HHS Public Access

Author manuscript

Methods Mol Biol. Author manuscript; available in PMC 2019 January 01.

Published in final edited form as:

Methods Mol Biol. 2018 ; 1685: 15–23. doi:10.1007/978-1-4939-7366-8_2.

Rational and Semirational Protein Design

Ivan V. Korendovych

Abstract

This mini review gives an overview over different design approaches and methodologies applied in rational and semirational enzyme engineering. The underlying principles for engineering novel activities, enantioselectivity, substrate specificity, stability, and pH optimum are summarized.

Keywords

Rational protein design; Computational enzyme design; De novo enzyme design; Molecular dynamics; Molecular docking; Enantioselectivity; Substrate specificity; Thermostability; pH optimum

---

The ability to produce desired molecules in a direct, inexpensive and efficient fashion is the ultimate goal of applied chemistry. Despite the abundance of easy and inexpensive sources of energy (e.g., heat, electricity, and light) the complex task of taking available chemical building blocks to drive thermodynamically allowed processes in one particular direction is far from solved. Nature has found many ways to accomplish this task through enzymatic catalysis, promoted by proteins and nucleic acids. Thus, it is hardly surprising that ever since the discovery of the first enzyme chemists attempt to replicate their amazing efficiency by creating proteins capable of producing chemicals of industrial relevance. Many different approaches have been explored with various degrees of success (Table 1). Existing catalysts were repurposed to change the substrate scope and reactions specificity. Proteins that have no enzymatic function adopted new catalytic functions. Catalysts have been prepared from protein scaffolds not present in nature and proteins that have no observable enzymatic activity for the reaction of interest—this I refer to as de novo design. Finally, catalysts for reactions that were not observed in nature until now could be created in protein scaffolds by mutagenesis: novel activities were designed by a careful placement of chemical functionalities that are provided by nature's menu of amino acids to stabilize transition states, enable proton transfers, facilitate the interaction of the substrate with the active site or with cofactors present in the protein, or modulate the chemical reactivity of natural cofactors. The spectrum of catalysis was further extended by introducing artificial cofactors or unnatural amino acids [28]. Table 1 gives examples for this large spectrum of design approaches.

Design tools have been very diverse: ranging from purely combinatorial [29] to highly rational [30]. Combinatorial methods (relying on random mutagenesis) have been successful in repurposing of existing proteins to adopt new functions and creation of new catalytic function from random sequences [29, 31, 32]. However, the enormity of sequence space to be explored in a design problem means that in practical terms some degree of rational input has to be made to limit the search space to a manageable size. Thus, a clear line between

rational and combinatorial approaches is hard, if not impossible, to draw. One crucial requirement in rational design is the necessity to understand the molecular basis of the protein's property that is the subject of the design study (structure–function relationship). Table 1 lists specific rational design techniques and how they are used to modify a well-defined property of an enzyme.

Many application-oriented enzyme-engineering projects focus on creating or adapting the substrate scope of an enzyme to gain access to (a class of) compounds of interest. This often also involves tuning enantioselectivity or regioselectivity in the desired direction. Increasing the stability of the biocatalyst under process conditions is an equally important goal. For all these questions, a rational understanding has become available during the past few decades.

1 Semirational Tools for Engineering Substrate Specificity and Enantioselectivity

Certain features of catalysts can be modified relatively easily: Substrate specificity and enantioselectivity are often governed by steric factors of the active site [33]. Thus, the easiest approach to guide a semirational design is to use structural visualization to identify hot spot residues that are then targeted in a site-saturation mutagenesis experiment. The active site must be shape-complementary to the transition state of the reaction to accelerate formation of the desired product [34]. A well-defined geometry allows the preferred binding and positioning of one enantiomeric form of the substrate, or the preferred creation of one configuration of the chiral product. On the contrary, binding poses that lead to undesired regio or stereo isomers have to be blocked. Additionally, selectivity towards different substrates is affected during their passage of the entrance tunnel of the enzyme: modifications of tunnel residues influence the access of different compounds to the active site and thus induce selectivity [35].

