β-lactamase are expected to display active Bla in the periplasm and grow on ampicillin-containing media (Bla^+ phenotype). Improper folding of the fusion within the cytoplasm should not be able to confer antibiotic resistance (Bla^- phenotype).

We tested the growth phenotypes of *E. coli* MC4100A cells (34) expressing ssDsbA-Bla (SRP-exported Bla; positive control) or the membrane protein fusions ssDsbA-Bla-YidC, ssDsbA-Bla-YegH, and ssDsbA-Bla-BR2 on LB agar plates containing 0.1% L-arabinose to induce protein synthesis from the *araBAD* promoter, as well as increasing concentrations of ampicillin. We observed that only cells expressing Bla fusions to membrane proteins with N_{out} topologies could grow on ampicillin-containing media (Fig. 1B). Expression of the gene encoding the ssDsbA-Bla fusion with YidC, a protein with a cytoplasmic N terminus, did not permit bacterial growth even at ampicillin concentrations as low as 40 μg/ml (Fig. 1B). These results are consistent with previous observations that Bla is a reliable reporter of membrane protein topology (2). Expression of the gene encoding exported Bla or the gene for the Bla fusion with YegH, a well-accumulated bacterial membrane protein (Fig. 1C), resulted in bacterial growth under all ampicillin concentrations tested, whereas expression of the

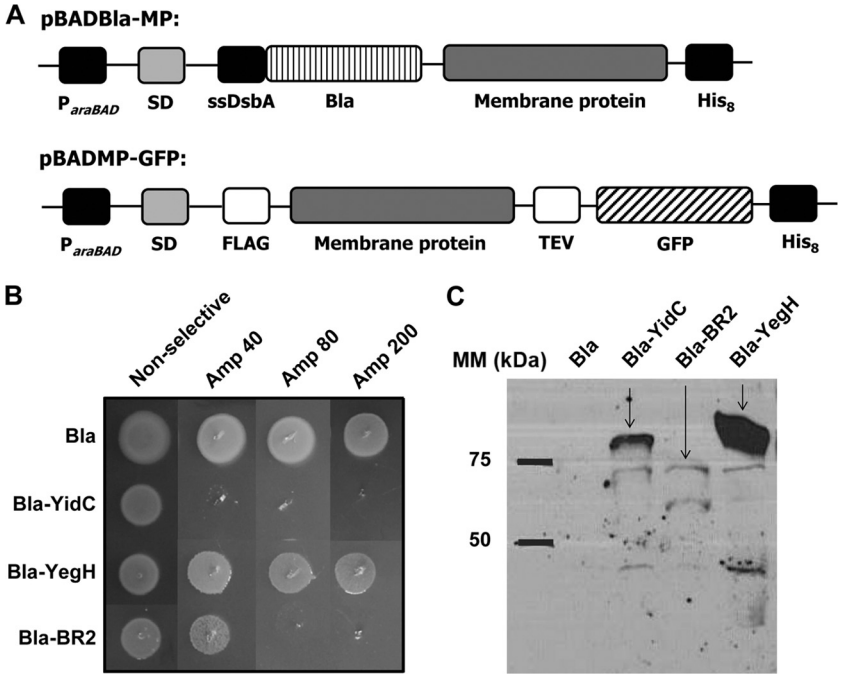

FIG. 1. (A) Constructs used in this study. (Top) The pBADBla-MP family of expression vectors includes a membrane protein (MP)-encoding gene flanked by the SRP-recognized DsbA signaling sequence (ssDsbA) and a C-terminal octahistidine tag under the control of the *araBAD* promoter, with an optimized Shine-Dalgarno (SD) sequence. (Bottom) The pBADMP-GFP plasmid family encodes membrane proteins with C-terminally fused GFP and a tobacco etch virus (TEV) cleavage site between the MP and GFP (no signaling sequence added). (B) Growth phenotypes of *E. coli* cells expressing ssDsbA-Bla (Bla) or ssDsbA-Bla-membrane protein fusions (Bla-YidC, Bla-YegH, and Bla-BR2) on LB agar media containing 0.1% L-arabinose with and without increasing concentrations of ampicillin (Amp; followed by a number indicating an amount in μg/ml). (C) Western blots of isolated total membrane fractions of MC4100A cells expressing ssDsbA-Bla, ssDsbA-Bla-YidC, ssDsbA-Bla-YegH, or ssDsbA-Bla-BR2, probed with an anti-Bla antibody. Extracts with equal numbers of cells were loaded in the lanes. MM, molecular mass.

