

# Computational tools for designing and engineering biocatalysts

Jiri Damborsky and Jan Brezovsky

Current computational tools to assist experimentalists for the design and engineering of proteins with desired catalytic properties are reviewed. The applications of these tools for *de novo* design of protein active sites, optimization of substrate access and product exit pathways, redesign of protein–protein interfaces, identification of neutral/advantageous/deleterious mutations in the libraries from directed evolution and stabilization of protein structures are described. Remarkable progress is seen in *de novo* design of enzymes catalyzing a chemical reaction for which a natural biocatalyst does not exist. Yet, constructed biocatalysts do not match natural enzymes in their efficiency, suggesting that more research is needed to capture all the important features of natural biocatalysts in theoretical designs.

## Address

Loschmidt Laboratories, Institute of Experimental Biology and National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Kamenice 5, Brno, Czech Republic

Corresponding author: Damborsky, Jiri ([jiri@chemi.muni.cz](mailto:jiri@chemi.muni.cz))

Current Opinion in Chemical Biology 2009, 13:26–34

This review comes from a themed issue on  
Biocatalysis and Biotransformation  
Edited by Romas Kazlauskas and Stefan Lutz

Available online 16th March 2009

1367-5931/\$ – see front matter  
Published by Elsevier Ltd.

DOI 10.1016/j.cbpa.2009.02.021

## Introduction

Enzymes are molecular machines that catalyze chemical reactions in living organisms. It is of great scientific interest and practical need to construct enzymes with new catalytic properties and enhanced stabilities. The methods of directed evolution, based on several rounds of mutagenesis in combination with efficient screening or selection, have been particularly successful in this effort owing to the high complexity of protein structures and our limited understanding of the protein structure–function relationships. Long-term efforts to assist directed evolution in focusing on the regions in protein structures relevant for the function as well as to design enzymes *de novo*, led to development of a large variety of computational tools. These theoretical tools are maturing and 2008 is to be remembered as the year when the grand challenge – *de novo* design of an enzyme catalyzing a chemical reaction for which a natural biocatalyst does not

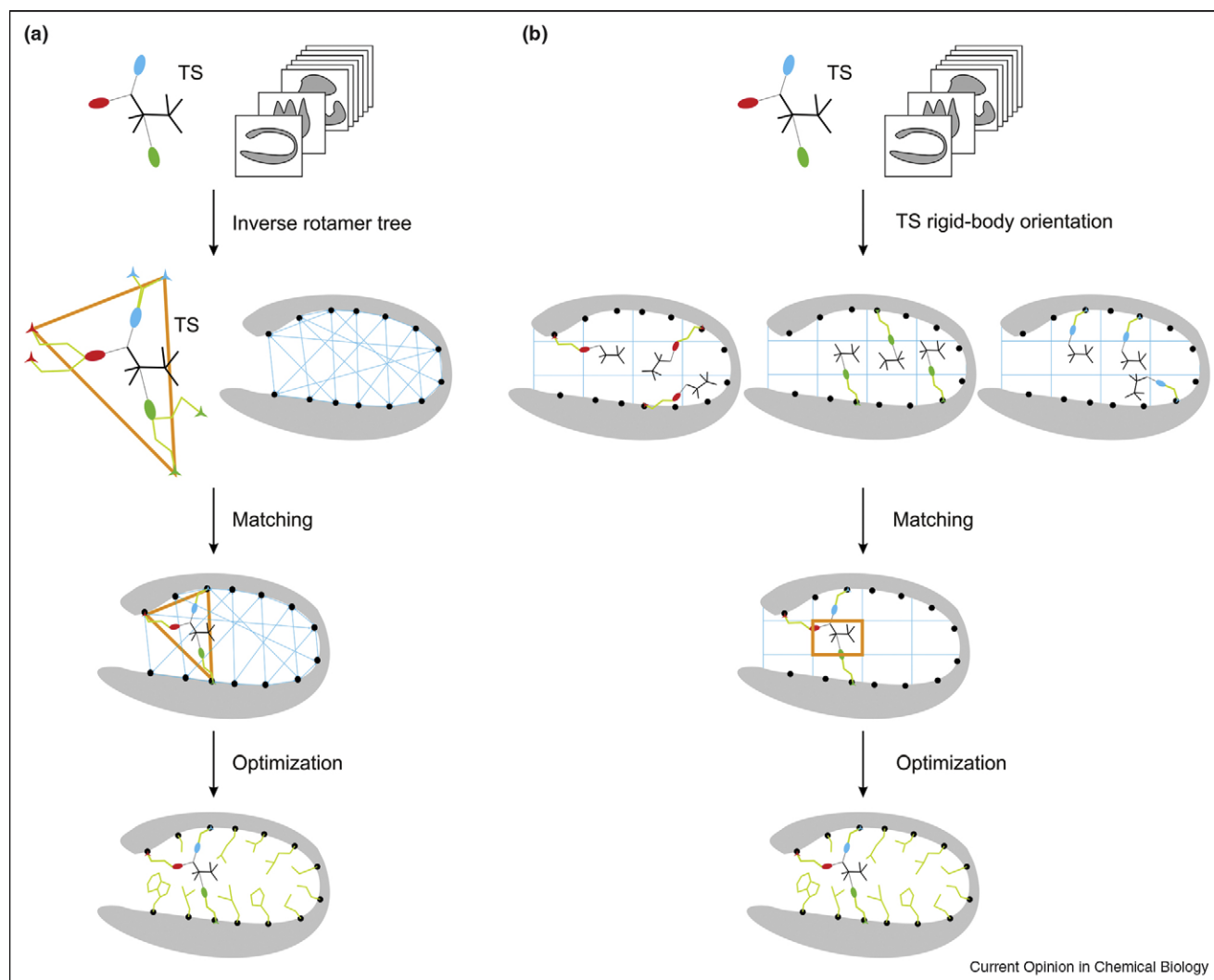
exist – has been met. In this review, we discuss progress in the development of tools for protein design and engineering in the past two years. Other recent reviews can provide additional background and viewpoints [1–8].

