# Computational design of biological catalysts

**Sergio Martí,[a] Juan Andrés,[a] Vicent Moliner,*[a] Estanislao Silla,[b] Iñaki Tuñón*[b] and Juan Bertrán[c]**

***Received 7th July 2008***
***First published as an Advance Article on the web 6th October 2008***
**DOI: 10.1039/b710705f**

The purpose of this *tutorial review* is to illustrate the way to design new and powerful catalysts. The first possibility to get a biological catalyst for a given chemical process is to use existing enzymes that catalyze related reactions. The second possibility is the use of immune systems that recognize stable molecules resembling the transition structure of the target reaction. We finally show how computational techniques are able to provide an enormous quantity of information, providing clues to guide the development of new biological catalysts.

## 1. Introduction

Scientists have always been looking at Nature in order to find a source of inspiration to mimic the great power and efficiency of its processes. One of these processes is reactions catalyzed by enzymes, very complex molecular machines capable of enhancing the rate constant of chemical reactions. The purpose of this review is to illustrate the way in which computational tools can be used to assist in the process of designing new and powerful catalysts, taking as reference natural enzymes. In this introduction we focus on two different kind of proteins that can be used as starting points to get new biological catalysts. The first possibility to get a biological catalyst for a given chemical process is to use existing enzymes that catalyze related reactions. As we discuss, enzymes present a typical feature of living organisms which is the evolvability, which means that they are not dead ends but their efficiency can be enhanced or new catalytic activities derived from existing enzymes. The second possibility is the use of immune antibodies, designed to recognize stable molecules resembling the transition structure of the target reaction. We show in this review how these materials can be engineered to tailor new activities using different strategies. As discussed in the second part of the paper, computational techniques are able to provide an enormous quantity of information about chemical reactions in complex media, providing then clues to guide the development of new and more efficient biological catalysts.

[a] *Departament de Química Física i Analítica, Universitat Jaume I, 12071 Castellón, Spain. E-mail: moliner@uji.es*
[b] *Departament de Química Física, Universitat de València, 46100 Burjasot, Spain. E-mail: ignacio.tunon@uv.es*
[c] *Departament de Química, Universitat Autónoma de Barcelona, 08193 Bellaterra, Spain*

### 1.1 Enzyme evolvability

Enzymes display enormous reaction rate enhancements in water at moderate pH values and mild pressures and temperatures, which correspond to the standard physiological conditions of living organisms. In fact, reactions with half times approaching the age of the Earth are accelerated by many orders of magnitude up to $10^{20}$, which is the exceptional case of arginine decarboxylase. Other examples can be found in the survival of paper documents or ancient ships for long periods under water which can be explained by the fact that the glycosidic bonds of cellulose are very resistant to hydrolysis in the absence of cellulases that catalyse their hydrolysis.[1]

The traditional view on enzymes holds that their catalytic activities, while optimized by evolution, also represent highly dead ends (one gene, one function).[2] However, it has been

**Vicent Moliner**

*Vicent Moliner obtained his PhD (1993) from the Universitat Jaume I, under the supervision of Juan Andrés. He did postdoctoral studies with Ian Williams (University of Bath, 1996–1997), and also with O. Tapia (Uppsala) and J. Krechl (Prague). He was appointed as Assistant Professor at the University Jaume I in 1994 and has been a Lecturer in Physical Chemistry since 1998. His research interest is mainly focused on theoretical studies of chemical reactivity in enzyme catalysis.*

**Iñaki Tuñón**

*Iñaki Tuñón obtained his PhD (1993) working on continuum models from the Universidad de Valencia, under the supervision of Prof. Silla. He followed with postdoctoral research in 1994 at the Université de Nancy with Prof. Rivail. After one year he got a postdoctoral position at the Universidad de Valencia and has been a Lecturer in Physical Chemistry since 1998. His research is focused on theoretical studies of chemical reactivity in enzymes and in solution.*

 

recently suggested that this paradigm, which has dominated thinking in this field in the past, could be too simplistic. In this sense, despite an enzyme being generally defined as a selective catalyst capable of differentiating between different substrates and speeding up the rate of a particular chemical reaction, some enzymes have been found to present promiscuous activity, accepting alternative substrates and catalyzing secondary reactions.[2–6] This promiscuity provides a raw starting point for the evolution of enzymes, as a new duplicated gene presenting low activity would be the germen for adaptive evolution.[3] In fact, new enzymatic functions can evolve over a period of years or even months, as recently happened as a response to new synthetic chemicals or drugs.[6] According to previous studies, promiscuous activities exhibit high plasticity as they can be readily increased by means of one or a few mutations, allowing reaching of the threshold value for being improved under selective pressure.[3] Instead, primary activity presents a large robustness against mutations.[6] As a consequence, the active sites of these existing enzymes provide obvious materials to engineer novel enzymes with new catalytic functions.[2]

Enzyme promiscuity can be classified in three different categories: substrate promiscuity or ambiguity (the enzyme accepts structurally distinct substrates but catalyzes the same chemical reaction), catalytic promiscuity (the enzyme accepts different substrates and catalyzes different overall reactions), and product promiscuity (the enzyme accepts a single substrate and uses similar chemical mechanisms to catalyze the formation of different products).[2]

An example of the substrate promiscuity can be illustrated with the cytochrome P450 (CYP), a vast superfamily of enzymes found in almost all life forms. One of the most striking characteristics of some CYPs is that individual enzymes can interact with numerous structurally diverse substrates. This broad specificity usually serves by its obvious benefits playing an important role in metabolism. In fact, it is thought that more than 90% of drug and chemical oxidations in human beings are mediated by these promiscuous enzymes, probably based on the fact that they are not restricted to a particular substrate.[7] In general, these microsomal enzymes determine the bioavailability of drug molecules by converting them to more soluble, often inactive products that are readily excreted. Another example of substrate promiscuity is the existence of antibiotic resistance mainly due to a family of enzymes called β-lactamases. The ancestral substrate of this enzyme was the penicillins but, after new generations of antibiotics were introduced, these enzymes have co-evolved, broadening their activity to even more elaborate antibiotics. β-Lactamases hydrolyze the β-lactam ring of antibiotics and thereby allow survival of pathogenic bacteria challenged by treatment with these agents. Metallo-β-lactamases, which contain one or two zinc ions bound in the active site, have become a severe clinical problem due to their especially broad substrate spectra and potential for horizontal transfer.[8,9]

