Perspective

Advances and Challenges in Membrane Protein Expression

A. James Link
Dept. of Chemical Engineering, and the Institute for Cellular and Molecular Biology, University of Texas, Austin, TX 78712

George Georgiou
Depts. of Chemical Engineering, Biomedical Engineering, Microbiology and Molecular Genetics, and the Institute for Cellular and Molecular Biology, University of Texas, Austin, TX 78712

DOI 10.1002/aic.11107
Published online February 12, 2007 in Wiley InterScience (www.interscience.wiley.com).

Keywords: structural biology, heterologous expression, G-protein coupled receptors (GPCRs)

Introduction

Cell membranes play a critical role in a plethora of functions essential for life, including, energy generation, nutrient transport, signaling, the maintenance of a stable environment (homeostasis), processing or synthesis of certain biomolecules, and the creation of compartments where particular biological processes take place. Far from being a simple mechanical barrier, cellular membranes are highly dynamic environments comprised of lipids, small molecules, and typically hundreds of proteins. Nearly all cell membranes are comprised of bilayers consisting of two surface-active molecules, typically lipids that assemble with their hydrophobic moieties in the interior of the bilayer and the polar groups exposed to the aqueous environment (Figure 1). Membrane proteins (MPs) perform a diverse set of tasks across the kingdoms of life including energy production,[1] signal transduction,[2] transport of solutes,[3] determination of cell shape,[4] and redox reactions.[5] Consistent with the diverse roles played by MPs, the genes coding for these proteins occupy a significant portion of an organism's genome. For example, at least 1000 of the bacterium *E. coli*'s 4252 genes are predicted to code for MPs (CyberCell database, http://redpoll.pharmacy.ualberta.ca/CCDB/, see Figure 2).[6] MPs can be divided into two classes depending on their predominant secondary structure. Proteins composed of mostly α-helices comprise the vast majority of MPs in eukaryotes, gram-positive bacteria, and the plasma membrane of gram-negative bacteria, while β-sheet based proteins are found in the outer membrane of gram-negative bacteria. Because of their prevalence, this review will focus on α-helical MPs. For the sake of simplicity and brevity, the abbreviation "MPs" will refer to α-helical membrane proteins for the remainder of this article.

The α-helices of MPs span the cell membrane, and, thus, are termed transmembrane helices (TMH) (Figure 1). The spatial organization of TMHs along with the location of the N- and C-termini of an MP determine its topology. Great progress has been made in the last several years in the prediction of MP topology with the advent of tools using hidden Markov models,[7] which are now freely available via the internet (http://www.cbs.dtu.dk/services/TMHMM-2.0/). The quality of these topology predictions increases even more when combined with experimental data such as the location of the N- and C-termini of the protein.[8] Combined with bioinformatics tools, topology predictors have been applied to the membrane proteomes of many different organisms.[9] While the topological information that these models provide is useful, understanding the molecular details of protein function requires a model at atomic resolution (Figure 1). Such models are currently only attainable via X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, or, at lower resolution (about 7 Å resolution), by cryo-electron microscopy (cryo-EM).

Although the number of available protein structures has increased dramatically to >35,000 in recent years, structures of MPs are seriously underrepresented, with fewer than 120 distinct structures present in the protein data bank[10] (PDB, see Figure 2). Furthermore, the majority of MPs with known structure are of bacterial origin with only about a dozen structures from higher organisms. There are several major hurdles preventing larger-scale structural biology efforts toward MPs. First and foremost, the production and purification of sufficient amounts of membrane proteins is a difficult undertaking. In some successful examples of MP structure determination, the protein was obtained from an abundant natural sources, such as rhodopsin from the bovine eye.[11] Such natural sources are not typically available for the vast majority of proteins, however, so it is desirable to be able to produce the protein of

Correspondence concerning this article should be addressed to G. Georgiou at gg@che.utexas.edu.

