Published in final edited form as:

Curr Opin Chem Biol. 2010 October ; 14(5): 676–682. doi:10.1016/j.cbpa.2010.08.010.

Conformational diversity and computational enzyme design

Jonathan K. Lassila

Department of Biochemistry, Stanford University, Stanford, CA 94305

Abstract

The application of computational protein design methods to the design of enzyme active sites offers potential routes to new catalysts and new reaction specificities. Computational design methods have typically treated the protein backbone as a rigid structure for the sake of computational tractability. However, this fixed-backbone approximation introduces its own special challenges for enzyme design and it contrasts with an emerging picture of natural enzymes as dynamic ensembles with multiple conformations and motions throughout a reaction cycle. This review considers the impact of conformational variation and dynamics on computational enzyme design and it highlights new approaches to addressing protein conformational diversity in enzyme design including recent advances in multistate design, backbone flexibility, and computational library design.

Introduction

The quest to design enzymatic activity tests both our understanding of the physical origins of catalysis and our ability to find creative strategies to confront the enormous size of protein sequence space. The application of computational protein design technology represents a compelling approach because it offers the benefit of surveying astronomically large numbers of possible amino acid sequences *in silico* and because it suggests the possibility of a predictive framework for relating protein sequence to catalytic function.

However, the requirement for reasonable calculation time imposes limitations on the accuracy of prediction. Computational protein design methods have typically employed two key conformational approximations to reduce calculation time: the conformation of the protein backbone is fixed and the amino acid side chains are modeled by a set of discrete rotameric configurations[1-5]. The use of fast energy functions to calculate pairwise interaction energies represents a further tradeoff between accuracy and speed [6-9], but the use of pairwise potentials allows search algorithms to efficiently identify optimal amino acid sequences with respect to these energy functions[10,11]. These procedures have proven successful for designing stable proteins[2-5], and they have been extended for enzyme design through incorporation of geometry-based interaction requirements[12-18].

The conformational limitations of the fixed-backbone and rotamer approximations can lead to unpredictable results from oversensitivity to small geometric changes, a problem that has been reduced for the general protein design problem through empirical adjustments to the energy function[19-21]. However, natural enzymes feature numerous interactions with precise

© 2010 Elsevier Ltd. All rights reserved.

Corresponding author: Lassila, Jonathan K (jklassila@gmail.com).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

geometry and tuned electrostatics, and a balance of conformational sampling and energetic terms chosen for protein stability may not be optimal for stabilization of reaction transition states. While computational enzyme design approaches have used additional constraints to enforce predefined catalytic interaction geometries[12-18], these constraints may contribute to, rather than alleviate, the problems of conformational discretization[14].

This themed issue on conformational diversity and catalytic mechanisms highlights an additional, fundamental limitation of the fixed-backbone approximation in computational enzyme design: evidence continues to emerge that natural enzymes do not behave as single static structures. Kinetic and single-molecule studies have suggested that some enzymes may have multiple long-lived, catalytically-competent conformational states[22-24]. Multiple motions and conformational states may be involved during a reaction cycle to allow substrates to enter, to position substrates and side chains for different reaction steps, and to release products[25,26]. A key element of enzymatic catalysis is the use of binding energy to limit the motions of enzymes and substrates to orientations and conformations where reaction transition states are stabilized[27], and discussion continues over the circumstances and extent to which dynamic motions play additional roles in the catalysis of individual reaction steps[28-30]. Variation of substrate specificity and allostery provide additional important motivations for considering conformational variation[31-33].

This review highlights recent efforts to understand how conformational diversity and dynamics impact computational enzyme design as well as new developments that offer promise for introducing conformational diversity into enzyme design calculations.

Conformation and dynamics of computationally designed proteins

Studies of the structural and dynamic properties of computationally designed proteins have revealed conformational variations—both intended and unintended—that illustrate current issues for control of conformation in protein design.