The catalytic promiscuity, the ability of an enzyme to catalyze, at low level, a reaction other than its cognate reaction that is maintained *via* selective pressure, provides a unique opportunity to dissect the origin of enzymatic rate enhancement *via* a comparative approach. The alkaline phosphatase (AP) superfamily is an ideal system to be used in making such comparison. There are enzymes belonging to this AP superfamily, also called AP, that catalyze the hydrolysis of phosphate monoesters but also present promiscuous activities towards sulfate monoesters, phosphate diesters and phosphite. Starting from this promiscuity activity, few mutations would be needed to reach a threshold value that could provide a selective advantage, allowing selection to optimize new enzymes in the superfamily. These observations explain the enzyme activity divergence of superfamilies from a common ancestor. This could be the case with nucleotide pyrophosphatase/phosphodiesterase (NPP) that presents highly common structural features with AP but preferentially hydrolyzes phosphate diesters.[10] Another example of catalytic promiscuity is illustrated by an enzyme that belongs to a new group of lactonases: phosphotriesterase (PTE) which appeared on this planet only several decades ago from a pre-existing hydrolase. This system provides a powerful demonstration of the evolvability of enzymes since it has evolved for the purpose of degradation of synthetic insecticides introduced in the 20th century. Although toxic to humans, bacteria have developed ways of degrading these compounds so that they do not accumulate in the environment. Interestingly, no naturally occurring substrate has been identified for PTE.[5,6,11]

An excellent model to illustrate the last case of promiscuity, product promiscuity, is γ-humulene synthase, a sesquiterpene synthase that is known to produce 52 different sesquiterpenes from a sole substrate, farnesyl diphosphate, through a wide variety of cyclization mechanisms. The product distribution can be very sensitive to changes in specific amino acid residues. Nevertheless, it is extremely difficult to predict the relationship between primary amino acid sequence and product distribution. In fact, closely related enzymes within or near species present functional disparity.[12]

All these previous examples of promiscuity illustrate the enzyme evolvability; this is, their ability to adapt to new functions. It is important to point out that improving an enzyme does not necessary mean increasing the rate constant of the catalytic process, instead it can be more interesting to improve the robustness of the protein structure against a wider range of temperatures and solvents. This target could be useful for certain industrial purposes. The practical procedure to modify an existing enzyme is the field of protein engineering, summarized in the next section.

### 1.2 Protein engineering

One day, we shall be able to *a priori* design amino acid sequences that will fold into proteins with desired functions. As this is not yet possible, scientists have used directed evolution to generate molecules with novel properties starting from natural enzymes.[13] Evolution, normally applied to animals and plants, requires the generation of variants and differential propagation of those with favourable features. Biologists and chemists have recently begun to use evolutionary strategies to tailor the properties of individual molecules instead of whole organisms. Random mutations or recombination can, in many cases, be done efficiently, leading

in this way to molecular evolution in the laboratory. The successful variants can be identified either by screening or by selection. While screening requires an active search of all variants, selection is based on the exclusive survival of organisms containing the desired variants of the protein, mimicking the true Darwinian evolution. This is an iterative process that requires, before starting new iterations of the process, the favourable variants to be amplified by cloning. The challenge is to collapse the time scale for evolution from millions of years to months or even weeks. Evolution does not work towards any particular direction, nor is there a goal; the underlying processes occur spontaneously during reproduction and survival. In contrast, the laboratory evolution experiments often have a defined goal, and the key processes (mutation, recombination, and screening or selection) are carefully controlled by the experimentalist. The general techniques of directed evolution mimic natural evolution processes such as random mutagenesis and sexual recombination. Thus, new proteins with new desired functions can be derived through mutations of a few residues or recombining fragments swapped between two parent sequences.[14] In this last technique it is possible to explore distant regions of sequence space, while this is not generally possible using random mutations. In both cases these techniques permit us to engineer enzymes without needing to understand them in great detail.

Another option is the rational design approach that consists of direct mutation of residues on selected specific positions of proteins.[15] The selection of residues to be mutated is often deduced from X-ray diffraction structures of the protein–substrate complex. Moreover, information on the molecular mechanism of the chemical reaction is required. Mutation of a few amino acids can render an important change in the active site of the enzyme, while the structure of the full protein remains almost invariant. As a consequence, significant catalytic effects can be derived. However, the lack of knowledge on the relationship between amino acid sequence, structure and function, together with the extreme sensitivity of catalytic activity to seemingly modest structural perturbations, make redesign of an enzyme in the laboratory so difficult. As we will show in detail below, computational study of the catalytic function of the protein can provide the information needed to design successful mutations directed to a particular purpose.

It is interesting to note that while rational design is usually focused on genetic selection that produces protein mutations close to the active site, directed evolution techniques produce random mutations that, most of the time, belong to regions of the protein far from the active site. So, both strategies can be satisfactorily combined to get an improved function.

### 1.3 Catalytic antibodies

Redesign of biological catalysts is not restricted to enzymes. In particular, almost three decades ago, immune-globulin proteins were used to produce catalytic antibodies (CA).[16] The immune system of living organisms can recognize and produce antibodies to a wide range of antigens. This enormous diversity arises from the ingenious reshuffling of the DNA sequences encoding components of the antibodies. Once a germline is selected from the pool based on its initial affinity for the antigen (also known as the hapten), additional structural diversity is generated by an affinity maturation process that can take some weeks. In this process, somatic mutations are introduced through changes in the CA protein amino acid sequence. This process leads to a high affinity monoclonal antibody. The most abundant of these antibodies, the immunoglobulin proteins, consist of two identical light chains and two identical heavy chains connected by disulfide bridges. Both the light and heavy chains have a constant domain and a variable domain, being the antigen recognizing site composed of the variable region of both chains.

In order to understand how an antibody can be used as a catalyst, the fundamentals of enzyme catalysis have to be remembered. The decrease in the activation barrier in an enzymatic process arises from the relatively larger binding energy of the transition state (TS) when compared to the reactants (Michaelis complex, MC). This idea can be translated to antibodies utilising the immune response to a stable molecule that mimics as closely as possible the presumed structural and electronic features of the TS of a particular reaction. This is the so called transition state analogue (TSA). Therefore the TSA used as the hapten is a stable molecule which resembles the TS. Just as a human hand can adapt to a large number of shapes, an antibody active site can change its conformation to complement different ligands.[16] Due to its intrinsic variability, a significant reorganization of the germline active site can occur upon ligand binding. Afterwards, in the maturation process of a monoclonal antibody, the optimal active site conformation is fixed, increasing the affinity for a particular antigen. The result should be the evolution of the antibody binding site with maximum complementarity to the TSA and, hopefully, to the real TS.