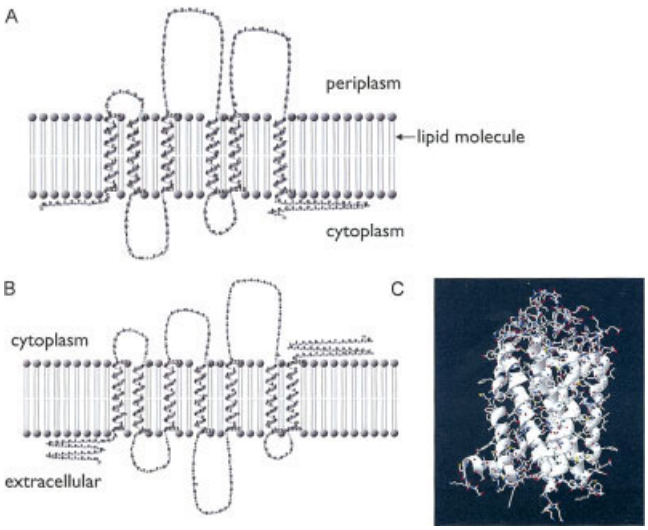

Figure 1. (a) Model a-helical membrane protein: the *E. coli* glycerol facilitator protein GlpF.

The transmembrane helices traverse a lipid bilayer which comprises the majority of the cell membrane. Balls on the lipids represent polar head groups while sticks represent the hydrophobic tails. The orientation of the protein with respect to the *E. coli* cytoplasm and periplasm is also shown, and comparison of (b) topological model of bovine rhodopsin, and (c) X-ray crystal structure of the same protein. While topology predictions can describe the orientation of transmembrane helices with respect the cell membrane, the X-ray crystal structure provides a picture of the protein at atomic resolution in three dimensions. A and B drawn with TMRPres2D software (http://biophysics.biol.uoa.gr/TMRPres2D/), C drawn from coordinates in PDB file 1F88.

interest in another organism.[12] This is accomplished by genetic engineering means, whereby, the gene encoding the membrane protein of interest is inserted and expressed in a host cell, typically a microbe, such as yeast or bacteria. This article will highlight several successful strategies for such heterologous expression of MPs, as well as discussing prospects of further engineering of hosts for improved MP expression. Chemical engineers are beginning to play a central role in the development of methods for more efficient production of membrane proteins and their crystallization in a form that can diffract X-rays in order to obtain structural information at atomic resolution.[13] In parallel, engineers are working closely with chemists in high-throughput screening programs aimed at the discovery of small molecule ligands with therapeutic potential.

Heterologous Membrane Protein Expression

The use of heterologous microorganisms for the expression of MPs affords several advantages compared to the isolation of MPs from natural sources. Since bacteria and yeast can be propagated in chemically-defined media, they are ideal hosts for labeling with the heavy-atom containing selenomethionine for X-ray structure determination.[14] Similarly, proteins produced in these hosts can be readily labeled with isotopically enriched amino acids for NMR studies. Furthermore, since heterologous protein expression relies on recombinant DNA manipulations, it is very simple to generate mutants of the protein under study, something which is impossible with proteins derived from natural sources.

One concern that arises when relying on a heterologous expression system is whether the protein produced is identical in composition and structure to that found in the natural host. The evidence accumulated so far indicates that membrane proteins produced by recombinant means do indeed fold into the same conformation as the authentic polypeptide derived from natural sources. For example, the structure of the ATP-driven calcium transporter SERCA1a was initially solved by X-ray crystallography using protein derived from the sarcoplasmic reticulum of rabbit cells.[15] Subsequently, the same protein was expressed in the yeast *Saccharomyces cerevisiae* with a biotin-based affinity tag to simplify purification. The yeast-derived SERCA1a, and the rabbit protein had identical enzymatic activity and formed similar crystals [16] validating the use of heterologously produced proteins for structural studies. However, caution will have to be exerted when analyzing MPs that contain posttranslational modifications including glycosylation, lipidation, disulfide-bond formation, and addition of phosphate or methyl groups. In general, microbial cells perform such modifications either poorly or in a different manner than the cells of higher organisms. Thus, even though the heterologous MP has the same amino acid composition as the naturally isolated material, differences in post-translational modifications may result in structural differences between the two proteins.

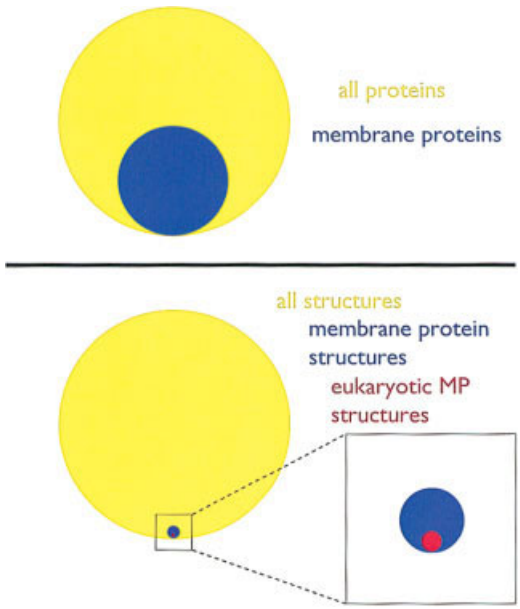

Figure 2. Comparison between fraction of membrane proteins in a typical genome and fraction of membrane protein structures in the Protein Data Bank (PDB).