Rational redesign of the active site is often easily possible, e.g., by blocking the productive binding of the undesired enantiomer by introducing a bulky residue. However, as enzymes are often more complex than it is apparent from the structural models, many effects cannot be predicted (due to protein dynamics or effects on protein folding). The more detailed the available information and knowledge of catalysis is, the better. While detailed structural information on the intermediates in the catalytic cycle *can* be obtained, most of X-ray and NMR structures present in the Protein Data Bank represent the structure *without* direct information about how substrate binds or is turned over. Additional studies that require crystallization of the enzyme with an appropriate inhibitor may require a long time without any definitive guarantee of success. Fortunately, several very successful algorithms have been developed to identify the location and possible poses of the substrate in the enzyme [36]. Cavity search and docking techniques give hints how and where the substrate might be bound. Even low-resolution information about how the substrate associates with the protein is often sufficient to make educated guesses in which positions mutagenesis needs to be done to achieve maximum desired effect. Especially when the active site or the substrate is large and can adopt multiple conformations, or when binding is based mainly on hydrophobic interactions, reliable predictions are not yet possible. Partial or complete

randomization of identified hot spots is therefore an efficient approach, which often leads to success. Iterative site saturation mutagenesis has become a very popular engineering tool [37].

The CAVER software is an easy to handle tool for identification and analysis of tunnels and channels in protein structures [38]. CAVER is used as a plugin in Pymol, a program, which is employed frequently for protein visualizing [39]. It predicts the location of “hot spot” residues, which can be mutated to enhance enzyme activity, stability, specificity, and enantioselectivity. Another commonly used program YASARA [39] provides the user with a graphic, user-friendly interface to detect hotspots and to perform molecular mechanics based simulations for rational protein engineering. If no structure is available for the protein of interest, YASARA has a tool for the computer-aided construction of a homology model. Some structural information, although the accuracy of the model might be limited, can be obtained from related proteins with sequence identities as low as 30%. On the other hand, if a reliable structure is available, computational docking–which is also integrated in YASARA–has shown enormous predictive power in identifying residues to be modified in order to alter the selectivity and improve the reactivity of the existing proteins. However, much care has to be taken when interpreting results of docking experiments that rely on homology models.

It is universally accepted that enzymes are far from static and rely on concerted movement of amino acids to achieve function [40]. Semiempirical molecular dynamics, (MD) approaches have been extremely useful in deciphering the intricate details of protein-catalyzed chemical reactions [41]. Owing to the continuous improvement of computational hardware MD techniques are becoming more and more available to solve protein design problems [42]. MD simulations have been useful in improving the enzyme activity and enantioselectivity. This is the most difficult and time consuming aspect of rational design and much needs to be learned before our methods are efficient and accurate enough to reliably predict mutations that are likely to improve enzymatic efficiency.

MD simulations generate an ensemble of possible conformations and conformational transitions, as compared to a static picture provided by X-ray crystallography. Combined with knowledge of the reaction mechanism (e.g., from quantum mechanical modeling), MD simulations determine how frequently geometries that will promote catalysis according to the model are observed, as compared to “unproductive conformations” [43]. MD simulations are also used to identify dynamic, flexible regions of a protein. Changes in these regions can affect protein stability and activity, because catalysis requires certain flexibility of critical residues or parts of the protein. Loop flexibility can also determine reaction specificity, as was demonstrated by reengineering a phenylalanine mutase into a phenylalanine ammonia lyase by introducing a single mutation in a loop near the active site [44].

2 Advanced Computational Engineering for Optimizing Enantioselectivity and Thermostability

Computational engineering creates large virtual libraries of variants in silico. Designs are then evaluated and ranked automatically, e.g., by energy scoring functions or geometric

restraints, and only a few hits (ten to some hundreds) are manually inspected and tested in the lab [30]. Different tools that often introduce several mutations at once are used for the creation of the libraries. Computational enzyme design allows for engineering highly enantioselective catalysts for a particular chemical transformation already catalyzed by the enzyme, complete redesign of active sites to fit substrate structures that are very different from the natural ones, and de novo design of enzymes, i.e., proteins catalyzing nonnatural reactions.