Following these strategies, CA have been produced for a plethora of chemical reactions, ranging from hydrolytic reactions, sigmatropic or cycloaddition rearrangements, carbon–carbon bond-forming reactions, redox reactions, *etc*... An unanticipated feature of some CA is their promiscuity. In particular, aldolase CA can catalyze over 100 different reactions, including aldehyde–aldehyde, ketone–aldehyde, and ketone–ketone condensations.[17] Another example is the CA 21D8, which was developed to accelerate a decarboxylation reaction, and has shown hydrolytic activity.[18] These observations suggest that certain antibody scaffolds are intrinsically predisposed towards catalysis, a property that can be enhanced and refined during the affinity maturation process in response to a TSA.[18] However, as a general feature, the catalytic power of CA is never as high as the one obtained in enzymes ($10^6$ in CA against $10^{20}$ in enzymes). Their modest efficiencies appear to be a direct consequence of the simple strategy used to generate them. Whereas the process of natural selection optimized enzymes on the basis of their catalytic activity, the microevolutionary mechanism of the immune systems selects antibodies for increased affinity to an ''imperfect'' TS. Although nature uses a wide variety of different protein scaffold to build enzyme active sites, there is a limited structural diversity inherent to the immune system. It is therefore conceivable that antibody structure itself places intrinsic limitations on the kind of reactions amenable to catalysis and on attainable efficiencies.[17,18]

 

Different strategies have been proposed to improve the catalytic efficiency of CAs. One is based on the search for a better hapten (TSA), as the TSA is never sufficiently similar to the TS. Moreover, it is difficult to find a proper TSA if the catalyzed reaction takes place through several steps. The other is the use of molecular engineering (based on site-directed mutagenesis). Keeping in mind that the starting point for rational design of CAs is frequently the X-ray structure of a complex between the protein and a stable molecule, both strategies present two serious drawbacks. First, the specific pattern of interactions established between the substrate, in its TS, and the residues of the active site does not exactly match to the one found in the protein–substrate crystal structure (whatever molecule was used as a substrate); and second, the static picture obtained from X-ray crystallography techniques does not reflect the flexibility and dynamics of the protein. The molecular engineering employed to improve the catalytic efficiency of CAs are thus similar to the ones applied to the redesign of enzymes. Computational design can be crucial to improve both kinds of biological catalysts.

## 2. Computational design

The methods and techniques of computational chemistry have become a promising complementary tool for the rational design of protein structures with catalytic activity. These methods could be classified into two classes; structure-based methods and reaction analysis-based methods. From a methodological point of view, the first class are based on the application of molecular mechanics force fields to structures deduced from X-ray crystallographic data or from quantum mechanical calculations on reduced active site models, while the second group of methods use hybrid quantum mechanics/molecular mechanics (QM/MM) techniques[19] that allow to follow the complete chemical reaction including the dynamics of the full system. These two types of methods are described in the next two sub-sections.

### 2.1 Structure-based methods

As mentioned before, it is not straightforward to deduce the three dimensional structure of a protein from the knowledge of its sequence of amino acids. In fact, this goal has only been reached in proteins with a small number of residues. An alternative strategy in protein design is to solve the problem in the inverse way; to compute an amino acid sequence that would fold into a target backbone structure.[20] These algorithms are based on searching for the global minimum energy conformation of the protein in its holoform within the frozen position of atoms belonging to the backbone. This search is carried out in the sequence space using a library of rotamers, which is usually based on molecular mechanics (MM) classical force fields.

It would be even more difficult to relate the structure of the protein with its function. Optimizing the enzymatic function is a much more subtle problem, since mutations of residues in the active site that presumably could improve the efficiency of the enzyme can decrease the stability of the full protein. Enzymes have evolved under selective pressure to both maintain the stability of the overall structure and the biochemical function. There are two opposing trends, on one hand enzymes fold into compact structures, on the other hand they must also be active to catalyze chemical reactions. The active site of an enzyme is highly strained because is designed to develop favourable interactions with the TS of the catalyzed reactions. This strain diminishes the stability of the global structure of the enzyme and thus a trade-off between stability and function is established.[9,21–23] Warshel suggested that charged groups and dipoles of the active site are preorganized to preferentially interact with the TS and, consequently, the active site of an enzyme itself is highly strained because of unfavourable interactions amongst these groups.[24]

According to the most accepted interpretation of enzyme catalysis, optimization of catalytic function would imply the stabilization of the TS of the catalyzed reaction by interactions with those residues of the active site.[25,26] Following these ideas, some successful approaches to computational enzyme design have been focused on the optimization of catalytic side-chain contacts with small molecules placed in active sites.[27] In this sense, combinatorial optimization algorithms that integrate ligand docking and placement of amino side-chain rotamer libraries have been used to identify sequences that form complementary ligand-binding surfaces. Obviously, these small molecules should resemble the true TS of the reaction in some electronic and geometrical features. This can be accomplished using high energy reaction intermediates as, for example, in the hydrolysis of *p*-nitrophenyl acetate using a catalytically inert protein scaffold.[28] A slightly different approach would imply the use of the TS of the reaction optimized in the gas phase instead of high energy pseudo-stable intermediates. This strategy has been used in the design of variants of *Escherichia coli* chorismate mutase (EcCM) in order to improve the catalytic effect of the wild type enzyme.[29,30] In this study, the target was to improve the catalytic activity of an existing enzyme: EcCM that catalyzes the Claisen rearrangement of chorismate to prephenate (see Fig. 1). A gas phase optimized TS structure was docked at the position of a TSA in the EcCM crystal structure. Then, amino acid identities and conformations were optimized for 18 active-site residues in the presence of this *ab initio* calculated TS structure. The calculations returned wild-type residues in

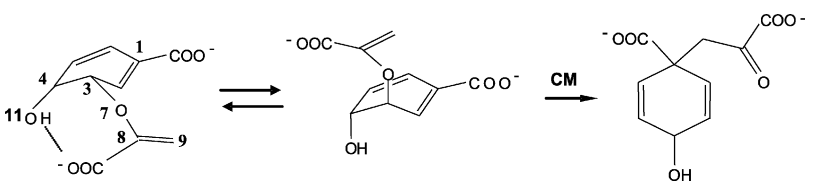


**Fig. 1** Claisen rearrangement of chorismate to prephenate showing the conformation pre-equilibrium of the former.

12 of the 18 designed positions, and the six mutations predicted by the calculations were to amino acids that did not make hydrogen bond contacts to the reacting molecule. The variants were tested experimentally and three active site mutants exhibited catalytic activity similar to or greater than wild-type enzyme. One of them, the mutant Ala32Ser, showed increased catalytic properties: larger $k_{cat}$ and also larger $k_{cat}/K_M$.

