

# COMPUTATIONAL STRATEGIES FOR THE DESIGN OF NEW ENZYMATIC FUNCTIONS

**K Świderek,<sup>1,2</sup> I. Tuñón,<sup>1</sup> V. Moliner,<sup>3</sup> J. Bertran<sup>4,\*</sup>**

1. Departament de Química Física, Universitat de València, 46100 Burjasot, (Spain)

2. Institute of Applied Radiation Chemistry, Lodz University of Technology, 90-924  
Lodz, (Poland)

3. Departament de Química Física i Analítica; Universitat Jaume I, 12071 Castellón  
(Spain)

4. Departament de Química; Universitat Autònoma de Barcelona, 08193 Bellaterra,  
(Spain)

Corresponding author:

Joan Bertran: [bertran@klinton.uab.es](mailto:bertran@klinton.uab.es)

## Highlights

- The design of new biocatalysts requires a detailed knowledge of the reaction to be catalysed.
- The combination of experimental and computational techniques is a very reasonable strategy.
- New enzymes have been designed by combination of *de novo* techniques with laboratory evolution.
- Kemp, Diels-Alder and retro-aldol reactions have been used to illustrate recent achievements.
- Flexible molecular models treated by QM/MM multiscale methods can be used in future studies.

**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.

## 1. 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 catalyse 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.<sup>1,2,3,4</sup>

Many reactions employed in the chemical and biochemical industry need of efficient catalysts to improve their rate, their selectivity and their environmental friendliness. 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 catalyse 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.<sup>5</sup>

Directed evolution<sup>6</sup> applies to strategies based in natural evolution to tailor the properties of molecules instead of organisms.<sup>7</sup> 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.<sup>8</sup> 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. In particular, when computational analysis of the enzymatic activity is used to guide the mutations, one usually refers to this strategy as computational design.<sup>9</sup>

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) more than thirty years ago;<sup>10</sup> 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.<sup>9,11</sup> The first group of proteins, CAs, are obtained using as hapten a Transition State Analogue (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.<sup>12</sup> This redesign process can take advantage from the different promiscuous behaviours 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.<sup>13</sup> 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 catalysed and, in particular, in the TS and the physicochemical principles governing

its stabilization (see Figure 1). This information is obtained from computational simulations. Mayo and co-workers pioneered the field of the *de novo* design of enzymes<sup>14</sup> converting the *E. 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.<sup>15</sup> This program explores simultaneously the conformational and sequence space of the designed biocatalysts.<sup>16</sup> 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,<sup>17</sup> the Diels-Alder reaction<sup>18</sup> and the retro-aldolase reaction.<sup>19</sup> 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’,<sup>20</sup> 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.<sup>21</sup> 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.<sup>22,23</sup> 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.<sup>24,25</sup> 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<sup>26</sup> although, for instance, Rosetta program was recently modified to consider flexible backbones.<sup>27</sup> The optimization in the sequence space can be expanded considering also non-canonical amino acids.<sup>28,29,30</sup> Algorithm improvements have been also achieved taking advantages of player strategies and social evolution.<sup>31</sup> 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,<sup>32</sup> Scaffold-Selection,<sup>33</sup> and PRODA\_MATCH.<sup>34,35</sup> A different strategy is employed by the SABER program.<sup>36</sup> 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.

