Minireview

A rational route to probing membrane proteins

Amy E Keating

Address: MIT Department of Biology, Massachusetts Avenue, Cambridge, MA 02139, USA. Email: keating@mit.edu

Published: 31 May 2007

Genome ***Biology*** 2007, **8:**214 (doi:10.1186/gb-2007-8-5-214)

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2007/8/5/214

Abstract

A recent report describes the design of short peptides that bind specifically to transmembrane regions of integrins, providing an exciting tool for probing the biology of membrane proteins.

Membrane proteins constitute around 20-30% of most proteomes. They carry out numerous critical functions and are significantly over-represented as drug targets compared with soluble proteins. However, membrane proteins present a host of practical challenges that have limited our understanding of their structure-function relationships. Methods that are standard for investigating the interactions among soluble proteins, such as phage display, yeast two-hybrid analysis, or any experiment that requires specific antibodies, are difficult or impossible to apply to transmembrane regions of membrane proteins. This makes it hard to probe the effects of specifically inhibiting or activating proteins that reside within the membrane. New reagents and approaches for deciphering membrane protein function could significantly advance our understanding.

Given the difficulties of experimentally selecting probes specific for membrane proteins, the rational design of such molecules is appealing. In particular, computational protein design holds promise for providing micro-scale tools appropriate for manipulating the molecular world. Successes in designing protein sequences that adopt desired folds, specifically recognize small molecules or catalyze reactions have raised hopes that rational design may provide a route to useful reagents and therapeutics [1-7]. The obstacles that confront the field are significant, however. In particular, the challenge of designing proteins or peptides to bind tightly and specifically to native protein targets is largely unmet, although this is arguably one of the areas where the impact of protein design could be greatest. Two big problems confront protein engineers. One is the vast sequence/structure space in which possible solutions lie (the 'search problem'). The other is the physics of molecular recognition, which is complex and has proved difficult to capture in computational methods that are fast enough to use for design (the 'energy problem').

There are theoretical reasons why membrane proteins may present easier targets for design than soluble ones. Both the search problem and the energy problem are simplified in membranes. Because of the hydrophobic environment, the amino-acid alphabet used by the intramembrane regions of proteins is restricted. The space of possible topologies is also limited, and the energy terms that are most important for folding and recognition in membranes are easier to model than those that are critical for soluble proteins. DeGrado and co-workers [8] have recently seized on these advantages to design the first peptide sequences that bind specifically to transmembrane helices. They designed three CHAMP peptides (computed helical anti-membrane proteins) that bind to the cell adhesion molecules integrin α_{IIb} or integrin α_v *in vitro*, as well as in mammalian cells. This success supports the idea that membrane proteins are particularly good targets for computational design, and suggests a bright future in which biophysical principles, captured in efficient design algorithms, will provide new opportunities to probe the biology of membrane proteins.

Challenges and successes in computational design

A series of remarkable results from the computational protein-design field over the past several years illustrates the power of a good match between problem and method. Although it is not yet possible to apply automated methods to provide any desired function, computational design is well suited to identifying combinations of amino acids that stabilize a specified backbone geometry. Sequences that adopt an impressive range of both native [1,2] and novel

[3,4] folds have been successfully engineered. Introducing function into these folds is more difficult, although Hellinga and co-workers [5] have developed dynamic receptors that recognize small molecules via steric complementarity and appropriate hydrogen bonding using computational methods. A small number of proteins with enzymatic activity have also been designed [6,7].

The very small number of successful design projects that have identified peptides or proteins that bind to native targets illustrates the difficulty of this problem for soluble proteins. Nearly a decade ago, Ghirlanda *et al.* [9] used computational methods to design a hairpin of helices to bind a soluble helix comprising the calmodulin-binding domain of calcineurin, forming a three-helix coiled coil. More recently, Reina *et al.* [10] redesigned a PDZ domain to change its peptide-ligand-binding specificity. And in work redesigning calmodulin, Mayo and colleagues [11] identified variants with greater specificity than wild type. In my laboratory, we have designed novel peptide ligands for the anti-apoptotic protein Bcl-x_L [12].

Part of the difficulty of protein design stems from the vast size of the search spaces. Even short peptides can span an astronomical sequence space (20^N, for a peptide of length N) and can adopt an essentially infinite number of conformations. In general, only a small fraction of possible sequences and structures can be considered computationally, and for soluble proteins this can be very limiting. For membrane proteins, however, restricting the structure and sequence space probably poses a less severe approximation. A growing set of membrane protein structures reveals that α-helical transmembrane regions pack against one another in a limited set of geometries; these geometries can be broken into subsets characterized by the sequence of the protein [13]. Thus, when Yin *et al.* [8] sought a template on which to design peptides to bind to integrin α_{IIb} or integrin α_v, both of which contain a small-X_3-small sequence motif, they were able to consider just 35 appropriate helix-helix pairings taken from structures in the Protein Data Bank. They tested five of these in the design of anti-α_{IIb} peptides and 15 for anti-α_v. Membrane proteins also use a limited amino-acid alphabet compared to soluble proteins, due to the hydrophobic nature of the lipid membrane in which they reside. In the CHAMP designs, most of the residues were selected from a set of just eight amino acids that comprise 75% of membrane-protein residues (Ala, Phe, Gly, Ile, Leu, Ser, Val and Thr). Thus, the search problem for this design application was restricted to sampling sequences, and optimizing side-chain conformations, for combinations of these residues.