Recently, a new strategy has been proposed that is based on several steps starting from modelling the gas phase TS of the chemical reaction using *ab initio* methods. Once structural and electronic features of the gas phase TS are known, additional functional groups are included in an enlarged model in an attempt to stabilize the TS. This truncated active site is afterwards grafted into an existing protein and then the amino acid sequence of the protein is altered to accommodate the changes.[31,32] Thus, Röthlisberger *et al.* have developed new enzyme catalysts for the Kemp elimination reaction for which no natural enzyme is known.[31] This reaction takes place through the deprotonation of a carbon atom by a general base, as shown in Fig. 2. Stabilization of the TS of this reaction was obtained by placing an adequate base close to the position of the proton donor carbon atom, placing a proton donor stabilizing the charged developed on the oxygen atom and also stacking groups favouring the electronic delocalization observed in the TS. In particular, these authors used a carboxylate group (glutamate or aspartate) or a catalytic dyad (histidine-aspartate) as feasible bases for the deprotonation reaction. The authors searched for protein backbones capable of supporting these active site models using a large set of protein scaffolds. For each candidate, MM was used to optimize the orientation of a rigid TS and the surrounding residues satisfying the geometrical constrains derived from the gas phase model. Afterwards, residues were redesigned to maximize interaction of the TS and the stability of the active site conformation. The authors obtained high-resolution X-ray diffraction structures of the designed proteins, showing that the model and the crystal structure were virtually superimposable, thus validating the accuracy of the computational method. Nevertheless, subtle deviations observed in the backbone indicated loop regions in which explicitly modelling backbone flexibility may yield improved designs. Obviously, limitations due to the use of a frozen protein backbone during the theoretical modelling, necessarily reducing the capability of the full protein scaffold that can hold, fix, or orient the catalytic residues in the active site, are responsible for these deficiencies. Anyway, Röthlisberger *et al.* show that several of their new designed enzymes revealed a promising catalytic activity with a maximum rate enhancement of about $10^5$.

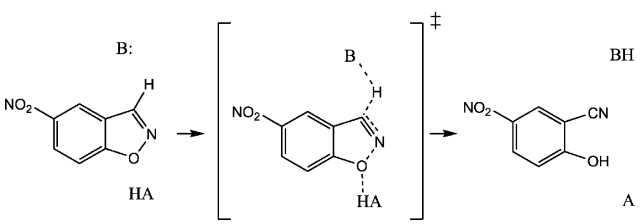


**Fig. 2** Schematic representation of the Kemp elimination process.

Afterwards, directed evolution carried out in the laboratory allowed them to arrive up to $10^6$.

A similar strategy has been used to design artificial enzymes with retro-aldol activity by Jiang *et al.*[32] In this case, the design is complicated due to the fact that this reaction takes place in several chemical steps and thus through different transition states that must be stabilized. A maximum rate enhancement of about $10^4$ was finally reached.

It is important to point out that the performances of the artificial enzymes obtained by Röthlisberger *et al.*[31] and by Jiang *et al.*[32] are quite modest compared with reaction rate enhancements achieved by natural enzymes. In fact, the chosen reactions, Kemp elimination and the retro-aldol reaction, respectively, are accelerated by several catalysts, including various synthetic compounds, catalytic antibodies, serum albumin or various peptides. Moreover, the methodology used in both studies presents the limitation that the designed enzymes have been designed trying to optimize only the chemical reaction steps, and not binding of substrates, product release, pH or thermal stability. Furthermore, from a computational point of view, there are some drawbacks in these methodologies: first, the TS of the catalyzed chemical reaction step has been obtained in the gas phase, not including the key residues of the active site, neither the rest of the protein environment; and second, the structure of the backbone of the protein remains frozen during the functional design modelling, not introducing its inherent flexibility and dynamics. The introduction of backbone flexibility and dynamics into functional design calculations is likely to be necessary to take into account the reorganization energy effects and thus improving activity.[26] All in all, the results of Röthlisberger *et al.*[31] and those by Jiang *et al.*[32] have to be considered a milestone in biochemistry as, for the first time, artificial enzymes have been designed for non-biological reactions.

### 2.2 Reaction analysis-based methods

An alternative computational protocol that, although containing some limitations, sorts out the difficulties present in previous strategies is based on the analysis of the dynamics of the chemical reaction in the protein active site. In this method, using a hybrid QM/MM scheme, the TS of the chemical reaction is obtained under the protein environment effect.[25] In order to carry out this kind of study, a small part of the system is described by quantum mechanics (QM), while the rest of the system is treated by classical force fields (MM). Then, the free energy profile of the transformation from reactants to products can be obtained as a potential of mean force (PMF) from biased molecular dynamics simulations using an adequate reaction coordinate. The maximum of this profile renders the TS of the reaction while the free energy difference between this maximum and the reactants will be related to the chemical rate constant, which can be compared with experimental data.[25]

In order to design an efficient and selective biological catalyst, this methodology can be applied first to an existing enzyme catalyzing the same or a similar reaction, with the aim of obtaining information on the key aspects that provide the clues to the catalysis. In particular, the reaction analysis-based

 

method allows detection of all the specific substrate–protein interactions that are established in the active site of a catalyst, both at the reactant complex (the Michaelis complex) and at the TS. It is very important to point out that this pattern of interactions is not necessary equal to the ones that can be deduced from static X-ray diffraction studies of stable TSA–protein complexes, nor are they necessarily constant along all the stages of the reaction. A deep insight into the evolution of the substrate–protein interactions along the chemical reaction profile can be used to deduce which residues of the active site are responsible for the preferential stabilization of the TS with respect to reactants. Information deduced from this analysis can be used to design a new catalyst capable of enhancing the rate constant of the chemical step.[33,34] Thus, one could think in a *de novo* design of a new biological catalyst or to select a proper protein scaffold as starting point. This latter alternative would be even less complex if an existing enzyme catalyzes this or a similar reaction, as one could try to mutate single residues in the active site searching for similar effects to the ones observed in this native enzyme. An adequate protein structure to be used can be a CA, which was synthesized against a TSA that resembled, as close as possible, the real TS of the reaction to catalyze, or a native enzyme that presented certain secondary activity catalysing the chemical reaction that is the subject of study. It is important to point out that, once the methodology was polished and properly optimized, the design of efficient biological catalysts would be particularly interesting in those reactions where no native efficient enzyme was known.

An application of this strategy is illustrated with the chorismate to prephenate rearrangement that is catalyzed by chorismate mutase (CM) native enzymes, by a synthetic CA (1F7), and also by a promiscuous enzyme, the isochorismate pyruvate lyase (IPL) from *Pseudomonas aeruginosa*, an enzyme that evolved over millions of years to catalyze isochorismate transformation into pyruvate and salicylate but also presents secondary activity catalyzing the transformation of chorismate into prephenate. Thus, the chorismate to prephenate rearrangement is an excellent candidate to show how to design a biological catalyst by a rational design that implies directed mutation on particular residues of the active site guided by the catalytic features observed in native CM enzymes.

