Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Chemical Biology

Computational tools for designing and engineering enzymes

Jiri Damborsky and Jan Brezovsky

Protein engineering strategies aimed at constructing enzymes with novel or improved activities, specificities, and stabilities greatly benefit from *in silico* methods. Computational methods can be principally grouped into three main categories: bioinformatics; molecular modelling; and *de novo* design. Particularly *de novo* protein design is experiencing rapid development, resulting in more robust and reliable predictions. A recent trend in the field is to combine several computational approaches in an interactive manner and to complement them with structural analysis and directed evolution. A detailed investigation of designed catalysts provides valuable information on the structural basis of molecular recognition, biochemical catalysis, and natural protein evolution.

Addresses
Loschmidt Laboratories, Department of Experimental Biology and Research Centre for Toxic Compounds in the Environment, Faculty of Science, Masaryk University, Kamenice 5/A13, 625 00 Brno, Czech Republic

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

Current Opinion in Chemical Biology 2014, **19**:8–16

This review comes from a themed issue on **Biocatalysis and biotransformation**

Edited by **Jeffrey C Moore** and **Uwe T Bornscheuer**

1367-5931/$ – see front matter,

http://dx.doi.org/10.1016/j.cbpa.2013.12.003

Introduction

Enzymes catalyse chemical reactions in living cells and are used in a wide range of practical applications. In the past, the applications had to be built around the limitations of the enzyme; today, the enzymes can be engineered to fit the process of interest [1••]. New enzymes were traditionally obtained by isolating them from native organisms. Later, enzymes were obtained from metagenomic libraries without the need to culture host organisms. DNA cloning technologies enabled enzyme production in heterologous hosts and changes to the genetic code to introduce modifications into the protein structure. The invention of directed evolution techniques opened new possibilities for massive or systematic mutagenesis.

More recently, focused directed evolution of selected regions and the use of restricted genetic code have become popular means of producing smaller and smarter mutant libraries. Structural biology techniques such as protein crystallography or NMR spectrometry allow the determination of protein structures to atomic resolutions and the employment of molecular modelling for identifying mutagenesis hot spots. The most recent driving force in the field stems from gene synthesis technology, which allows the synthesis of gene coding for putative enzymes from genetic databases, as well as the production of computationally designed proteins.

In silico methods, ranging from bioinformatic analysis of primary sequences, through computer simulations of tertiary structures, to the prediction of novel structures by *de novo* design, wind through the platforms aimed at constructing optimal biocatalysts. We discuss the computational tools and their applications in protein design that have been published in the past two years. We do not cover the tools suitable for designing smart libraries for focused directed evolution since we have reviewed them recently elsewhere [2]. The article is structured in three parts according to the purpose of the design: Firstly, engineering enzyme activity; secondly, engineering enzyme specificity; and finally, engineering enzyme stability.

A number of excellent reviews have been published recently on related topics. Davids and co-workers overviewed the methods suitable for designing focused libraries and high-throughput screening [3]. Progress in *de novo* protein design was discussed in the reviews by Davey and Chica [4], Hilvert [5], Khare and Fleishman [6], Kries and co-workers [7], and Wijma and Janssen [8]. Approaches for protein stabilization were covered by the reviews of Wijma and co-workers [9], Bommarius and Paye [10], Socha and Tokuriki [11], and Stepankova and co-workers [12].

Designing and engineering enzyme activity

Bioinformatics

Suplatov and co-workers developed the web server *ZEBRA* for analysing enzyme functional subfamilies [13]. The server attempts to systematically identify and analyse adaptive mutations. These subfamily specific positions (SSPs) are conserved within the subfamily, but should differ among them. The implemented statistical analysis evaluates the significance of SSPs, which can then be modified by rational design or focused directed evolution. The method has been tested with the α/β-hydrolase superfamily [14]. SSPs calculated for the amidases were integrated into the sequence of the lipase CALB and the library of mutants was constructed. *In silico* screening of the library for the reactive

enzyme–substrate complexes resulted in the selection of lipases with significantly improved amidase activity.

The *JANUS* method analyses multiple-sequence alignments to predict mutations required for interconversion of structurally related but functionally distinct enzymes [15]. The method has been validated by the interconversion of aspartate aminotransferase into tyrosine aminotransferase. The incorporation of 35 mutations resulted in a protein with the desired specificity but low catalytic activity, which had to be optimized by DNA back-shuffling.

Yang and co-workers presented a computational approach for engineering an allosteric regulation [16]. The authors conducted a statistical comparison of catalytic and allosteric binding sites, which revealed that allosteric sites are evolutionarily more variable and comprise more hydrophobic residues than the catalytic sites. The approach was applied to the deregulation of the allostery in fructose-1,6-bisphosphate, but it remains to be seen whether the methodology will work for other enzyme families.