## Computational tools for *de novo* design of active sites

Computational *de novo* design relies on the introduction of amino acid residues essential for catalysis into the existing scaffolds. The underlining idea is that enzymes enhance chemical reactions by lowering an activation barrier due to stabilization of the transition state by the residues of the active site [9]. Initially, the transition state of the reaction and the idealized active site geometry is modeled using quantum mechanics. The library of protein scaffolds is then searched to identify potential binding pockets that bind tightly to the transition state and retain the desired geometry of the functional groups (Figure 1). Using geometry-based identification, the transition state is matched with the binding site and the position of the transition state and the catalytic side chains are optimized. Finally, the remaining residues for tight binding of the transition state are designed and the designs are ranked on the basis of transition state binding energy and catalytic geometry.

The *de novo* design of enzymes has been pioneered by the programs DEZYMER [10], ORBIT [11] and ROSETTA [12]. Current computational designs can provide the close to atomic resolution predictions. Nevertheless, constructs based on computational designs need to be subjected to several rounds of directed evolution for fine-tuning of structures and improvement of catalytic efficiency. Design of enzymes for the chemical reactions that are not catalyzed by naturally occurring biocatalysts is particularly challenging. Röthlisberger *et al.* developed an enzyme that enhances the rate of Kemp elimination up to  $10^5$  [13]. Kemp elimination is a model reaction for proton transfer from carbon. Seven rounds of *in vitro* evolution, consisting of random mutagenesis and shuffling, enhanced computational designs and produced an enzyme showing a >200-fold increase in catalytic efficiency and multiple turnovers (>1000 catalytic cycles). Some of the mutations introduced during directed evolution have been localized in the residues adjacent to designed positions, some changed flexibility of the region neighboring the active site or adjusted  $pK_a$  of the catalytic residues. The position and functional role of these mutations provide important insight into strength and shortcomings of current designs, which need to be understood in order to match efficiency of natural catalysts in the future (see Section Outlook). Mutational analysis

Figure 1



*De novo* design of active sites in existing scaffolds. (a) In the 'inside-out' method, an inverse rotamer tree is built up from the active site description, and the backbone coordinates of all the rotamer combinations are compared with backbone coordinates of the set of scaffolds using a geometric-hashing based algorithm. (b) In the 'outside-in' method, side-chain rotamers and the transition state (TS) models are sequentially placed at all scaffold positions using a rigid body searching algorithm. The position of the TS model is recorded in a hash table, which is scanned for the TS positions that are found when placing each of the catalytic side chains independently. After putative active sites have been identified by one of the two methods, the remaining residues in the pocket around the docked TS model are redesigned and optimized.

further confirmed that catalysis was due to the computationally designed active site, and the high-resolution crystal structure confirmed that designs had close to atomic accuracy. Improved algorithms that rely on hashing techniques implemented in ROSETTA enabled *de novo* design of biocatalysts for multistep reactions [14<sup>••</sup>]. The target biocatalyst was retro-aldolase, breaking a carbon-carbon bond in a non-natural substrate. The designs spanned a broad range of protein folds and 32 out of 72 experimentally characterized designs showed detectable retro-aldolase activity. Zhu and Lai developed a new *de novo* protein design method that enables the user to search for suitable scaffolds directly in the Protein Data

Bank [15]. The method has been validated theoretically with triose phosphate isomerase. The novel dead-end elimination algorithm, guaranteeing that no rotamers belonging to the minimized global minimum energy conformation are pruned, can also be of interest to the community working on *de novo* protein design [16]. The algorithm is useful for identifying minimized global minimum energy conformations and for the filtering of ensemble-based scoring. Tynan-Connolly and Nielsen developed a fast and accurate algorithm for redesign of the  $pK_a$  values of catalytic residues placed into the putative catalytic sites [17]. The algorithm identifies point mutations that change the net charge of the

enzyme, but maintain its fold and activity, and could be included as an additional step in *de novo* designs. The initial *de novo* designs could be additionally fine-tuned by simulations employing transition states [18,19].

### Computational tools for design of ligand exchange pathways

The traditional Emil Fisher's 'lock-key' model uses analogy between enzyme (lock) and substrate (key) to describe the need for a matching shape of a substrate in order to fit to the active site of an enzyme [20]. The preference of an enzyme for given substrates is attributed to the quality of the match between enzyme active site and transition states of individual substrates. Lock-key model, or its modified version, the induced-fit model [21], explains catalysis by an enzyme with an easily accessible active site, while it is less appropriate for the enzymes with active sites buried in the protein core. The activity and specificity of such proteins is determined by not only the geometry and properties of the active site but also tunnels (keyholes) connecting the active site with bulk solvent. The recently proposed 'keyhole-lock-key' model reflects importance of substrate entry and product exit pathways for catalysis by the enzymes with buried active sites (J Damborsky, abstract, HHMI International Research Scholars Meeting, Ashburn, Virginia, USA, September 2006, p. 50). Anatomy, physico-chemical properties and dynamics of protein tunnels can lead to discrimination of potential substrates and can also determine kinetics of substrate entry and products exit. Modification of these tunnels, rather than the active sites, is a new paradigm in protein engineering [22<sup>••</sup>], creating demand for development of new computational tools.