Let us first analyze the reaction in native enzymes. As depicted in Fig. 1, a conformational equilibrium between reactive and non-reactive conformers precedes the conversion of chorismate to prephenate. We call reactive reactants those chorismate conformers closer to the TS conformation: they present a pseudo-diaxial character (regarding the relative orientation of the ring substituents) and a significant closer distance between the two carbon atoms to be bonded in the rearrangement (C1–C9). The non-reactive reactants present a pseudo-equatorial conformation with a longer C1–C9 distance, far from the TS. It is obvious that in gas phase the pseudo-diequatorial chorismate conformer is the most stable due to the presence of an intramolecular hydrogen bond. In solution, the formation of solute–solvent hydrogen bond interactions can dramatically change this behaviour. The pioneering QM/MM work of Richards *et al.* on *Bacillus subtilis* chorismate mutase (BsCM)[35] demonstrated that the inclusion of the enzyme environment in the calculations facilitates the stabilisation of a minimum energy structure presenting a distorted geometry compared to the ground state structure of chorismate in gas phase: the chorismate–enzyme complex is displaced towards the structure of the TS. In this way, these results provided an example of the geometrical changes induced in the substrate by the enzyme. PMF profiles can be obtained for the diequatorial–diaxial and for the short–long C1–C9 distance equilibria, in aqueous and protein environments demonstrating also the role of the enzyme in increasing the population of reactive reactant conformers.[25] Then, using the antisymmetric combination of the forming (C1–C9) and breaking (C3–O7) bond distances as a distinguished reaction coordinate, the PMF profiles for the chemical step of the reaction can be obtained in aqueous solution and in the enzyme. The resulting PMF profiles obtained in the CMs (EcCM and BsCM, respectively) together with the result obtained in water are depicted in Fig. 3a. According to these profiles both enzymes are able to diminish the activation free energy of the chemical reaction by about 8–9 kcal $mol^{-1}$,

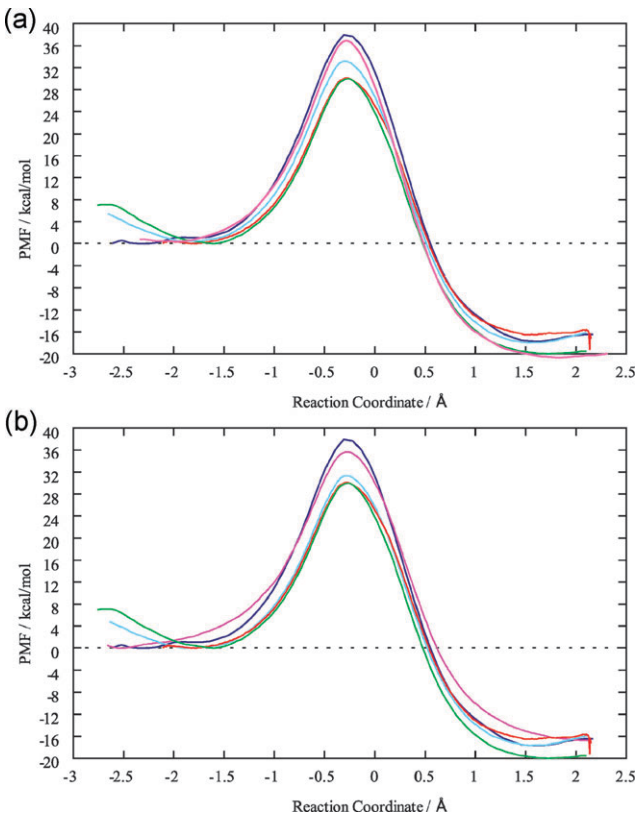


**Fig. 3** AM1/MM free energy profiles (in terms of PMFs) for the chorismate to prephenate rearrangement obtained in different environments: (a) EcCM (red line), BsCM (green line), aqueous solution (dark blue line), 1F7 (purple line) and mutated 1F7-N33S (light blue line) CAs; (b) EcCM (red line), BsCM (green line), aqueous solution (dark blue line), IPL (purple line) and mutated IPL A38I (light blue line). The reaction coordinate is, in all cases, the antisymmetric combination of the inter-atomic distances of the breaking and forming bonds, C3···O7 and C1···C9, respectively. Profiles of Fig. 3a are obtained from the data of ref. 33, 34 and 36, while Fig. 3b is adapted in part with permission from ref. 34, Copyright 2008 ACS.

a result that is in very good agreement with the experimental data.

Once obtained, these promising results, the substrate–protein averaged interactions, have to be analyzed for the TS, reactive reactants and non-reactive reactants. This analysis is of particular importance as far as charge distribution of the substrate, and as a consequence the pattern of electrostatic interactions with protein amino acids is not necessarily constant along the reaction progress. The conclusions derived from this comparative study are that while the patterns of interactions in the TS and the reactive reactants are very similar, this is not the case for the non-reactive reactant complex. This implies that while the enzyme structure remains essentially unaltered during the reaction progress, it must be considerably deformed to accommodate reactant conformations very different from the TS structure. As the enzyme deformation has an associated energy penalty, this suggests that an enzyme site prepared to accommodate the TS should also favour those reactants more similar to the TS structure.[25] Thus, substrate geometry deformation and TS stabilization are in fact closely related concepts having a common origin in the enzyme structure. Moreover, from the configurations obtained in the QM/MM MD simulation one can determine which interactions favour the stabilization of the TS relative to the reactant state, contributing thus to a diminution of the activation free energy. In both enzymes, BsCM and EcCM, the TS stabilizing interactions take place through the positively charged residues with the two negatively charged carboxylate groups of the substrate and the negative charge developing in the ether oxygen.[25] In BsCM, Arg7, Arg63 and Arg116 make stronger interactions with the carboxylate group in the TS while Arg90 stabilizes the charge on the ether oxygen. The same role is played in EcCM by Arg11, Arg28, Arg51 and Lys39. As pointed out in a previous review on enzyme catalysis,[25] it seems that the favourable relative substrate–protein interactions in the TS, with respect to reactants, are more dramatic in the enzyme than in solution. As mentioned above, these conclusions on the origin of enzyme catalysis in the CMs are being applied in the next step of this protocol: to use a CA as protein scaffold to develop catalysts. This CA has to present certain activity in this reaction.

The free energy profile obtained for the chorismate to prephenate rearrangement in the 1F7 CA[33] is depicted in Fig. 3a together with previously commented upon PMF profiles. The resulting profile is in accordance with the expected results: the catalytic efficiency of the 1F7 appearing between the catalytic power of both enzymes and the reaction studied in solution. The analysis of the substrate–protein interactions established in the TS of 1F7 shows how their magnitudes are dramatically smaller than in the natural enzymes, except for the interaction established with ArgH95 (see Fig. 4), which presents similar values to the enzymatic ones.[33] It seems that the 1F7 does not properly interact with the two carboxylate groups of the substrate, leaving them partially exposed to the solvent, which can be corroborated with the strong interactions established with the water molecules accessible to the cavity, much smaller than in any of the two CMs. This result suggests that the substrate fits better in the enzyme active sites than in the CA pocket.

