

Computational Design of Anti-flu Proteins

Understanding the physical underpinnings of how proteins interact specifically with one another and not with the myriad other molecules that coexist in every cellular compartment is a major goal of molecular biology. The broad outlines of an answer were suggested by Linus Pauling in the 1940's: the aggregate effect of numerous weak and nonspecific van der Waals, hydrogen-bonding, and electrostatic interactions underlie high specificity and affinity. Since Pauling's days thousands of co-crystal structures have provided concrete examples for how molecular recognition is achieved in different biological contexts. Yet, the ultimate proof for understanding a natural phenomenon lies in recapitulating it; in the words of Thomas Edison, 'until man duplicates a single blade of grass, Nature laughs at his so-called scientific knowledge'. In recent papers, a group of researchers led by David Baker (University of Washington) provided the first high-affinity and high-specificity interactions designed entirely from scratch. X-ray diffraction data collected at the Stanford Synchrotron Radiation Lightsource (SSRL) played a key role in the validation of the two best designs. In the process of developing the computational methods, the team made breakthroughs in understanding the design principles of natural functional sites. These will lead, it is hoped, to more advances in the design of novel functionalities in the future.

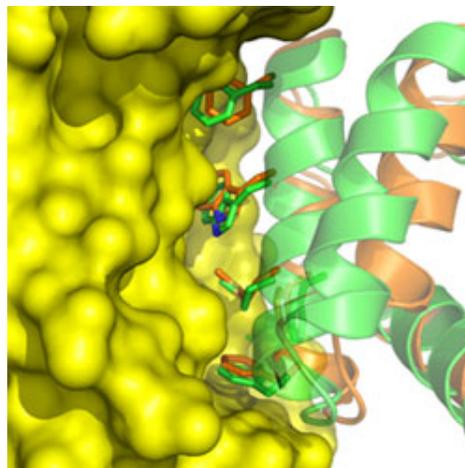


Figure legend: Computationally designed protein bound to the surface of Spanish flu (H1N1) hemagglutinin [1]. The model (green ribbon) and experimentally determined crystal structure (gold) are virtually indistinguishable, providing atomic-level validation of the designed interaction. PDB entry 3r2x.

The team targeted a surface on the influenza hemagglutinin protein that is crucial for flu virus attachment and invasion of cells that line the human respiratory tract[1]. This site is so crucial that it is almost fully conserved among flu strains as remote as the 1918 Spanish and 2009 swine flus (H1N1), Asian flu (H2N2), and bird flu (H5N1), and binding to this site is known to neutralize the virus' ability to infect human cells. The team used massively parallel computing to sift through numerous structural substates of a set of more than 800 natural proteins, the structures of which were solved and deposited in the Protein Databank (PDB). The computations searched for those few structural substates that were predicted to form, as first outlined by Pauling, numerous weak interactions with the target hemagglutinin surface. Binding experiments then isolated two designed proteins, which bound specifically to the target site, and these were further optimized using *in vitro* evolution. The resulting proteins bound Spanish and avian flu hemagglutinin with very high affinity and blocked the replication of H1N1 flu viruses in human cell cultures. Crystal structures of these proteins bound to Spanish flu hemagglutinin were solved by the team's collaborators in Ian Wilson's lab at Scripps, and were virtually indistinguishable from the designed models, providing crucial atomic-level validation for the computational methods. These proteins are now being developed as potential therapeutics against a wide range of pathogenic flu viruses. If successful, this would mark the first example of de novo designed proteins with therapeutic applications. The authors predict that other viruses, which have to date been recalcitrant to small-molecule or antibody therapeutics, may succumb to this new approach to the design of antiviral therapies.

Design of molecular recognition has been a long-standing challenge in molecular biology[2]. To achieve success the authors needed to make progress in both the computational and experimental aspects of molecular recognition. The team compared structures of natural protein-protein complexes solved using x-ray crystallography with a set of computationally designed proteins predicted to bind at high affinity and found that natural binding surfaces tended to be conformationally more rigid than those designed by the computer; ensuring that these binding sites cannot reconfigure to form other conformations that are incompatible with binding the target is therefore an important design principle of natural binding sites[3]. The team developed a computational method that constrains the plasticity of residues at the core of the interaction site. With this insight into molecular recognition, computational design now produced binding surfaces with predicted rigidities that were comparable to those of natural binding surfaces. However, the team found that the designed proteins bound their targets with dissociation constants (K_D) of roughly 10nM, even following standard affinity maturation techniques. They hypothesized that designed proteins, which are unlike natural ones that undergo numerous generations of mutation and selection, could be improved by a combination of substitutions. Many of these individual mutations were predicted to improve binding by a measure too small to detect using standard strategies, but in combination with other mutations they boosted affinity substantially. They therefore used a deep sequencing approach to follow the in vitro selection process of each possible single-point substitution in the designed anti-flu binders[4]. This procedure uncovered 9 substitutions that individually contributed little to binding affinity, but when put together, improved affinity over 100-fold, rivaling and surpassing the affinity of antibodies currently under development as anti-flu therapies.