In cases where x-ray or NMR structures have been obtained for full-sequence computationally designed proteins, there has typically been good overall agreement between design and model, but some variation in side chain placement and local backbone structure[2,34-37]. Dynamic properties of designed proteins have been investigated with NMR methods. Two studies concluded that wild-type and designed proteins displayed similar side-chain and backbone dynamics[35,38], while two other studies attributed differences to imperfect core packing and rotameric strain[39,40]. A recent NMR study of a designed protein G variant concluded that backbone motions on the μ s-ms timescale were increased in the designed sequence relative to the natural sequence as a result of overpacking of the hydrophobic core[41].

Large-scale changes in conformation of computationally designed proteins have also been observed. Design methods were used to produce a single protein sequence that could switch between a zinc-binding motif and a helical bundle with changes in pH and Zn^{2+} concentration [42]. This example represents a successful use of simultaneous design for multiple conformational states (see *Multi-state computational design*) on the scale of the full three-dimensional fold.

In another case, unintended large-scale switching in three-dimensional fold may also have been observed. In this work, two similar thermostable variations of a full-sequence protein redesign were tested experimentally and the solution NMR structure of one showed the intended fold [37]. The other variant was crystallized and surprisingly showed a different fold, a tetrameric four-helix bundle[43]. The crystallization conditions contained 25% dioxane, which was found to increase both helicity and molecular weight[43]. While it remains untested whether a full switch of three-dimensional fold occurred as a result of crystallization solvent, this example

illustrates a key complexity of protein design energy functions: they must optimize for fold *specificity* in addition to thermal stability. As design efforts move toward functional interactions including enzymatic activity, additional energetic and conformational complexities arise.

Conformation and dynamics in *de novo* designed enzymes

Efforts to use computational design to produce new catalytic sites in previously inactive protein scaffolds have yielded a few successes. Ester hydrolysis, Kemp elimination, and retroaldol cleavage reactions catalyzed by computationally designed active sites have been reported[13, 16,17]. Most recently, catalysis of a Diels-Alder reaction has been reported[18]. These enzymes show the potential of computational design, but they also illustrate how much remains unknown. Second-order rate constants (k_{cat}/K_M) for these enzymes range from about 0.01 to 100 M⁻¹s⁻¹, much lower than those of natural enzymes (10⁵ to 10⁹ M⁻¹s⁻¹) (Figure 1).

Four recent studies have investigated these designed enzymes experimentally[44,45] and computationally[46,47]. For the Kemp elimination designs, directed evolution was used to further improve activity [17,44] (Figure 1), and crystallographic structures were determined for improved variants [44]. These unliganded structures showed changes to the network of interactions that surround the designed catalytic residues as well as some rearrangement of an active site loop relative to the original design, suggesting possible structural origins for the overall 200-fold improvement in k_{cat}/K_M . A QM/MM computational analysis of the Kemp eliminase design models also suggested possible structural explanations for differences in reactivity among the designs[46]. Both studies of the Kemp eliminases identified that a key area for improvement lies in optimizing the network of side chains that surround and properly position the designed catalytic base.

A computational study performed molecular dynamics simulations of a designed retroaldolase and concluded that the dynamic motions of the enzyme were insufficiently constrained to optimally promote the intended reactive geometries of the proposed catalytic residues and bound substrate[47].

An experimental mechanistic study of the designed retroaldolases compared the designs to related small molecule catalysts in solution[45]. This work suggested that most of the rate acceleration by the best designed enzyme could be accounted for through a combination of shifting the lysine pK_a and hydrophobic binding of the substrate[45]. Specific side-chain contacts intended to stabilize a bound water molecule were found not to contribute to catalysis, and in accord with the molecular dynamics study, investigation of regio- and enantio-selectivity suggested a lack of precision and specificity in binding interactions[45].

These analyses suggest that a major current challenge for computational enzyme design is in *limiting* conformational diversity of the enzyme-substrate complex. In other words, achieving specificity and precision in binding small-molecule ligands is a preliminary step to design of efficient enzymatic activity that has not yet been established. While an earlier work raised high expectations for computational design of new ligand binding sites, the results from that study have been called into question[48,49]. Rigorous structural and biochemical evaluation of the capabilities of computational design methods to produce specific new ligand-binding sites is needed.