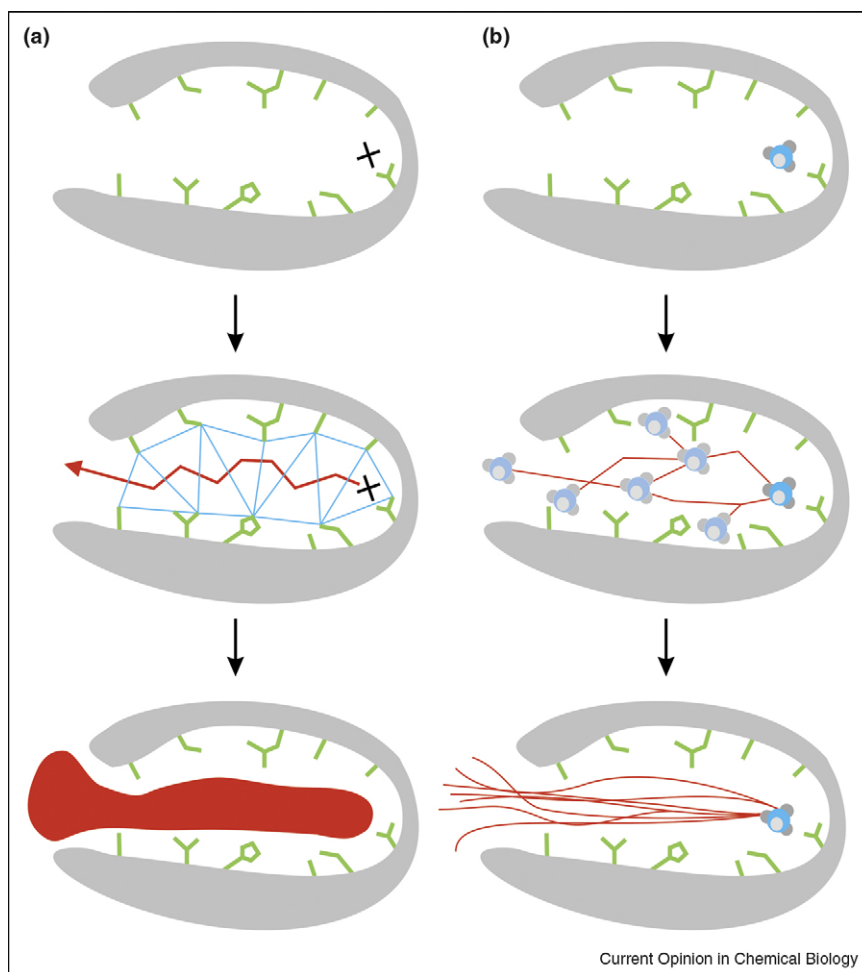
The program CAVER was developed specifically for identification of tunnels in proteins with buried active sites [23<sup>•</sup>,24]. The original algorithm searched for the tunnels on the grid, while the current version employs a faster and more accurate algorithm based on decomposition of a metric space using Voronoi diagrams and Delaunay triangulation (Figure 2). Voronoi diagram is a decomposition of a metric space determined by distances to a specified discrete set of objects in the space, while Delaunay triangulation is a collection of edges satisfying an 'empty circle' property. CAVER has been applied to available structures of haloalkane dehalogenases, providing interpretation for functionally important mutations localized on the protein surface [23<sup>•</sup>,25]. Application of CAVER for the design of protein variants with modified tunnels resulted in the mutant showing 32 times higher activity toward the non-natural substrate 1,2,3-trichloropropane (Pavlova *et al.*, unpublished). Other tools for calculation of pathways in the proteins named MOLE [26], MOLAXIS [27], HOLLOW [28] and CHUNNEL [29] were recently developed. Individual programs differ in speed, accuracy,

requirements for input data and in visualization of calculated results. A very novel and interesting approach for computation of pathways based on an efficient path-planning algorithm from robotics has been introduced by Guieysse *et al.* [30<sup>•</sup>] (Figure 2). The algorithm was able to semi-quantitatively reproduce the effect of substrate access on enantioselectivity of lipase from *Burkholderia cepacia* and enabled the identification of target residues for engineering enantioselectivity. Saturated mutagenesis of the residues surrounding the mouth openings or bottlenecks of tunnels is a very powerful strategy for the engineering of ligand exchange pathways in enzymes with buried active sites [31] (Pavlova *et al.*, unpublished).