In the first step, optimal geometries of possible active site residues that stabilize the transition state of the reaction are predicted using QM simulations. The resulting arrangements of amino acid residues, called theozymes, are placed in suitable protein scaffolds identified by RosettaMatch. Finally, RosettaDesign optimizes the complete active site pocket to allow the precise positioning of the catalytic residues and the transition state. Designs are then evaluated and ranked in silico, and only a few (ten to some hundreds) are manually inspected and tested in the lab. This strategy was used for de novo engineering Kemp eliminases, retroaldolase, Diels-Alderases, but also to generate highly enantioselective epoxide hydrolases [13–17]. For the latter study, in silico variants were screened using high-throughput multiple independent MD simulations [45], a technique that leads to a more complete sampling of protein conformational space in a shorter time (compared to long single-run MD simulations) and showed an improved correlation between predicted and observed enantioselectivity. This helped to reduce library size that had to be actually screened. Moreover, computational approaches can assist in improving enzymes using directed evolution: semirationally developed libraries produced up to 4–5-fold higher hit rate as compared to a full coverage libraries thus greatly limiting effort to identify productive mutations [46]. While many programs have been developed and successfully used for performing MD simulations, YASARA provides a user-friendly interface for a beginner.

A second very important engineering target is enzyme thermostability. It became clear early on that practical applicability (and evolvability!) of an enzymatic catalyst is related to its stability [47]. Enzymes from thermophilic organisms are commonly used in many different applications, but what if the catalyst to be repurposed/improved has no obvious thermophilic analog? Homology modeling and rational evaluation of the structure has been very productive in identification of mutations to improve stability. This sometimes also leads to improvement in the yield of recombinant expression of soluble enzymes, although the evidence is somewhat anecdotal.

Several approaches have been successfully used to predict and improve thermostability [42, 48–50]. Most often protein stability is increased by rigidification of flexible sites. Analysis of B-factors in crystal structures (B-Fit Method) [5], high temperature unfolding MD simulations, and comparative MD simulations of homologous proteins from mesophilic and thermophilic organisms at different temperatures unravel flexible regions of the protein that are susceptible to unfolding to guide reengineering [42].

Alternatively, protein stability can be increased by improving hydrophobic packing of the protein core [51] and/or creating a favorable network of positive and negative charges at the protein surface [52]. Scoring of variants is then performed by evaluating differences in the

free energy of folding using specifically parameterized energy functions [49, 53] such as one included in the FoldX suite. Finally, stabilizing disulfide bridges can be engineered into the protein using the FRESCO algorithm in YASARA.

In summary, the path to developing the ability to create *functional* proteins for a particular purpose has been long, windy, and full of obstacles. Decades of research in biochemistry, enzymology, and biotechnology produced a number of exciting discoveries that advance our understanding of enzymatic catalysis, nonetheless we still fall short from being able to create a single unique tool that will allow us to create efficient protein catalysts from scratch [54]. Despite the disappointment with the overall progress of the field, fueled in part by the overzealous promises that could not be fulfilled thrown around so easily, many amazing stories of success that apply rational principles to (re)design of proteins have emerged through the years [55–59]. Advances in computation led to an explosive growth of structural information and the development of robust tools for building protein structures of predefined fold. Creating a crucial link between a (re)designed well-defined structure and catalytic function is the next major milestone for the field.

