Published in final edited form as:

Science. 2016 May 6; 352(6286): 657–658. doi:10.1126/science.aaf7599.

Programming protein shapes and specificities by modular design

Ravit Netzer and Sarel J Fleishman

Department of Biomolecular Science, Weizmann Institute of Science, Rehovot, Israel

Exciting progress was made over the past decade in designing protein folds entirely on the computer. Dozens of computationally designed proteins with geometric properties not observed in nature, and even a novel topology, all showed atomic accuracy in experiment (1–5). Not only were these designer proteins precise, they were also very stable. Although our understanding and control of biomolecular structure are far from complete, it is clear that the high precision and stability observed in these experiments are essential ingredients for realizing the grand challenge of designing enzymes and binders completely from scratch. But natural proteins have an important property that designed ones have so far lacked – they contain non-canonical structural features that predispose them to their molecular function. While computer algorithms optimize surface polarity, core hydrophobicity, and backbone regularity – in a nutshell, they optimize stability – evolution selects proteins for their ability to perform a vital molecular function, often at the expense of stability. In fact, numerous molecular structures of natural proteins tell a similar story: the active site is the protein's most energetically perturbed region, concentrating multiple same-charge chemical groups, buried polar atoms, hydrophobic surfaces exposed to water, or broken and twisted α helices and β sheets. Could design algorithms be extended to encode such features, and if so, would the resulting proteins still show high stability and precision? In back-to-back papers in this issue of *Science* two groups use the Rosetta software suite for biomolecular modeling to program into designed proteins features typically seen in natural ones, and in both cases show that high precision and stability are retained.

On pg. ### Boyken et al. (6) ask how to design buried hydrogen bonds in protein-protein interfaces starting from designed helical coiled coils. In nature, polar interactions encode high-precision binding specificity, but they typically require supporting networks of polar groups on both binding partners for accurate positioning (panel A). Due to this complexity designed binders have so far been devoid of elaborate polar networks (7). To overcome this limitation Boyken et al. designed a large repertoire of coiled coils, including novel topologies, and developed an algorithm called HBNet to search for constellations of residues, where all polar atoms are connected through stabilizing hydrogen-bond networks. The majority of the designed coiled coils showed atomic accuracy, formed only the intended oligomers, and were stable at temperatures as high as 95°C. Furthermore, these networks are reminiscent of the simplicity and elegance of the DNA double-helix, where every base on one strand is paired to a complementary base on the other through buried hydrogen bonds.

Inspired by the double-helix, Boyken et al. designed long coiled coils built from modular parts each with its own constellation of polar side chains (panel B). An exciting prospect is that these modular coiled coils may provide the basis for a new generation of protein-based molecular structures of programmable shape, similar to DNA origami, but unlike DNA these assemblies could be interfaced with proteins of desired function.

On pg. ### Jacobs et al. (8) address a complementary question of how to construct new proteins that exhibit non-canonical features such as those seen in protein active sites, such as cavities, long unstructured regions, and kinked helices (panel C). Since computational modeling favors stable, canonical protein features, Jacobs et al. turn to nature. Their algorithm, called SEWING, starts by generating a ‘parts-list’ – thousands of backbone fragments that are observed in natural proteins. They then define structural rules that determine which pairs of backbone fragments can be joined, and run computer simulations where 3-4 fragments are combined, and subjected to sequence optimization. This strategy of modular design allowed Jacobs et al. to tap into an enormous space of potential backbones ($>10^{16}$), every fragment of which has actually been tested and retained by natural evolution and is therefore inherently stable. The authors experimentally tested 21 SEWING designs with diverse geometries, including some with cavities that could allow small-molecule binding (panel D). Three designs were hyper-stable and molecular structures showed atomic accuracy. Notably, modular design has natural and protein-engineering parallels; indeed, gene recombination is the primary means of diversification in natural protein families and is regularly used by protein engineers (9). The work by Jacobs et al. extends the reach of modular design to fragment combinations from non-homologous proteins, for which genetic recombination is unlikely.

The marvelous molecular activities we see in nature require a balance between stabilizing features that specify molecular structure and functional features, which are often destabilizing. Although the two papers do not demonstrate new molecular activities, the higher level of control over biomolecular shape and interactions brings us a step closer to realizing this goal. Future studies will show how this delicate balance between stabilizing and functional features could be leveraged to design new binding specificities and activities completely on the computer.

References

1. Kuhlman B, et al. Design of a novel globular protein fold with atomic-level accuracy. *Science*. 2003; 302:1364–8. [PubMed: 14631033]
2. Koga N, et al. Principles for designing ideal protein structures. *Nature*. 2012; 491:222–7. [PubMed: 23135467]
3. Huang P-S, et al. High thermodynamic stability of parametrically designed helical bundles. *Science*. 2014; 346:481–5. [PubMed: 25342806]
4. Brunette TJ, et al. Exploring the repeat protein universe through computational protein design. *Nature*. 2015; 528:580–4. [PubMed: 26675729]
5. Doyle L, et al. Rational design of α -helical tandem repeat proteins with closed architectures. *Nature*. 2015; 528:585–8. [PubMed: 26675735]
6. Boyken S, et al. De novo design of protein homo-oligomers with modular hydrogen bond network-mediated specificity. *Science*. 2016

7. Fleishman SJ, et al. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science*. 2011; 332:816–21. [PubMed: 21566186]
8. Jacobs T, et al. Design of structurally distinct proteins using strategies inspired by evolution. *Science*. 2016
9. Khersonsky O, Fleishman SJ. Why reinvent the wheel? Building new proteins based on ready-made parts. *Protein Sci*. 2016; doi: 10.1002/pro.2892

Successful fold designs and future challenges for designing new folds for function.

(A) In nature, protein-protein interactions are often dominated by hydrogen-bonding networks that involve both side-chain and backbone atoms for accurate preorganization (PDB entry: 1BGS; protein partners in light blue and deep teal). (B) Boyken *et al.* designed intricate side-chain mediated hydrogen-bond networks (dashed gray lines) on *de novo* designed coiled coils. (C) A cavity on the surface of influenza hemagglutinin (wheat, PDB entry: 4YY1) is used by the virus to attach to glycosylated receptors on the cell surface

(sugar moiety in sticks). **(D)** Jacobs *et al.* designed novel proteins that contained non-canonical features, such as surface cavities (green surface).