### Computational tools for design of protein-protein interfaces

Protein-protein interactions are key components of all signal transduction processes, and the methods to alter these interactions represent important tools for dissecting function of connectivities in signal networks. The program ORBIT, referred to above, has been used for computational design of ubiquitous messenger protein calmodulin [32<sup>•</sup>]. Calmodulin responds to the different levels of  $\text{Ca}^{2+}$  and interacts with calmodulin-dependent protein kinase II and calcineurin in the cells. Using a modified energy function, emphasizing intermolecular interactions in the sequence selection procedure, the authors designed a protein exhibiting a 900-fold increase in binding specificity toward two binding partners of calmodulin. The authors concluded that this is the highest specificity switch achieved in any protein-protein interface by computational protein design. Fajardo-Sanchez *et al.* used the FOLD-X algorithm for design of specific heterodimer interfaces between two meganucleases [33]. Constructed functional endonucleases cut various non-palindromic DNA sequences, which is an essential step toward safer targeted genome engineering. Tomic *et al.* recently conducted comparative binding energy (COMBINE) analysis of 122 different wild type and mutant complexes of small guanosine triphosphate-binding proteins, Ras and Rap, and their effectors Raf and RalGDS [34]. Ras-family proteins are involved in cellular pathways leading to cell growth, differentiation and apoptosis and their engineering has potential for anticancer therapies. The COMBINE model decomposed the binding affinities of interacting partners on a residue-wise basis and revealed three negatively charged residues on the interface of Ras as hot-spots for its interaction with Raf and Ral-GDS. The concept called alanine scanning systematically estimates the effect of substitution of individual amino acids by an alanine and is well suited for the analysis of interactions in the protein-protein complexes. Theoretical alanine scanning was introduced by Massova and Kollman [35], adopted for ROSETTA by Kortemme *et al.* [36] and recently improved by Moreira *et al.* [37].

Figure 2



Identification of ligand exchange pathways. (a) CAVAR algorithm searches for the pathways from a point located inside the active site to the outside solvent using Voronoi diagrams and Delaunay Triangulation. (b) Path-planning algorithm starts from the enzyme-substrate complex and incrementally constructs a tree structure by expanding branches toward unexplored regions of the search-space, while satisfying motion constraints.

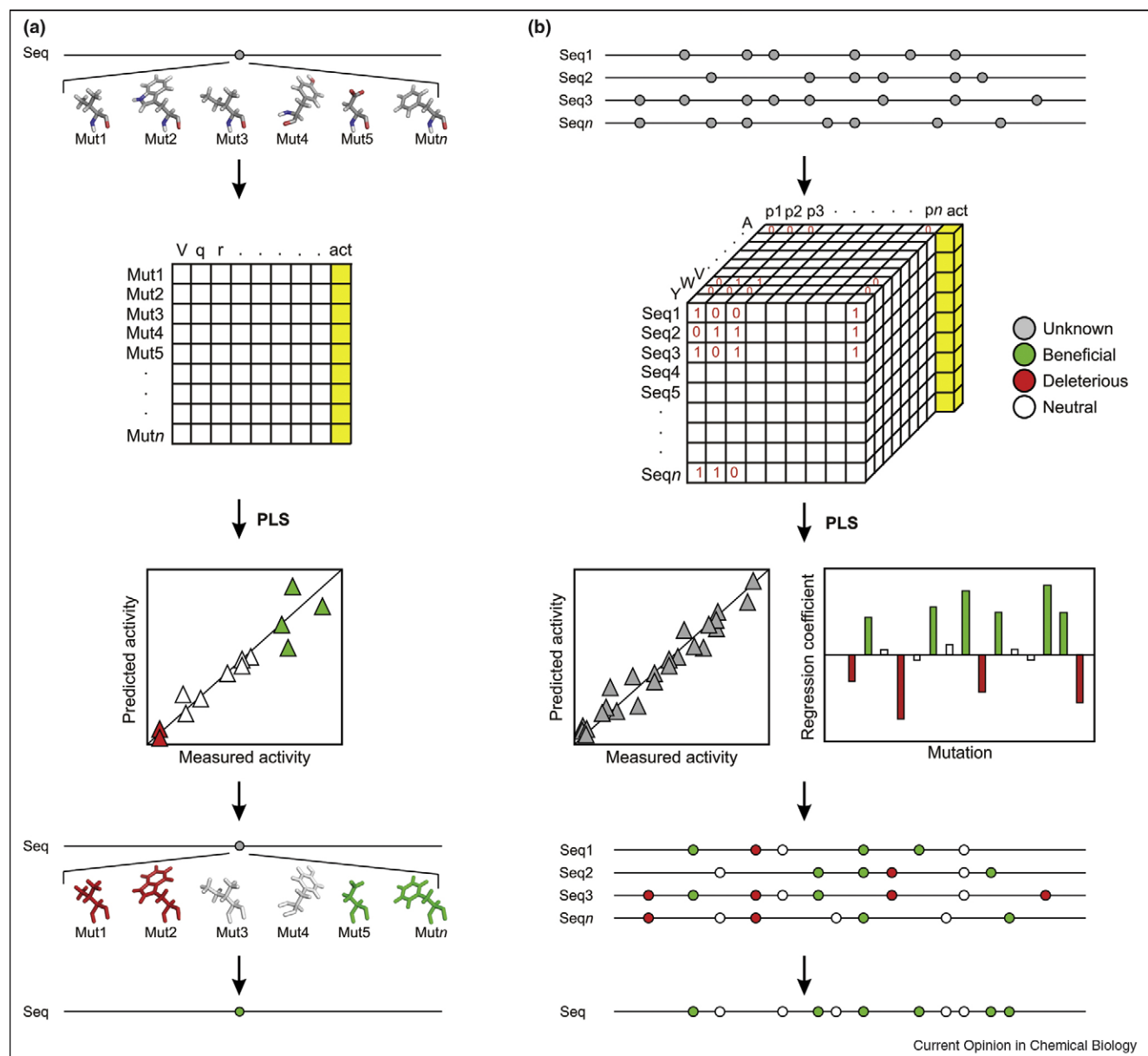
### Computational tools for the assignment of neutral, advantageous and deleterious mutations

Directed evolution methods developed over the past 15 years are powerful in generating molecular diversity, which must be screened or selected for desired catalytic activity or new properties. It is becoming generally accepted that blind generation of large libraries and laborious screening is not an efficient way of obtaining a good biocatalyst [5]. Narrowing the sequence space using the structural information or identification of useful mutations by statistical analysis and their intentional incorporation into the new generation significantly improves efficiency of directed evolution experiments. Quantitative Structure-Activity Relationships (QSAR) is a traditional approach used for establishing quantitative relationships between the structure of small molecules and their biological activity in drug

design and ecotoxicology [38]. Multivariate statistics are often employed for construction of QSAR models since they are particularly suitable for the analysis of large data matrices made of many parameters and few objects.