References

1. Ghislieri D, Green AP, Pontini M, et al. Engineering an enantioselective amine oxidase for the synthesis of pharmaceutical building blocks and alkaloid natural products. J Am Chem Soc. 2013; 135:10863–10869. [PubMed: 23808566]
2. Kille S, Zilly FE, Acevedo JP, et al. Regio- and stereoselectivity of P450-catalysed hydroxylation of steroids controlled by laboratory evolution. Nat Chem. 2011; 3:738–743. [PubMed: 21860465]
3. Liskova V, Bednar D, Prudnikova T, et al. Balancing the stability-activity trade-off by fine-tuning dehalogenase access tunnels. ChemCatChem. 2015; 7:648–659.
4. Pavlova M, Klvana M, Prokop Z, et al. Redesigning dehalogenase access tunnels as a strategy for degrading an anthropogenic substrate. Nat Chem Biol. 2009; 5:727–733. [PubMed: 19701186]
5. Reetz MT, Carballeira JD, Vogel A. Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability. Angew Chem Int Ed. 2006; 45:7745–7751.
6. Raza S, Fransson L, Hult K. Enantios-electivity in *Candida antarctica* lipase B: a molecular dynamics study. Protein Sci. 2001; 10:329–338. [PubMed: 11266619]
7. Rotticci D, Rotticci-Mulder JC, Denman S, et al. Improved enantioselectivity of a lipase by rational protein engineering. Chem-BioChem. 2001; 2:766–770.
8. Tynan-Connolly BM, Nielsen JE. Redesigning protein pK(a) values. Protein Sci. 2007; 16:239–249. [PubMed: 17189477]
9. Pokhrel S, Joo JC, Yoo YJ. Shifting the optimum pH of *Bacillus circulans* xylanase towards acidic side by introducing arginine. Biotechnol Bioprocess Eng. 2013; 18:35–42.
10. Pokhrel S, Joo JC, Kim YH, et al. Rational design of a *Bacillus circulans* xylanase by introducing charged residue to shift the pH optimum. Process Biochem. 2012; 47:2487–2493.
11. Xu H, Zhang F, Shang H, et al. Alkalophilic adaptation of XynB endoxylanase from *Aspergillus niger* via rational design of pKa of catalytic residues. J Biosci Bioeng. 2013; 115:618–622. [PubMed: 23290994]
12. Wijma HJ, Floor RJ, Jekel PA, et al. Computationally designed libraries for rapid enzyme stabilization. Protein Eng Des Sel. 2014; 27:49–58. [PubMed: 24402331]
13. Wijma HJ, Floor RJ, Bjelic S, et al. Enantioselective enzymes by computational design and *in silico* screening. Angew Chem Int Ed. 2015; 54:3726–3730.
14. Rothlisberger D, Khersonsky O, Wollacott AM, et al. Kemp elimination catalysts by computational enzyme design. Nature. 2008; 453:190–195. [PubMed: 18354394]