Downloaded from http://aem.asm.org/ on July 24, 2015 by Univ of Texas Austin

TABLE 2. Plasmids used in this work

Plasmid	Protein expressed[a]	Marker	Origin of replication	Source
pBADBla-YidC	ssDsbA-Bla-YidC	Cm^r	ACYC	This work
pBADBla-YegH	ssDsbA-Bla-YegH	Cm^r	ACYC	This work
pBADBla-BR2	ssDsbA-Bla-BR2	Cm^r	ACYC	This work
pBADSmRBla-BR2	ssDsbA-Bla-BR2-His_8	Sm^r	ACYC	This work
pBADBR2-GFP	FLAG-BR2-TEV-GFP-His_8	Amp^r	ACYC	This work
pBADCB1-GFP	FLAG-CB1-TEV-GFP-His_8	Amp^r	ACYC	This work
pBADNKR1-GFP	FLAG-NKR1-TEV-GFP-His_8	Amp^r	ACYC	This work
pBADYidC-GFP	YidC-TEV-GFP-His_8	Amp^r	ACYC	This work
pBADCstA-GFP	FLAG-CstA-TEV-GFP-His_8	Amp^r	ACYC	This work
pBADSCD-GFP	FLAG-SCD-TEV-GFP-His_8	Amp^r	ACYC	This work
pBADBR2	FLAG-BR2-His_8	Amp^r	ACYC	This work
ASKA library	All known *E. coli* proteins	Cm^r	ColE1	Kitagawa et al. (21)

[a] TEV, tobacco etch virus.

gene encoding the Bla fusion with BR2, a very poorly accumulating human membrane protein (Fig. 1C), allowed bacterial growth only at low ampicillin concentrations (Fig. 1B). These results demonstrate that the ssDsbA-Bla fusions can serve as a reliable reporter for the accumulation of properly localized membrane proteins.

Selection for single *E. coli* genes that enhance bacterial membrane protein production. Having constructed a reliable and rapid selection system for membrane protein production, we elected to isolate genes that upon overexpression would enhance the yields of inner membrane protein production in *E. coli*. For this, we utilized the ASKA library, an ordered library of plasmids encoding all known *E. coli* open reading frames (ORFs) under the control of the T5*lac* promoter (21). As our model membrane protein, we chose to study the human GPCR BR2, a membrane protein whose expression in *E. coli* results in minute amounts of membrane-embedded polypeptide (Fig. 1C) (27). GPCRs constitute the most important class of drug targets in pharmaceutical discovery (19), and they are notoriously difficult to express, typically yielding very small amounts of membrane-embedded protein (36).

E. coli MC4100A cells were transformed with the expression vector pBADSmRBla-BR2 (Table 2) and the ASKA library. Approximately 10^6 cells were plated on agar plates containing 0.1% L-arabinose, 50 μg/ml spectinomycin, 30 μg/ml chloramphenicol, and 80 μg/ml ampicillin in the presence and absence of 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG). A number of Bla^+ colonies appeared on plates both with and without IPTG after 2 days of growth at room temperature. Since ampicillin resistance can arise from a number of reasons not necessarily related to enhanced Bla-BR2 production, such as altered outer membrane composition or overexpression of a gene encoding bacterial protein with β-lactam hydrolytic activity (e.g., the chromosomal *ampC* gene), it was necessary to use a secondary screen to further evaluate the phenotypes of the isolated clones.

Drew et al. have shown that the fluorescence of bacterial cells expressing GFP fusions to the C termini of C_{in} membrane proteins correlates well with the amount of membrane-integrated and properly folded protein expressed (10). ASKA plasmids isolated from colonies that grew on ampicillin plates were transformed into MC4100A cells also containing the vector pBADBR2-GFP, in which the gene encoding BR2 is fused to the 3′ end of *gfpmut2* (5) and expressed under the control of the *araBAD* promoter (Fig. 1A, bottom, and Table 2). Cell fluorescence was analyzed by flow cytometry and compared to the fluorescence of MC4100A cells overexpressing the gene encoding BR2-GFP alone. Most of the clones carrying the plasmids isolated based on ampicillin resistance exhibited reproducibly higher BR2-GFP fluorescence. Three plasmids were found to mediate 3-fold or higher BR2-GFP fluorescence relative to that of controls not containing an ASKA plasmid. Sequencing led to the identification of the genes *ybaB*, which encodes a soluble protein of unknown function, *yciQ*, which encodes a putative inner membrane protein also of unknown function, and *glpQ*, which encodes the periplasmic glycerophosphoryl diester phosphodiesterase enzyme GlpQ (Fig. 2). Cells overexpressing *ybaB* exhibited BR2-GFP fluorescence that was an order of magnitude higher than MC4100A's (Fig. 2), and therefore this gene was selected for further study.