QSAR methodology has been applied to protein variants of haloalkane dehalogenase, subtilisin, T4 lysozyme and tryptophan synthase and used for the first time in the context of protein engineering in so called quantitative structure-function relationships (QSFR) analysis [39]. In QSFR analysis, structural and physico-chemical properties of the amino acid residues are correlated with functional data for the set of proteins carrying systematic substitutions at the certain positions (Figure 3). The QSFR model enables quantitative interpretation of the effect of mutations on protein function or stability, and prediction of the properties for the variants that have

Figure 3



Quantitative structure–function and sequence–activity relationships. **(a)** QSFR analysis quantitatively describes mutations (Mut) by the set of molecular descriptors (V, q, r) and uses a multivariate statistical method, Partial-Least Squares regression (PLS), for establishing a relationship between these descriptors and activity (act) of mutant enzymes. The established relationship, PLS model, can be used to make predictions of activities for mutants, which have not been constructed and experimentally tested. **(b)** ProSAR analysis quantifies the effects of various mutations in the combinatorial library of mutant sequences (Seq) and classifies them as beneficial (green), deleterious (red) or neutral (white). The most active variant is then used as a parent template, into which beneficial mutations are incorporated by oligonucleotide-based semi-synthetic shuffling and deleterious mutations are removed by reverting to the previous template codons.

never been constructed and tested experimentally. Fox *et al.* combined the QSAR approach with recombination-based directed evolution in ProSAR analysis, facilitating mutation-oriented enzyme optimization by permitting to capture information contained in the sequence–activity data [40<sup>••</sup>]. Sequence–activity data are used for building

a statistical model that assigns the regression coefficient to each mutation (Figure 3). The multivariate optimization strategy then evaluates the regression coefficients to decide whether a mutation should be retained, discarded or tested again in new combinatorial libraries. ProSAR enabled a remarkable 4000-fold improvement in



the volumetric productivity of biocatalytic synthesis of the cholesterol-lowering drug atorvastatin by bacterial halohydrin dehalogenase. Ahle and Estell developed a method for the engineering of multiple properties based on a systematic evaluation of sequence–activity relationships using specially designed libraries [41<sup>••</sup>]. Their method provides means to evaluate all positions in a protein for all properties of interest by building site-evaluation libraries at each site. These libraries contain 9–19 substitutions at each position and are used to distinguish ‘productive’ and ‘unproductive’ sites. Productive sites are those that have at least one substitution that is better than the parent enzyme, while unproductive sites have no mutant that is better than parent enzyme. A simplified protein engineering strategy according to Ahle and Estell is to build site-evaluation libraries and screen using a simple activity and/or stability screen. The rationale for this strategy is that practical applications do not require finding the best sequence in the protein space, but identification of at least one sequence that meets or exceeds the minimum value required for a number of properties. Masso and Vasiman introduced another approach in which they calculate environmental changes in the protein structure relative to the wild type by Delaunay triangulation and then use a machine-learning approach for inferring the activity levels of protein variants [42]. The novelty of this approach comes from the use of statistical contact potentials, rather than simple sequence, structure or evolutionary information, as the descriptors for model development. By leveraging the power of machine learning on as few as seven energy-based attributes, the method outperformed techniques that utilize a much larger number of predictors. Barak *et al.* used the statistical analysis of 7000 mutants constructed by directed evolution, yielding an initial 230-fold improvement in activity, for prediction of only 11 mutants yielding >1500-fold improvement in activity relative to the wild-type enzyme [43].

### Computational tools for design of protein stability

Stability of proteins has been of great interest to protein engineers as well as practitioners implementing the use enzymes in industrial processes. Despite their many favorable properties, the marginal stability of enzymes in reaction media often has prevented or delayed their implementation for the industrial-scale synthesis of fine chemicals and pharmaceuticals [2]. A number of simple rules defining the stability of protein structures have been formulated over the years and can be found in biochemistry textbooks. Here we focus on current computational tools for the design of mutations for stabilizing protein structures.

Optimization of the distribution of surface charge–charge interactions for stabilization of protein structures

