

Commentary

Minimized protein structures: A little goes a long way

Maria N. Nedwidek* and Michael H. Hecht*†

Departments of *Molecular Biology and †Chemistry, Princeton University, Princeton, NJ 08544

In this issue of the *Proceedings*, Starovasnik, Braisted, and Wells (1) present the solution structure of synthetic peptides that are half the size of their natural counterpart, but nonetheless maintain well ordered structures and bind tightly to their natural target site. Small protein domains that are structurally stable and functionally active serve as attractive models for studies of protein folding and as starting points for drug design.

Minimized protein domains were selected by Braisted and Wells (2) solely by the domains' ability to bind to the target site, the constant region of G-type immunoglobulins (IgG). The current paper demonstrates that despite their small size, these miniature binding proteins form well ordered structures that closely resemble the binding motif in the larger natural protein. The similarity between the observed structures and the intended target structure validates the protein minimization strategy and demonstrates that proteins can be designed to do more with less.

The sequences studied by Starovasnik, Braisted, and Wells are derived from combinatorial libraries evolved *in vitro* (1, 2). Combinatorial methods recently have revolutionized the discovery of new molecules in the laboratory (3). However, combinatorial approaches have been used by nature for eons to evolve new sequences capable of folding into stable structures and performing essential functions. With time and resources that are virtually unlimited, nature can successfully select folded proteins from combinatorial libraries generated entirely at random (4). Compared with nature, laboratory scientists are at a decided disadvantage: With limited resources and far less patience, we can neither sample sequence diversity as extensively nor wait as long. To succeed in the laboratory, we must depart from nature's random approach; we must incorporate elements of rational design to constrain our synthetic libraries to those regions of sequence space most likely to yield desired molecules. For example, merely constraining a library of *de novo* sequences to conform to the polar/nonpolar periodicity of known secondary structures (5) vastly increases the likelihood of isolating folded, water-soluble proteins (6, 7).

Ultimately, the combinatorial libraries most likely to yield molecules with the intended properties are those that are constrained by the best designs. The libraries designed by Braisted and Wells encoded sequences of 38 residues (2). A hypothetical combinatorial library generated completely at random would contain 20^{38} sequences. This number ($\approx 3 \times 10^{49}$) is so large that synthesis of merely one molecule of each sequence would require reagents with a mass larger than that of the earth. Moreover, within such a random collection, sequences capable of folding into well ordered helical hairpins possessing high binding affinities for particular target sites would be exceedingly rare. To circumvent these issues, Wells and colleagues constrained their combinatorial libraries of 38-mers to smaller and more productive regions of sequence space. By using structure-based design, they vastly increased the probability of finding successful sequences.

However, even within well designed combinatorial libraries, those sequences capable of folding and functioning in the desired manner will be relatively rare. Finding these successful sequences amidst a large excess of failures is a difficult task. This endeavor has been made possible by the development of phage display technology (8–10). By fusing peptide sequences to the coat protein of M13 bacteriophage, libraries of amino acid sequences can be displayed on the surface of bacteriophage, where they can be subjected to screens for desired binding properties. Binding is the “hook” that “fishes” out the selected amino acid sequences. Because each peptide is physically linked to the phage particle, and hence to the gene encoding its unique sequence, the selected sequences can readily be propagated in bacteria and prepared for detailed analysis.

Whereas function (i.e., binding) is the basis of the selection, phage display libraries also can be used to study structural properties. If the side chains responsible for binding are close in space, but distant in the linear sequence—if they form a discontinuous epitope—then formation of a functional binding site necessitates folding into a specific three-dimensional structure. Because phage libraries facilitate exploration of extensive sequence diversity, phage display technology can be used to probe the sequence determinants of such structures (11). Moreover, as demonstrated by the NMR structures reported in this issue of the *Proceedings* (1), appropriately designed phage libraries can be used to guide the discovery of new structures that are both well ordered and stable. The authors have shown us that selecting for high affinity to an appropriate ligand is a powerful method for isolating a desired structure from a sea of unstructured sequences.