This research shows that the design of atomically accurate interactions is now feasible, but does it stand up to Edison's exacting demand of 'duplicating a single blade of grass'? In a follow-up analysis the team asked why only a handful of designed proteins bound their target out of a total of more than 80 proteins tested. To ensure that this question is answered as comprehensively as possible they enlisted 28 research teams that specialize in predicting bound protein configurations[5]. In a research publication written by 96 authors, they concluded that a major difference between the design models and natural binding proteins is that unstructured regions (lacking α helices and β strands) in natural proteins are stabilized by other structural elements within their host protein, whereas in design models unstructured regions often lack such stabilization. Indeed, many classes of natural proteins involved in molecular recognition, including enzymes and immune-system antibodies, use unstructured regions in their binding surfaces, yet computational design has not succeeded in designing such functional segments from scratch. This failure of computational design suggests that we still lack clear understanding of the intimate relationships between protein conformation, stability, and function[2]. Further work is therefore needed to shed new light on how unstructured regions in proteins are stabilized. This new understanding will doubtless be followed by an effort to duplicate Nature's abilities, generating new proteins for use in research, technology, and medicine.

Lead authors on this research in David Baker's lab are: Sarel Fleishman (now at the Weizmann Institute of Science, Israel), Timothy Whitehead (now at Michigan State University), and Aaron Chevalier, and in Ian Wilson's lab: Damian Ekiert (now at UCSF) and Cyrille Dreyfus.

Primary Citation

[1] Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. Fleishman SJ, Whitehead TA, Ekiert DC, Dreyfus C, Corn JE, Strauch EM, Wilson IA, Baker D. *Science*. 2011 May 13;332(6031):816-21.

Additional References

[2] Role of the biomolecular energy gap in protein design, structure, and evolution. Fleishman SJ, Baker D. *Cell*. 2012 Apr 13;149(2):262-73.

[3] Restricted sidechain plasticity in the structures of native proteins and complexes. Fleishman SJ, Khare SD, Koga N, Baker D. *Protein Sci*. 2011 Apr;20(4):753-7

[4] Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. Whitehead TA, Chevalier A, Song Y, Dreyfus C, Fleishman SJ, De Mattos C, Myers CA, Kamisetty H, Blair P, Wilson IA, Baker D. *Nat Biotechnol*. 2012 May 27;30(6):543-8

[5] Community-wide assessment of protein-interface modeling suggests improvements to design methodology. Fleishman SJ, Whitehead TA, Strauch EM, Corn JE, Qin S, Zhou HX, Mitchell JC, Demerdash ON, Takeda-Shitaka M, Terashi G, Moal IH, Li X, Bates PA, Zacharias M, Park H, Ko JS, Lee H, Seok C, Bourquard T, Bernauer J, Poupon A, Azé J, Soner S, Ovali SK, Ozbek P, Tal NB, Haliloglu T, Hwang H, Vreven T, Pierce BG, Weng Z, Pérez-Cano L, Pons C, Fernández-Recio J, Jiang F, Yang F, Gong X, Cao L, Xu X, Liu B, Wang P, Li C, Wang C, Robert CH, Guharoy M, Liu S, Huang Y, Li L, Guo D, Chen Y, Xiao Y, London N, Itzhaki Z, Schueler-Furman O, Inbar Y, Potapov V, Cohen M, Schreiber G, Tsuchiya Y, Kanamori E, Standley DM, Nakamura H, Kinoshita K, Driggers CM, Hall RG, Morgan JL, Hsu VL, Zhan J, Yang Y, Zhou Y, Kastiris PL, Bonvin AM, Zhang W, Camacho CJ, Kilambi KP, Sircar A, Gray JJ, Ohue M, Uchikoga N, Matsuzaki Y, Ishida T, Akiyama Y, Khashan R, Bush S, Fouches D, Tropsha A, Esquivel-Rodríguez J, Kihara D, Stranges PB, Jacak R, Kuhlman B, Huang SY, Zou X, Wodak SJ, Janin J, Baker D. *J Mol Biol*. 414:289-302.

SSRL is primarily supported by the DOE Offices of Basic Energy Sciences and Biological and Environmental Research, with additional support from the National Institutes of Health, National Institute of General Medical Sciences and the National Center for Research Resources.