20–30% of an organism's genome codes for membrane proteins while only 0.3% of the structures in the PDB are of membrane proteins. Eukaryotic membrane protein structure comprise less than 0.03% of the structures in the PDB.

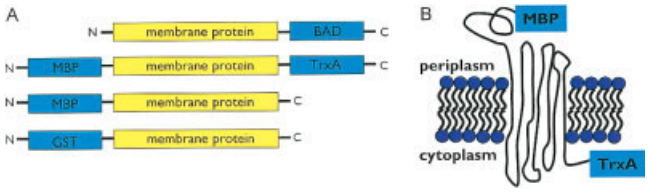

Figure 3. (A) Fusion partners used for heterologous expression of mammalian membrane proteins. BAD: oxaloacetate decarboxylase (biotinylated), MBP: maltose binding protein, TrxA: thioredoxin, GST: glutathione-S-transferase, and (B) the orientation of the MBP-neurotensin receptor-TrxA fusion in the *E. coli* inner membrane.

One strategy that has proved successful for the heterologous expression of MPs is the genetic fusion of the MP to another well-expressing protein (Figure 3). In conjunction with the proteome-wide topology studies discussed earlier Drew et al.[8] constructed fusions of green fluorescent protein (GFP) to the C-terminus of *E. coli* MPs as a robust platform for protein expression in *E. coli*.[17] The physical linkage of a well-folded fusion partner often serves to enhance folding of the MP. In addition, GFP renders the fusion-protein fluorescent, simplifying the detection of the protein during downstream processing steps, especially during solubilization of the protein in detergent micelles and subsequent chromatographic purification. Furthermore, misfolding of the MP interferes with the folding of the GFP and prevents fluorescence. Consequently, fluorescence can be used as a relatively accurate predictor of the expression (i.e., production) level of the MP-GFP fusion.

We and others have used the fusion protein strategy for the expression and purification of several G-protein coupled receptors (GPCRs) in *E. coli*. GPCRs are the most ubiquitous signaling molecules in mammalian cells, responding to signals ranging from light and small molecules to peptides and proteins. Genes for GPCRs account for more than 5% of the genome of the model organism *C. elegans*, with more than 1,100 distinct genes known.[18] Up to 2/3 of currently marketed drugs are believed to interact with GPCRs,[2] and GPCRs remain some of the most promising targets for future drug development. Despite the prevalence and biotechnological importance of these proteins, the only GPCR crystal structure available is that of the bovine rhodopsin.[11] Rhodopsin is exceptional among GPCRs in that it can be isolated in large amounts from natural sources, while other GPCRs, especially those of interest for pharmacological purposes, are present in minute amounts. Moreover, for human GPCRs, the use of natural material is not an option, underscoring the need for heterologous expression systems.

Grisshammer and colleagues have performed exhaustive studies on the expression of the neurotensin receptor (NTR) in *E. coli*.[19] They constructed a tripartite fusion protein in which *E. coli* maltose binding protein (MBP) is fused to the N-terminus of NTR and *E. coli* thioredoxin (TrxA) is fused to its C-terminus. GPCRs are composed of seven transmembrane helices with an N_{out}-C_{in} topology, so the MBP moiety of the fusion protein resides in the *E. coli* secretory compartment (periplasm), while TrxA resides in the cytoplasm (Figure 3). Recently, the MBP-NTR-TrxA fusion protein has been expressed on a large scale, and purified in its active form using a neurotensin affinity matrix.[20] This protein has been used in solid-state NMR experiments to examine the ligand binding dynamics of neurotensin to the receptor.[21] A similar tripartite fusion scheme has been used for the expression of the peripheral human cannabinoid receptor (CB2) in *E. coli*.[22] Earlier reports demonstrated that CB2 could be produced using only an N-terminal MBP fusion partner.[23] The N-terminal fusion of MBP also facilitated the expression of the adenosine A2a receptor.[24] Glutathione-S-transferase (GST) has also been used successfully as fusion partner for the expression of the olfactory receptor OR 5 in *E. coli*.[25] Tian et al. report the expression and purification of a fusion partner-free vasopressin receptor (V2R).[26] The yield of purified GPCRs obtained via expression in *E. coli* is typically on the order of 1 mg/L of culture, which is significantly lower than yields expected from well-behaved soluble proteins (from 10–100 mg/L). Moreover, the yield of functional GPCR as judged by radioligand binding studies is often only in the range of 1–10 μg/L, underscoring the difficulty of producing these proteins heterologously.