has been tested by Schweiker *et al.* [44<sup>•</sup>]. Computational analysis of the distribution of surface charges of the Fyn SH3 domain showed that the increase in favorable energy per substitution, levels off at five substitutions. Experimentally characterized construct containing four charge reversals, and one newly introduced charge, showed an increase in melting temperature. The application of VOLSURF, a tool for calculation of 2D molecular descriptors from 3D molecular interaction energy grid maps, for quantitative prediction of protein thermostability, has been described by Braiuca *et al.* [45]. Authors gathered data for 22 thermostable and hyperthermostable organisms and established a multivariate model that identified structural requirements correlating with protein stability. The calculation of interaction energy grid maps using VOLSURF is straightforward and fast, making this approach generally applicable for protein engineering. The method may find their use in rationalization of structural features important for protein stability as well as in computer-assisted design of thermostable biocatalysts. Cabrita *et al.* described design and construction of TEV protease mutants with improved stability and solubility [46]. The mutants have been designed using the tool POPMUSIC, which estimates changes in free energies by means of database-derived potentials. The potentials of POPMUSIC are derived from observed frequencies of sequence and structure patterns in a dataset consisting of 141 non-homologous protein structures. The different linear combinations of the database-derived potentials are used on the basis of the solvent accessibility of mutated residues. Bae *et al.* developed a bioinformatics method called improved configurational entropy (ICE) for the stabilization of proteins [47<sup>••</sup>]. Using ICE, the authors were able to design more stable variants of a mesophilic adenylate kinase using the sequence information from one psychrophilic homologue as the only input data. The method does not require a three-dimensional structure or a large number of homologous sequences, making it broadly applicable. There is growing number of methods employing machine-learning tools for predicting protein stability. For example, Montanucci *et al.* developed a method employing the support vector machines for prediction of mutations, including insertions and deletions, which enhance thermostability of a given protein sequence [48]. Their method achieved 88% accuracy and correctly classified 12 out of 14 mutants with enhanced thermostability. The approach of Masso and Vaisman is based on a four-body statistical contact potential and predicts an empirical normalized measure of the ensuing environmental perturbation at every residue position with 84% accuracy [49]. The field of predicting protein stability would benefit from the assessment of newly developed methods as it is usual for example in the community-wide experiment of protein structure

prediction called Critical Assessment of Techniques for Protein Structure Prediction.

## Outlook

In spite of the remarkable progress in *de novo* design, developed biocatalysts do not match natural catalysts in their efficiency. What is missing in current designs? Designs employing site-directed mutagenesis may not achieve precise stabilization of the transition states, where the differences in the distances at the picometer scale can be important [50<sup>••</sup>]. Protein backbone dynamics can also play a very important role in catalysis and are not included in current design methods. Existing *de novo* design methods focus only on chemical steps, while exchange of ligands between the active site and the solvent accompanied by conformational changes in the protein are not considered, even though they could be rate-determining.

The major limitation of existing tools dealing with identification of ligand exchange pathways is that they cannot efficiently analyze and visualize data derived for the large ensembles of structures, which is much needed for the description of tunnel dynamics. Clustering of tunnels identified in individual structures of the ensemble is difficult when some of the tunnels disappear and some are newly formed in time. More research is needed on mechanisms of substrate entry or product release in order to develop parameters specifying biochemically relevant tunnels and channels.

Design of protein–protein interfaces could potentially benefit from better integration of time-demanding energy-based tools, calculating change in the stability of a complex upon introduction of a mutation, with fast sequence- and structure-based tools, analyzing distribution of sequence conservation and physico-chemical properties at the interface.

The application of statistical analyses and QSAR-like approaches in directed evolution experiments can be highly recommended and will be seen more often in future projects focused on optimization of biocatalysts. Statistical analyses reduce the sequence space and reduce demand on laborious screening of large mutant libraries. Moreover, biochemical interpretation of developed multivariate models broadens our understanding of structure–function relationships.

A large number of new *in silico* methods predicting mutations for stabilization of proteins corresponds with the importance of stability for the application of biocatalysts to industrial-scale processes. Development of methods predicting stabilization of protein structures by various additives would be highly desirable to avoid case-by-case experimentation. However, fundamental studies of the effect of additives on protein

structures are needed for development of reliable predictive models.

## Acknowledgements

Research work of the authors is supported by the research grants from the Czech Ministry of Education (LC06010, MSM0021622412 and MSM0021622413) and the Czech Science Foundation (201/07/0927 and 203/08/0114). Financial support is greatly appreciated.

## References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Lippow SM, Tidor B: **Progress in computational protein design.** *Curr Opin Biotech* 2007, **18**:305-311.
2. Polizzi KM, Bommarius AS, Broering JM, Chaparro-Riggers JF: **Stability of biocatalysts.** *Curr Opin Chem Biol* 2007, **11**:220-225.
3. Qian Z, Fields CJ, Yu Y, Lutz S: **Recent progress in engineering alpha/beta hydrolase-fold family members.** *Biotech J* 2007, **2**:192-200.
4. Chaparro-Riggers JF, Polizzi KM, Bommarius AS: **Better library design: data-driven protein engineering.** *Biotech J* 2007, **2**:180-191.
5. Jäckel C, Kast P, Hilvert D: **Protein design by directed evolution.** *Annu Rev Biophys* 2008, **37**:153-173.
6. Leisola M, Turunen O: **Protein engineering: opportunities and challenges.** *Appl Microbiol Biotechnol* 2007, **75**:1225-1232.
7. Fox RJ, Huisman GW: **Enzyme optimization: moving from blind evolution to statistical exploration of sequence–function space.** *Trends Biotechnol* 2008, **26**:132-138.
8. Gruber J, Zawaira A, Saunders R, Barrett CP, Noble ME: **Computational analyses of the surface properties of protein–protein interfaces.** *Acta Crystallogr D Biol Crystallogr* 2007, **63**:50-57.
9. 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.
- Garcia-Viloca *et al.* reviewed the current understanding of enzyme catalysis in the light of generalized transition state theory.
10. Hellinga HW, Richards FM: **Construction of new ligand binding sites in proteins of known structure.** *J Mol Biol* 1991, **222**:763-785.
11. Dahiyat BI, Mayo SL: **Protein design automation.** *Prot Sci* 1996, **5**:895-903.
12. Zanghellini A, Jiang L, Wollacott AM, Cheng G, Meiler J, Althoff EA, Röthlisberger D, Baker D: **New algorithms and an *in silico* benchmark for computational enzyme design.** *Prot Sci* 2006, **15**:2785-2794.
13. Röthlisberger D, Khersonsky O, Wollacott AM, Jiang L, Dechncie 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.
- Röthlisberger *et al.* described the computational design of eight enzymes that used two different catalytic motifs to catalyze the Kemp elimination – a model reaction for proton transfer from carbon – with measured rate enhancements of up to 10<sup>5</sup> and multiple turnovers. Designed model matched the crystal structure with atomic accuracy.
14. Jiang L, Althoff EA, Clemente FR, Doyle L, Röthlisberger 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.
- Jiang *et al.* designed *de novo* and constructed the enzyme catalyzing multistep reaction. Constructed retro-aldolase used four different catalytic motifs to catalyze the breaking of a carbon–carbon bond in a non-natural substrate. Designs based on explicit water were more successful than those involving charged side-chain networks.

