

An exciting but challenging road ahead for computational enzyme design

David Baker*

Department of Biochemistry and Howard Hughes Medical Institute, University of Washington, Seattle, Washington

Received 30 July 2010; Accepted 30 July 2010

DOI: 10.1002/pro.481

Published online 17 August 2010 proteinscience.org

Recent work has demonstrated that computational enzyme design can generate active catalysts.^{1–3} Although the progress is encouraging, the future will be brighter for this new field if its current limitations and the challenges which must be overcome are as broadly understood as its promise. This essay compares the activities of *de novo* designed enzymes to those of naturally occurring enzymes and highlights the considerable challenges which must be overcome for computational design to produce enzymes with levels of activity similar to those of naturally occurring enzymes.

Naturally occurring enzymes are exceptional catalysts. For example, arginine decarboxylase, alkaline phosphatase, and staphylococcal nuclease enhance the rates of the reactions they catalyze by more than 10^{14} fold.⁴ The effective second order rate constants for naturally occurring enzymes are typically within three orders of magnitude of diffusion control. In contrast, most computationally designed enzymes to date provide rate enhancements of less than 10^6 and are more than six orders of magnitude from the diffusion limit. Furthermore, only a small fraction of computational designs have even these very modest levels of catalytic prowess; the majority have no detectable activity.^{1–3} A final caution is that the levels of activities that have been achieved are not much higher than those of catalytic antibodies developed 15–25 years ago.⁵ Clearly, computational

enzyme design has a long way to go to consistently achieve native like levels of catalytic activity.

Why are the activities of computationally designed enzymes and the overall design success rate so low? *De novo* design of enzyme catalysts requires models of ideal active sites that can catalyze the reaction of interest, and design methods that can create stable proteins which contain these sites. A design can fail at three different levels: first, the hypothesis represented by a proposed model of an ideal active site can be incorrect (perfect structural recapitulation in this case would not produce an active catalyst), second, the desired active site geometry may not be structurally realized by the actual design, and third, even with a correct active site description and perfect structural recapitulation, a designed enzyme can have little activity if the surrounding protein context, for example the long range electrostatics or dynamics, is not compatible with catalysis. Determining the reasons for the very low activity of designed enzymes could provide insight into important issues in both enzymology and protein design. Is the low activity because the original conception of the ideal active site was incorrect? Because the designed site was only in part realized? What is the influence of protein context—are protein elements not part of the original ideal active site impeding catalysis? Answering these questions will also provide the basis for iterative improvement of designed catalysts which will likely be critical to achieving high catalytic activity.

Much work needs to be done to understand why current computationally designed enzymes are not

*Correspondence to: David Baker, J557 Health Sciences Building, Box 357350, Seattle 98195, WA.
E-mail: dabaker@u.washington.edu.

better catalysts, and to make possible the robust design of much more active enzymes. To begin with, mechanistic studies of designed enzymes are needed to determine the contributions of designed interactions to catalysis and to compare these contributions to those of analogous catalytic elements in naturally occurring enzymes. Mechanistic studies will also be important to identify the rate limiting steps in designed enzyme catalyzed reactions and to guide the incorporation of additional catalytic elements to help overcome the barriers.

Structural characterization of designed enzymes is essential to determine the extent to which the design process succeeds in producing the target active site geometries, and to provide a starting point for iterative improvement of the designs. Crystal structures have been reported for designed enzymes in the absence of bound ligand and generally show quite good agreement of the protein structure with the design model, but co-crystal structures are necessary to determine the extent to which the designed interactions with the transition state are indeed being made. Such structures will help determine whether the low level of activity is because the desired transition state binding geometry was not fully achieved, or whether the original hypothesis about how to construct an active site was missing key elements. Crystal structures of inactive designs are also important to shed light on why such a large fraction of designs fail; one possibility is that the sequence changes introduced in the design process cause changes in backbone and/or sidechain conformations not accurately modeled in the calculations. Finally, studies of the dynamics of active site residues and loops by NMR⁶ should provide valuable information on conformational changes important for catalysis, for example loop movements to allow substrate entry and product release, as well as the mobility of designed catalytic residues.

Directed evolution⁷ of designed enzymes is needed to identify sequence features missing from the computational designs that confer increased activity. Amino acid substitutions that increase activity should have been incorporated in the original design process, and feedback on these missing elements and on deviations between the designed and actual structure will guide improvement in the computational design process. Directed evolution can also help determine the fraction of nascent enzymes for which there is a direct evolutionary path to more native like levels of activity—an issue of considerable relevance not only to enzyme design but also to theories about the evolution of naturally occurring enzymes.