***ybaB* coexpression dramatically increases the production of membrane-bound and properly folded BR2 in *E. coli*.** The level of accumulation of BR2-GFP in the membrane fractions of cells coexpressing *ybaB* was analyzed by Western blotting and compared to BR2-GFP production in wild-type MC4100A. While wild-type *E. coli* cells accumulated miniscule quantities of membrane-incorporated BR2-GFP, which were barely detectable by Western blotting (Fig. 3A), coexpression of *ybaB* resulted in a dramatic increase in the accumulation of the BR2-GFP fusion (Fig. 3A).

It has recently been shown that correctly folded membrane protein-GFP fusions retain their fluorescence following sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) of nonboiled protein samples and that the exhibited in-gel fluorescence is a quantitative measure of the amount of properly folded membrane protein (14). Protein samples derived from the total membrane fraction of MC4100A cells overexpressing the gene encoding BR2-GFP without and with *ybaB* coexpression were separated by SDS-PAGE and analyzed for in-gel fluorescence. This analysis showed that *ybaB* coexpression results in the production of a highly fluorescent BR2-GFP band, despite the fact that wild-type bacterial cells cannot produce any BR2-GFP fluorescence (Fig. 3B). These results indicate that *ybaB* coexpression leads to a large enhancement of the accumulation of the membrane-integrated and well-folded BR2 receptor that can be produced in bacteria.

Downloaded from http://aem.asm.org/ on July 24, 2015 by Univ of Texas Austin

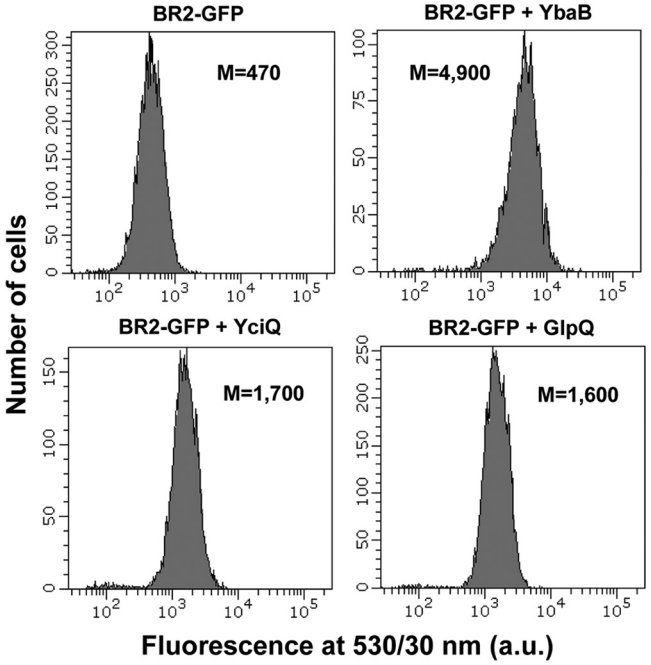

FIG. 2. Comparison of the BR2-GFP fluorescence of wild-type MC4100A cells with the BR2-GFP fluorescence of cells overexpressing *ybaB*, *yciQ*, or *glpQ*. BR2-GFP was expressed at room temperature for approximately 5 h. Fluorescence histograms correspond to a population of 10,000 cells. M, arithmetic mean; a.u., arbitrary units.

Similar results were also obtained for the other two identified genes, *glpQ* and *yciQ*, although their enhancing effect on BR2 yield were less pronounced than those observed for *ybaB* (Fig. 3A and B and data not shown).

ybaB coexpression resulted in a dramatic enhancement in the production of membrane-associated receptor also when BR2 was expressed alone, i.e., without the presence of the GFP fusion partner (Fig. 3C).

Coexpression of *ybaB* broadly confers increased production of prokaryotic and eukaryotic membrane proteins in *E. coli*. The effect of *ybaB* coexpression on the production of five additional integral membrane proteins of bacterial or human origin in *E. coli* was examined. Human central cannabinoid receptor 1 (CB1), human neurokinin (substance P) receptor 1 (NKR1), the human stearoyl-coenzyme A (CoA) desaturase (SCD), the *E. coli* membrane integrase YidC, and the *E. coli* putative peptide transporter CstA (Table 1) were produced as C-terminally fused to GFP (Fig. 1B, bottom; Table 2) in *E. coli* MC4100A. Membrane protein production with and without *ybaB* coexpression was monitored by flow cytometry. *ybaB* conferred increased GFP fluorescence to all the membrane proteins tested (Fig. 4). Interestingly, the observed increases in fluorescence were much more pronounced in the cases of the GPCRs (900% increase for BR2, 750% for CB1, and 430% for NKR1) than in the membrane proteins not belonging to the GPCR superfamily (30% increase for the *E. coli* YidC, 70% for *E. coli* CstA, and 190% for the human SCD). Thus, *ybaB* appears to be a general enhancer of membrane protein production in *E. coli*, which is particularly effective for GPCRs.