Molecular modelling

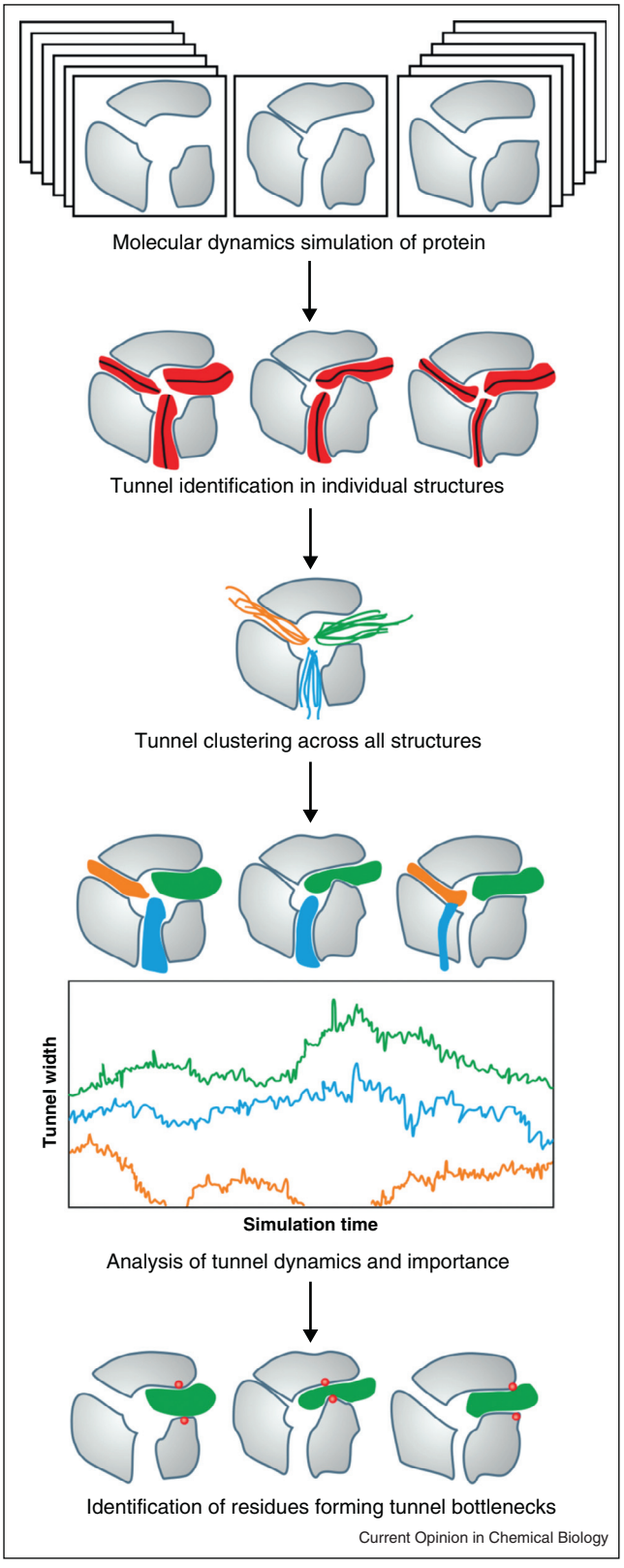
Biedermannova and co-workers combined several molecular modelling methods to study the effect of tunnel mutations on kinetics and reaction mechanisms of haloalkane dehalogenase [17]. The software tool *CAVER* [18•] was used to analyse tunnel dynamics in trajectories obtained by molecular dynamic simulations (Figure 1) and complemented with an analysis of products egressing from buried active sites using *Random Accelerated Molecular Dynamics* (RAMD). The energy barriers of the product release, calculated by the *Adaptive Biasing Force* (ABF) method, were in good agreement with the data from transient kinetic experiments. A redesign of protein tunnels and gates [19] using dedicated software tools [20] provides a useful strategy for engineering enzyme activity.

De novo design

De novo protein design has become mainstream, with more than half of the reviewed articles employing this approach to some extent. This unprecedented research activity makes *de novo* design more robust, accurate, and reliable [6]. The software suites *ROSETTA* and *ORBIT* are the most widely used, and web-based applications were recently developed (see below). Designed enzymes can catalyse non-biological reactions, including multistep retroaldol transformation, Diels–Alder cycloaddition, and proton transfer [5,7]. They typically do not meet the efficiencies of natural enzymes, but can be improved by directed evolution [21,22,23•,31••].

The methodology of computational protein design is being continuously improved by the integration of novel protocols. Hallen and co-workers introduced an algorithm

Figure 1



Workflow for analysis of tunnels in dynamic protein structures using CAVER [18•].

called *Dead-End Elimination with Perturbations* (DEEPer) for identifying global minimum-energy conformation of structures with large backbone perturbations [24]. The algorithm is expected to provide more realistic modelling of backbone flexibility. Nivón and co-workers developed the *Pareto-Optimal Refinement Method* for the efficient design of initial scaffold libraries [25]. Keedy and co-workers proposed a novel algorithm for modelling local backrub motions, which are subtle backbone adjustments taking place during amino acid substitutions [26]. These motions participate in natural protein evolution and their implementation in computational design algorithms improves model accuracy. The incorporation of 114 non-canonical amino acids into *ROSETTA* by building necessary backbone-dependent rotamer libraries and the parameterization and construction of a scoring function were described by Renfrew and co-workers [27•].

A computational method to redesign the active site for catalysing new reactions was developed by Khare and co-workers [28•]. Using this method, the authors engineered an organophosphate hydrolase starting from a functionally diverse set of mononuclear zinc-containing metalloenzyme scaffolds. This redesigned enzyme showed the catalytic efficiency k_{cat}/K_m of 10^4 M^{-1} s^{-1} after several rounds of saturation and random mutagenesis, representing an impressive increase in activity, greater than 10^7-fold.

Nosrati and Houk developed the software tool *Selection of Active/Binding Sites for Enzyme Redesign* (SABER) to analyse the functional sites of the proteins stored in the Protein Data Bank [29]. The tool identifies the active sites amendable to computational redesign by locating potential catalytic residues in pre-defined spatial arrangements. The software was thoroughly validated by its identification of enzymes possessing the catalytic residues of o-succinyl benzoate synthase and designed Kemp eliminase.

Privett and co-workers presented an *iterative approach* for constructing a highly active enzyme catalysing Kemp elimination reactions [30•]. The initial design was analysed by protein crystallography and molecular dynamics, which revealed inactivity due to the presence of water molecules in the active site and the high flexibility of the active site residues. The mutagenesis focused deeper into the interior of the protein and resulted in the enzyme binding the transition state in an orientation flipped in relation to the design model.

Blomberg and co-workers applied error-prone polymerase chain reactions and DNA shuffling to identify hot spots along the gene of designed Kemp eliminase, which were subsequently mutagenized by focused mutagenesis [31••,32]. The designed enzyme was evolved to accelerate the chemical reaction 6×10^8-fold, approaching the catalytic efficiency of highly optimized natural enzymes. The crystal structure of the evolved enzyme was determined to a 1.09 Å resolution and analysed using the software tool *CAVER* [18•]. Future efforts to turn evolved enzymes into perfect catalysts will focus on optimizing protein dynamics.

Designing and engineering enzyme selectivity

Bioinformatics

The application of traditional bioinformatic tools for the design of selective enzymes is difficult because primary sequences do not contain sufficient information for describing the spatial interactions of a substrate molecule with the active site residues. Bioinformatic tools can be used to identify interaction hot spots and restrict the alphabet of substitutions for the design of smart libraries in focused directed evolution [2], which is outside the scope of this review.