The energy problem in protein design is to determine which of many possible sequence-structure combinations is lowest in energy (or has some other desired characteristic). This is typically very daunting. The physics of protein folding and association is determined by a delicate balance of enthalpic and entropic terms, and includes contributions from van der Waals, electrostatic and solvation energies. All of these are difficult to model accurately under the approximations that are typically used in design calculations. Solvation and electrostatic effects are particularly hard to model in an aqueous environment [14]. Yin *et al.* [8] were able to simplify their membrane design problem by making three assumptions. The first was that they did not need to accurately compute interactions between backbone atoms, for example, interhelical C-H••••O=C hydrogen bonds, because they restricted their backbone sampling to a few naturally occurring geometries where these interactions were already built in. Thus, they did not rely on a computational energy function to correctly position the helices with respect to one another. This approach is also common in the design of soluble proteins. Their second assumption was membrane-protein-specific, and posited that a simplified statistical model could be used to capture solvation effects, as a function of depth in the membrane. Finally, they assumed that good packing of the side chains would be sufficient to achieve both affinity and specificity; given the hydrophobic nature of the side chains and their environment, electrostatic interactions were not treated explicitly. This assumption is also more realistic for membrane proteins than for soluble ones. Yin *et al.* [8] used computational analyses guided by these principles and visual inspection to choose final sequences. Remarkably, this strategy succeeded in three out of three attempts.

Specificity without specific design?

The most notable feature of the designed CHAMP peptides is that they are specific for their intended targets. This is true despite the fact that specificity was not explicitly modeled in the design procedure. The designed peptides did interact with themselves, as homodimers, but the anti-α_{IIb} peptide did not bind to integin α_v, and the anti-α_v peptide did not bind to integrin α_{IIb}. This was tested in a bacterial dominant-negative assay and also in a single-molecule assay for the adhesion of platelets to beads coated with fibrinogen (testing for activation of α_{IIb}) or osteopontin (testing for activation of α_v). The specificity is notable, because the sequences of α_{IIb} and α_v are quite similar (both bind integrin β_3), and also because steric patterning is not a reliable strategy for engineering specificity into soluble proteins. Specificity is essential, however, if reagents such as the CHAMP peptides are to be useful for cellular applications. For example, the authors point out that their anti-α_v peptide had to recognize α_v amid large amounts of α_{IIb} on the cell surface in order to be effective.

A critical question going forward will be the extent to which specificity against other classes of transmembrane alpha helices has also been achieved 'for free' using this design procedure. Self-association of the designs suggests that some

improvements in specificity may be necessary for optimal efficacy. However, even if it turns out that additional steps are necessary, such as the explicit consideration of undesired states in the modeling procedure, this work has demonstrated the potential of short designer peptides for providing valuable probes for use in studying membrane protein function. It has also highlighted the good match between computational design and membrane targets, which will no doubt be exploited further in future.

Acknowledgements

I thank Gevorg Grigoryan and James Apgar for thoughtful comments on the manuscript.

References

1. Dahiyat BI, Mayo SL: ***De novo* protein design: fully automated sequence selection.** *Science* 1997, **278:**82-87.
2. Dantas G, Kuhlman B, Callender D, Wong M, Baker D: **A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins.** *J Mol Biol* 2003, **332:**449-460.
3. Harbury PB, Plecs JJ, Tidor B, Alber T, Kim PS: **High-resolution protein design with backbone freedom.** *Science* 1998, **282:** 1462-1467.
4. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D: **Design of a novel globular protein fold with atomic-level accuracy.** *Science* 2003, **302:**1364-1368.
5. Looger LL, Dwyer MA, Smith JJ, Hellinga HW: **Computational design of receptor and sensor proteins with novel functions.** *Nature* 2003, **423:**185-190.
6. Bolon DN, Mayo SL: **Enzyme-like proteins by computational design**. *Proc Natl Acad Sci USA* 2001, **98:**14274-14279.
7. Dwyer MA, Looger LL, Hellinga HW: **Computational design of a biologically active enzyme.** *Science* 2004, **304:**1967-1971.
8. Yin H, Slusky JS, Berger BW, Walters RS, Vilaire G, Litvinov RI, Lear JD, Caputo GA, Bennett JS, DeGrado WF: **Computational design of peptides that target transmembrane helices.** *Science* 2007, **315:**1817-1822.
9. Ghirlanda G, Lear JD, Lombardi A, DeGrado WF: **From synthetic coiled coils to functional proteins: automated design of a receptor for the calmodulin-binding domain of calcineurin.** *J Mol Biol* 1998, **281:**379-391.
10. Reina J, Lacroix E, Hobson SD, Fernandez-Ballester G, Rybin V, Schwab MS, Serrano L, Gonzalez C: **Computer-aided design of a PDZ domain to recognize new target sequences.** *Nat Struct Biol* 2002, **9:**621-627.
11. Shifman JM, Mayo SL: **Exploring the origins of binding specificity through the computational redesign of calmodulin.** *Proc Natl Acad Sci USA* 2003, **100:**13274-13279.
12. Fu X, Apgar JR, Keating AE: **Modeling backbone flexibility to achieve sequence diversity: The design of novel alpha-helical ligands for Bcl-xL.** *J Mol Biol* 2007, in press. doi:10.1016/j.jmb.2007.04.069.
13. Walters RF, DeGrado WF: **Helix-packing motifs in membrane proteins.** *Proc Natl Acad Sci USA* 2006, **103:**13658-13663.
14. Boas FE, Harbury PB: **Potential energy functions for protein design.** *Curr Opin Struct Biol* 2007, **17:**199-204.