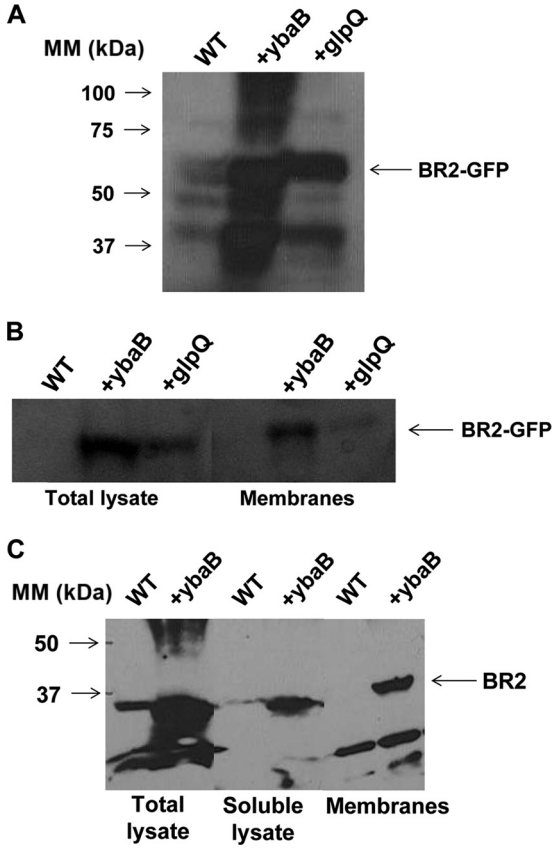

FIG. 3. (A) Comparison of levels of production of membrane-integrated BR2-GFP fusion in MC4100A cells, with and without coexpression of *ybaB* or *glpQ*, by Western blotting with anti-GFP antibody. (B) Comparison of the levels of in-gel fluorescence of the BR2-GFP fusion in the total lysate and total membrane fraction of MC4100A cells, with and without coexpression of *ybaB* or *glpQ*. (C) Comparison of the levels of production of the GFP-free BR2 receptor in different fractions of wild-type MC4100A cells and in MC4100A cells overexpressing *ybaB* by Western blotting with an anti-FLAG antibody. The total lysate comprises the soluble fraction (which under the conditions used here includes the membrane fraction) and the insoluble fraction (comprising cell debris and insoluble protein inclusion bodies [data not shown]). (A and C) The observed lower-molecular-mass bands likely correspond to protein degradation products, presumably due to the low proteolytic stability of BR2-GFP and BR2. Extensive degradation is typical for membrane-embedded GPCRs in *E. coli* (27). The full-length receptor, however, appears to be the major protein product in both panels. In all panels, BR2 and BR2-GFP were expressed at 25°C for approximately 5 h. WT, wild type.

Based on the fluorescence exhibited by CB1-GFP and BR2-GFP upon coexpression of *ybaB* and comparison of this fluorescence with that demonstrated by previous studies where these receptors were detergent solubilized and purified by affinity chromatography (27), we estimate that MC4100A cells overexpressing CB1 and BR2 under the herein-described conditions can produce 100 μg and 1 mg of detergent-soluble CB1 and BR2 receptor, respectively, per liter of shake flask culture.

DISCUSSION

In this work, we sought to identify overexpressed genes that confer enhanced recombinant membrane protein production

Downloaded from http://aem.asm.org/ on July 24, 2015 by Univ of Texas Austin

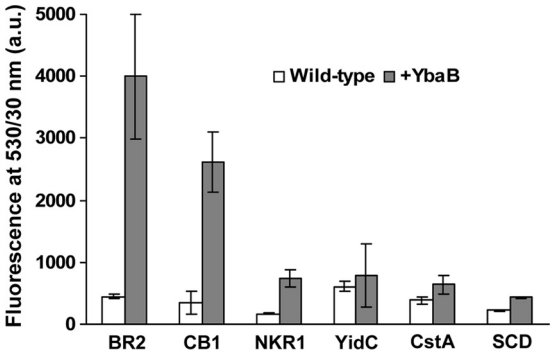

FIG. 4. Comparison of the levels of fluorescence of different membrane protein-GFP fusions expressed in MC4100A cells without (wild type) or with overexpression of *ybaB* (+YbaB). The reported values correspond to the mean values from four replica experiments. Membrane protein-GFP fusions were expressed at room temperature for approximately 5 h. a.u., arbitrary units.