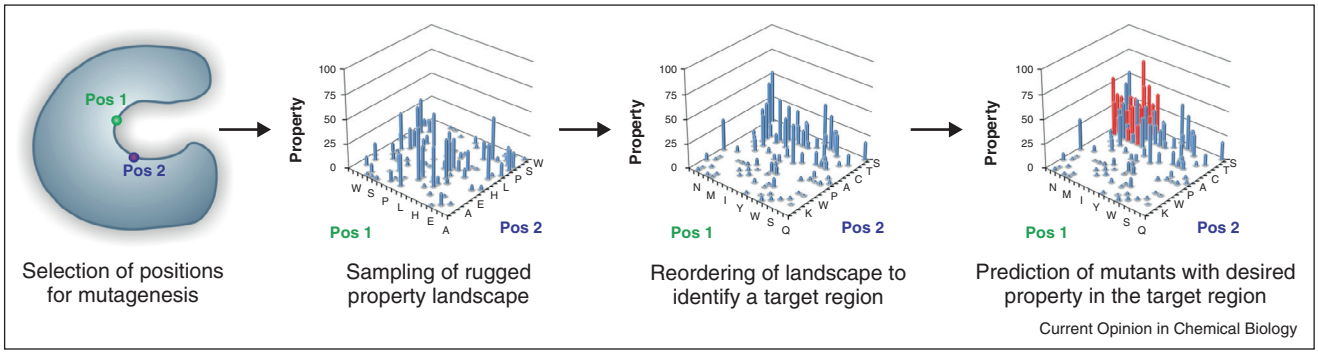
Directed evolution experiments can be rationalized by the development of predictive models trained on experimental data, similar to Quantitative Structure-Activity Relationships. Feng and co-workers developed the *Adaptive Substituent Reordering Algorithm* (ASRA) which can be employed in combination with many directed evolution methods [33••]. ASRA identifies the underlying regularity of the protein property landscape and makes predictions about the properties of uncharacterized proteins (Figure 2). Importantly, ASRA does not require assumptions of linearity, additivity, or any functional form of structure–property relationships. The application of ASRA was demonstrated with epoxide hydrolases, for which the method provided reliable predictions of multiple mutants with improved enantioselectivity. This approach should also be applicable to other properties, such as enzyme activity and stability.

Molecular modelling

Pratterr and co-workers applied molecular docking, molecular dynamics, and quantum mechanical calculations for the rational redesign of the substrate ligand at the metal centre of mononuclear non-heme iron(II)-dependent hydrolase. Construction of only 10 enzyme variants resulted in a novel enzyme showing a 9300-fold enantioselectivity switch [34•]. Durmaz and co-workers employed molecular docking, classical molecular dynamics, and steered molecular dynamics for engineering the chain-length specificity of lipase [35]. The single-point mutation L360F lowered the activation barrier for hydrolysis of C4 substrates, while the same mutation increased this barrier for C8 substrates.

The *empirical valence bond* modelling technique was used for a quantitative analysis of enantioselectivity of lipase by Frushicheva and Warshel [36•]. The authors evaluated both k_{cat} and k_{cat}/K_m for individual enantiomers of 4-nitrophenyl 2-methylheptanoate and concluded that extensive sampling is essential for obtaining converging

Figure 2



Workflow for adaptive substituent reordering algorithm identifying mutants with desired properties [33••].

results. They further outlined the use of the *linear response approximation* approach to analyse the contributions of each residue to the free energy corresponding to enzyme enantioselectivity.

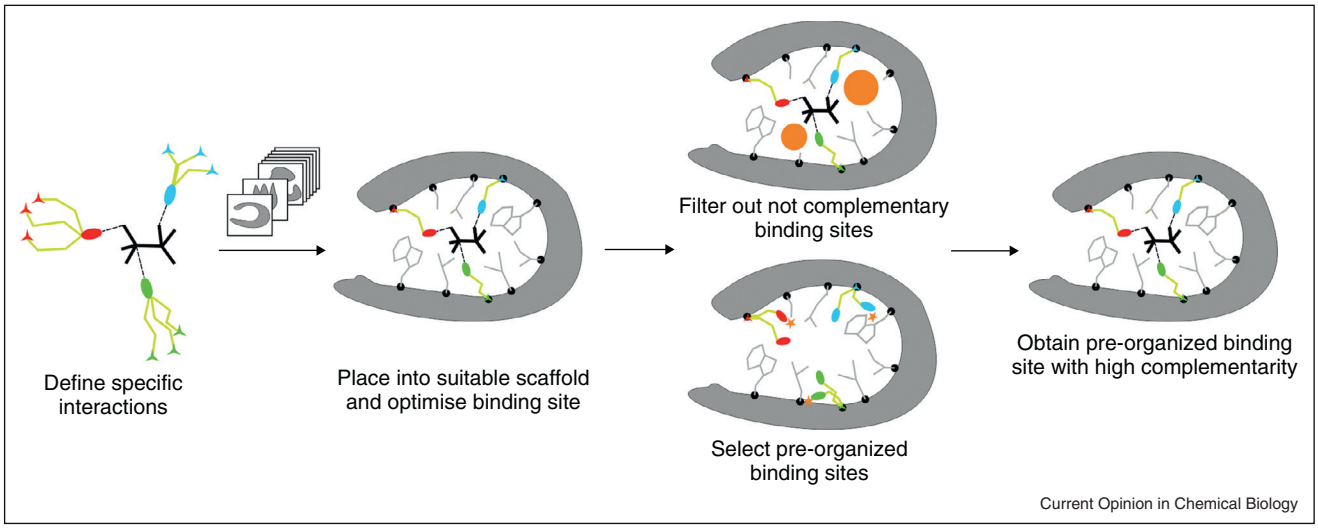
De novo design

The design of biocatalysts binding small ligands with good affinity is a very challenging problem, requiring precise calculation of rather weak protein–ligand interactions. An optimized binding site must provide: Firstly, specific energetically favourable hydrogen-bonding and van der Waals interactions with the ligand; secondly, high overall shape complementarity to the ligand; and finally, structural pre-organization in the unbound protein state to minimize entropy loss upon ligand binding. Tinberg and co-workers [37••] demonstrated the first successful *de novo* design of ligand-binding protein with a low-to-mid micromolar range (Figure 3). The designed binder of digoxin could be further optimized by site-saturation mutagenesis and selections using yeast surface display and fluorescence-activated cell sorting, reaching a picomolar level of binding affinity and high selectivity.

