Finding the determinants of outer membrane protein insertion is a task of intrinsic scientific interest, as a missing piece in a mechanistic model of bacteria. It is also of potentially great technological value. GFP would be a great tool for neuroscience experiments if it could be made to insert into membranes (Blunck et al., 2005). GFP is a β-barrel, a structure it shares with bacterial outer membrane proteins. Understanding why outer membrane β-barrels insert may allow the design of transmembrane mutants of water-soluble β-barrels such as GFP.

Outer membrane proteins (OMPs) are insoluble in water, and if unfolded can spontaneously refold and insert into vesicles (Surrey and Jähnig, 1992). Together this suggests that at equilibrium, OMPs are in membranes, rather than only being held there by some kinetic barrier. It is likely, then, then negative of folding is a necessary condition for insertion.

Under the hypothesis that for a whole OMP is a sum of contributions from solvent-exposed residues on its surface, the problem of estimating for an arbitrary OMP becomes the problem of finding free energies of transfer for individual amino acids. Once this is achieved, the calculation of is as simple as summing the transfer energies. There have been three broad categories of approaches to this problem.

One is experimental. 's for each amino acid have been derived through a mutation study (Moon and Fleming, 2011). Another is through simulation. Molecular dynamics simulations have been used to derive 's which are very close to those estimated from experiment (Gumbart and Roux, 2012).

This study is concerned with the development and application of the knowledge-based approach to the problem. In a knowledge-based approach, one desires to know the *energy* of a particular state; and, from a structure database, one knows the *frequency* of that state. Under the hypothesis that the frequency and the energy are related by Boltzmann's law, the energy of each state is derived (Sippl, 1993). The resulting energy function is called a *knowledge-based potential*.

This work builds upon the knowledge-based Ezβ potential (Hsieh et al., 2012). Unlike the above referenced experiment- and simulation- based values, the Ezβ potential is *depth-dependent*: it estimates the energy of transfer to a given depth in the membrane, not just to the center. Depth is represented by a number, z, that represents the distance from the center: at z=0, Ezβ should, and does, correlate with the experimental values. This depth dependence is likely to increase accuracy because the membrane has been experimentally shown to have varying hydrophobicity; and because each residue has a distinct depth-dependent frequency profile, sometimes with a smooth transition or even a peak partway through, that seem to reflect a dependence of insertion energy upon depth that is more complicated than a simple distinction between "in" and "out" (Hsieh et al., 2012).

Membrane protein structures are scarce, though. Using only the currently solved structures,

---

*Interpretation of the potential*

The justification for using the Boltzmann law to derive energies from frequencies is difficult and technical, and it may not, in fact, be justified in the case of an insertion potential (Borg et al., 2012). To a certain degree, it does not matter. Documenting the depth-dependence of frequencies has its own interest, and this knowledge can guide design without being formalized as a scoring function.

There are certain questions that seem to demand an energy function. For example: are the surfaces of OMPs naturally selected to keep the OMPs steady in the membrane, so they do not fall out? This can be translated to: is the increase in insertion energy as the protein moves up or down seem steeper than it would be by chance? However, even in this case, it seems that some heuristic scoring rule loosely related to the energy, combined with a statistical test, can do the job.

However, whatever a heuristic scoring rule might be capable of, it is hard to compare to other predictors of insertion, and hard to fit into our physical understanding of membrane proteins. For example, one might wonder, can we sum the score with an energy, calculated in another way? Is summation the appropriate operation, and do we double-count information by doing this? Also, it would be interesting to compare the scoring rule to other measures of insertion, such as the experimentally derived free energies. But are we looking for a linear relationship, or a parabolic or some other relationship?

There is another reason to look for an energetic interpretation, or at least some other interpretation. Humans are analogical reasoners, not calculators; and it is hard to use math that we cannot relate to something familiar. I am bothered that I do not have a good analogy, and I suspect that without one, I will make mistakes when I try to reason about them. For example, I used to mistakenly think of Ezβ as an energy of interaction with solvent, rather than a change in interaction energy after transfer from water. Because of this, I thought that the Ezβ moment would be sensitive to the hydrophobic residues found at the top and bottom of the ScrY interface, when in fact their Ezβ is little different than it would be if they were all aspartates. **(true?)**

We could be making such errors of interpretation in our application of Ezβ. In fact, I think I have a plausible candidate. Jim used Ezβ to calculate changes in energy as a protein moves up and down in a membrane, designing so that these changes would be steep. However, Ezβ does not really measure frequency purely as a function of the depth dependent nature of the membrane. It confounds influences due to the varying membrane environment and the varying protein environment. For example, it may be the case that while tyrosine is enriched in the aromatic belt due to the energetic favorability of the snorkeling position, phenylalanine only appears to be enriched in the aromatic belt because it is suppressed in the hydrophobic core by the strong FV antimotif (Jackups and Liang, 2005). Then, if the protein were moved so that the aromatic belt were in the center of the membrane, the energy of the tyrosines would go up, because they left their lowest energy spot in the membrane; but the energy of the phenylalanines would go down, because they are in a more hydrophobic environment but still free of valines. The protein environment moves with the protein; the lipid environment does not; the two are inseparable in Ezβ, and therefore the potential is actually far less trustworthy when used to model proteins that are not in their natural orientations. However, this is obscured by our interpretation of it as a transfer energy.

So aside from easy comparison with other energies, I do not think that there is anything that can be done with a well justified energy function that cannot be done with a heuristic scoring rule. But I do expect a well justified energy function to be easier to use appropriately. It would be a tool superior not in its cutting edge but in its handle.