Although we have focused on *E. coli* thus far as an expression host for GPCRs, other microorganisms have been successfully used for preparative expression of these proteins. In one notable example, Robinson and coworkers expressed the A2a adenosine receptor described ealier at very high-functional yields as a GFP fusion[27] in *Saccharomyces cerevisiae*. According to radioligand binding assays, approximately 120,000 active receptor molecules were produced per cell for a total yield of active protein of approximately 4 mg/L of culture. This represents the highest yield of active GPCR reported in the literature for any heterologous expression system.[28] Also, the gram-positive bacterium *Lactococcus lactis* has also been used as a host for the expression of GPCRs and other eukaryotic MPs with some success.[17,29] As *L. lactis* has only a single cell membrane (compared to two membranes in *E. coli*), cell lysis and isolation of MPs may be simpler in this organism.

Applying Physiological Information to MP Expression

The earlier examples of successful membrane protein expression all relied on the selection of an appropriate fusion partner, or affinity tag, that improves the expression of the MP. Another approach, which has the potential of being more generalizable, involves engineering of the expression host for the specific task of MP expression. Recent biochemical, genetic, and structural data have greatly enhanced our understanding of the synthesis of MPs, and their subsequent insertion into the membrane (collectively called membrane protein biogenesis).

Whereas most soluble proteins are synthesized completely in the cytoplasm of the cell, the synthesis of membrane proteins is halted after the emergence of a short "signal anchor" from the ribosome. This nascent polypeptide binds to a riboprotein complex, the signal recognition particle (SRP). The ribosome-SRP complex is concomitantly transported to the plasma membrane of bacteria or the endoplasmic reticulum in eukaryotes.[30,31] The nascent protein next interacts with the signal-particle receptor, [32] and is delivered to the secretion apparatus (called the translocon), Sec61 in eukaryotes and SecYEG in bacteria. Elegant photocrosslinking studies on the bacterial

export system have thoroughly established the pathway of a nascent MP through the ribosome, and into the Sec translocon.[33] In addition, a recent cryo-EM structure of the SecYEG translocon bound to a translating ribosome provides some insight into how the translocon may assist in the insertion of MPs into the membrane.[34] The putative membrane integrase YidC is also emerging as an important factor for MP biogenesis.[35] Fröderberg et al. performed a systematic study on the effects of depletion of translocon components, SRP components, and YidC on the biogenesis of several model MPs in *E. coli*.[36] Elimination of either the SRP, core components of the SecYEG translocon, or YidC in cells led to severely disrupted MP biogenesis. It remains to be seen if overproduction of any of these factors will lead to more efficient MP biogenesis.

Recombinant protein expression is often stressful for the host cell.[37] Understanding in detail the stress responses that arise from heterologous expression of MPs may permit rational strain engineering to eliminate or combat these stress responses, and enhance the production of membrane proteins. Steps toward this goal have been taken in the yeast *Saccharomyces cerevisiae*. Expression of the P2 adenosine transporter from *Trypanosoma equiperdum* in yeast results in accumulation of misfolded secreted protein, and in turn stimulates the unfolded protein response (UPR).[38] Induction of the UPR coincided with low-functional yields of the P2 transporter. To overcome this limitation, the expression level of the P2 transporter was reduced in order to avoid the elicitation of the UPR, ultimately leading to higher yield of functional P2 transporter. Further cataloging of the stress responses induced by MP expression may be facilitated by genomics techniques, and by emerging proteomic technologies aimed at elucidating newly-synthesized proteomes in cells.[39]

Outlook

Careful choice of fusion partners and affinity tags can lead to vastly improved yields of complex MPs. Among the challenges remaining in the expression of MPs is a thorough determination of the cell's response to MP expression. The combination of a "rationalization" of heterologous MP expression[40] with continued improvements in protein engineering should usher in serious progress toward the goal of enabling MP production for structural biology and other efforts. Protein structures provide information that is not obtainable by any other analysis method, and often provide a window into the inner workings of a protein. Many human membrane proteins are pharmaceutically relevant, and the structure determination of these proteins bound to drugs or ligands will improve our understanding the mechanisms of drug action. The ability to reliably produce MPs will also facilitate future drug discovery via high-throughput bioassay screens. Although progress has been made toward the expression and production of MPs, much more effort is required in order to fully understand their structure and function.

