Archives of Biochemistry and Biophysics xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Archives of Biochemistry and Biophysics

journal homepage: www.elsevier.com/locate/yabbi

Review

Computational strategies for the design of new enzymatic functions

K. Świderek [a,b], I. Tuñón [a], V. Moliner [c], J. Bertran [d,*]

[a] Departament de Química Física, Universitat de València, 46100 Burjasot, Spain
[b] Institute of Applied Radiation Chemistry, Lodz University of Technology, 90-924 Lodz, Poland
[c] Departament de Química Física i Analítica, Universitat Jaume I, 12071 Castellón, Spain
[d] Departament de Química, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain

ARTICLE INFO

Article history:
Received 26 November 2014
and in revised form 9 March 2015
Available online xxxx

Keywords:
Protein design
De novo design
QM/MM
Molecular dynamics
Kemp elimination
Diels–Alder
Retro-Aldolase

ABSTRACT

In this contribution, recent developments in the design of biocatalysts are reviewed with particular emphasis in the *de novo* strategy. Studies based on three different reactions, Kemp elimination, Diels–Alder and Retro-Aldolase, are used to illustrate different success achieved during the last years. Finally, a section is devoted to the particular case of designed metalloenzymes. As a general conclusion, the interplay between new and more sophisticated engineering protocols and computational methods, based on molecular dynamics simulations with Quantum Mechanics/Molecular Mechanics potentials and fully flexible models, seems to constitute the bed rock for present and future successful design strategies.

Introduction

Enzymes are biological catalysts that speed up chemical reactions making them compatible with life. Often, these catalysts show important advantages with respect to non-natural catalysts such as their chemo-, regio- and stereoselectivity and the ability to work under mild conditions of temperature and pressure. Although the knowledge about the origin of enzymatic efficiency to catalyze chemical reactions is still not complete, there have been numerous studies that have provided a solid understanding about some of the key factors in biocatalysis [1–4].

Many reactions employed in the chemical and biochemical industry need of efficient catalysts to improve their rate, their selectivity and their environmental friendship. Obviously enzymes are excellent candidates to fulfil these requirements but, in order to extend their applicability to different purposes under different conditions, they must be modified using protein engineering techniques. First, enzymes need to be stable enough in different conditions of pH, temperature, solvent composition and substrate concentration. In addition, many of the chemical reactions carried out in the industry do not have a natural counterpart and then new enzymes with new functions should be developed. During the last years, different protein designs have been proposed to catalyze new chemical reactions. These designs can be classified, according to the strategy used in their development, into those based on directed evolution, rational design, or a combination of them named as semi-rational design [5].

Directed evolution [6] applies to strategies based in natural evolution to tailor the properties of molecules instead of organisms [7]. Random mutations or recombination can be done to evolve proteins in the laboratory identifying the successful variants by screening or by selection. New proteins with new desired functions can be obtained after some mutations or recombining protein fragments. The advantage of this strategy is that no structural information is needed and that distant regions of the sequence space can be explored. However, in most cases changes introduced far from the active site provoke minor effects in the kinetic properties of the protein. Another limitation of this strategy is that a minimum threshold of activity is required to start the cycles of mutations and screening.

Although in the next future we should be able to design a particular amino acid sequence that will fold into a particular structure with the desired function, rational design refers, in these days, to the introduction of direct mutations of selected residues on specific positions of an already existing protein [8]. Mutations of few residues can lead to important changes in the active site while the global structure of the protein remains essentially unaltered. These mutations are driven by the analysis of the data obtained from different sources, ranging from X-ray to simulations.

* Corresponding author.
E-mail address: bertran@klingon.uab.es (J. Bertran).

http://dx.doi.org/10.1016/j.abb.2015.03.013
0003-9861/

In particular, when computational analysis of the enzymatic activity is used to guide the mutations, one usually refers to this strategy as computational design [9].

These two strategies can be combined into an efficient procedure for the design of new proteins with new functions. In the semi-rational approach a new function is obtained from rational design and the activity of the resulting protein is improved by means of several directed evolution cycles.