The protein minimization strategy is general and will likely be applied to shrink a variety of functional proteins to smaller peptide binding domains. The first protein minimized by Wells and colleagues was atrial natriuretic peptide (ANP), a peptide hormone that regulates blood pressure. Li *et al.* (12) used alanine scanning mutagenesis to demonstrate that residues important for binding of ANP to its extracellular receptor form a discontinuous functional epitope. To minimize ANP, they repositioned a disulfide bond and optimized binding to the target receptor by phage display. Ultimately, they reduced the size of the original ANP by half, while maintaining binding to the ANP receptor (12).

After the successful minimization of ANP, Braisted and Wells (2) focused on reducing the size of the IgG-binding Z domain (13) derived from the B subunit of staphylococcal protein A. The natural domain forms a three-helix bundle. The “front” face of two of these α -helices contains the side chains that form the receptor binding site (13, 14). The starting sequence for protein minimization was a truncated version of the Z domain containing only the sequence of the first two helices. This 38-residue peptide was fused to the M13 gene III coat protein and displayed on the surface of M13 phage. The truncated sequence does not fold and has no detectable binding to IgG. Conversion of this sequence into a well folded and functional mini-protein was accomplished in three successive stages (2). In the first stage, a combinatorial library was designed to adjust the “exoface” (the face that formerly

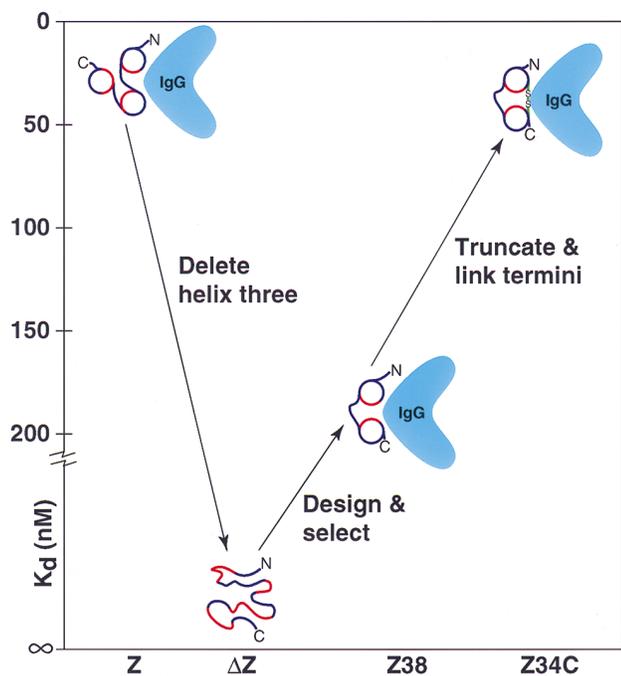


FIG. 1. Protein evolution *in vitro*. Minimization of the three-helix bundle Z domain into a two-helix bundle that maintains equivalent function with half the size. Z, natural three-helix domain. ΔZ , starting 38-residue truncated peptide. Z38, selected two-helix binding domain. Z34C, further truncated 34-residue peptide with engineered disulfide bond. Blue, hydrophilic amino acids; red, hydrophobic amino acids; green, disulfide bond.

contacted the now missing third helix). With the goal of either realigning or replacing the hydrophobic contacts that previously existed with helix three, four positions on the exoface were randomized to all 20 amino acids. Exoface selectants converged to replace two of these hydrophobics with charged residues and maintained the others as hydrophobic to stabilize the peripheral interface between helices one and two. In the second stage of the minimization, the best exoface selectant was used as the starting sequence for adjustment of the “intraface” (the interaction face between the two helices). With the goal of repacking the hydrophobic core and aligning it between two helices, five positions were targeted for random mutagenesis. Three of the five residues converged to non-wild-type residues. In the third stage of the minimization, the best exoface/intraface selectant was used as the starting sequence for the construction of libraries designed to probe for improved binding at the “interface” (the face containing the binding epitope). In this final stage, 19 residues at or near the binding interface with IgG were randomized in groups of four at a time.

