Membrane transport proteins: surprises in structural sameness

Douglas L Theobald & Christopher Miller

An X-ray crystal structure of an organic anion transporter identifies it as an ion channel instead. Its similarity to an unrelated family of water channels raises evolutionary questions that have been recently bubbling up around membrane proteins.

Although the first X-ray crystal structure of a membrane protein is 25 years behind us[1], in a sense it's still early days. Of the ~50,000 entries in the protein structure database, fewer than 0.5% represent unique membrane proteins, which themselves make up roughly one-third of cellular proteomes. For the past two decades, these structures have been climbing the exponential foot of the growth curve, their numbers doubling every 3 years. Many of these proteins move solutes across cell membranes, and such proteins fall into two broad classes—channels and transporters—that work by distinct mechanisms. Channels catalyze passive diffusion across the membrane through a watery pore spanning the membrane-embedded protein, whereas transporters work via a cycle of conformational changes that expose substrate-binding sites alternately to the two sides of the membrane. Because diffusion is fast compared to typical protein conformational rearrangements, channels typically have far higher solute transport rates (millions of ions per second) than do transporters (hundreds of solutes per second). In this issue, Waight and colleagues[2] report high-resolution structures of *Vibrio cholerae* FocA, a novel membrane protein in the widespread but heretofore structurally unexplored FNT family of Formate and Nitrite Transporters, which move various polyatomic anions across the membranes of bacteria as well as those of certain pathogenic fungi and parasites.

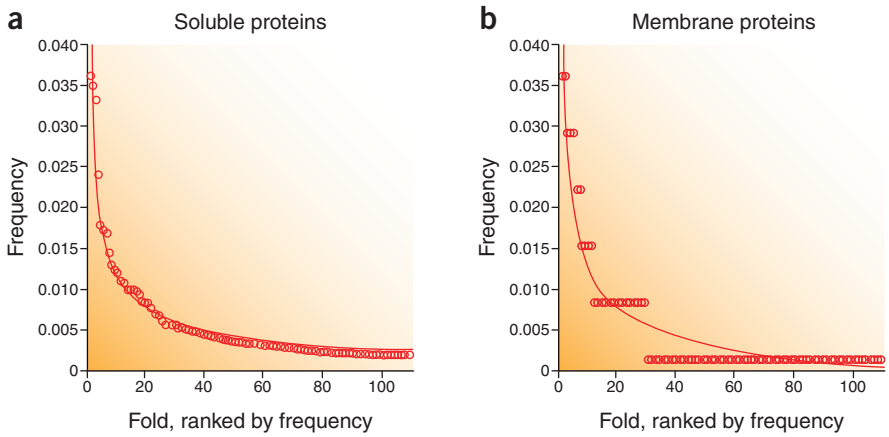

Figure 1 Fold statistics are similar for sequence-heterogeneous membrane proteins and soluble proteins. (**a**) Soluble proteins. The Structural Classification of Proteins (SCOP) domain database was filtered by *E* value (<0.01) to obtain a set of protein domains lacking any significant sequence similarities. The first 110 most common superfamily folds (comprising 4,573 domains) are shown ranked from most frequent to least. The ranked fold distribution is fit to a power law model: $y = Ax^B$, where *A* is the *y* intercept, *B* is the exponent, *x* is the rank of the fold, and *y* is the fold's frequency among all the sequence-dissimilar domains. As expected, the folds show a good fit to a power law, with coefficients of correlation *R* near 1. (**b**) Membrane proteins. The same plot for membrane proteins (164 domains). The power law for membrane protein domains is remarkably similar to that for soluble proteins.

The structure is delightfully arresting in several ways. First, by displaying a formate-occupied pore running across the protein's entire transmembrane width, it shows that the protein is not a transporter at all, as had been supposed from familiar sequence-based transmembrane topology. Rather, it's a channel—indeed, the first organic ion–selective channel known. (To validate this conclusion, future work must examine its formate conduction rate, anion selectivity, and other biophysical properties.) Second, the FocA fold is uncannily similar to that of the aquaporin (AqP) family of water channels—another big surprise, as these two kinds of channels show no similarity in amino acid sequence. Finally, whereas all known AqP channels are homotetramers with a pore in each subunit, FocA is a five-pore homopentamer. These multiple unexpected results, which are reprised in *Nature* by another group working on a close homolog from *Escherichia coli*[3], raise fundamental questions about structural

Douglas L. Theobald and Christopher Miller are in the Department of Biochemistry, Brandeis University, Waltham, Massachusetts, USA.
e-mail: dtheobald@brandeis.edu, cmiller@brandeis.edu

constraints placed on evolving membrane proteins by the heterogeneous environments in which they reside.