A remarkable 116-fold improvement in the catalytic activity of α-gliadin peptidase with gluten tetrapeptide and an 877-fold switch in the enzyme's substrate specificity were achieved by identifying an enzyme with pre-existing activity and optimizing it with computational protein design [38]. The engineered enzyme can be produced with a good yield of 30 mg/L and can serve as a potential oral therapeutic for celiac disease.

A new method for the computational design of binding pockets for small molecules, *POCKETOPTIMIZER*, was developed by Malisi and co-workers [39•]. The tool can be used to modify the residues making up the protein

Figure 3



Workflow for *de novo* design of ligand binding protein [37••].

binding pocket to improve or newly establish the binding of a small ligand. The programme employs a protein–ligand scoring function for estimating the free energy of binding using the molecular modelling programs *AUTODOCK VINA* and *CADDSUITE*, and *AMBER* to analyse protein packing. The method was tested with a benchmark set consisting of proteins with available mutants showing different binding affinities with their ligands and known structures. The tool predicts correctly increased affinity in 66% and 69% using *CADDSUITE* and *AUTODOCK VINA*, respectively.

Designing and engineering enzyme stability

Bioinformatics

A *consensus-based design* of thermostable proteins uses evolutionary information from multiple sequence alignments to predict the most suitable — most often naturally occurring — amino acids at a particular position. Anbar and co-workers successfully applied a consensual approach to engineering endoglucanase with a 14-fold improved half-life at 85 °C [40]. The regions responsible for improved thermostability could be identified by subsequent molecular dynamic simulations. Blum and co-workers combined *structure-guided consensus* with the *B-FIT* method to engineer thermostable α-amino ester hydrolase with improved T_{50} by 7 °C and 1.3-fold improved activity compared to wild-type enzyme [41].

Sullivan and co-workers addressed a problem with the low reliability of the *consensual approach* by distinguishing stabilizing from destabilizing mutations [42]. Consensus mutations at more conserved positions were more likely to be stabilizing in the model protein triosephosphate isomerase, while mutations at highly correlated positions were destabilizing. The authors suggest the exclusion of the sites with high statistical correlations to other sites and nearly invariant positions from the consensus design. The application of this approach to the model protein improved its thermostability by 8 °C. Wang and co-workers developed the *combinatorial coevolving-site saturation mutagenesis* (CCSM) method for identifying hotspots for mutagenesis [43]. The method targets functionally correlated variation sites. It was validated with α-amylase, for which thermal stability was improved by 8 °C. We note that the underlying principle of CCSM is opposite to the approach proposed by Sullivan and co-workers [42].

Construction of chimeric proteins is another well-established approach of protein stabilization, mimicking the process of DNA recombination. Romero and co-workers demonstrated that the protein fitness landscape can be efficiently inferred from experimental data using *Gaussian processes* [44••]. The fitness landscape describes how protein contributes to organismal fitness, or it may represent its biophysical properties, such as stability, catalytic activity, and specificity. The authors developed two different sequence design algorithms based on *Bayesian decision theory* (Figure 4). The first algorithm identifies small sets of sequences that are informative about the landscape, and the second algorithm identifies highly optimized sequences. Using these algorithms and the data set of 261 sequences with known properties, the authors engineered chimeric P450 enzymes that were more thermostable than any mutants previously prepared by chimeragenesis, rational design, or directed evolution. The method should be applicable to modelling other enzyme properties as well. A related approach was employed to identify seven functional chimeras of arginases from the *SCHEMA*-based library [45].

Shanmugaratnam and co-workers developed an interesting approach allowing construction of chimeric proteins combining the fragments from different protein folds [46]. Using proteins belonging to flavodoxin-like and $(\beta\alpha)_8$-barrel folds as a model system, the authors demonstrated the power of recombination for diversifying protein structures.