15. Siegel JB, Zanghellini A, Lovick HM, et al. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. Science. 2010; 329:309–313. [PubMed: 20647463]
16. Blomberg R, Kries H, Pinkas DM, et al. Precision is essential for efficient catalysis in an evolved Kemp eliminase. Nature. 2013; 503:418–421. [PubMed: 24132235]
17. Jiang L, Althoff EA, Clemente FR, et al. De novo computational design of retro-aldol enzymes. Science. 2008; 319:1387–1391. [PubMed: 18323453]
18. Korendovych IV, Kulp DW, Wu Y, et al. Design of a switchable eliminase. Proc Natl Acad Sci U S A. 2011; 108:6823–6827. [PubMed: 21482808]
19. Moroz YS, Dunston TT, Makhlynets OV, et al. New tricks for old proteins: single mutations in a nonenzymatic protein give rise to various enzymatic activities. J Am Chem Soc. 2015; 137:14905–14911. [PubMed: 26555770]
20. Raymond EA, Mack KL, Yoon JH, et al. Design of an allosterically regulated retroaldolase. Protein Sci. 2015; 24:561–570. [PubMed: 25516403]
21. Burton AJ, Thomson AR, Dawson WM, et al. Installing hydrolytic activity into a completely de novo protein framework. Nat Chem. 2016; 8:837–844. [PubMed: 27554410]
22. Pan T, Liu Y, Si C, et al. Construction of ATP-switched allosteric antioxidant selenoenzyme. ACS Catalysis. 2017; 7(3):1875–1879.
23. Renfrew PD, Choi EJ, Bonneau R, et al. Incorporation of noncanonical amino acids into Rosetta and use in computational protein-peptide interface design. PLoS One. 2012; 7:e32637. [PubMed: 22431978]
24. Mills JH, Khare SD, Bolduc JM, et al. Computational design of an unnatural amino acid dependent metalloprotein with atomic level accuracy. J Am Chem Soc. 2013; 135:13393–13399. [PubMed: 23924187]
25. Reetz MT, Jiao N. Copper–phthalocyanine conjugates of serum albumins as enantioselective catalysts in Diels–Alder reactions. Angew Chem Int Ed. 2006; 45:2416–2419.
26. Key HM, Dydio P, Clark DS, et al. Abiological catalysis by artificial haem proteins containing noble metals in place of iron. Nature. 2016; 534:534–537. [PubMed: 27296224]
27. Lo C, Ringenberg MR, Gnandt D, et al. Artificial metalloenzymes for olefin metathesis based on the biotin-(strept)avidin technology. Chem Commun. 2011; 47:12065–12067.
28. Bornscheuer UT, Huisman GW, Kazlauskas RJ, et al. Engineering the third wave of biocatalysis. Nature. 2012; 485:185–194. [PubMed: 22575958]
29. Seelig B, Szostak JW. Selection and evolution of enzymes from a partially randomized non-catalytic scaffold. Nature. 2007; 448:828–831. [PubMed: 17700701]
30. Kiss G, Celebi-Olcum N, Moretti R, et al. Computational enzyme design. Angew Chem Int Ed. 2013; 52:5700–5725.
31. Reetz MT. Biocatalysis in organic chemistry and biotechnology: past, present and future. J Am Chem Soc. 2013; 135:12480–12496. [PubMed: 23930719]
32. Reetz MT. Laboratory evolution of stereoselective enzymes: a prolific source of catalysts for asymmetric reactions. Angew Chem Int Ed. 2011; 50:138–174.
33. Reetz MT. Controlling the enantioselectivity of enzymes by directed evolution: practical and theoretical ramifications. Proc Natl Acad Sci U S A. 2004; 101:5716–5722. [PubMed: 15079053]
34. Schramm VL. Enzymatic transition states: thermodynamics, dynamics and analogue design. Arch Biochem Biophys. 2005; 433:13–26. [PubMed: 15581562]
35. Butler CF, Peet C, Mason AE, et al. Key mutations alter the cytochrome P450 BM3 conformational landscape and remove inherent substrate bias. J Biol Chem. 2013; 288:25387–25399. [PubMed: 23828198]
36. Sousa SF, Fernandes PA, Ramos MJ. Protein–ligand docking: current status and future challenges. Proteins Struct Funct Bioinf. 2006; 65:15–26.
37. Gumulya Y, Sanchis J, Reetz MT. Many pathways in laboratory evolution can lead to improved enzymes: how to escape from local minima. ChemBioChem. 2012; 13:1060–1066. [PubMed: 22522601]