Over the past few years, the steady increase of published membrane protein structures has turned up numerous unexpected examples of structural similarity in apparently unrelated protein families, as seen here with FocA and AqP. Another striking instance is the close match of transmembrane helix packing in no less than five sequence-unrelated membrane transporters whose structures were recently solved[4]. Likewise, the structure of a glutamate receptor[5] impressively overlays K^+ channel pore domains, despite the marginality of sequence alignments between these two ion-channel families[6]. This structural sameness has provoked two types of diametrically opposite, evolution-directed mumblings among membrane-transport investigators: does it reflect divergence of common ancestors so ancient that sequence similarities have been wiped out by accumulated drift, or is it a manifestation of the restricted folding space available to membrane proteins, which live in something like 2½ dimensions (quasi-2D) rather than in the 3 dimensions enjoyed by soluble proteins?

This question is an old one in biology: nearly two centuries ago, in a pre-Darwinian epoch, the two most influential biologists in Europe staged a series of public debates to hash out this very issue. Are the widespread similarities in biological structures due to "homology," as Geoffroy contended, or to functional necessity, as Cuvier argued[7]? Although Cuvier allegedly won the 1830 debate, evolutionary biology eventually vindicated Geoffroy, at least in confirming the homology of metazoan body plans.

At the molecular level, however, the vexing question still remains unresolved for proteins possessing similar topology yet lacking detectable sequence similarity. When we do see striking congruence between molecular sequences, conventional wisdom says we have good reason to conclude homology—descent from a common molecular ancestor. The traditional argument is premised on two facts. First, sequence space is, for all practical purposes, infinite. Second, no particular sequence is strictly necessary to perform a given function or to adopt a certain conformation. Proteins are notoriously resilient with respect to substitutions, and we have seen time and again that vastly different sequences and structures do perform similar functions. Thus, because of the extreme improbability of independently evolving similar stretches of amino acids, significant sequence similarity is widely regarded as sufficient evidence for homology.

But what can be said when structural similarity is all we have? Unfortunately, the argument above breaks down when applied to structures. Unlike sequence space, 'fold space' is patently finite and small, with roughly 1,000 different protein folds available to soluble proteins (according to the molecular typologists at the Structural Classification of Proteins (SCOP) and CATH Protein Structure Classification databases). To arrive independently at similar folds, then, is nowhere near as implausible as in the case of sequences. And at first blush, it would appear that the problem is even more acute for membrane proteins. Fold space is surely more constrained within a membrane than it is in water. Soluble proteins experience a single solvent, but membrane proteins must deal with five: two aqueous and one greasy, separated by two transition zones[8]. Moreover, membrane proteins use a small subset of residues to find comfort in the membrane's hydrophobic interior, and an arginine/lysine bias restrains their aqueous surfaces[9]. Fold space in quasi-2D is further restricted by transmembrane topology—after all, there are only so many ways to connect a handful of transmembrane helices, all of comparable lengths. So, particular structures could be necessary, or nearly so, to perform certain membrane-related functions. Taken together, these considerations give good reason to expect that true structural convergence should be more prevalent among membrane proteins.

If these thought experiments are correct, and convergence really is more likely among membrane proteins, then we might expect a greater frequency of structurally similar domains that lack sequence similarity. But when we survey the prevalence of membrane protein folds, similar statistics are seen with soluble proteins, albeit with many more data points (compare **Fig. 1a**,**b**). At the dawn of the great explosion of structural biology in the mid-1970s, when the structural database was as small as it is now for membrane proteins, crystallographers were already wondering at structural similarities among soluble proteins that had no clear sequence similarity (globin and Rossmann folds, Ig domains, TIM barrels, etc.). Nevertheless, if we confine our view to just the transporters and channels (rather than the entire structurally known membrane universe), we see that a few transmembrane folds seem to pop up roughly ten times more often than would be expected from the global statistics in **Figure 1b**. Is this because these domains are modern descendants of an ancient, functionally critical channel, where structure has been retained as sequences drift? Or are these folds especially suited to their specific functions (namely, getting small molecules across a membrane), such that evolution has funneled various proteins from different origins to these particular topologies multiple times? Cuvier would likely enjoy knowing that the debate continues yet.

1. Deisenhofer, J. *et al. J. Mol. Biol.* **180**, 385–398 (1984).
2. Waight, A.B., Love, J. & Wang, D.-N. *Nat. Struct. Mol. Biol.* **17**, 31–37 (2010).
3. Wang, Y. *et al. Nature* **462**, 467–472 (2009).
4. Fang, Y. *et al. Nature* **460**, 1040–1043 (2009).
5. Sobolevsky, A.I., Rosconi, M.P. & Gouaux, E. *Nature* **462**, 745–756 (2009)
6. Wo, Z.G. & Oswald, R.E. *Trends Neurosci.* **18**, 161–168 (1995).
7. Appel, T.A. *The Cuvier-Geoffroy Debate: French Biology in the Decades before Darwin* (Oxford University Press, Oxford, UK, 1987).
8. Yau, W.M. *et al. Biochemistry* **37**, 14713–14718 (1998).
9. von Heijne, G. & Gavel, Y. *Eur. J. Biochem.* **174**, 671–678 (1988).