As pointed out in our previous paper,[33] it is important to emphasize that the pattern of interactions obtained in the TS of the 1F7 is not equal to the TSA–CA structure coming from the X-ray diffraction study (see Fig. 4). The experimentally obtained structure of the CA–TSA complex is appreciably different to the TS structures located in the CA active site. Thus, for instance, AsnH33 presents a noticeable different orientation in the TS–CA than in the TSA–CA complex: while in the latter the hydrogen atoms of the amino group interact with the hydroxyl group of the inhibitor, in the TS–CA complex there are esteric interactions between this residue and the aliphatic hydrogen atoms of the substrate hindering the optimum positioning of the substrate into the cavity, and thus preventing strong interactions between the carboxylate group and amino acid residues of the inner part of the cavity. The low capability of the 1F7 to enhance the rate constant of the chorismate to prephenate rearrangement can be understood from this analysis. The strong stabilizing interactions observed in the enzyme between both carboxylate groups and the protein are not reproduced by the immune system process when eliciting antibodies against a stable molecule that resembles, but is not equal to, the TS of the desired chemical transformation. Thus, it was proposed changing the AsnH33 by a serine

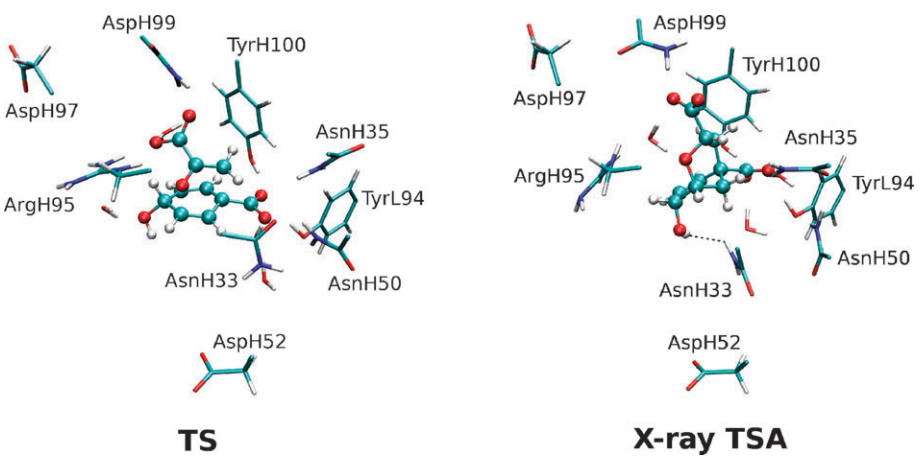


**Fig. 4** Snapshot of the TS of the chorismate rearrangement in 1F7 and X-ray structure of the TSA in the same active site.

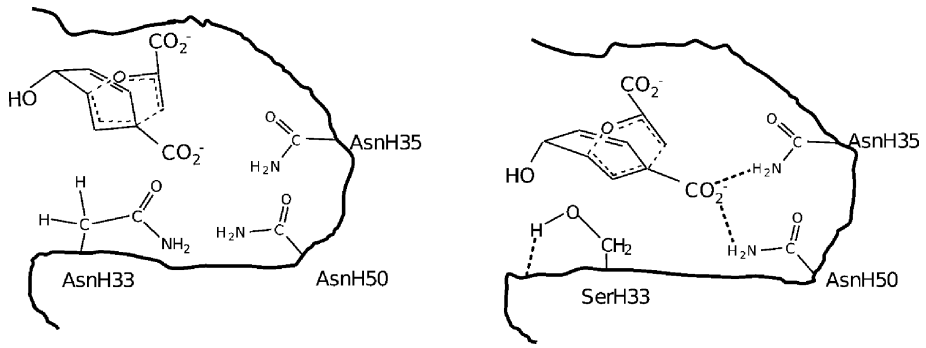


**Fig. 5** Schematic representation of the substrate–protein interactions in the active site of the 1F7 and the 1F7-N33S CAs at their respective TSs showing the effect of AsnH33Ser mutation. Adapted in part with permission from ref. 33. Copyright 2007 Wiley-VCH.

residue as schematically depicted in Fig. 5. The effect of this smaller amino acid is the generation of extra space in the cavity, allowing the ring of the substrate to slightly rotate and its carboxylate groups to optimize the interactions with the available residues placed in the inner part of the catalytic cavity. In particular, the interactions established between the carboxylate group and residues such as AsnH35 and AsnH50 are stronger in the mutated CA than in 1F7. Simultaneously, the interactions of the substrate with the solvent molecules are reduced, approaching to the situation observed in the CMs. The steric hindrance of the AsnH33 with the substrate in 1F7 prevents this movement, while in the mutated CA a combination of the smaller size of the residue and the weaker interaction established with the substrate facilitates a more favourable relative orientation in the CA cavity. The free energy profile of the new 1F7-N33S compared to the 1F7 and previous paths obtained in the enzyme and aqueous environment is presented in Fig. 3a. This mutation yields a noticeable decreasing of the free energy barrier, by comparison with the PMF obtained in the 1F7. The corresponding activation free energy is 4.5 kcal $mol^{-1}$ lower than the original 1F7 CA and only 2.4 kcal $mol^{-1}$ above the most efficient BsCM enzyme. This diminution would imply an increasing of the rate constant by a factor of $10^3$ at room temperature, compared with the 1F7 CA.[33]

Although the present work shows how our methodology, combined with other experimental strategies, may be used to generate an improved catalyst from a relatively modestly efficient CA, the limited structural diversity of the immune system together with the relatively weak substrate–CA binding interactions that can be reached with this protein impose inherent limits on catalysts designed from CA scaffolds. Thus, the next step in this methodology is to use a different protein scaffold as starting point: a promiscuous IPL enzyme. The free energy profile, in terms of PMF, for the chorismate to prephenate reaction carried out in the active site of the *Pseudomonas aeruginosa* IPL is shown in Fig. 3b, together with previous profiles obtain in CM and aqueous solution.[34] As expected, the catalytic power of the CMs is much higher than that of IPL, which is reasonable keeping in mind that the chorismate to prephenate rearrangement is a secondary reaction for the latter enzyme. In fact, IPL provides a free energy barrier lowering of only 2.3 kcal $mol^{-1}$ with respect to the reaction in aqueous solution, while CMs are able to reach a diminution of the free energy barrier about 4 times larger.

From the analysis of the reaction process carried out in the active site, it can be concluded that the catalytic active site of this IPL does not perfectly fit to the chorismate substrate. In particular we have analyzed the diaxial or diequatorial character of the substrate located in the IPL active site along the reaction profile. The evolution of the dihedral angle defining the diaxial character of the substrate, as a function of the reaction coordinate, obtained in the CMs, water and IPL is presented in Fig. 6. It is observed how the pseudo-diaxial character, required in the TS, is lost in the reactant state obtained in IPL and in water solution, while the active sites of EcCM and BsCM constrain the substrate avoiding a complete relaxation of the substrate that would render a non-reactive reactant conformation. It seems that IPL is not able to retain this favourable conformation from TS to reactants, undoubtedly due to the fact that the active site is not perfectly suited to accommodate the substrate of this reaction.