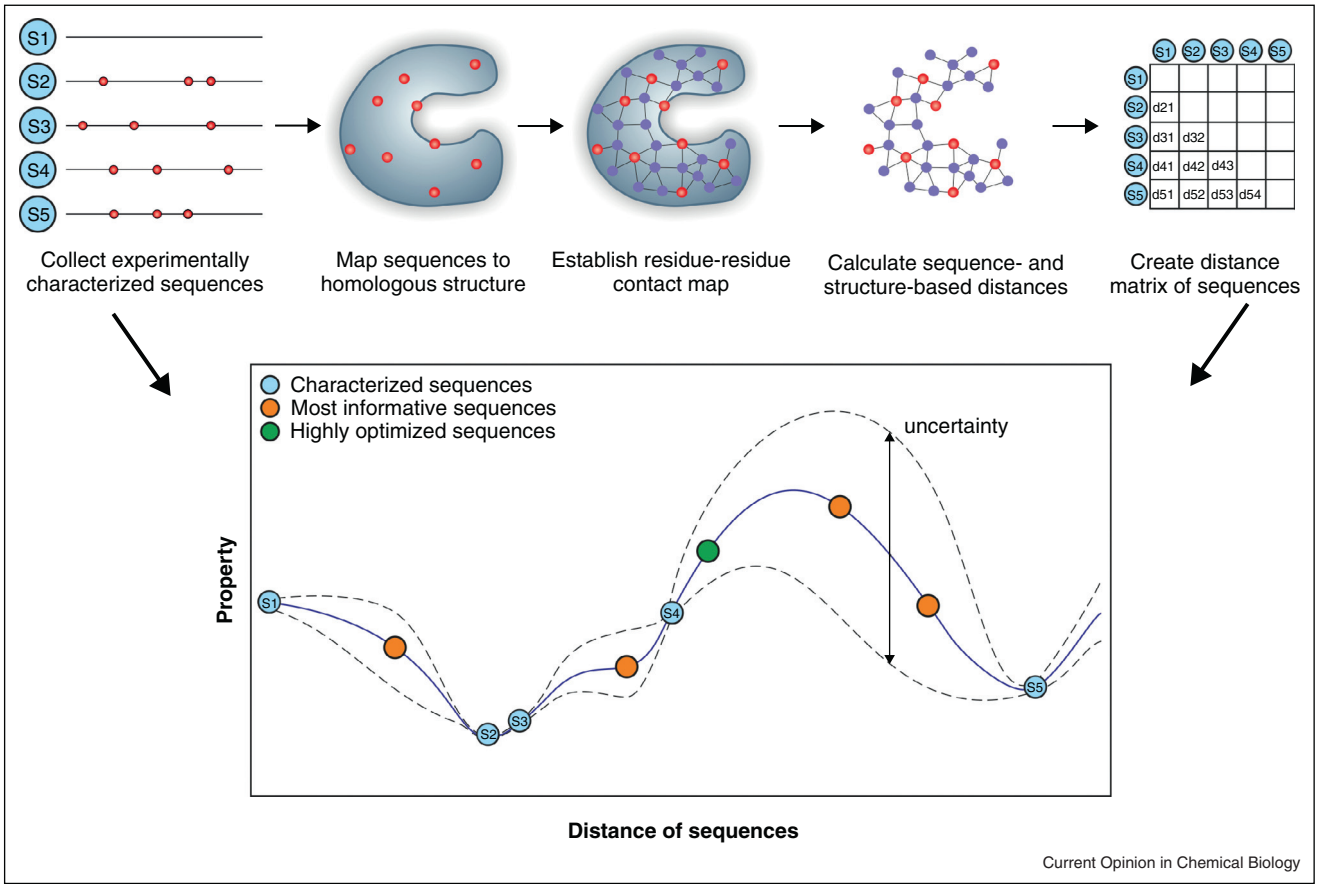
Molecular modelling

Tian and co-workers applied their software tools *PRETHERMUT* to evaluate the free energy of unfolding and *PAREPRO* to calculate the site evolutionary entropy of methyl parathion hydrolase [47]. Seven selected positions were saturated, leading to the identification of six positions capable of enhancing protein stability. A step-wise recombination of these mutations resulted in a four-point mutant with improved melting temperature (T_m) by 11.7 °C. A similar method, combining a consensual approach with *FOLDX* calculation, provided cellobiohydrolase I with T_{50} improved by 4.7 °C [48]. Jacobs and co-workers combined a consensus-based approach with *POPMUSIC* calculations to design stable FN3 domains of two different proteins [49]. The final constructs showed increased stability, high expression levels, and good solubility in *Escherichia coli*.

Raghunathan and co-workers proposed engineering surface charges to modify protein stability and aggregation properties [50]. The authors co-introduced stabilizing and surface-charge modifying mutations and increased the thermostability of a green fluorescent protein without compromising its functional properties.

Kim and co-workers applied network analysis to the protein structure for identifying hydrophobic interaction clusters [51]. A network parameter of structural hierarchy, k of k-clique, was used to predict stabilizing mutations of xylanase. The melting temperature of the triple mutant was increased by 8 °C, while retaining 74% of enzyme activity. The ensemble-based *Constraint Network Analysis* (CNA) was used to investigate the relationship between flexibility and thermostability of five citrate synthases by Rathi and co-workers [52]. The authors showed a good ability of CNA to identify a set of weak spots in protein

Figure 4



Workflow for property landscape analysis using the Gaussian process [44••].

structures that are suitable targets for stability pursuing mutagenesis.

Koudelakova and co-workers demonstrated that as few as four mutations introduced into the access tunnel of haloalkane dehalogenase increased its thermostability by 19 °C and extended the half-life in 40% dimethyl sulfoxide from minutes to weeks [53•]. *CAVER* [18•] calculations were used to analyse protein tunnels and molecular dynamic simulations were used to analyse protein accessibility for the molecules of organic co-solvents. *FOLDX* analysis carried for more than 220 000 mutations in 26 enzymes from all six enzyme classes confirmed that the concept is generally applicable to enzymes with buried active sites.

De novo design

Borgo and Havranek developed an automated protocol *ROSETTA VIP* — void identification and packing — to improve poorly packed protein cores [54]. The protocol uses the *ROSETTA HOLES* analysis module and a simple geometric scoring function to identify a small set of mutations that may yield improved packing. The protocol is applicable for stabilizing both designed and native proteins against chemicals and thermal denaturation.

Protein *WISDOM* is a web-based tool integrating methods for various protein design problems, including *de novo* design [55]. The tool enables searching for templates, designing optimized sequences with stability, analysing fold specificity and binding affinity, and quantitative assessment of the designs by ranking of sequences as well as structures. *EVODESIGN* is another web-based tool for designing optimal protein sequences of given scaffolds while predicting multiple sequence and structure-based features for design ranking [56•]. The tool uses *Metropolis Monte-Carlo* search of profiles constructed for homologous structure families in the Protein Data Bank. The set of local structure features, including secondary structures, torsion angles, and solvation, are predicted by neural-network training for optimization of structure packing. These tools will make *de novo* protein design accessible to a wider community due to their user-friendly web interfaces. Analogously, *PYROSETTA* Toolkit [57•] provides a

graphical user interface for preparing and running protocols of *ROSETTA* and for data analysis.

Conclusions and outlook

Protein engineering is one of the most dynamically developing scientific fields. The way proteins are being engineered has changed dramatically over the last few decades, primarily due to novel experimental technologies such DNA cloning, high throughput and deep sequencing, directed evolution methods, fluorescence-based sorting technologies, and gene synthesis. *In silico* approaches assisting protein design and engineering are being developed back to back with experimental techniques.

The *bioinformatics* approaches are most successfully used for engineering protein stability. This is because low resolution data and evolutionary information are sufficient for identifying stabilizing mutations, but less suitable for predicting mutations determining enzyme specificity or activity. Predictions of stable consensual, ancestral, and chimeric sequences are particularly popular. The great challenge for bioinformatics analysis is the continuous growth of sequences in genomic databases and the urgent need for predicting a protein function from a sequence alone. Analysis of sequences within enzyme subfamilies and identification of SSPs, developed for protein engineering applications, can possibly find a use in these functional assignments. Newly developed methods resembling quantitative-structure activity relationships should allow the prediction of optimized sequences from experimental data collected during directed evolution experiments.