The challenge of designing specific, precise ligand-binding sites is likely to require that design calculations improve upon the fixed-backbone methods used in enzyme designs to date. New methods that seek to introduce conformational diversity into design calculations have great potential to improve the success and reliability of active site and ligand-binding site design.

Multi-state computational design

Multi-state computational design methods allow selection of a single amino acid sequence that simultaneously satisfies multiple design objectives. A typical application of multi-state design has been the variation of interaction specificity through a combination of positive design for the desired state and negative design against an undesired state. A few examples include computational alteration of specificity for coiled-coil dimers[50], protein-ligand binding interactions[51], and protein-protein interfaces[52]. Multi-state computational design methods are well suited to designing substrate specificity, stereospecificity, or even catalytic specificity through the combination of positive and negative design.

New directions for treating conformational diversity have been opened up by the recent development of search algorithms for efficient optimization with respect to large numbers of objective states (Figure 2). A multi-state design algorithm based on the efficient FASTER method[11,53] was tested in experiments that designed sequences with respect to ensembles of 128 different target backbone templates[54]. In this work, the use of multi-state design in conjunction with a template backbone ensemble constructed from constrained molecular dynamics simulations yielded improved results in a large-scale experimental test of protein stabilities. In a different approach, a cluster-expansion method[55] was used for simultaneous design with many states to introduce backbone conformational diversity[56] and to perform negative design for specificity[57], although *de novo* active site design may be a more difficult application of this approach. By allowing simultaneous design for large ensembles of backbone conformations, these new methods offer promising routes to relieving problems of the fixed-backbone approximation and they may enable new strategies, such as designing sequences consistent with moveable active-site loops for substrate entry.

Multi-state design methods offer additional potential for addressing conformational variation in enzyme design through improved modeling of multistep reaction mechanisms. To date, reported successes in *de novo* computational enzyme design have used only single-state design with composite structural motifs representing all transition states and ground states during multistep reaction pathways[13,16-18] (Figure 3). In the future, the use of multistate design to simultaneously model transition states for key individual reaction steps may allow for greater precision in defining specific interactions. A multi-state approach with negative design could offer additional benefit for relieving tight product inhibition as has been observed in computationally designed retroaldolase enzymes[16,45].

Incorporation of backbone flexibility

Flexible-backbone methods have also been developed to introduce conformational complexity into protein design calculations. These include the use of iterative cycles of backbone relaxation and side-chain design (Figure 2) as well as generation of local backbone perturbations. These approaches can add improved reliability to modeling and they allow utilization of backbone conformation for enhanced binding specificity[58,59]. Computational *de novo* active site design examples to date have not made use of flexible-backbone methods[13,16-18], but flexibility has been included in efforts to vary substrate specificity in existing natural enzymes. Flexible-backbone methods have recently been reviewed[58,59], so only the most recent applications to enzyme specificity are highlighted here.

Computational design methods were used to alter the substrate specificity of gramicidin S synthetase A phenylalanine adenylation domain, yielding a 2200-fold increase in specificity for the target Leu substrate[60]. About 5-fold of this specificity increase arose from steps to vary selected residues beyond the active site in the context of backbone flexibility. In another approach, an iterative process including backbone perturbation was used to identify mutations

to three amino acid positions of a xylose reductase to increase specificity for NADH over NADPH by 10^4 -fold[61].

Two new approaches were described for handling backbone loop remodeling in the context of the types of geometrical restrictions needed for binding and catalytic interactions. One approach defines a library of specific, favorable ligand-sidechain interaction motifs, and then performs cycles of backbone loop variation and motif incorporation to find compatible combinations of backbone loop structures and interaction motifs[62]. A related approach defines required interaction geometries between sidechains and ligands[63]. A series of loop-building steps are performed to find backbone conformations compatible with these contacting side chains, and then fixed-backbone design is used to populate the loop with side chains. This latter procedure was used to produce a change in specificity of 10^6 -fold in a guanine deaminase for ammeline over guanine[63].