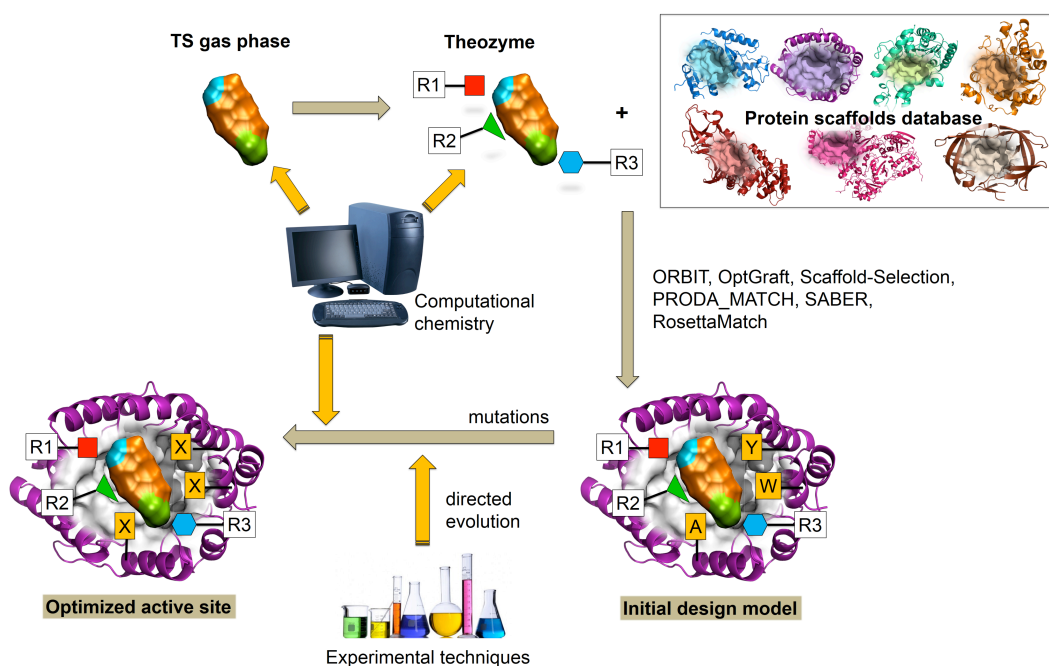


Figure 1. Schematic representation of the *de novo* design of new enzymes

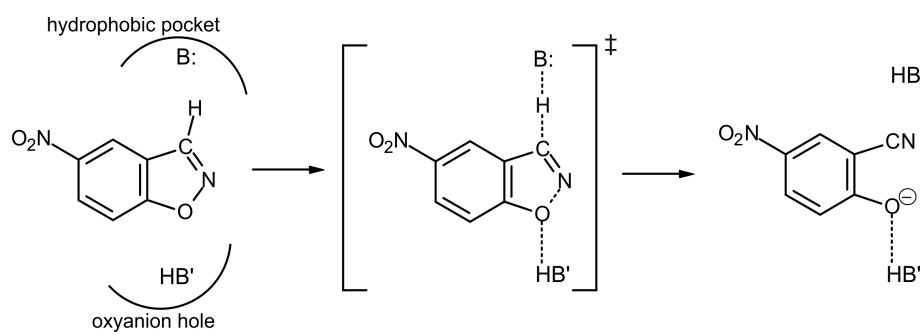
In spite of all these improvements the *de novo* design strategy still suffers from limitations, as demonstrated by the fact that the rate of successful designs is quite low and, in addition, the best ones display reaction rate enhancements far from those of natural enzymes. As pointed out by Baker,<sup>37</sup> 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”.<sup>7</sup> 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 Mechanics/Molecular Mechanics (QM/MM) potentials, is much higher and then it has to be used in combination with the first one.

Many excellent reviews about the design of new biocatalysts have been recently published.<sup>5,7,9,11,38,39,40,41,42,43,44,45</sup> This review will focus mainly on the advances obtained in the field of the *de novo* design during the last years, emphasizing the new achievements of the different computational strategies employed in the design process. The three aforementioned reactions, Kemp, Diels-Alder and retro-aldolase, will be used to illustrate these points. An additional section will be devoted to the particularities of the design of metal-containing enzymes.

## 2. Kemp elimination

The Kemp elimination, that consists in the conversion of benzisoxazoles into salicylonitriles, is a well-studied reaction that takes place by means of an E2 elimination mechanism (see Scheme 1). The reaction implies a proton transfer from a carbon atom of the substrate to a base, thus creating an anion species with the concomitant opening of the ring.<sup>46,47,48</sup> The reaction appear to be very sensitive to the base strength, having tested hydroxide ion,<sup>46,47</sup> amines<sup>46,47</sup> or the conjugated base of acids.<sup>48</sup> Since no naturally occurring enzyme has been identified to catalyse the reaction, and together with the fact that it is not a very complex reaction, it has been used as a benchmark of different protocols to design new enzymes.



Scheme 1. Representation of the base-catalyzed Kemp elimination

## 2.1. Antibodies and other catalysts

The first designed biocatalysts of Kemp elimination (34E4 and 35F10), with rate acceleration greater than  $10^8$ , were catalytic antibodies generated in response to a positively charged hapten mimicking the TS geometry.<sup>49</sup> Other cationic haptens have been used to get catalytic antibodies that catalyze this reaction, such as 4B2 and 6C2,<sup>50</sup> or the 13G5<sup>51</sup>, being the latter, together with the previously mentioned 34E4, the ones showing the highest catalytic activity. Following a similar experimental protocol, an antibody with an active site consisting of a hydrophobic pocket with a conserved lysine residue was generated against a reactive diketone hapten. This antibody, 38C2, showed activity for a variety of reactions that proceed via an enamine intermediate and, interestingly, also present promiscuous activity for the base-catalysed Kemp elimination.<sup>52</sup>