The best selectant, called Z38, is α -helical and binds IgG only 12-fold more weakly than the natural Z domain. NMR studies demonstrate that Z38 forms a helical hairpin, but the ends of the peptide are unstructured (see fig. 4 in ref. 1). This mobility may account for the reduced binding affinity of Z38 relative to the natural three-helix bundle. The authors thus tailored the molecule to have a more stable structure and tighter binding to IgG. Because the five N-terminal residues of

the natural domain are unstructured by NMR (15, 16) and unnecessary for binding to IgG (2), these residues were deleted. In addition, a disulfide bond was engineered to connect the two helices by introducing cysteines near the N- and C-termini of the peptide. The resulting disulfide-linked derivative, called Z34C, is more stable and has much greater chemical shift dispersion and a more precisely determined NMR structure (1). This improvement in structural stability is borne out by tighter binding to IgG (see Fig. 1).

The natural Z domain of protein A binds to IgG with a K_d of 15 nM (1). The initial two-helix construct, before redesign, is unstable and has no detectable affinity for IgG. Z38 binds with a K_d of 185 nM, and Z34C binds 9-fold more tightly than Z38, with a K_d of 20 nM, which is virtually the same as the natural affinity of Z for IgG (see Fig. 1). (1). The solution structure of this functional mimic is essentially identical to the crystal structure of the first two helices of a single natural binding domain of protein A complexed with the constant region of IgG (14). Determination of the NMR structure of this two-helix hairpin (1) proves conclusively that binding affinity can be used to successfully select for well folded structures.

Smaller proteins are desirable as models for theoretical studies of protein folding and as a starting point for the design of new drugs (17). However, sequences less than 40 residues are typically incapable of folding into unique structures (18). Starovasnik, Braisted, and Wells’ proof that proteins minimized by evolution *in vitro* can indeed form structures that are both well ordered and capable of function shows us that a little can go a long way.

We gratefully acknowledge Scott R. Harris for assistance in preparing Fig. 1.

1. Starovasnik, M. A., Braisted, A. C. & Wells, J. A. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 10080–10085.
2. Braisted, A. C. & Wells, J. A. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 5688–5692.
3. Ellman, J., Stoddard, B. & Wells, J. A. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 2779–2782.
4. Luria, S. E. & Delbrück, M. (1943) *Genetics* **28**, 491–511.
5. West, M. W. & Hecht, M. H. (1995) *Protein Sci.* **4**, 2032–2039.
6. Kamtekar, S., Schiffer, J. M., Xiong, H., Babik, J. M. & Hecht, M. H. (1993) *Science* **262**, 1680–1685.
7. Roy, S., Ratanaswamy, G., Boice, J. A., Fairman, R., McLendon, G. & Hecht, M. H. (1997) *J. Amer. Chem. Soc.* **119**, 5302–5306.
8. Smith, G. P. (1985) *Science* **228**, 1315–1317.
9. Scott, J. K. & Smith, G. P. (1990) *Science* **249**, 386–390.
10. Lowman, H. B. & Wells, J. A. (1991) *Methods Enzymol.* **3**, 205–216.
11. Gu, H., Yi, Q., Bray, S. T., Riddle, D. S., Shiau, A. K. & Baker, D. (1995) *Protein Sci.* **4**, 1108–1117.
12. Li, B., Tom, J. Y. K., Oare, D., Yen, R., Fairbrother, W. J., Wells, J. A. & Cunningham, B. C. (1995) *Science* **270**, 1657–1660.
13. Nord, K., Nilsson, J., Nilsson, B., Uhlén, M. & Nygren, P.-Å. (1995) *Protein Eng.* **8**, 601–608.
14. Diesenhofer, J. (1981) *Biochemistry* **20**, 2361–2370.
15. Gouda, H., Torigoe, H., Saito, A., Sato, M., Arata, Y. & Shimada, I. (1992) *Biochemistry* **31**, 9665–9672.
16. Tashiro, M. & Montelione, G. T. (1995) *Curr. Opin. Struct. Biol.* **5**, 471–481.
17. Gallup, M. A., Barrett, R. W., Dower, W. J., Fodor, S. P. A. & Gordon, E. M. (1994) *J. Med. Chem.* **37**, 1233–1251.
18. Degrado, W. F. & Sosnick, T. R. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 5680–5681.