Multi-state design and flexible-backbone methods offer complementary routes for incorporation of conformational diversity into enzyme design calculations (Figure 2). These methods show potential for improved active site design through better sampling of conformational space, and they allow new strategies for increased binding specificity. In the future, these methods might enable more dramatic approaches to enzyme design, including consideration of protein motions, explicit design for catalytic promiscuity, or designed catalysis of multiple individual reaction steps.

Computational library design: Reducing the impact of limitations

Given the current limitations imposed by both limited conformational sampling and energy functions, a reasonable question is the extent to which present design methods can quantitatively predict energetic changes. This question was recently highlighted in a comparison of experimental and calculated free energy changes for folding upon mutation ($\Delta\Delta G$) with several energy functions[64]. Comparing the effects of 2156 single mutations, the study found that the general trend of the data could be predicted, but correlation between experimental and calculated energies with design energy functions was poor, with correlation coefficients as low as 0.26. While there are necessary differences between energy functions for protein design and for energetic calculations[6-9,19-21], the reality is that there is little expectation of strong quantitative correlation between design energies and catalytic activities with current enzyme design methods.

Acknowledging these limitations in quantitative predictive power, design groups have developed procedures to use the sequence and energetic information from protein design to produce libraries of designed sequences for screening and selection[54,65-69]. This approach takes the view that computational design can efficiently *enrich* sequence space for functional variants even when quantitative energetic prediction is limited. In recent work, a computational library design algorithm was described that optimizes the extent to which the library reflects the list of energetically ranked sequences, allowing a more direct linkage between calculated energies and the output library[54]. The procedure was used in a large-scale test of stabilities of proteins designed using multi-state methods.

Given the special challenges of conformational sampling and energetic modeling for active site design, computational library design may prove especially valuable in future computational enzyme design efforts.

Conclusions

Many crucial questions remain unanswered in computational protein design, including the optimal balances of computational speed and accuracy in energetics and conformational

sampling. Recent successes in *de novo* design of enzymatic activity show the potential of computational enzyme design methods, but these successes do not indicate that the methods have reached a high level of quantitative predictive capability for enzymatic activity. Investigations of computationally designed proteins and enzymes suggest that a critical challenge is in imparting specificity and precision in binding interactions with substrate, as well as in the network of interactions that position catalytic side chains.

New methods that incorporate conformational diversity are likely to help in achieving this goal, first, by relieving problems associated with fixed-backbone protein design, and second, by allowing more precise definition of interactions with transition states relative to products and substrates. Negative design and backbone flexibility can allow better design for substrate specificity or promiscuity. Methods for computational design of targeted protein libraries can help to reduce the impact of existing limitations. Along with energy function improvements, these new developments are important steps toward the long-term goal of reliable enzyme design.

Acknowledgments

Funding was provided by a National Institutes of Health postdoctoral fellowship (F32 GM080865).

References and recommended reading

1. Ponder JW, Richards FM. Tertiary templates for proteins: Use of packing criteria in the enumeration of allowed sequences for different structural classes. *Journal of Molecular Biology* 1987;193:775–791. [PubMed: 2441069]
2. Dahiyat BI, Mayo SL. De Novo Protein Design: Fully Automated Sequence Selection. *Science* 1997;278:82–87. [PubMed: 9311930]
3. Pokala N, Handel TM. Protein Design-- Where We Were, Where We Are, Where We're Going. *Journal of Structural Biology* 2001;134:269–281. [PubMed: 11551185]
4. Lippow SM, Tidor B. Progress in computational protein design. *Current Opinion in Biotechnology* 2007;18:305–311. [PubMed: 17644370]
5. Suarez M, Jaramillo A. Challenges in the computational design of proteins. *Journal of the Royal Society Interface* 2009;6:S477–S491.
6. Gordon DB, Marshall SA, Mayo SL. Energy Functions for Protein Design. *Current Opinion in Structural Biology* 1999;9:509–513. [PubMed: 10449371]
7. Mendes J, Guerois R, Serrano L. Energy estimation in protein design. *Current Opinion in Structural Biology* 2002;12:441–446. [PubMed: 12163065]
8. Vizcarra CL, Mayo SL. Electrostatics in computational protein design. *Current Opinion in Chemical Biology* 2005;9:622–626. [PubMed: 16257567]
9. Boas FE, Harbury PB. Potential energy functions for protein design. *Current Opinion in Structural Biology* 2007;17:199–204. [PubMed: 17387014]
10. Voigt CA, Gordon DB, Mayo SL. Trading accuracy for speed: A quantitative comparison of search algorithms in protein sequence design. *Journal of Molecular Biology* 2000;299:789–803. [PubMed: 10835284]
11. Desmet J, Spriet J, Lasters I. Fast and Accurate Side-Chain Topology and Energy Refinement (FASTER) as a New Method for Protein Structure Optimization. *Proteins: Structure, Function, and Genetics* 2002;48:31–43.
12. Hellinga HW, Richards FM. Construction of new ligand binding sites in proteins of known structure. *Journal of Molecular Biology* 1991;222:763–785. [PubMed: 1749000]
13. Bolon DN, Mayo SL. Enzyme-like proteins by computational design. *Proceedings of the National Academy of Sciences of the USA* 2001;98:14274–14279. [PubMed: 11724958]
14. Lassila JK, Privett HK, Allen BD, Mayo SL. Combinatorial methods for small-molecule placement in computational enzyme design. *Proceedings of the National Academy of Sciences of the USA* 2006;103:16710–16715. [PubMed: 17075051]