The *molecular modelling* approaches greatly benefit from growing computational power and parallelized calculations on graphical cards. Larger molecular systems can be studied by quantum mechanics and longer simulation times can be achieved by molecular dynamics. Molecular modelling studies often combine several *in silico* methods, including bioinformatics analysis, to describe structure–function properties and predict beneficial mutations. A typical example of this is the prediction of thermostable proteins by combining the calculation of Gibbs free energies with evolutionary analyses. Current challenges include the quantitative modelling of selectivities and activities, which require the precise estimation of binding energies and reaction activation barriers. The quality of force fields properly describing the polarizability of atoms and conformational behaviour of amino acid residues as well as extensive sampling can be important for obtaining quantitative results. We expect even closer integration of molecular modelling methods with *de novo* design in near future.

The *de novo protein design* approach made the greatest progress of the three reviewed categories over the last two years. The approach was successfully used for the design of enzymes catalysing single step and multi-step chemical reactions, and for the design of a ligand binding site. Graphical user interfaces and web servers are being developed, making the *de novo* design accessible to a wider community. The current challenges are sorting successful designs from non-successful ones and designing biocatalysts to match the catalytic efficiencies of natural enzymes. *De novo* design should be also demonstrated for a wider range of chemical reactions, including those with more complex mechanisms. The design of active sites for chemical reactions can be complemented by the design of ligand transport pathways. Iterative cycles of *de novo* design, structural analysis, and molecular modelling can be further extended by transient kinetics and studies of kinetic isotopic effects, dissecting individual reaction steps, and providing valuable mechanistic information for further improvement of modelling protocols. Designed enzymes need to be improved by many rounds of directed evolution, and this will not change in the near future.

Acknowledgements

The authors are supported by grants from the European Regional Development Fund (CZ.1.05/2.1.00/01.0001), the Grant Agency of the Czech Republic (P503/12/0572), and the European Union Framework Programme (NEWPROT 289350). The work of JB is supported by the 'Employment of the Best Young Scientists for International Cooperation Empowerment' Programme (CZ1.07/2.3.00/30.0037), co-financed from European Social Fund and the state budget of the Czech Republic.

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. •• Bornscheuer UT, Huisman GW, Kazlauskas RJ, Lutz S, Moore JC, Robins K: **Engineering the third wave of biocatalysis**. *Nature* 2012, **485**:185-194.

This review describes the major developments in protein engineering, including examples of successfully engineered proteins. New concepts are highlighted and further challenges are outlined.

2. Sebestova E, Bendl J, Brezovsky J, Damborsky J: **Computational tools for designing smart libraries**. In *Directed Evolution Library Creation: Methods and Protocols*. Edited by Gillam EMJ, Copp JN, Ackerley DF. New York: Humana Press; 2013.

3. Davids T, Schmidt M, Böttcher D, Bornscheuer UT: **Strategies for the discovery and engineering of enzymes for biocatalysis**. *Curr Opin Chem Biol* 2013, **17**:215-220.

4. Davey JA, Chica RA: **Multistate approaches in computational protein design**. *Protein Sci* 2012, **21**:1241-1252.

5. Hilvert D: **Design of protein catalysts**. *Annu Rev Biochem* 2013, **82**:447-470.

6. Khare SD, Fleishman SJ: **Emerging themes in the computational design of novel enzymes and protein–protein interfaces**. *FEBS Lett* 2013, **587**:1147-1154.

7. Kries H, Blomberg R, Hilvert D: ***De novo* enzymes by computational design**. *Curr Opin Chem Biol* 2013, **17**:221-228.

8. Wijma HJ, Janssen DB: **Computational design gains momentum in enzyme catalysis engineering**. *FEBS J* 2013, **280**:2948-2960.

9. 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.

10. Bommarius AS, Paye MF: **Stabilizing biocatalysts**. *Chem Soc Rev* 2013, **42**:6534-6565.

11. Socha RD, Tokuriki N: **Modulating protein stability – directed evolution strategies for improved protein function**. *FEBS J* 2013, **280**:5582-5595.

12. Stepankova V, Bidmanova S, Koudelakova T, Prokop Z, Chaloupkova R, Damborsky J: **Strategies for stabilization of enzymes in organic solvents**. *ACS Catal* 2013 http://dx.doi.org/10.1021/cs400684x.

13. Suplatov D, Kirilin E, Takhaveev V, Svedas V: **Zebra: a web server for bioinformatic analysis of diverse protein families**. *J Biomol Struct Dyn* 2013 http://dx.doi.org/10.1080/07391102.2013.834514.

14. Suplatov DA, Besenmatter W, Svedas VK, Svendsen A: **Bioinformatic analysis of α/β-hydrolase fold enzymes reveals subfamily-specific positions responsible for discrimination of amidase and lipase activities**. *Protein Eng Des Sel* 2012, **25**:689-697.

15. Addington TA, Mertz RW, Siegel JB, Thompson JM, Fisher AJ, Filkov V, Fleischman NM, Suen AA, Zhang C, Toney MD: **Janus: prediction and ranking of mutations required for functional interconversion of enzymes**. *J Mol Biol* 2013, **425**:1378-1389.

16. Yang J-S, Seo SW, Jang S, Jung GY, Kim S: **Rational engineering of enzyme allosteric regulation through sequence evolution analysis**. *PLoS Comput Biol* 2012, **8**:e1002612.

17. Biedermannova L, Prokop Z, Gora A, Chovancova E, Kovacs M, Damborsky J, Wade RC: **A single mutation in a tunnel to the active site changes the mechanism and kinetics of product release in haloalkane dehalogenase LinB**. *J Biol Chem* 2012, **287**:29062-29074.

18. • Chovancova E, Pavelka A, Benes P, Strnad O, Brezovsky J, Kozlikova B, Gora A, Sustr V, Klvana M, Medek P *et al.*: **CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures**. *PLoS Comput Biol* 2012, **8**:e1002708.
The article introduced a new version of the software tool suitable for analysis of tunnels and channels in dynamical protein structures. The software implements several new algorithms for tunnel identification and clustering.