in *E. coli*. First, we developed a bacterial genetic selection system for correct membrane protein insertion by utilizing fusions with the enzyme β-lactamase and testing bacterial growth on ampicillin-containing media. Screening of the ASKA library led to the identification of three genes whose coexpression resulted in a marked increase in the production of the human GPCR BR2: *ybaB*, *glpQ*, and *yciQ*. Among these, *ybaB* had the most pronounced effect and was found to have a broad effect on enhancing membrane protein accumulation, especially that of human GPCRs.

YbaB is a small, 12-kDa protein whose homologues are widespread among bacteria (4). The function of YbaB is unknown. A *ybaB* homologue appears to be essential for *Mycobacterium genitalium*, a bacterial species with one of the smallest known genomes (12). However, *ybaB* does not appear to be essential in gammaproteobacteria. In a number of bacteria, including *E. coli*, *Haemophilus influenzae*, and *Bacillus subtilis*, *ybaB* is cotranscribed with *recR*, a gene involved in the recovery of replication after DNA damage (4). For this reason, YbaB has been thought to be involved in DNA repair. Recent observations, however, such as the fact that *recR* alone is able to complement the recombination defects of a double *ybaB recR* deletion mutant of *Streptomyces coelicolor* (31), suggest otherwise (4). The crystal structures of YbaB from *E. coli* and from *H. influenzae* have been solved, revealing a novel protein fold where a YbaB homodimer forms an α-helical tweezers-like structure (25). This structure appears not to correspond to an enzyme or to a DNA-binding protein and was proposed to play a regulatory role by mimicking DNA and competing for binding to DNA-binding proteins (25). Very recently, however, it was found that *E. coli* and *H. influenzae* YbaBs exhibit specific DNA binding activity (4).

An important characteristic of *ybaB* is that it is one of the *E. coli* genes whose expression is regulated by the extracytoplasmic sigma factor σ^E (33). Apart from being involved in periplasmic and outer membrane stress responses, σ^E has been found to be activated in response to the production of misfolded inner membrane proteins (38). It is likely that certain σ^E-regulated genes are involved in the folding/assembly/quality control pathways of inner membrane proteins in bacteria and that *ybaB* may have such a function.

In addition to *ybaB*, we identified two enhancers of membrane protein expression, *yciQ* and *glpQ*. *yciQ* encodes a putative inner membrane protein of unknown function with a predicted N_{out}-C_{in} topology and five transmembrane domains (http://www.cbs.dtu.dk/services/TMHMM-2.0/). It is interesting that the overexpression of such a large inner membrane protein can assist the production of a second membrane protein, considering the constraint imposed by the very limited capacity and the molecular crowding of the bacterial membrane. *glpQ* encodes the periplasmic glycerophosphoryl diester phosphodiesterase enzyme GlpQ, which hydrolyzes glycerophosphodiesters into an alcohol and *sn*-glycerol-3-phosphate (24). *sn*-Glycerol-3-phosphate can subsequently be transported into the cytoplasmic space, where it can be utilized for lipid biosynthesis, such as the synthesis of cardiolipin. Therefore, one possibility why BR2 production is increased upon *glpQ* coexpression is that the membrane lipid composition is altered in a manner that favors the folding of BR2 and, possibly, of other mammalian GPCRs as well.

The availability of a genetic selection using Bla fusions can provide several advantages over the currently available screening/selection systems that rely on C-terminal fusions to GFP or cytoplasmic antibiotic resistance markers (8, 10, 27, 28, 39). Compared to screens with GFP fusions, genetic selections such as the one reported here allow for a much higher number of clones to be interrogated, and they do not require expensive instrumentation, such as a fluorescence-activated cell sorting (FACS) device for library screening. Furthermore, it has been observed that cytoplasmic reporter proteins, such as fluorescent proteins and enzymes, can sometimes retain their activity even when they aggregate and form insoluble inclusion bodies (13). This effect has also been observed when certain cytoplasmic reporters are expressed as part of fusions with membrane proteins (our unpublished results). Thus, a selection system utilizing a translocated reporter, such as Bla, may be less prone to false positives than selections/screens that use cytoplasmic reporter proteins.