15. Zanghellini A, Jiang L, Wollacott AM, Cheng G, Meiler J, Althoff EA, Rothlisberger D, Baker D. New algorithms and an in silico benchmark for computational enzyme design. *Protein Science* 2006;15:2785–2794. [PubMed: 17132862]
- **16. Jiang L, Althoff EA, Clemente FR, Doyle L, Rothlisberger D, Zanghellini A, Gallaher JL, Betker JL, Tanaka F, Barbas CF 3rd, et al. De novo computational design of retro-aldol enzymes. *Science* 2008;319:1387–1391. [PubMed: 18323453] [Using fixed-backbone design with geometry-based interaction requirements, new active sites were constructed for catalysis of a retroaldol cleavage reaction through imine catalysis with the primary amine of the lysine side chain]
- **17. Rothlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, Gallaher JL, Althoff EA, Zanghellini A, Dym O, et al. Kemp elimination catalysts by computational enzyme design. *Nature* 2008;453:190–195. [PubMed: 18354394] [Catalysts for Kemp elimination were constructed using fixed-backbone design and geometry-based requirements specifying a general base, a hydrogen-bond donor, and aromatic stacking]
- **18. Siegel JB, Zanghellini A, Lovick HM, Kiss G, Lambert AR, StClair JL, Gallaher JL, Hilvert D, Gelb MH, Stoddard B, Houk KN, Michael FE, Baker D. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science* 2010;329:309–313. [PubMed: 20647463] [Active sites for catalysis of a Diels-Alder cycloaddition were designed with the fixed-backbone approach using geometric requirements for a hydrogen bond acceptor to the diene and a donor to the dienophile]
19. Dahiyat BI, Mayo SL. Probing the role of packing specificity in protein design. *Proceedings of the National Academy of Sciences of the United States of America* 1997;94:10172–10177. [PubMed: 9294182]
20. Kuhlman B, Baker D. Native protein sequences are close to optimal for their structures. *Proceedings of the National Academy of Sciences of the United States of America* 2000;97:10383–10388. [PubMed: 10984534]
21. Grigoryan G, Ochoa A, Keating AE. Computing van der Waals energies in the context of the rotamer approximation. *Proteins* 2007;68:863–878. [PubMed: 17554777]
22. Schmid FX, Blaschek H. A native-like intermediate on the ribonuclease A folding pathway. *European Journal of Biochemistry* 1981;114:111–117. [PubMed: 6260489]
23. Lu HP, Xun L, Xie S. Single-molecule enzymatic dynamics. *Science* 1998;282:1877–1882. [PubMed: 9836635]
24. Flomenbom O, Velonia K, Loos D, Masuo S, Cotlet M, Engelborghs Y, Hofkens J, Rowan AE, Nolte RJM, Van Der Auweraer M, de Schryver FC, Klafter J. Stretched exponential decay and correlations in the catalytic activity of fluctuating single lipase molecules. *Proceedings of the National Academy of Sciences of the United States of America* 2005;102:2368–2372. [PubMed: 15695587]
25. Hammes-Schiffer S, Benkovic SJ. Relating protein motion to catalysis. *Annual Review of Biochemistry* 2006;75:519–541.
26. Boehr DD, Dyson HJ, Wright PE. An NMR perspective on enzyme dynamics. *Chemical Reviews* 2006;106:3055–3079. [PubMed: 16895318]
27. Jencks WP. Binding Energy, Specificity, and Enzymic Catalysis: Circe Effect. *Advances in Enzymology* 1975;43:219–410.
28. Benkovic SJ, Hammes-Schiffer S. A perspective on enzyme catalysis. *Science* 2003;301:1196–1202. [PubMed: 12947189]
29. Garcia-Viloca M, Gao J, Karplus M, Truhlar DG. How enzymes work: Analysis by modern rate theory and computer simulations. *Science* 2004;303:186–195. [PubMed: 14716003]
30. Klinman JP. An integrated model for enzyme catalysis emerges from studies on hydrogen tunneling. *Chemical Physics Letters* 2009;471:179–193. [PubMed: 20354595]
31. Hedstrom L, Szilagyi L, Rutter WJ. Converting trypsin to chymotrypsin: the role of surface loops. *Science* 1992;255:1249–1253. [PubMed: 1546324]
32. Ma B, Shatsky M, Wolfson HJ, Nussinov R. Multiple diverse ligands binding at a single protein site: A matter of pre-existing populations. *Protein Science* 2002;11:184–197. [PubMed: 11790828]
33. James LC, Tawfik DS. Conformational diversity and protein evolution - a 60-year-old hypothesis revisited. *Trends in Biochemical Sciences* 2003;28:361–368. [PubMed: 12878003]