19. Gora A, Brezovsky J, Damborsky J: **Gates of enzymes**. *Chem Rev* 2013, **113**:5871-5923.

20. Brezovsky J, Chovancova E, Gora A, Pavelka A, Biedermannova L, Damborsky J: **Software tools for identification, visualization and analysis of protein tunnels and channels**. *Biotechnol Adv* 2013, **31**:38-49.

21. Khersonsky O, Kiss G, Roethlisberger D, Dym O, Albeck S, Houk KN, Baker D, Tawfik DS: **Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59**. *Proc Natl Acad Sci U S A* 2012, **109**:10358-10363.

22. Althoff EA, Wang L, Jiang L, Giger L, Lassila JK, Wang Z, Smith M, Hari S, Kast P, Herschlag D *et al.*: **Robust design and optimization of retroaldol enzymes**. *Protein Sci* 2012, **21**:717-726.

23. • Giger L, Caner S, Obexer R, Kast P, Baker D, Ban N, Hilvert D: **Evolution of a designed retro-aldolase leads to complete active site remodeling**. *Nat Chem Biol* 2013, **9**:494-498.
Random mutagenesis applied to the designed retro-aldolase resulted in 4.4×10^3-fold activity improvement. Structural analysis of initial design and evolved variants revealed the importance of loop mobility and supporting functional groups.

24. Hallen MA, Keedy DA, Donald BR: **Dead-end elimination with perturbations (DEEPer): a provable protein design algorithm with continuous sidechain and backbone flexibility**. *Proteins* 2013, **81**:18-39.

25. Nivón LG, Moretti R, Baker D: **A Pareto-optimal refinement method for protein design scaffolds**. *PLoS ONE* 2013, **8**:e59004.

26. Keedy DA, Georgiev I, Triplett EB, Donald BR, Richardson DC, Richardson JS: **The role of local backrub motions in evolved and designed mutations**. *PLoS Comput Biol* 2012, **8**:e1002629.

27. • Renfrew PD, Choi EJ, Bonneau R, Kuhlman B: **Incorporation of noncanonical amino acids into Rosetta and use in computational protein–peptide interface design**. *PLoS ONE* 2012, **7**:e32637.
The article describes the incorporation of 114 noncanonical amino acids into the Rosetta modelling suite by building backbone-dependent rotamer libraries and the parameterization and construction of a scoring function.

28. • Khare SD, Kipnis Y, Greisen PJ, Takeuchi R, Ashani Y, Goldsmith M, Song Y, Gallaher JL, Silman I, Leader H *et al.*: **Computational redesign of a mononuclear zinc metalloenzyme for organophosphate hydrolysis**. *Nat Chem Biol* 2012, **8**:294-300.
The authors converted a zinc-dependent adenosine deaminase into an organophosphate hydrolase by *de novo* protein design and evolutionary evolution. The very low activity of the scaffold protein was increased $>10^7$-fold.

29. Nosrati GR, Houk KN: **SABER: a computational method for identifying active sites for new reactions**. *Protein Sci* 2012, **21**:697-706.

30. • Privett HK, Kiss G, Lee TM, Blomberg R, Chica RA, Thomas LM, Hilvert D, Houk KN, Mayo SL: **Iterative approach to computational enzyme design**. *Proc Natl Acad Sci U S A* 2012, **109**:3790-3795.
The article describes the power of an iterative approach combining *de novo* design, molecular modelling, and protein crystallography. The structural analysis revealed binding of the transition state in the orientation flipped to the design model.

31. •• Blomberg R, Kries H, Pinkas DM, Mittl PRE, Grütter MG, Privett HK, Mayo SL, Hilvert D: **Precision is essential for efficient catalysis in an evolved Kemp eliminase**. *Nature* 2013 http://dx.doi.org/10.1038/nature12623.
This study demonstrates that effective placement of appropriate catalytic groups in the right active site environment can bridge the efficiency gap between natural and designed enzymes. The evolved enzyme had k_{cat}/K_m of 230 000 M^{-1} s^{-1} and k_{cat} of 700 s^{-1}, which is above the activity of many natural enzymes.

32. Höhne M, Bornscheuer UT: **Protein engineering from 'scratch' becoming mature**. *Angew Chem Int Ed Engl* 2013 http://dx.doi.org/10.1002/anie.201309591.

33. •• Feng X, Sanchis J, Reetz MT, Rabitz H: **Enhancing the efficiency of directed evolution in focused enzyme libraries by the adaptive substituent reordering algorithm**. *Chem Eur J* 2012, **18**:5646-5654.
The authors present a new method, complementing a directed evolution, which identifies the underlying regularity of the protein property landscape and makes predictions about the property of uncharacterized proteins.

34. • Pratter SM, Konstantinovics C, Di Giuro CML, Leitner E, Kumar D, de Visser SP, Grogan G, Straganz GD: **Inversion of enantioselectivity of a mononuclear non-heme iron(II)-dependent hydroxylase by tuning the interplay of metal-center geometry and protein structure**. *Angew Chem Int Ed Engl* 2013, **52**:9677-9681.
Using molecular docking, molecular dynamics, and quantum mechanics, the authors redesigned the substrate ligand at the metal centre of a non-heme iron(II)-dependent hydrolase and achieved 9.3×10^3-fold change in enantioselectivity.

35. Durmaz E, Kuyucak S, Sezerman UO: **Modifying the catalytic preference of tributyrin in *Bacillus thermocatenulatus* lipase through in-silico modeling of enzyme–substrate complex**. *Protein Eng Des Sel* 2013, **26**:325-333.

36. • Frushicheva MP, Warshel A: **Towards quantitative computer-aided studies of enzymatic enantioselectivity: the case of *Candida antarctica* lipase A**. *ChemBioChem* 2012, **13**:215-223.
The study describes the use of the empirical valence bond method and linear response approximation for analysing the enantioselectivity of lipase. It is concluded that extensive sampling is essential for obtaining converging results.