15. Zhu X, Lai L: **A novel method for enzyme design.** *J Comput Chem* 2009, **30**:256-267.
16. Georgiev I, Lilien RH, Donald BR: **The minimized dead-end elimination criterion and its application to protein redesign in a hybrid scoring and search algorithm for computing partition functions over molecular ensembles.** *J Comput Chem* 2008, **29**:1527-1542.
17. Tynan-Connolly BM, Nielsen JE: **Redesigning protein  $pK_a$  values.** *Prot Sci* 2007, **16**:239-249.
18. Pan Y, Gao D, Yang W, Cho H, Zhan CG: **Free energy perturbation (FEP) simulation on the transition states of cocaine hydrolysis catalyzed by human butyrylcholinesterase and its mutants.** *J Am Chem Soc* 2007, **129**:13537-13543.
19. Zheng F, Yang W, Ko MC, Liu J, Cho H, Gao D, Tong M, Tai HH, Woods JH, Zhan CG: **Most efficient cocaine hydrolase designed by virtual screening of transition states.** *J Am Chem Soc* 2008, **130**:12148-12155.
20. Fischer E: **Einfluss der configuration auf die wirkung der enzyme.** *Ber Dtsch Chem Ges* 1894, **27**:2985-2993.
21. Koshland DE: **Application of a theory of enzyme specificity to protein synthesis.** *Proc Natl Acad Sci USA* 1958, **44**:98-104.
22. Chaloupkova R, Sykora J, Prokop Z, Jesenska A, Monincova M, Pavlova M, Tsuda M, Nagata Y, Damborsky J: **Modification of activity and specificity of haloalkane dehalogenase from *Sphingomonas paucimobilis* UT26 by engineering of its entrance tunnel.** *J Biol Chem* 2003, **278**:52622-52628.
- Chaloupkova *et al.* conducted saturated mutagenesis of the residue located at the mouth opening of the tunnel connecting the buried active site with the protein surface. The residue has been selected for mutagenesis based on structural comparisons and phylogenetic interference, and its modification resulted in significant changes in both catalytic activity and substrate specificity. The protocol has been implemented in the HOTSPOT WIZARD: <http://loschmidt.chemi.muni.cz/hotspotwizard/>.
23. Petrek M, Otyepka M, Banas P, Kosinova P, Koca J, Damborsky J: **CAVER: a new tool to explore routes from protein clefts, pockets and cavities.** *BMC Bioinf* 2006, **7**:316.
- Petrek *et al.* developed the computational tool for identification and visualization of pathways connecting the buried active site with the bulk solvent. The tool allows calculation of tunnels, channels or pores in the biomolecules, inorganic materials and molecular ensembles.
24. Damborsky J, Petrek M, Banas P, Otyepka M: **Identification of tunnels in proteins, nucleic acids, inorganic materials and molecular ensembles.** *Biotech J* 2007, **2**:62-67.
25. Silberstein M, Damborsky J, Vajda S: **Exploring the binding sites of the haloalkane dehalogenase DhIA from *Xanthobacter autotrophicus* GJ10.** *Biochemistry* 2007, **46**:9239-9249.
26. Petrek M, Kosinova P, Koca J, Otyepka M: **MOLE: a Voronoi diagram-based explorer of molecular channels, pores, and tunnels.** *Structure* 2007, **15**:1357-1363.
27. Yaffe E, Fishelovitch D, Wolfson HJ, Halperin D, Nussinov R: **MolAxis: a server for identification of channels in macromolecules.** *Nucl Acid Res* 2008, **36**:W210-W215.
28. Ho BK, Gruswitz F: **HOLLOW: Generating accurate representations of channel and interior surfaces in molecular structures.** *BMC Struct Biol* 2008, **8**:49.
29. Coleman RG, Sharp KA: **Finding and characterizing tunnels in macromolecules with application to ion channels and pores.** *Biophys J* 2009, **96**:632-645.
30. Guieysse D, Cortes J, Puech-Guenot S, Barbe S, Lafaquiere V, Monsan P, Simeon T, Andre I, Remaud-Simeon M: **A structure-controlled investigation of lipase enantioselectivity by a path-planning approach.** *ChemBioChem* 2008, **9**:1308-1317.
- Guieysse *et al.* applied a new approach based on the efficient path-planning algorithm from robotics for the prediction of enzyme enantioselectivity.
31. Hidalgo A, Schliessmann A, Molina R, Hermoso J, Bornscheuer UT: **A one-pot, simple methodology for cassette randomisation and recombination for focused directed evolution.** *Prot Eng Des Sel* 2008, **21**:567-576.
32. Yosef E, Politi R, Choi MH, Shifman JM: **Computational design of calmodulin mutants with up to 900-fold increase in binding specificity.** *J Mol Biol* 2009, **385**:1470-1480.
- Yosef *et al.* designed calmodulin mutants exhibiting up to 900-fold increase in binding specificity towards calmodulin-dependent protein kinase II peptide. These results show that computational redesign of protein-protein interfaces becomes a reliable method for altering protein binding affinity and specificity.
33. Fajardo-Sanchez E, Stricher F, Paques F, Isalan M, Serrano L: **Computer design of obligate heterodimer meganucleases allows efficient cutting of custom DNA sequences.** *Nucl Acid Res* 2008, **36**:2163-2173.
34. Tomic S, Bertosa B, Wang T, Wade RC: **COMBINE analysis of the specificity of binding of Ras proteins to their effectors.** *Proteins* 2007, **67**:435-447.
35. Massova I, Kollman PA: **Computational alanine scanning to probe protein-protein interactions: a novel approach to evaluate binding free energies.** *J Am Chem Soc* 1999, **121**:8133-8143.
36. Kortemme T, Kim DE, Baker D: **Computational alanine scanning of protein-protein interfaces.** *Sci STKE* 2004, **2004**:pl2.
37. Moreira IS, Fernandes PA, Ramos MJ: **Computational alanine scanning mutagenesis—an improved methodological approach.** *J Comput Chem* 2007, **28**:644-654.
38. Kubinyi H: **QSAR and 3D QSAR in drug design 1. Methodology.** *Drug Discov Today* 1997, **2**:457-467.
39. Damborsky J: **Quantitative structure-function and structure-stability relationships of purposely modified proteins.** *Prot Eng* 1998, **11**:21-30.
40. Fox RJ, Davis SC, Mundorff EC, Newman LM, Gavrilovic V, Ma SK, Chung LM, Ching C, Tam S, Muley S *et al.*: **Improving catalytic function by ProSAR-driven enzyme evolution.** *Nat Biotechnol* 2007, **25**:338-344.
- Fox *et al.* developed ProSAR that augments recombination-based directed evolution by incorporating a strategy for multivariate statistical analysis of protein sequence activity relationships. This combination facilitates mutation-oriented enzyme optimization by permitting the capture of information contained in the sequence-activity data.
41. Ahle W, Estell DA: **Systematic evaluation of sequence and activity relationships using site evaluation libraries for engineering multiple properties.** World Patent 2008, no 2008002472.
- Aehle and Estell patented construction of site-evaluation libraries for distinguishing 'productive' and 'unproductive' sites in the protein structure. Their strategy for engineering of multiple properties is based on the building of site-evaluation libraries and screening using activity and/or stability screens.
42. Masso M, Vaisman II: **Accurate prediction of enzyme mutant activity based on a multibody statistical potential.** *Bioinformatics* 2007, **23**:3155-3161.
43. Barak Y, Nov Y, Ackerley DF, Matin A: **Enzyme improvement in the absence of structural knowledge: a novel statistical approach.** *ISME J* 2008, **2**:171-179.
44. Schweiker KL, Zarrine-Afsar A, Davidson AR, Makhatazde GI: **Computational design of the Fyn SH3 domain with increased stability through optimization of surface charge charge interactions.** *Prot Sci* 2007, **16**:2694-2702.
- Schweiker *et al.* applied computational design of surface charge-charge interactions to increase stability of proteins. They demonstrated the seminal role of surface charge-charge interactions in determining protein stability suggesting that the optimization of surface interactions can be a good strategy to complement the optimization of a protein core to further enhance protein stability.
45. Braiuca P, Buthe A, Ebert C, Linda P, Gardossi L: **Vol surf computational method applied to the prediction of stability of thermostable enzymes.** *Biotech J* 2007, **2**:214-220.
46. Cabrita LD, Giliis D, Robertson AL, Dehouck Y, Rooman M, Bottomley SP: **Enhancing the stability and solubility of TEV protease using *in silico* design.** *Prot Sci* 2007, **16**:2360-2367.
47. Bae E, Bannen RM, Phillips GN Jr: **Bioinformatic method for protein thermal stabilization by structural entropy optimization.** *Proc Natl Acad Sci USA* 2008, **105**:9594-9597.