In order to understand this different behaviour of IPL with respect to CMs, the interactions of the enol pyruvyl moiety of the substrate in the active sites of IPL and EcCM (which is the most closely related structure) were compared. A deeper insight into the substrate–protein pattern of interaction in EcCM reveals that there is a hydrophobic valine residue (Val35) that constrains the position of the ether bridge.[34]

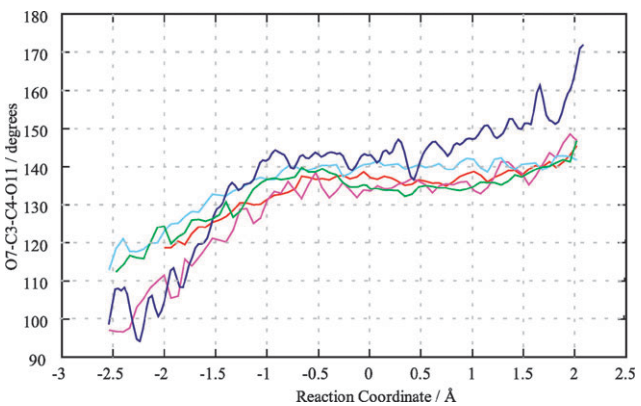


**Fig. 6** O7C3C4O11 dihedral angle evolution (in degrees) as a function of the reaction coordinate. EcCM (red line), BsCM (green line), IPL (purple line), IPL-A38I (light blue line) and in aqueous solution (dark blue line). Adapted in part with permission from ref. 34. Copyright 2008 ACS.

The equivalent residue in IPL is a smaller alanine (Ala38) that cannot perform the same role. Thus, we decided to carried out an *in silico* mutation of this residue to a larger one (from Ala to Ile) and to repeat the PMF for the chorismate to prephenate reaction.[34] The result of this mutation is that the mentioned dihedral angle evolution on the mutated IPL enzyme is closer to the CMs than to the native IPL or water (see Fig. 6). The active site of the engineered enzyme is better suited for the chorismate to prephenate rearrangement than the original one. In fact we have computed the contribution of the electrostatic interactions to catalysis as the difference between the enzyme–substrate interaction free energy in the TS and in the reactant state. This contribution is $-8.6$ kcal mol$^{-1}$ in IPL and $-11.8$ kcal mol$^{-1}$ in the mutated enzyme. This change is reflected in the energetics of the chemical reaction; as shown in Fig. 3b the free energy barrier in the mutated IPL is 4.4 kcal mol$^{-1}$ lower than in the native IPL, being only 2.0 kcal mol$^{-1}$ above the primary reaction catalyzed by BsCM.[34,36]

As mentioned in the structure-based methods section, Mayo and co-workers carried out an extensive study to improve the catalytic activity of CMs. After performing 19 possible amino acid substitutions applied over 6 different positions of the engineered chorismate mutase domain of EcCM, a Val35Ile mutation was obtained that renders an increase in the $k_{cat}$ of about 1.5 times.[30] Moreover, mutation of Val to Ala (the same residue present in IPL) reduces the $k_{cat}$ by a factor of 2. Thus, the overall effect for a Ala35Ile mutation in EcCM is an increase in the $k_{cat}$ of a factor of 3. Obviously, the predicted effect for the same mutation in the equivalent position in IPL (Ala38Ile) is much larger, which is consistent with the fact that this enzyme is not specialized in the catalysis of the chorismate to prephenate rearrangement. Our results are in agreement with this study of Mayo *et al.*; by analyzing the full reaction mechanism the method can render both a prediction about the effect of a mutation on IPL and an interpretation about the success of a Val35Ile mutation in EcCM. According to Fig. 6 a mutation of Val to a similar but larger amino acid at position 35 would keep the enol pyruvyl moiety in a diaxial conformation closer to the TS geometry and reducing then the free energy barrier. In this sense, QM/MM MD simulations of the Val35Ile variant of EcCM have been carried out, verifying that the proposed mutation increases the diaxial character of reactants by about 3 degrees. The experimental data of Mayo and co-workers give credence to the theoretical predictions obtained for IPL and allows confidence in this reaction analyzed-based method. Moreover, although the method has been applied to a reaction that is also catalyzed by a native enzyme, there is no need in having a reference catalyzed reaction to obtain the key features that can provoke catalysis. In fact, this method will be of maximum interest for those reactions where no enzyme is known.

## 3. Future perspectives

In this review we have presented different strategies that can be used to assist the development of new biological catalysts supported with computational chemistry techniques: structure-based methods and reaction-analysis based methods. The first group of methods are mainly based on the knowledge of the X-ray diffraction structure of a protein complexed with a substrate resembling the properties of the real TS of the chemical reaction or modelled from a rigid gas phase TS structure. Thus, these kinds of methods are focused on the optimization of catalytic side-chain contacts with molecular structures placed in active sites. These molecules, which have to resemble the true TS of the chemical reaction, are high energy pseudo-stable intermediates or TS structures optimized in gas phase or in reduced active site models. Then, combinatorial optimization algorithms that integrate ligand docking and placement of amino side-chain rotamer libraries can be carried out in order to optimize the interactions. Thus, these methods can be considered as a computational screening algorithm that dramatically reduces the number of putative mutations to be performed in the laboratory. Although impressive results have been obtained, some inherent drawbacks are present in these methodologies: first, the real TS of the catalyzed chemical reaction step is not considered and, second, the structure of the backbone of the protein remains frozen during the functional design modelling, not introducing its inherent flexibility and dynamics. Some progress may be achieved in introducing the flexibility of the backbone in the simulations,[37] but an upper limit will be always present; as far as the TS of the chemical reaction, which implies breaking and forming bonds, it cannot be described by using classical MM force fields and without including the influence of a protein environment able to adapt to the changes taking place during the reaction progress. Moreover, focusing the study on just the TS of the chemical reaction implies that reorganization energy, which is one of the key terms of catalysis,[26] is not considered in these molecular modelling methods.

The second strategy, based on the analysis of the reaction process in a realistic environment, can be considered as a step forward from the previous one and tries to overcome its limitations. The chemical reaction is studied by means of hybrid QM/MM methods in the presence of the protein. Afterwards, with a deep insight into the origin of the relative stabilization of the TS, and the analysis of those residues of the active site playing a key role in the reaction, several mutations can be proposed and checked *in silico*.

The improvement in the catalytic power of experimentally synthesized CAs or native enzymes that presented a threshold catalytic efficiency in a certain chemical reaction can be obtained after single mutations chosen by computational protein engineering. These methods can be considered as the benchmark, providing a powerful *in silico* test for guiding improvements in computational enzyme design. Nevertheless, it must be kept in mind that these methods are applied exclusively to the chemical step, trying to optimize the performance of the protein by reducing its free energy barrier, while other steps of the full catalytic process could be affected by the single mutations performed on the active site of the protein. It would be desirable to extend these methodologies to analyze the influence of mutations on steps like the Michaelis complex formation or the product release in order to improve the substrate affinity or the catalytic turnover, respectively. It must be taken into account that the theoretical study of these two steps presents additional difficulties related to the large

 

protein conformational changes inherent to the reactant complex formation and product release steps. Further improvements in simulation methods, focused to perform much larger conformational searches and/or the use of higher levels of theory in the description of the QM region of the system, will arrive in the near future. These provisions are supported by the dramatic increase in the computer capabilities and the efficiency of new algorithms. Anyway, improved and simplified theoretical protocols are desirable to reduce the time required to predict successful mutations. In this sense, the use of interaction energies and/or free energies seems to be a promising tool provided the real TS of the reaction, under the influence of a flexible environment, is known. It would be also desirable in the future for an improvement in the theoretical protocol that would allow reducing the number of steps in the theoretical screening and speeding up the predictions.