37. •• Tinberg CE, Khare SD, Dou J, Doyle L, Nelson JW, Schena A, Jankowski W, Kalodimos CG, Johnsson K, Stoddard BL *et al.*: **Computational design of ligand-binding proteins with high affinity and selectivity**. *Nature* 2013, **501**:212-216.
The authors demonstrate the *de novo* design of digoxenin-binding protein with micromolar affinity. The designed protein was further optimized by

saturation mutagenesis and yeast surface display, reaching a picomolar level of binding affinity.

38. Gordon SR, Stanley EJ, Wolf S, Toland A, Wu SJ, Hadidi D, Mills JH, Baker D, Pultz IS, Siegel JB: **Computational design of an alpha-gliadin peptidase**. *J Am Chem Soc* 2012, **134**:20513-20520.

39. • Malisi C, Schumann M, Toussaint NC, Kageyama J, Kohlbacher O, Höcker B: **Binding pocket optimization by computational protein design**. *PLoS ONE* 2012, **7**:e52505.
The article introduces a benchmark-set for binding pocket optimization. The set is used to develop a new method correctly predicting affinity-changing mutations with 69% reliability.

40. Anbar M, Gul O, Lamed R, Sezerman UO, Bayer EA: **Improved thermostability of *Clostridium thermocellum* endoglucanase Cel8A by using consensus-guided mutagenesis**. *Appl Environ Microbiol* 2012, **78**:3458-3464.

41. Blum JK, Ricketts MD, Bommarius AS: **Improved thermostability of AEH by combining B-FIT analysis and structure-guided consensus method**. *J Biotechnol* 2012, **160**:214-221.

42. Sullivan BJ, Nguyen T, Durani V, Mathur D, Rojas S, Thomas M, Syu T, Magliery TJ: **Stabilizing proteins from sequence statistics: the interplay of conservation and correlation in triosephosphate isomerase stability**. *J Mol Biol* 2012, **420**:384-399.

43. Wang C, Huang R, He B, Du Q: **Improving the thermostability of alpha-amylase by combinatorial coevolving-site saturation mutagenesis**. *BMC Bioinformatics* 2012, **13**:263.

44. •• Romero PA, Krause A, Arnold FH: **Navigating the protein fitness landscape with Gaussian processes**. *Proc Natl Acad Sci U S A* 2013, **110**:E193-E201.
The authors developed novel algorithms for modelling protein fitness landscapes using Gaussian processes, using them to predict chimeric P450 with T_{50} of 69.7 °C, which is 8.7 °C more stable than the variants obtained by directed evolution.

45. Romero PA, Stone E, Lamb C, Chantranupong L, Krause A, Miklos AE, Hughes RA, Fechtel B, Ellington AD, Arnold FH *et al.*: **SCHEMA-designed variants of human arginase I and II reveal sequence elements important to stability and catalysis**. *ACS Synth Biol* 2012, **1**:221-228.

46. Shanmugaratnam S, Eisenbeis S, Höcker B: **A highly stable protein chimera built from fragments of different folds**. *Protein Eng Des Sel* 2012, **25**:699-703.

47. Tian J, Wang P, Huang L, Chu X, Wu N, Fan Y: **Improving the thermostability of methyl parathion hydrolase from *Ochrobactrum* sp. M231 using a computationally aided method**. *Appl Microbiol Biotechnol* 2013, **97**:2997-3006.

48. Komor RS, Romero PA, Xie CB, Arnold FH: **Highly thermostable fungal cellobiohydrolase I (Cel7A) engineered using predictive methods**. *Protein Eng Des Sel* 2012, **25**:827-833.

49. Jacobs SA, Diem MD, Luo J, Teplyakov A, Obmolova G, Malia T, Gilliland GL, O'Neil KT: **Design of novel FN3 domains with high stability by a consensus sequence approach**. *Protein Eng Des Sel* 2012, **25**:107-117.

50. Raghunathan G, Sokalingam S, Soundrarajan N, Madan B, Munussami G, Lee S-G: **Modulation of protein stability and aggregation properties by surface charge engineering**. *Mol Biosyst* 2013, **9**:2379-2389.

51. Kim T, Joo JC, Yoo YJ: **Hydrophobic interaction network analysis for thermostabilization of a mesophilic xylanase**. *J Biotechnol* 2012, **161**:49-59.

52. Rathi PC, Radestock S, Gohlke H: **Thermostabilizing mutations preferentially occur at structural weak spots with a high mutation ratio**. *J Biotechnol* 2012, **159**:135-144.

53. • Koudelakova T, Chaloupkova R, Brezovsky J, Prokop Z, Sebestova E, Hesseler M, Khabiri M, Plevaka M, Kulik D, Kuta Smatanova I *et al.*: **Engineering enzyme stability and resistance to an organic cosolvent by modification of residues in the access tunnel**. *Angew Chem Int Ed Engl* 2013, **52**:1959-1963.
The article introduces a new concept of protein stabilization by engineering protein access tunnels. Four substitutions in the tunnel of dehalogenase increased its thermostability by 19 °C and extended the half-life in 40% DMSO from minutes to weeks.

54. Borgo B, Havranek JJ: **Automated selection of stabilizing mutations in designed and natural proteins**. *Proc Natl Acad Sci U S A* 2012, **109**:1494-1499.

55. Smadbeck J, Peterson MB, Khoury GA, Taylor MS, Floudas CA: **Protein WISDOM: a workbench for *in silico de novo* design of biomolecules**. *J Vis Exp* 2013 http://dx.doi.org/10.3791/50476.

56. • Mitra P, Shultis D, Zhang Y: **EvoDesign: *de novo* protein design based on structural and evolutionary profiles**. *Nucleic Acids Res* 2013, **41**:W273-W280.
The authors developed a web-based tool for designing optimal protein sequences of given scaffolds with predictions of multiple sequences and structure-based features for design ranking.

57. • Adolf-Bryfogle J, Dunbrack RL Jr: **The PyRosetta Toolkit: a graphical user interface for the Rosetta software suite**. *PLoS ONE* 2013, **8**:e66856.
The article presents a graphical user interface for preparing and running protocols of Rosetta modelling suite. The toolkit makes *de novo* design accessible to a wider scientific community.