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

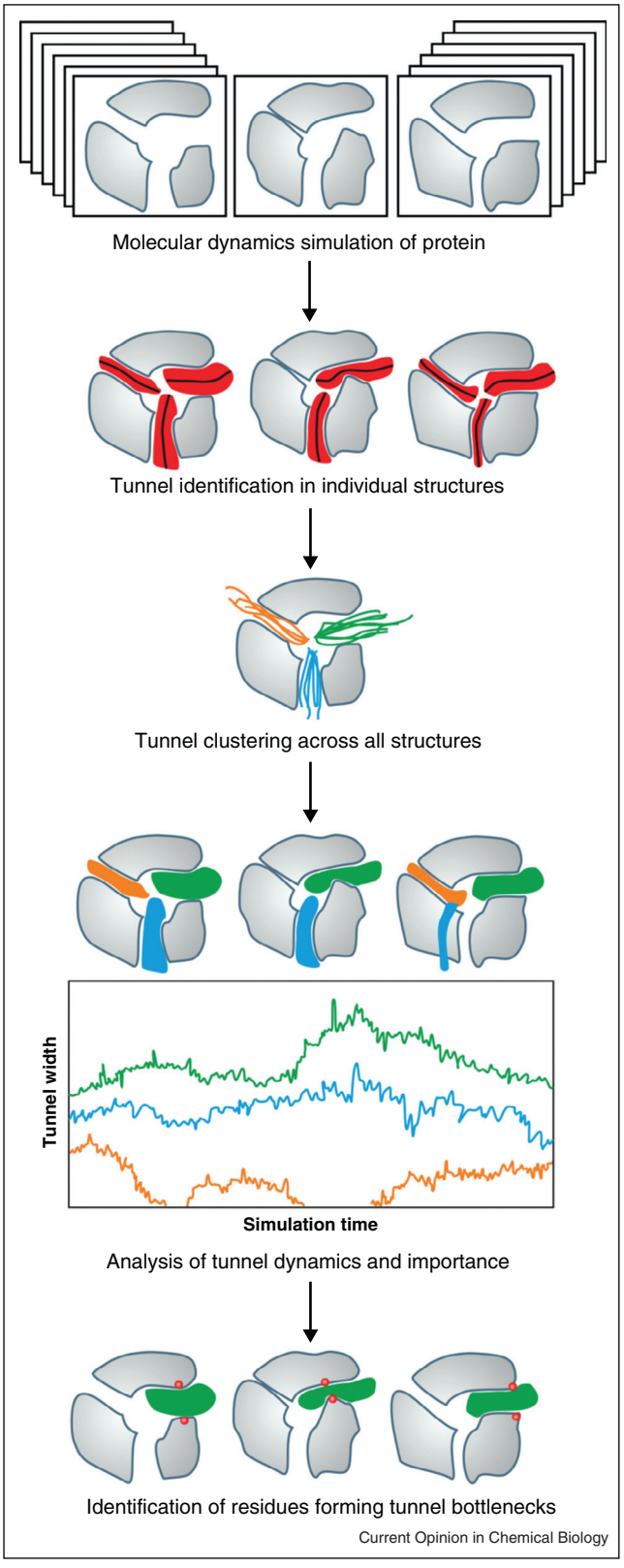
Biedermannova and co-workers combined several molecular modelling methods to study the effect of tunnel mutations on kinetics and reaction mechanisms of haloalkane dehalogenase [17]. The software tool *CAVER* [18•] was used to analyse tunnel dynamics in trajectories obtained by molecular dynamic simulations (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•].