Improvements in molecular force fields⁸ will also be important. Achieving precise control of active site side chain conformations, loop conformations, and transition state binding orientation requires accurate modeling of the subtle tradeoffs between electrostatic

interactions, interactions with solvent, and entropy loss.^{9,10} For example, the carboxylic acid group that acts as a general base in a subset of the Kemp eliminase designs¹ to function as in the design must lose favorable interactions with water when desolvated by the non polar substrate, and entropy if it has more conformational freedom in the unbound state. If the cost of desolvation or entropy loss is underestimated, the designed general base may instead swing wide of the substrate and not carry out catalysis; these unfavorable contributions must be more than balanced by favorable electrostatic and van der Waals interactions with the rest of the protein to hold the catalytic residue in place. Hence, accurate calculation of the balance between these competing effects is essential. Likewise, improved force fields would help discriminate among alternative transition state binding modes. Molecular simulation methods such as molecular dynamics with explicit solvent can also play an important role by assessing the conformational stability of designed sidechains and loops and the population and orientation of the transition state in the designed active site. Finally, quantum chemistry and QM/MM hybrid methods can play an important role in directly determining the magnitude of transition state energy barriers in the context of the designed active site and ultimately in the context of the designed protein.^{11,12}

Finally, improvements in the algorithms used for computational protein design will also be essential. A likely shortcoming in the current approach is that the computational method which incorporates an ideal active site onto a protein scaffold in the first step of the de novo design process can only handle 3–4 catalytic elements, for example, a general base, a general acid, and a pi stacking residue. The remaining interactions must come from the subsequent sequence optimization step, which is strongly constrained by the backbone of the selected scaffold. Native enzymes frequently have six or more residues at the active site which make important contributions to catalysis. The difference is clear in a comparison of the designed retroaldolases² to the naturally occurring DERA aldolase¹³: the former achieve catalysis primarily through a buried Schiff base forming lysine residue and a non polar substrate binding pocket, whereas the latter has an extensive network of charged residues supporting the catalytic lysine and promoting proton transfer. To overcome this limitation in catalytic site complexity, protein design methodology must be able to remodel the backbone of loops and other structural elements in the vicinity of the active site with a high level of precision to introduce the needed additional catalytic elements. Precise control of designed conformations has been demonstrated previously in the design of novel folds with atomic level accuracy,¹⁴ but this

could be achieved by focusing on core hydrophobic packing interactions which are easier to model than the more polar and more surface exposed interactions involved in catalysis. The improvements in forcefields discussed in the previous paragraph will facilitate precise remodeling of loop conformations. Equally important are conformational and sequence space sampling methodologies which can identify amino acid sequence changes, including insertions and deletions, which stabilize new backbone conformations introducing new catalytic elements relative to all other possible conformations for these regions. Feedback from experimental structural characterization will be important for evaluating and refining such backbone redesign methodologies.

We believe computational enzyme design has tremendous potential for a wide range of important applications and to illuminate fundamental issues in catalysis. To achieve these ends, advances in understanding in all of the areas described above will be important. The ascent from the very low activities of current de novo designed enzymes to the orders of magnitude higher activity levels typical of naturally occurring enzymes will likely require a concerted effort by structural biologists, mechanistic enzymologists, directed evolution practitioners, force field developers, and quantum chemists in addition to protein designers, and we encourage researchers from all of these areas to join on the exciting but challenging road ahead.

References

1. Rothlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, Gallaher JL, Althoff EA, Zanghellini A, Dym O, Albeck S, Houk KN, Tawfik DS, Baker D (2008) Kemp elimination catalysts by computational enzyme design. *Nature* 453:190–195.
2. Jiang L, Althoff EA, Clemente FR, Doyle L, Rothlisberger D, Zanghellini A, Gallaher JL, Betker JL, Tanaka F, Barbas CF, III, Hilvert D, Houk KN, Stoddard BL, Baker D (2008) De novo computational design of retro-aldol enzymes. *Science* 319:1387–1391.
3. Siegel JB, Zanghellini A, Lovick HM, Kiss G, Lambert AR, St Clair JL, Gallaher JL, Hilvert D, Gelb MH, Stoddard BL, Houk KN, Michael FE, Baker D (2010) Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science* 329:309–313.
4. Wolfenden R, Snider MJ (2001) The depth of chemical time and the power of enzymes as catalysts. *Acc Chem Res* 34:938–945.
5. Hilvert D (2000) Critical analysis of antibody catalysis. *Annu Rev Biochem* 69:751–793.
6. Bershtein S, Tawfik DS (2008) Advances in laboratory evolution of enzymes. *Curr Opin Chem Biol* 12:151–158.
7. Mittermaier AK, Kay LE (2009) Observing biological dynamics at atomic resolution using NMR. *Trends Biochem Sci* 34:601–611.
8. Ponder JW, Case DA (2003) Force fields for protein simulations. *Adv Protein Chem* 66:27–85.
9. Chen J, Im W, Brooks CL, III (2006) Balancing solvation and intramolecular interactions: toward a consistent generalized Born force field. *J Am Chem Soc* 128:3728–3736.
10. Hnizdo V, Tan J, Killian BJ, Gilson MK (2008) Efficient calculation of configurational entropy from molecular simulations by combining the mutual-information expansion and nearest-neighbor methods. *J Comput Chem* 29:1605–1614.
11. Smith AJ, Müller R, Toscano MD, Kast P, Hellinga HW, Hilvert D, Houk KN (2008) Structural reorganization and preorganization in enzyme active sites: comparisons of experimental and theoretically ideal active site geometries in the multistep serine esterase reaction cycle. *J Am Chem Soc* 130:15361–15373.
12. Acevedo O, Jorgensen WL (2010) Advances in quantum and molecular mechanical (QM/MM) simulations for organic and enzymatic reactions. *Acc Chem Res* 43:142–151.
13. Heine A, Luz JG, Wong CH, Wilson IA (2004) Analysis of the class I aldolase binding site architecture based on the crystal structure of 2-deoxyribose-5-phosphate aldolase at 0.99Å resolution. *J Mol Biol* 343:1019–1034.
14. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science* 302:1364–1368.