38. Chovancova E, Pavelka A, Benes P, et al. CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. PLoS Comput Biol. 2012; 8:e1002708. [PubMed: 23093919]
39. Available from: http://www.pymol.org
40. Eisenmesser EZ, Bosco DA, Akke M, et al. Enzyme dynamics during catalysis. Science. 2002; 295:1520–1523. [PubMed: 11859194]
41. Adcock SA, McCammon JA. Molecular dynamics: survey of methods for simulating the activity of proteins. Chem Rev. 2006; 106:1589–1615. [PubMed: 16683746]
42. Childers MC, Daggett V. Insights from molecular dynamics simulations for computational protein design. Mol Sys Des Eng. 2017; 2:9–33.
43. Wijma HJ, Floor RJ, Janssen DB. Structure- and sequence-analysis inspired engineering of proteins for enhanced thermostability. Curr Opin Struct Biol. 2013; 23:588–594. [PubMed: 23683520]
44. Bartsch S, Wybenga GG, Jansen M, et al. Redesign of a phenylalanine aminomutase into a phenylalanine ammonia lyase. Chem-CatChem. 2013; 5:1797–1802.
45. Wijma HJ, Marrink SJ, Janssen DB. Computationally efficient and accurate enantioselectivity modeling by clusters of molecular dynamics simulations. J Chem Inf Model. 2014; 54:2079–2092. [PubMed: 24916632]
46. Chen MMY, Snow CD, Vizcarra CL, et al. Comparison of random mutagenesis and semi-rational designed libraries for improved cytochrome P450 BM3-catalyzed hydroxylation of small alkanes. Protein Eng Des Sel. 2012; 25:171–178. [PubMed: 22334757]
47. Bloom JD, Labthavikul ST, Otey CR, et al. Protein stability promotes evolvability. Proc Nat Acad Sci U S A. 2006; 103:5869–6874.
48. Reetz MT, Soni P, Fernández L. Knowledge-guided laboratory evolution of protein thermolability. Biotechnol Bioeng. 2009; 102:1712–1717. [PubMed: 19072845]
49. Seeliger D, de Groot BL. Protein thermostability calculations using alchemical free energy simulations. Biophys J. 2010; 98:2309–2316. [PubMed: 20483340]
50. Zeiske T, Stafford KA, Palmer AG. Thermostability of enzymes from molecular dynamics simulations. J Chem Theory Comput. 2016; 12:2489–2492. [PubMed: 27123810]
51. Borgo B, Havranek JJ. Automated selection of stabilizing mutations in designed and natural proteins. Proc Nat Acad Sci U S A. 2012; 109:1494–1499.
52. Gribenko AV, Patel MM, Liu J, et al. Rational stabilization of enzymes by computational redesign of surface charge–charge interactions. Proc Natl Acad Sci U S A. 2009; 106:2601–2606. [PubMed: 19196981]
53. Van Durme J, Delgado J, Stricher F, et al. A graphical interface for the FoldX forcefield. Bioinformatics. 2011; 27:1711–1712. [PubMed: 21505037]
54. Korendovych IV, DeGrado WF. Catalytic effciency of designed catalytic proteins. Curr Opin Struct Biol. 2014; 27:113–121. [PubMed: 25048695]
55. Wijma HJ, Janssen DB. Computational design gains momentum in enzyme catalysis engineering. FEBS J. 2013; 280:2948–2960. [PubMed: 23647554]
56. Yeung N, Lin YW, Gao YG, et al. Rational design of a structural and functional nitric oxide reductase. Nature. 2009; 462:1079–1082. [PubMed: 19940850]
57. Kazlauskas RJ, Bornscheuer UT. Finding better protein engineering strategies. Nat Chem Biol. 2009; 5:526–529. [PubMed: 19620988]
58. Höhne M, Schätzle S, Jochens H, et al. Rational assignment of key motifs for function guides in silico enzyme identification. Nat Chem Biol. 2010; 6:807–813. [PubMed: 20871599]
59. Yin H, Slusky JS, Berger BW, et al. Computational design of peptides that target transmembrane helices. Science. 2007; 315:1817–1822. [PubMed: 17395823]

Table 1

Representative examples of proteins designed using various approaches

Design principles, methods[a]	Parameters introduced/optimized	Representative citations
Substitution of amino acids by rational design		
Visual inspection, Docking, ISM	Substrate specificity Stereoselectivity	[1]
CAVER, ISM	Activity, Stabiliy	[3, 4]
B-Fit, ISM	Thermostability	[5]
MD-simulations	Enantioselectivity	[6, 7]
Prediction of pK_a	pH Optimum	[8–11]
Computational design		
FRESCO	Thermostability	[12]
CASCO Rosetta Design	Enantioselectivity	[13]
Rosetta Design/Rosetta match	Introducing new chemical activities	[14–17]
Minimalist design	Introducing new chemical activities	[18–20]
De novo *design of protein folds*		
Semiempirical computation	Introducing catalysis	[21]
Introduction of noncanonical amino acids		
Rational, substrate docking	Introducing new chemical activities	[22]
Rosetta	Protein–peptide interface, metal cofactor binding	[23, 24]
Redesign of the existing or introduction of new cofactors		
Introducing metal cofactors into proteins		[25]
Substitution of metal ions in existing cofactors	Introducing new chemical activities	[26]
Transition metal complexes anchored by biotin conjugation		[27]

[a]Note the list is by no means exhaustive

ISM iterative saturation mutagenesis, *MD* molecular dynamics, *FRESCO* framework for rapid enzyme stabilization by computational libraries, *CASCO* catalytic selectivity by computational design