Producing membrane proteins one simulation at a time

DOI 10.1074/jbc.H117.813469

James C. Gumbart

From the Department of Physics, Georgia Institute of Technology, Atlanta Georgia 30332

Edited by Karen G. Fleming

Integral membrane proteins are studied with a number of structural and biophysical techniques, many requiring protein overexpression to reach sufficient quantities. However, achievement of the overexpression of membrane proteins is not necessarily straightforward, and the mechanisms and factors that influence expression are not clearly understood. A new study has now broken through this uncertainty by demonstrating the capability of coarse-grained simulations of membrane protein insertion to predict protein expression levels in Escherichia coli.

Membrane proteins comprise 20–30% of the genes of most organisms and are over-represented as drug targets (as much as 70%). Yet, obtaining sufficient quantities of protein for in vitro studies, especially for eukaryotic α-helical membrane proteins, is challenging for a number of reasons, including (but not limited to) their low natural expression levels. Escherichia coli is typically used as a model system to overexpress these proteins, but expression levels in this context vary widely, and the reasons why are not always apparent. Although different conditions can be explored to improve expression, there are no general guidelines for success. The authors of a new study now take a critical step in that direction in their report of a more rational approach for designing overexpression constructs that incorporates the details of the cellular machinery responsible for membrane insertion (1).

Membrane proteins are often co-translationally trafficked to the Sec translocon, through which they are inserted into the membrane to avoid buildup of aggregation-prone hydrophobic sequences. Two landmark studies, one in 2004 providing the first crystal structure of a Sec translocon (2), and one in 2005 cracking its code for transmembrane helix recognition (3), laid the groundwork for a number of experimental and computational investigations that further teased out the details of the insertion process (4). One result was the development by Zhang and Miller (5) of an in silico coarse-grained (CG) simulation approach that replicates co-translational membrane protein insertion on realistic time scales. Despite its simplicity, this approach has been demonstrated to reproduce numerous experimental observations while providing a molecular-scale depiction of their underlying events.

Niesen et al. (1) have now pushed their CG simulations out of the computer and into the lab (Fig. 1). The authors made the initial simulation–experiment connection for the six-transmembrane–domain (TM) protein TatC last year (6). In the present study, they created 140 variants, predicted the insertion efficiency of each from simulation, and then measured expression levels in E. coli. The variants ranged from single-point mutations to the swapping of entire loops. A statistical analysis of the results showed that mutations that enhanced insertion in the CG simulations were fourfold more likely to enhance expression in experiments over random selection, demonstrating the simulations’ predictive power. Furthermore, the authors found that the effects of point mutations on insertion were largely independent and cumulative, illuminating a path forward for greatly increased protein expression yields.

A surprising finding of the work from Niesen et al. (1) is the strong dependence of their model’s predictive capability on the simulated protein topology of the C-terminal tail, i.e. its placement with respect to the membrane. The authors found multiple other measures to predict expression efficiency poorly, including the fraction of simulations in which the topologies of all cytosolic/periplasmic loops were correct, the fraction in which only the mutated loop was correct, and the fraction in which each one of the seven loops/termini was correct. One possible explanation is that the C-terminal tail “aggregates” all previous errors and thus will report on topological errors anywhere in the protein. However, the lack of correlation of the C-terminal tail topology with any other measure undermines this explanation. Instead, the authors suspect that the C-terminal–tail topology is especially sensitive to mutations anywhere in TatC, an explanation also supported indirectly by an ampicillin resistance assay.

TatC is not the first protein in which the C-terminal tail has been found to play an outsized role in its topology. The four-TM protein EmrE, notorious for its ability to adopt opposite topologies in the membrane, has been shown to flip its topology based on even the charge of the very C-terminal residue (7). Although EmrE may be a relative outlier, it indicates that membrane protein folding is capable of a high degree of plasticity, depending on a variety of factors, many uncontrollable, a fact anathema to the careful design and production of de novo membrane proteins. However, the CG simulations have previously been demonstrated to correctly predict the topology of EmrE mutants, which were observed to reorient in the membrane through “kinetic annealing” (8). Although it remains to be seen if other proteins are as sensitive to C-terminal–tail topology as EmrE or TatC, Niesen et al. (1) already provide a general strategy for selecting the measure from the CG simula-

This work was supported by National Institutes of Health Grant R01-GM123169. The author declares that he has no conflicts of interest with the contents of this article. The content is solely the responsibility of the National Institutes of Health.