Bae *et al.* developed a method incorporating sequence alignments to redesign of stable proteins through optimization of local structural entropy. The variants of adenylate kinase showing melting temperature increased up to 12.5 °C were designed on the basis of the sequence of a single homologue. This result highlights the importance of entropy in the stability of protein structures.

48. Montanucci L, Fariselli P, Martelli PL, Casadio R: **Predicting protein thermostability changes from sequence upon multiple mutations.** *Bioinformatics* 2008, **24**:i190-i195.
49. Masso M, Vaisman II: **Accurate prediction of stability changes in protein mutants by combining machine learning with structure based computational mutagenesis.** *Bioinformatics* 2008, **24**:2002-2009.
50. Sigala PA, Kraut DA, Caaveiro JM, Phybus B, Ruben EA, Ringe D, Petsko GA, Herschlag D: **Testing geometrical discrimination within an enzyme active site: constrained hydrogen bonding in the ketosteroid isomerase oxanion hole.** *J Am Chem Soc* 2008, **130**:13696-13708.

Sigala *et al.* demonstrate that packing and binding interactions within the ketosteroid isomerase active site can constrain local side-chain reorientation and prevent hydrogen bond shortening by 0.1 Å or less. This constraint has substantial energetic effects on ligand binding and stabilization of the negative charge within the oxanion hole.