The orientations of the N and C termini of integral membrane proteins can be determined experimentally using membrane protein fusions with cytoplasmic or extracytoplasmic reporter proteins (9, 11). Furthermore, the number of transmembrane helices can be predicted with a high degree of accuracy using bioinformatic tools such as TMHMM (23). By combining these methods and applying them at a genome-wide scale, Daley and coworkers have determined the topology of the entire membrane proteomes of *E. coli* (6) and the yeast *Saccharomyces cerevisiae* (20). These studies indicated that (i) out the 541 *E. coli* membrane proteins with more than two transmembrane helices for which the localization of the C-terminal tail could be experimentally determined, 146 (27%) were found to have their N terminus exposed to the periplasm (6), a topology which is compatible with the use of the β-lactamase screen for expression; (ii) out of these 395 inner membrane proteins with cytoplasmically oriented N termini, 71 (13% of the total 541) have their C termini facing the periplasmic space; and (iii) the experimentally determined topology of the yeast membrane proteome (20) as well as the predicted membrane protein topologies of other genomes based on se-

Downloaded from http://aem.asm.org/ on July 24, 2015 by Univ of Texas Austin

quence homology (15) has revealed that the distribution of the extracytoplasmically localized termini of membrane proteins is very similar to that found in *E. coli* (6). Thus, a large fraction of the known or predicted membrane proteins may be compatible with the herein-described approach.

We have estimated that the yields of isolated CB1 and BR2 receptors in detergent-solubilized form under the conditions described here are approximately 100 μg and 1 mg per liter of shake flask culture, respectively. However, for the vast majority of bacterially expressed mammalian GPCRs, the fraction of the total protein corresponding to active receptor is typically very low (36). Since yields of at least 1 mg of isolated protein per 5 liters of culture in its native functional state are required for a particular host to become a suitable expression vehicle for the production of membrane proteins for structural studies (36), wild-type or engineered *E. coli* strains are not yet suitable to serve as expression hosts for most mammalian GPCRs. Establishing systems with the ability to produce greatly enhanced amounts of membrane-integrated receptors, such as the one described here, however, will assist further studies focused on increasing the ligand-binding activities of bacterially produced GPCRs. It must be noted that for eukaryotic membrane proteins, which are not as problematic as GPCRs and which can adopt their native conformation and exhibit activity when produced in *E. coli*, the yields reported here are sufficient to provide adequate amounts of protein for biophysical studies.

Bacterial strain engineering appears to be a very promising approach for generating specialized membrane protein-producing *E. coli* strains. Random or targeted modifications of the genetic background of the expression host coupled with reporter assays or genetic selection systems, such as the one described here, will likely yield efficient cell hosts for the production, isolation, and characterization of membrane proteins of outstanding pharmacological interest. Importantly, these approaches can also provide useful and unanticipated insights into how simple microbial hosts can be optimized for the production of complex recombinant proteins.

ACKNOWLEDGMENTS

This work was supported by the Welch Foundation and by National Institutes of Health grant NIH GM 55090.

We thank Dan Tawfik for providing the combined ASKA library and John F. Hunt for bringing to our attention the model membrane proteins CstA, YegH, and SCD and for providing plasmids with the coding sequence of SCD.