1 To whom correspondence should be addressed: 837 State St., Atlanta, GA 30332. E-mail: gumbart@physics.gatech.edu.

2 The abbreviations used are: CG, coarse-grained; TM, transmembrane.
tions that is most predictive of expression efficiency. Using sub-
sets of their full 140-mutant data set, the authors determined
that a training set with fewer than 20 mutants was already suf-
ficient to clearly identify C-terminal–tail topology as the most
predictive measure for TatC.

Looking beyond improving expression for existing proteins,
the method of Niesen et al. (1) will also aid in the creation of
entirely new membrane proteins, an area of research potentially
poised to usher in a new age of biotechnology (9). Membrane
proteins are relatively under-represented in this area, although
there have been some successes, e.g. a four-helix–bundle Zn$^{2+}$
transporter (10). However, this transporter was simple enough
to insert directly into the membrane; more complex creations
will require accounting for the ribosome–translocon system as
well.

Although designer membrane proteins are not yet some-
thing that can be ordered out of a catalogue, the work of Niesen
et al. (1) has taken us one step closer. By developing a strategy
to reliably determine expression efficiency for a variety of mutants
of a known membrane protein, they have overcome one major
technological hurdle, namely how to get a protein into the
membrane. Future work is now needed to expand this strategy
to other existing membrane proteins and, ultimately, to com-
pletely de novo ones.

References
1. Niesen, M. J. M., Marshall, S. S., Miller, T. F., III, and Clemons, W. M., Jr.
(2017) Improving membrane protein expression by optimizing integra-
tion efficiency. *J. Biol. Chem.* 292, 19537–19545
2. van den Berg, B., Clemons, W. M., Jr., Collinson, I., Modis, Y., Hartmann,
E., Harrison, S. C., and Rapoport, T. A. (2004) X-ray structure of a protein-
conducting channel. *Nature* 427, 36–44
3. Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H.,
Nilsson, I., White, S. H., and von Heijne, G. (2005) Recognition of trans-
membrane helices by the endoplasmic reticulum translocon. *Nature*
433, 377–381
4. Gumbart, J. C., and Chipot, C. (2016) Decrypting protein insertion
through the translocon with free-energy calculations. *Biochim. Biophys.
Acta* 1858, 1663–1671
5. Zhang, B., and Miller, T. F., III (2012) Long-timescale dynamics and reg-
ulation of Sec-facilitated protein translocation. *Cell Rep.* 2, 927–937
6. Marshall, S. S., Niesen, M. J. M., Müller, A., Tiemann, K., Saladi, S. M.,
Gailmidi, R. P., Zhang, B., Clemons, W. M., Jr, and Miller, T. F., III (2016)
A link between integral membrane protein expression and simulated in-
tegration efficiency. *Cell Rep.* 16, 2169–2177
7. Seppälä, S., Slusky, J. S., Lloris-Garcera, P., Rapp, M., and von Heijne, G.
(2010) Control of membrane protein topology by a single C-terminal resi-
due. *Science* 328, 1698–1700
8. Van Lehn, R. C., Zhang, B., and Miller, T. F., III (2015) Regulation of
multispanning membrane protein topology via post-translational anneal-
ing. *eLife* 4, e08697
9. Huang, P.-S., Boyken, S. E., and Baker, D. (2016) The coming of age of de
ovo protein design. *Nature* 537, 320–327
10. Joh, N. H., Wang, T., Bhatte, M. P., Acharya, R., Wu, Y., Grabe, M., Hong,
M., Grigoryan, G., and DeGrado, W. F. (2014) *De novo* design of a trans-
membrane Zn$^{2+}$–transporting four-helix bundle. *Science* 346, 1520–1524