34. 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. [PubMed: 14631033]
35. Dobson N, Dantas G, Baker D, Varani G. High-resolution structural validation of the computational redesign of human U1A protein. *Structure* 2006;14:847–856. [PubMed: 16698546]
36. Dantas G, Corrent C, Reichow SL, Havranek JJ, Eletr ZM, Isern NG, Kuhlman B, Varani G, Merritt EA, Baker D. High-resolution structural and thermodynamic analysis of extreme stabilization of human procarboxypeptidase by computational protein design. *Journal of Molecular Biology* 2007;366:1209–1221. [PubMed: 17196978]
37. Shah PS, Hom GK, Ross SA, Lassila JK, Crowhurst KA, Mayo SL. Full-sequence computational design and solution structure of a thermostable protein variant. *Journal of Molecular Biology* 2007;372:1–6. [PubMed: 17628593]
38. Johnson EC, Lazar GA, Desjarlais JR, Handel TM. Solution structure and dynamics of a designed hydrophobic core variant of ubiquitin. *Structure* 1999;7:967–976. [PubMed: 10467150]
39. Johnson EC, Handel TM. Effect of hydrophobic core packing on sidechain dynamics. *Journal of Biomolecular NMR* 1999;15:135–143. [PubMed: 20872109]
40. Walsh STR, Lee AL, DeGrado WF, Wand AJ. Dynamics of a de novo designed three-helix bundle protein studied by ¹⁵N, ¹³C, and ²H NMR relaxation methods. *Biochemistry* 2001;40:9560–9569. [PubMed: 11583155]
41. Crowhurst KA, Mayo SL. NMR-detected conformational exchange observed in a computationally designed variant of protein GB1. *Protein Engineering, Design, and Selection* 2008;21:577–587.
42. Ambroggio XI, Kuhlman B. Computational design of a single amino acid sequence that can switch between two distinct folds. *J. Am. Chem. Soc* 2006;128:1154–1161. [PubMed: 16433531]
43. Hom GK, Lassila JK, Thomas LM, Mayo SL. Dioxane contributes to the altered conformation and oligomerization state of a designed engrailed homeodomain variant. *Protein Science* 2005;14:1115–1119. [PubMed: 15741348]
- *44. Khersonsky O, Rothlisberger D, Dym O, Albeck S, Jackson CJ, Baker D, Tawfik DS. Evolutionary optimization of computationally designed enzymes: Kemp eliminases of the KE07 series. *J. Mol. Biol* 2010;396:1025–1042. [PubMed: 20036254] [Directed evolution was used to improve designed Kemp elimination enzymes and crystallographic structures were determined for improved variants]
- **45. Lassila JK, Baker D, Herschlag D. Origins of catalysis in computationally designed retroaldolase enzymes. *Proceedings of the National Academy of Sciences of the United States of America* 2010;107:4937–4942. [PubMed: 20194782] [Much of the rate acceleration of the most active retroaldolase from reference ¹⁶ was estimated to arise from shifting the pK_a of the imine-forming lysine and binding in a hydrophobic pocket. Designed water interactions were found not to be catalytic, and evidence for enantioselectivity or regioselectivity was not found, suggesting limited specificity and precision in binding]
46. Alexandrova AN, Rothlisberger D, Baker D, Jorgensen WL. Catalytic mechanism and performance of computationally designed enzymes for Kemp elimination. *J. Am. Chem. Soc* 2008;130:15907–15915. [PubMed: 18975945]
47. Ruscio JZ, Kohn JE, Ball KA, Head-Gordon T. The influence of protein dynamics on the success of computational enzyme design. *Journal of the American Chemical Society* 2009;131:14111–14115. [PubMed: 19788332]
48. Schreier B, Stump C, Wiesner S, Hocker B. Computational design of ligand binding is not a solved problem. *Proceedings of the National Academy of Sciences of the United States of America* 2009;106:18491–18496. [PubMed: 19833875]
49. Hayden EC. Key protein-design papers challenged. *Nature* 2009;461:859. [PubMed: 19829341]
50. Havranek JJ, Harbury PB. Automated design of specificity in molecular recognition. *Nature Structural Biology* 2003;10:45–52.
51. Boas FE, Harbury PB. Design of protein-ligand binding based on the molecular-mechanics energy model. *Journal of Molecular Biology* 2008;380:415–424. [PubMed: 18514737]
52. Mandell DJ, Kortemme T. Computer-aided design of functional protein interactions. *Nature Chemical Biology* 2009;5:797–807.
- **53. Allen BD, Mayo SL. An efficient algorithm for multistate protein design based on FASTER. *Journal of Computational Chemistry* 2010;31:904–916. [PubMed: 19637210] [A method for multistate