REFERENCES

1. **Beha, D., S. Deitermann, M. Muller, and H. G. Koch.** 2003. Export of beta-lactamase is independent of the signal recognition particle. J. Biol. Chem. **278:**22161–22167.
2. **Broome-Smith, J. K., M. Tadayyon, and Y. Zhang.** 1990. Beta-lactamase as a probe of membrane protein assembly and protein export. Mol. Microbiol. **4:**1637–1644.
3. **Chen, Y., J. Song, S. F. Sui, and D. N. Wang.** 2003. DnaK and DnaJ facilitated the folding process and reduced inclusion body formation of magnesium transporter CorA overexpressed in Escherichia coli. Protein Expr. Purif. **32:**221–231.
4. **Cooley, A. E., S. P. Riley, K. Kral, M. C. Miller, E. DeMoll, M. G. Fried, and B. Stevenson.** 2009. DNA-binding by Haemophilus influenzae and Escherichia coli YbaB, members of a widely-distributed bacterial protein family. BMC Microbiol. **9:**137.
5. **Cormack, B. P., R. H. Valdivia, and S. Falkow.** 1996. FACS-optimized mutants of the green fluorescent protein (GFP). Gene **173:**33–38.
6. **Daley, D. O., M. Rapp, E. Granseth, K. Melen, D. Drew, and G. von Heijne.** 2005. Global topology analysis of the Escherichia coli inner membrane proteome. Science **308:**1321–1323.
7. **Demain, A. L., and J. L. Adrio.** 2008. Contributions of microorganisms to industrial biology. Mol. Biotechnol. **38:**41–55.
8. **Drew, D., M. Lerch, E. Kunji, D. J. Slotboom, and J. W. de Gier.** 2006. Optimization of membrane protein overexpression and purification using GFP fusions. Nat. Methods **3:**303–313.
9. **Drew, D., D. Sjostrand, J. Nilsson, T. Urbig, C. N. Chin, J. W. de Gier, and G. von Heijne.** 2002. Rapid topology mapping of Escherichia coli inner-membrane proteins by prediction and PhoA/GFP fusion analysis. Proc. Natl. Acad. Sci. U. S. A. **99:**2690–2695.
10. **Drew, D. E., G. von Heijne, P. Nordlund, and J. W. de Gier.** 2001. Green fluorescent protein as an indicator to monitor membrane protein overexpression in Escherichia coli. FEBS Lett. **507:**220–224.
11. **Edelman, A., L. Bowler, J. K. Broome-Smith, and B. G. Spratt.** 1987. Use of a beta-lactamase fusion vector to investigate the organization of penicillin-binding protein 1B in the cytoplasmic membrane of Escherichia coli. Mol. Microbiol. **1:**101–106.
12. **Freiberg, C., B. Wieland, F. Spaltmann, K. Ehlert, H. Brotz, and H. Labischinski.** 2001. Identification of novel essential Escherichia coli genes conserved among pathogenic bacteria. J. Mol. Microbiol. Biotechnol. **3:**483–489.
13. **Garcia-Fruitos, E., N. Gonzalez-Montalban, M. Morell, A. Vera, R. M. Ferraz, A. Aris, S. Ventura, and A. Villaverde.** 2005. Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. Microb. Cell Fact. **4:**27.
14. **Geertsma, E. R., M. Groeneveld, D. J. Slotboom, and B. Poolman.** 2008. Quality control of overexpressed membrane proteins. Proc. Natl. Acad. Sci. U. S. A. **105:**5722–5727.
15. **Granseth, E., D. O. Daley, M. Rapp, K. Melen, and G. von Heijne.** 2005. Experimentally constrained topology models for 51,208 bacterial inner membrane proteins. J. Mol. Biol. **352:**489–494.
16. **Grisshammer, R., and C. G. Tate.** 1995. Overexpression of integral membrane proteins for structural studies. Q. Rev. Biophys. **28:**315–422.
17. **Guzman, L. M., D. Belin, M. J. Carson, and J. Beckwith.** 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J. Bacteriol. **177:**4121–4130.
18. **Hopkins, A. L., and C. R. Groom.** 2002. The druggable genome. Nat. Rev. Drug Discov. **1:**727–730.
19. **Jacoby, E., R. Bouhelal, M. Gerspacher, and K. Seuwen.** 2006. The 7 TM G-protein-coupled receptor target family. ChemMedChem **1:**761–782.
20. **Kim, H., K. Melen, M. Osterberg, and G. von Heijne.** 2006. A global topology map of the Saccharomyces cerevisiae membrane proteome. Proc. Natl. Acad. Sci. U. S. A. **103:**11142–11147.
21. **Kitagawa, M., T. Ara, M. Arifuzzaman, T. Ioka-Nakamichi, E. Inamoto, H. Toyonaga, and H. Mori.** 2005. Complete set of ORF clones of Escherichia coli ASKA library (a complete set of E. coli K-12 ORF archive): unique resources for biological research. DNA Res. **12:**291–299.
22. **Korepanova, A., F. P. Gao, Y. Hua, H. Qin, R. K. Nakamoto, and T. A. Cross.** 2005. Cloning and expression of multiple integral membrane proteins from Mycobacterium tuberculosis in Escherichia coli. Protein Sci. **14:**148–158.
23. **Krogh, A., B. Larsson, G. von Heijne, and E. L. Sonnhammer.** 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. **305:**567–580.
24. **Larson, T. J., M. Ehrmann, and W. Boos.** 1983. Periplasmic glycerophosphodiester phosphodiesterase of Escherichia coli, a new enzyme of the glp regulon. J. Biol. Chem. **258:**5428–5432.
25. **Lim, K., A. Tempczyk, J. F. Parsons, N. Bonander, J. Toedt, Z. Kelman, A. Howard, E. Eisenstein, and O. Herzberg.** 2003. Crystal structure of YbaB from Haemophilus influenzae (HI0442), a protein of unknown function coexpressed with the recombinational DNA repair protein RecR. Proteins **50:**375–379.
26. **Link, A. J., and G. Georgiou.** 2007. Advances and challenges in membrane protein expression. AIChE J. **53:**752–756.
27. **Link, A. J., G. Skretas, E. M. Strauch, N. S. Chari, and G. Georgiou.** 2008. Efficient production of membrane-integrated and detergent-soluble G protein-coupled receptors in Escherichia coli. Protein. Sci. **17:**1857–1863.
28. **Massey-Gendel, E., A. Zhao, G. Boulting, H. Y. Kim, M. A. Balamotis, L. M. Seligman, R. K. Nakamoto, and J. U. Bowie.** 2009. Genetic selection system for improving recombinant membrane protein expression in E. coli. Protein Sci. **18:**372–383.
29. **Miroux, B., and J. E. Walker.** 1996. Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. J. Mol. Biol. **260:**289–298.
30. **Molina, D. M., T. Cornvik, S. Eshaghi, J. Z. Haeggstrom, P. Nordlund, and M. I. Sabet.** 2008. Engineering membrane protein overproduction in Escherichia coli. Protein Sci. **17:**673–680.
31. **Pelaez, A. I., R. M. Ribas-Aparicio, A. Gomez, and M. R. Rodicio.** 2001. Structural and functional characterization of the recR gene of Streptomyces. Mol. Genet. Genomics **265:**663–672.
32. **Raine, A., R. Ullers, M. Pavlov, J. Luirink, J. E. Wikberg, and M. Ehrenberg.**