Literature Cited

1. Muller V, Gruber G. ATP synthases: structure, function and evolution of unique energy converters. *Cellular and Molecular Life Sciences*. 2003;60:474–494.
2. Bartfai T, Benovic JL, Bockaert J, Bond RA, Bouvier M, Christopoulos A, Civelli O, Devi LA, George SR, Inui A, Kobilka B, Leurs R, Neubig R, Pin JP, Quirion R, Roques BP, Sakmar TP, Seifert R, Stenkamp RE, Strange PG. The state of GPCR research in 2004. *Nature Reviews Drug Discovery*. 2004;3:574–626.
3. Nikaido H. Molecular basis of bacterial outer membrane permeability revisited. *Microbiology and Molecular Biology Reviews*. 2003;67:593–656.
4. Michie KA, Lowe J. Dynamic filaments of the bacterial cytoskeleton. *Annual Review of Biochem*. 2006;75:467–492.
5. Kadokura H, Katzen F, Beckwith J. Protein disulfide bond formation in prokaryotes. *Annual Review Of Biochemistry*. 2003;72:111–135.
6. Sundararaj S, Guo A, Habibi-Nazhad B, Rouani M, Stothard P, Ellison M, Wishart DS. The CyberCell Database (CCDB): a comprehensive, self-updating, relational database to coordinate and facilitate in silico modeling of Escherichia coli. *Nucleic Acids Res*. 2004;32:D293–D295.
7. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J of Molecular Biology*. 2001;305:567–580.
8. Daley DO, Rapp M, Granseth E, Melen K, Drew D, von Heijne G. Global topology analysis of the Escherichia coli inner membrane proteome. *Science*. 2005;308:1321–1323.
9. Granseth E, Daley DO, Rapp M, Melen K, von Heijne G. Experimentally constrained topology models for 51,208 bacterial inner membrane proteins. *J of Molecular Biology*. 2005;352:489–494.
10. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. *Nucleic Acids Res*. 2000;28:235–242.
11. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science*. 2000;289:739–745.
12. Grisshammer R, Tate CG. Overexpression of Integral Membrane-Proteins for Structural Studies. *Quarterly Reviews of Biophysics*. 1995;28:315–422.
13. Berger BW, Gendron CM, Lenhoff AM, Kaler EW. Effects of additives on surfactant phase behavior relevant to bacteriorhodopsin crystallization. *Protein Sci*. 2006;15:2682–2696.
14. Hendrickson WA, Horton JR, Lemaster DM. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (mad) – a vehicle for direct determination of 3-dimensional structure. *EMBO J*. 1990;9:1665–1672.
15. Toyoshima C, Nakasako M, Nomura H, Ogawa H. Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 angstrom resolution. *Nature*. 2000;405:647–655.
16. Jidenko M, Nielsen RC, Sorensen TLM, Moller JV, le Maire M, Nissen P, Jaxel C. Crystallization of a mammalian membrane protein overexpressed in Saccharomyces cerevisiae. *Proceedings of The National Academy of*