Another aspect that should also be incorporated in the future is the thermal stability of the enzyme or their robustness against pressure or pH conditions, which are obviously affected by mutations. One of the goals that could be envisaged in the near future is the prediction of mutations that enhance the stability of a given biological catalyst (native or engineered).

As a final summary, theoretical tools are able to provide an enormous amount of detailed knowledge that is not always available from experimental techniques, offering then alternative ways to analyze and understand the chemical process under study. The theoretical strategies presented here can be considered as a numerical experiment helping experimentalists in active-site redesign by means of a few directed mutations. We are sure that some efforts in this direction will be presented with impressive results in the next few years, probably combining the advantages of the two main methods presented in this review.

## References


1 R. Wolfenden and M. J. Snider, *Acc. Chem. Res.*, 2001, **34**, 938–945.
2 M. D. Toscano, K. J. Woycechowsky and D. Hilvert, *Angew. Chem., Int. Ed.*, 2007, **46**, 3212–3236.
3 P. J. O'Brien and D. Herschlag, *Chem. Biol.*, 1999, **6**, R91–R105.
4 S. D. Copley, *Curr. Opin. Chem. Biol.*, 2003, **7**, 265–272.
5 A. Aharoni, L. Gaidukov, O. Khersonsky, S. M. Gould, C. Roodveldt and D. S. Tawfik, *Nat. Genet.*, 2005, **37**, 73–76.
6 O. Khersonsky, C. Roodveldt and D. S. Tawfik, *Curr. Opin. Chem. Biol.*, 2006, **10**, 498–508.
7 J.-D. Marechal and M.-J. Sutcliffe, *Curr. Top. Med. Chem.*, 2006, **6**, 1619–1626.
8 M. W. Crowder, J. Spencer and A. J. Vila, *Acc. Chem. Res.*, 2006, **39**, 721–728.
9 X. Wang, G. Minasov and B. K. Shoichet, *J. Mol. Biol.*, 2002, **320**, 85–95.
10 J. G. Zalatan, T. D. Fenn, A. T. Brunger and D. Herschlag, *Biochemistry*, 2006, **45**, 9788–9803.
11 L. Afriat, C. Roodveldt, G. Manco and D. S. Tawfik, *Biochemistry*, 2006, **45**, 13677–13685.
12 Y. Yoshikuni, T. E. Ferrin and J. D. Keasling, *Nature*, 2006, **440**, 1078–1082.
13 F. H. Arnold, *Acc. Chem. Res.*, 1998, **31**, 125–131.
14 M. N. Carbone and F. H. Arnold, *Curr. Opin. Struct. Biol.*, 2007, **17**, 454–459.
15 K. L. Morley and R. J. Kazlauskas, *Trends Biotechnol.*, 2005, **23**, 231–237.
16 *Catalytic Antibodies*, ed. E. Keinan, Wiley-VCH Verlag, Weinheim, 2005.
17 D. Hilvert, *Annu. Rev. Biochem.*, 2000, **69**, 751–793.
18 A. C. Backes, K. Hotta and D. Hilvert, *Helv. Chim. Acta*, 2003, **86**, 1167–1174.
19 A. Warshel and M. Levitt, *J. Mol. Biol.*, 1976, **103**, 227.
20 C. A. Voigt, D. B. Gordon and S. L. Mayo, *J. Mol. Biol.*, 2000, **299**, 789–803.
21 B. M. Beadle and B. K. Shoichet, *J. Mol. Biol.*, 2002, **321**, 285–296.
22 R. A. Nagatani, A. González, B. K. Shoichet, L. S. Brinen and P. C. Babbitt, *Biochemistry*, 2007, **46**, 6688–6695.
23 C. Cheng, B. Qian, R. Samudrala and D. Baker, *Nucleic Acids Res.*, 2005, **33**, 5861–5867.
24 A. Warshel, *Proc. Natl. Acad. Sci. U. S. A.*, 1978, **75**, 5250–5254.
25 M. Roca, S. Martí, J. Andrés, V. Moliner, E. Silla, I. Tuñón and J. Bertrán, *Chem. Soc. Rev.*, 2004, **33**, 98–107.
26 A. Warshel, *Chem. Rev.*, 2006, **106**, 3210–3235.
27 J. K. Lassila, H. K. Privett, B. D. Hallen and S. L. Mayo, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 16710–16715.
28 D. N. Bolon and S. L. Mayo, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 14274–14279.
29 J. K. Lassila, J. R. Keeffe, P. Oelschlaeger and S. L. Mayo, *Protein Eng., Des. Sel.*, 2005, **18**, 161–163.
30 J. K. Lassila, J. R. Keeffe, P. Kast and S. L. Mayo, *Biochemistry*, 2007, **46**, 6883–6891.
31 D. Röthlisberger, O. Khersonsky, A. M. Wollacott, L. Jiang, J. DeChancie, J. Betker, J. L. Gallaher, E. A. Althoff, A. Zanghellini, O. Dym, S. Albeck, K. N. Houk, D. S. Tawfik and D. Baker, *Nature*, 2008, **453**, 190–195.
32 L. Jiang, E. A. Althoff, F. R. Clemente, L. Doyle, D. Röthlisberger, A. Zanghellini, J. L. Gallaher, J. L. Betker, F. Tanaka, C. F. Barbas, D. Hilvert, K. N. Houk, B. L. Stoddard and D. Baker, *Science*, 2008, **319**, 1387–1391.
33 S. Martí, J. Andrés, V. Moliner, E. Silla, I. Tuñón and J. Bertrán, *Angew. Chem., Int. Ed.*, 2007, **46**, 286–290.
34 S. Martí, J. Andrés, V. Moliner, E. Silla, I. Tuñón and J. Bertrán, *J. Am. Chem. Soc.*, 2008, **130**, 2894–2895.
35 P. D. Lyne, A. J. Mulholland and W. G. Richards, *J. Am. Chem. Soc.*, 1995, **117**, 11345–11350.
36 S. Martí, J. Andrés, V. Moliner, E. Silla, I. Tuñón, J. Bertrán and M. J. Field, *J. Am. Chem. Soc.*, 2001, **123**, 1709–1712.
37 C. Wang, P. Bradley and D. Baker, *J. Mol. Biol.*, 2007, **373**, 503–519.