protein design is described that allows simultaneous design against large numbers of objective states such as different protein backbone conformations. Clear implementation details and computational test results are presented]

- *54. Allen BD, Nisthal A, Mayo SL. Computational protein design of structural ensembles: Validation by automated screening of combinatorial libraries. *Proceedings of the National Academy of Sciences of the United States of America*. in press. [The multi-state design method from reference ⁵³ was used to design against 128 backbone conformations simultaneously in tests of different types of backbone structural ensembles. Protein stabilities were measured using a high-throughput approach, and a new method was described for optimizing the correlation between computationally designed protein libraries and calculated energies]
- 55. Grigoryan G, Zhou F, Lustig SR, Ceder G, Morgan D, Keating AE. Ultra-fast evaluation of protein energies directly from sequence. *PLoS Computational Biology* 2006;2:551–563.
- 56. Apgar JR, Hahn S, Grigoryan G, Keating AE. Cluster expansion models for flexible-backbone protein energetics. *Journal of Computational Chemistry* 2009;30:2402–2413. [PubMed: 19360809]
- 57. Grigoryan G, Reinke AW, Keating AE. Design of protein-interaction specificity gives selective bZIP-binding peptides. *Nature* 2009;458:859–864. [PubMed: 19370028]
- *58. Mandell DJ, Kortemme T. Backbone flexibility in computational protein design. *Current Opinion in Biotechnology* 2009;20:420–428. [PubMed: 19709874] [A thorough review of methods for inclusion of backbone flexibility in computational protein design]
- 59. Butterfoss GL, Kuhlman B. Computer-based design of novel protein structures. *Annual Review of Biophysics and Biomolecular Structure* 2006;35:49–65.
- 60. Chen CY, Georgiev I, Anderson AC, Donald BR. Computational structure-based redesign of enzyme activity. *Proceedings of the National Academy of Sciences of the United States of America* 2009;106:3764–3769. [PubMed: 19228942]
- 61. Khoury GA, Fazelina H, Chin JW, Pantazes RJ, Cirino PC, Maranas CD. Computational design of *Candida boidinii* xylose reductase for altered cofactor specificity. *Protein Science* 2009;18:2125–2138. [PubMed: 19693930]
- *62. Havranek JJ, Baker D. Motif-directed flexible backbone design of functional interactions. *Protein Science* 2009;18:1293–1305. [PubMed: 19472357] [A method was described for iteratively combining backbone movement with the design of specific interactions between side chains and DNA. The underlying principle of substituting from a large library of possible side chain-ligand interaction types is applicable to the problem of active site design]
- *63. Murphy PM, Bolduc JM, Gallaher JL, Stoddard BL, Baker D. Alteration of enzyme specificity by computational loop remodeling and design. *Proceedings of the National Academy of Sciences of the United States of America* 2009;106:9215–9220. [PubMed: 19470646] [Substrate specificity of human guanine deaminase was altered through the use of a flexible-backbone procedure. The new substrate was modeled with a desired new side-chain contact, and a backbone segment of 2-5 residues was remodeled to accommodate the new side chain]