Sciences of The United States of America. 2005;102: 11687–11691.
17. Drew D, Slotboom DJ, Friso G, Reda T, Genevaux P, Rapp M, Meindl-Beinker NM, Lambert W, Lerch M, Daley DO, Van Wijk KJ, Hirst J, Kunji E, De Gier JW. A scalable, GFP-based pipeline for membrane protein overexpression screening and purification. *Protein Sci*. 2005;14:2011–2017.
18. Bargmann CI. Neurobiology of the caenorhabditis elegans genome. *Science*. 1998;282:2028–2033.
19. Tucker J, Grisshammer R. Purification of a rat neurotensin receptor expressed in Escherichia coli. *Biochem J*. 1996;317:891–899.
20. White JF, Trinh LB, Shiloach J, Grisshammer R. Automated large-scale purification of a G protein-coupled receptor for neurotensin. *Febs Letters*. 2004;564:289–293.
21. Luca S, White JF, Sohal AK, Filippov DV, van Boom JH, Grisshammer R, Baldus M. The conformation of neurotensin bound to its G protein-coupled receptor. *Proceedings of The National Academy of Sciences of The United States of America*. 2003;100:10706–10711.
22. Yeliseev AA, Wong KK, Soubias O, Gawrisch K. Expression of human peripheral cannabinoid receptor for structural studies. *Protein Sci*. 2005;14:2638–2653.
23. Calandra B, Tucker J, Shire D, Grisshammer R. Expression in Escherichia coli and characterisation of the human central CB1 and peripheral CB2 cannabinoid receptors. *Biotechnol Letts*. 1997;19:425–428.
24. Weiss HM, Grisshammer R. Purification and characterization of the human adenosine A(2a) receptor functionally expressed in Escherichia coli. *Euro J of Biochemistry*. 2002;269:82–92.
25. Kiefer H, Krieger J, Olszewski JD, vonHeijne G, Prestwich GD, Breer H. Expression of an olfactory receptor in Escherichia coli: Purification, reconstitution, and ligand binding. *Biochemistry*. 1996;35:16077–16084.
26. Tian CL, Breyer RM, Kim HJ, Karra MD, Friedman DB, Karpay A, Sanders CR. Solution NMR spectroscopy of the human vasopressin V2 receptor, a G protein-coupled receptor. *J of The A Chem Soc*. 2005;127:8010–8011.
27. Niebauer RT, Robinson AS. Exceptional total and functional yields of the human adenosine (A2a) receptor expressed in the yeast Saccharomyces cerevisiae. *Protein Expression and Purification*. 2006;46:204–211.
28. Sarramegna V, Talmont R, Demange P, Milon A. Heterologous expression of G-protein-coupled receptors: comparison of expression systems from the standpoint of large-scale production and purification. *Cellular and Molecular Life Sciences*. 2003;60:1529–1546.
29. Monne M, Chan KW, Slotboom DJ, Kunji ERS. Functional expression of eukaryotic membrane proteins in Lactococcus lactis. *Protein Sci*. 2005;14:3048–3056.
30. Walter P, Johnson AE. Signal Sequence Recognition and protein targeting to the endoplasmic-reticulum membrane. *Annual Rev of Cell Biology*. 1994;10:87–119.
31. Luirink J, von Heijne G, Houben E, de Gier JW. Biogenesis of inner membrane proteins in Escherichia coli. *Annual Rev of Microbiology*. 2005;59:329–355.
32. Miller JD, Bernstein HD, Walter P. Interaction of escherichia-coli ffh/4.5s ribonucleoprotein and ftsy mimics that of mammalian signal recognition particle and its receptor. *Nature*. 1994;367:657–659.
33. Houben ENG, Zarivach R, Oudega B, Luirink J. Early encounters of a nascent membrane protein: specificity and timing of contacts inside and outside the ribosome. *J of Cell Biology*. 2005;170:27–35.
34. Mitra K, Schaffitzel C, Shaikh T, Tarna F, Jenni S, III CLB, Ban N, Frank J. Structure of the *E. coli* protein-conduction channel bound to a translating ribosome. *Nature* 2005;438:318–324.
35. Samuelson JC, Chen MY, Jiang FL, Moller I, Wiedmann M, Kuhn A, Phillips GJ, Dalbey RE. YidC mediates membrane protein insertion in bacteria. *Nature*. 2000;406:637–641.
36. Froderberg L, Houben E, Samuelson JC, Chen MY, Park SK, Phillips GJ, Dalbey R, Luirink J, de Gier JWL. Versatility of inner membrane protein biogenesis in Escherichia coli. *Molecular Microbiology*. 2003;47:1015–1027.
37. Hoffmann F, Rinas U. Kinetics of heat-shock response and inclusion body formation during temperature-induced production of basic fibroblast growth factor in high-cell-density cultures of recombinant Escherichia coli. *Biotechnol Progress*. 2000;16:1000–1007.
38. Griffith DA, Delipala C, Leadsham J, Jarvis SM, Oesterhelt D. A novel yeast expression system for the overproduction of quality-controlled membrane proteins. *Febs Letts*. 2003;553:45–50.
39. Dieterich DC, Link AJ, Graumann J, Tirrell DA, Schuman EM. Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proceedings of the National Academy of Sciences of the United States of America*. 2006;103:9482–9487.
40. Wagner S, Bader ML, Drew D, de Gier JW. Rationalizing membrane protein overexpression. *Trends In Biotechnol*. 2006;24:364–371.