The strategies for the computational design of new enzymes can start from different protein scaffolds. These can be (i) immune-globulins, proteins that were already used to produce catalytic antibodies (CAs)[1] more than thirty years ago [10]; (ii) a promiscuous protein showing catalytic properties for a secondary reaction that is redesigned to improve or to change this activity; and (iii) a protein without specific catalytic properties that is used as a scaffold to support a design from scratch, or *de novo* design, of a new activity [9,11]. The first group of proteins, CAs, are obtained using as hapten a Transition State Analog (TSA), a stable molecule resembling the Transition State (TS) structure of the reaction. While showing certain catalytic activity, their efficiency is low compared to that of natural enzymes, which can be attributed to either the nature of immune-globulins, to the lack of adequacy of the TSA to the real TS of the target reaction, or to the fact that these proteins bind the products so tightly that is the major flaw in their ability to rival enzymes for catalytic efficiency. The use of these proteins in the design process seems to decline in the last years. Instead, other protein structures seem more adequate for the process of designing new biocatalysts. The active sites of natural enzymes are an obvious starting point for this process. These active sites, composed by few residues that promote a particular chemical reaction, can be redesigned for a new function after few mutations [12]. This redesign process can take advantage from the different promiscuous behaviors that can appear in a natural enzyme. Enzymes promiscuity with respect to different substrates, different catalytic activities or different reaction conditions provides the raw material for the redesign of new properties [13]. The promiscuous activity in natural enzymes, which could be already known or discovered by chance, can be also induced by rational design or directed evolution. Finally, *de novo* design is based in the knowledge of the chemical reaction to be catalyzed and, in particular, in the TS and the physicochemical principles governing its stabilization (see Fig. 1). This information is obtained from computational simulations. Mayo and co-workers pioneered the field of the *de novo* design of enzymes [14] converting the *Escherichia coli* protein thioredoxin into a primitive esterase. In particular, a histidine nucleophile was introduced into an engineered surface active site reaching a 180-fold rate acceleration. The process made use of the ORBIT program to identify enzyme-like active sites within a pre-existing protein scaffold [15]. This program explores simultaneously the conformational and sequence space of the designed biocatalysts [16]. The actual boom in *de novo* design starts with three significant works about three different reaction for which no natural enzymes are known: the Kemp elimination [17], the Diels–Alder reaction [18] and the Retro-Aldolase reaction [19]. In all the three cases, which will be described in the following sections, the design process employed the Rosetta software developed at University of Washington. A previous step for the design process is the knowledge of the reaction TS and the design of a minimalist active site to stabilize its charge distribution. This active site, often known as a 'theozyme' [20], consists of disembodied amino acids placed at adequate positions around the substrate to promote the reaction. At this stage, quantum mechanical methods are needed to properly localize the TS in this reduced model, as far as changes happen at the electronic level. In the next step the minimal active site model must be placed into an existing protein structure, what is done using the RosettaMatch module [21]. This module tests different protein scaffolds grafting the 'theozyme' at each of the possible active-site positions. After trying different rotamers, a match is obtained when the substrate can be positioned without clashing to the protein backbone and key interactions observed in the 'theozyme' are conserved. The third step of the process is the design of the resulting candidates to optimize the surrounding residues for catalysis [22,23]. This step consists in several cycles of sequence design and protein optimization. For this purpose protein residues are classified as designable or repackable, according to their distance to the ligand. The optimization or packing procedure is based in a simple repulsive potential. The last step consists in the ranking or scoring of the designed sequences [24,25]. The hundreds or thousands of resulting designs must be evaluated before production. With this purpose each of the resulting designs is scored using different criteria that include ligand affinity and protein stability. Only those better placed in the final ranking will be expressed and tried experimentally.

Different improvements have been incorporated to this strategy during the last years. Usually, programs for protein design consider fixed the backbone of the protein during the design and optimization steps [26] although, for instance, Rosetta program was recently modified to consider flexible backbones [27]. The optimization in the sequence space can be expanded considering also non-canonical amino acids [28–30]. Algorithm improvements have been also achieved taking advantages of player strategies and social evolution [31]. Obviously, although Rosetta is probably the best-known suite of programs for enzyme design, there exist alternatives. For instance, the matching of the ideal 'theozyme' into an existing protein scaffold can be also carried out with OptGraft [32], Scaffold-Selection [33], and PRODA_MATCH [34,35]. A different strategy is employed by the SABER program [36]. Instead of placing the ideal active site into selected proteins, this program searches for proteins already presenting the catalytic residues in the adequate relative positions, thus requiring the introduction of a lower number of mutations.

In spite of all these improvements the *de novo* design strategy still suffers from limitations, as demonstrated by the fact that the rate of successful designs is quite low and, in addition, the best ones display reaction rate enhancements far from those of natural enzymes. As pointed out by Baker [37], a design can fail at different levels: (i) the proposed 'theozyme' can be not a perfect model for the real TS of the enzymatic reaction, (ii) the active site designed into a given protein scaffold can be distorted with respect to the ideal geometry, and (iii) the effect of the long range electrostatic interactions and/or protein dynamics can be incompatible with catalysis. In addition, a more complete design strategy should not consider exclusively the TS stabilization but the whole reaction process and, in particular, the differential stabilization of the TS with respect to the reactants. In other words, the enzyme must be designed considering not only the preorganization of the active site around the TS but also the reorganization that occurs during the chemical process and that can impose an energy penalty during the barrier climbing process. In a review published in 2008, we already distinguished between "structure-based strategies" and "reaction analysis-based strategies" [7]. Enzyme designs based in a 'theozyme' would fall in the first category while the limitations pointed out in current strategies could be overcome with methods that make use of the analysis of reaction profiles. Obviously the computational cost of this second strategy, which is based in the use of molecular dynamics (MD) with hybrid Quantum

[1] *Abbreviations used:* CAs, catalytic antibodies; QM/MM, Quantum Mechanics/Molecular Mechanics; TSA, Transition State Analog; TS, Transition State; MD, molecular dynamics; FEP, free-energy perturbation; MC, Monte Carlo; DFT, density functional theory; EVB, Empirical Valence Bond; NAC, near attack-conformation; LmrR, Lactococcal multidrug resistance Regulator; NOR, nitric oxide reductase; HCO, heme copper oxidase.

Please cite this article in press as: K. Świderek et al., Arch. Biochem. Biophys. (2015), http://dx.doi.org/10.1016/j.abb.2015.03.013