- **64. Popatov V, Cohen M, Schreiber G. Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. *Protein Engineering, Design, and Selection* 2009;22:553–560. [Several energy functions for protein design and modeling were tested for their ability to reflect changes in stability of 2156 single mutations. Poor correlations were observed between calculated and experimentally determined $\Delta\Delta G$ values, providing a reminder that successful application of energy functions to protein design need not indicate quantitative energetic modeling of physical interactions]
- 65. Kono H, Saven JG. Statistical theory for protein combinatorial libraries. Packing interactions, backbone flexibility, and the sequence variability of a main-chain structure. *Journal of Molecular Biology* 2001;306:607–628. [PubMed: 11178917]
- 66. Hayes RJ, Bentzien J, Ary ML, Hwang MY, Jacinto JM, Vielmetter J, Kundu A, Dahiyat BI. Combining computational and experimental screening for rapid optimization of protein properties. *Proceedings of the National Academy of Sciences of the United States of America* 2002;99:15926–15931. [PubMed: 12446841]
- 67. Bolon DN, Voigt CA, Mayo SL. De novo design of biocatalysts. *Current Opinion in Chemical Biology* 2002;6:125–129. [PubMed: 12038994]

68. Mena MA, Daugherty PS. Automated design of degenerate codon libraries. *Protein Engineering, Design, and Selection* 2005;18:559–561.
69. Treynor TP, Vizcarra CL, Nedelcu D, Mayo SL. Computationally designed libraries of fluorescent proteins evaluated by preservation and diversity of function. *Proceedings of the National Academy of Sciences of the United States of America* 2007;104:48–53. [PubMed: 17179210]

Figure 1.

Activities for computationally designed active sites catalyzing ester hydrolysis[13], retroaldol cleavage[16], Kemp elimination[17], and Diels-Alder cycloaddition (a bimolecular reaction; reported pseudo-second order rate constants for diene with apparent saturation of dienophile are shown)[18]. Kemp eliminases were improved through directed evolution[17,44]; the highest resulting activity is indicated with an asterisk.

Figure 2.

Approaches for computational protein design. Multi-state design methods allow design of a single optimized sequence with respect to multiple backbone conformations. Iterative approaches to backbone flexibility alternate between design steps, where side chain identities are allowed to vary, and backbone optimization steps, where conformation is optimized for a given amino acid sequence.

Figure 3.

Example of the use of a single structural motif to represent key ground states and transition states. The multistep reaction pathway is catalyzed by retroaldolase enzymes that were designed using the composite structural motif shown[16].