Downloaded from http://aem.asm.org/ on July 24, 2015 by Univ of Texas Austin

2003. Targeting and insertion of heterologous membrane proteins in E. coli. Biochimie **85:**659–668.
33. **Rezuchova, B., H. Miticka, D. Homerova, M. Roberts, and J. Kormanec.** 2003. New members of the Escherichia coli sigmaE regulon identified by a two-plasmid system. FEMS Microbiol. Lett. **225:**1–7.
34. **Santini, C. L., A. Bernadac, M. Zhang, A. Chanal, B. Ize, C. Blanco, and L. F. Wu.** 2001. Translocation of jellyfish green fluorescent protein via the Tat system of Escherichia coli and change of its periplasmic localization in response to osmotic up-shock. J. Biol. Chem. **276:**8159–8164.
35. **Sarkar, C. A., I. Dodevski, M. Kenig, S. Dudli, A. Mohr, E. Hermans, and A. Pluckthun.** 2008. Directed evolution of a G-protein coupled receptor for expression, stability, and binding selectivity. Proc. Natl. Acad. Sci. U. S. A. **105:**14808–14813.
36. **Sarramegna, V., F. Talmont, P. Demange, and A. Milon.** 2003. Heterologous expression of G-protein-coupled receptors: comparison of expression systems from the standpoint of large-scale production and purification. Cell. Mol. Life Sci. **60:**1529–1546.
37. **Schierle, C. F., M. Berkmen, D. Huber, C. Kumamoto, D. Boyd, and J. Beckwith.** 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.
38. **Shimohata, N., S. Nagamori, Y. Akiyama, H. R. Kaback, and K. Ito.** 2007. SecY alterations that impair membrane protein folding and generate a membrane stress. J. Cell Biol. **176:**307–317.
39. **Skretas, G., and G. Georgiou.** 2009. Genetic analysis of G protein-coupled receptor expression in Escherichia coli: inhibitory role of DnaJ on the membrane integration of the human central cannabinoid receptor. Biotechnol. Bioeng. **102:**357–367.
40. **Wagner, S., L. Baars, A. J. Ytterberg, A. Klussmeier, C. S. Wagner, O. Nord, P. A. Nygren, K. J. van Wijk, and J. W. de Gier.** 2007. Consequences of membrane protein overexpression in Escherichia coli. Mol. Cell Proteomics **6:**1527–1550.
41. **Wagner, S., M. L. Bader, D. Drew, and J. W. de Gier.** 2006. Rationalizing membrane protein overexpression. Trends Biotechnol. **24:**364–371.
42. **Wagner, S., M. M. Klepsch, S. Schlegel, A. Appel, R. Draheim, M. Tarry, M. Hogbom, K. J. van Wijk, D. J. Slotboom, J. O. Persson, and J. W. de Gier.** 2008. Tuning Escherichia coli for membrane protein overexpression. Proc. Natl. Acad. Sci. U. S. A. **105:**14371–14376.

Downloaded from http://aem.asm.org/ on July 24, 2015 by Univ of Texas Austin