The use of X-ray diffraction has allowed obtaining details, at atomic level, of the structures of the originally designed catalytic antibodies, as well as variants generated by mutation of key residues. Analysis of these structures together with kinetic studies, has contributed to the elucidation of the key features responsible of the catalytic activity of these new enzymes. Thus, the active site of 34E4 shows conformational plasticity that can be responsible of a reduced efficiency since only one of the possible orientations of the substrate is reactive. An optimal microenvironment must be present in the hydrophobic pocket to impose the proper positioning of the substrate to interact with the conjugated base, usually Glu or Asp, and to modulate the  $pK_a$  of the base, since water molecules could stabilize their non-protonated form. It has been also observed that buried water molecules found in the active site can stabilize the negatively charged developed in the phenoxide oxygen atom of the substrate. Finally,  $\pi$  stacking interactions between a Trp residue and the highly polarizable TS has been also proposed

as the origin of catalysis in 34E4.<sup>53,54,55</sup> In the case of 13G5 antibody, solved X-ray structures suggest that Asp and Glu would act as a base to abstract the proton from the substrate and to orient a water molecule to stabilize the leaving group, respectively.<sup>56</sup> Kemp elimination is dramatically sensitive to the medium, as illustrated by the fact that the reaction with acetate is over  $10^7$  folds slower in water than in acetonitrile, as a consequence of the stabilization of carboxylate anions in aqueous environments.<sup>57</sup> Theoretical studies based on quantum chemical calculations on model systems in gas phase have been used to propose the most appropriate base and acid moieties to predict catalytic functionalities and to design haptens to produce improved catalytic antibodies.<sup>58</sup>

Two independent investigations showed that familiar “off-the-shelf” proteins, serum albumins, catalyse the same Kemp elimination at similar rates than the previous catalytic antibodies. In this case, the base was a lysine side-chain amino group located in a hydrophobic pocket.<sup>59,60,61</sup> It has been demonstrated that these albumin based catalysts, that uses a lysine as a base, are less sensitive to the environment. Based on quantum mechanical optimized geometries on gas phase and in solution represented by continuum models, this different environmental dependent behaviour has been rationalized.<sup>62</sup>

A different scaffold employed to design a catalyst for the Kemp elimination has been the polyethyleneimine, highly branched polymers with primary, secondary and tertiary amines linked by ethylene units that can be modified by alkylation or acylation of their amine groups. These enzyme-like catalysts were named as synzymes.<sup>63</sup> More recently, a new Kemp bio-catalysts has been designed by a single mutation in a small regulatory protein calmodulin (CaM). Native CaM, upon binding Ca(II), opens the hydrophobic pockets in each of its domains which, when introducing a dehydrated carboxylate, allows its accommodation to productively interact with the Michaelis complex and stabilize the TS of the reaction.<sup>64,65</sup> Additionally, this new enzyme serves as a calcium ion sensor. Engineering of model protein cavities to catalyse the Kemp elimination has been also obtained by a single mutation of Met to His in the artificial L99A cavity in T4 Lysozyme.<sup>66</sup> The presence of the new catalytic base, His, that abstract the proton from the substrate in a hydrophobic buried cavity is enough to catalyse the process, demonstrating that protein model cavities may provide templates for studying protein design in simplified environments.

These studies demonstrate the possibility of designing new enzymes by a reduced number of mutations in the native protein scaffold. Nevertheless, the catalytic efficiency of these new biocatalysts is still far from the most efficient enzymes.

## 2.2. *de novo* design of new enzymes

The first example of *de novo* protein design for Kemp elimination was published in 2008 by Houk, Tawfik, Baker and co-workers using the strategy described in the introduction.<sup>17</sup> The significance of this new protocol based on cutting-edge computational methods was emphasized by Nanda, who stressed the importance of combining experimental and computational tools.<sup>67</sup> In their work, different non-immunoglobulin proteins were used as starting scaffolds selected by means of Rosetta software.<sup>24,25,26</sup> In particular, from previously designed theozymes computed by means of quantum mechanical calculations in reduced models, two active-site motifs were considered differing in the choice of catalytic base; one utilized a carboxylate group for benisoxazole deprotonation, whereas the other exploited a histidine side chain backed up by an aspartate or glutamate to regulate its pK<sub>a</sub> and tautomeric state. In addition, a hydrogen-bond donor was provided to stabilize the developing negative charge on the phenolic oxygen atom, and  $\pi$ -stacking interactions were included for TS stabilization. Fifty-nine designs, each containing 10 to 15 mutations, were experimentally characterized and eight of them resulted to be active. The X-ray diffraction analysis of one of the newly produced enzymes, KE07, was solved and demonstrated to superimpose well with the computational model. Interestingly, although the best computationally designed enzyme was more than one order of magnitude less efficient than comparable antibody catalysts, their activity was significantly improved by directed evolution.<sup>17</sup> KE07 was used as a template for further incorporation of up to 8 mutations providing a 10<sup>6</sup>-fold rate acceleration, but with a marked decrease in the overall stability.<sup>68</sup> In a different study using the KE70 as template, new computationally guided optimizations were carried out including backbone flexibility. This, combined with further directed evolution render a 400-fold increased activity.<sup>69</sup> The highest active new catalyst originally obtained by Tawfik and co-workers, KE59, presented the inconvenience of being the least stable. The strategy to improve the performance of this template was to carry out a number of fold-stabilizing mutations prior to directed evolution cycles. Sixteen rounds of mutation and selection led to more than 2,000-fold increase in catalytic efficiency.<sup>70</sup>

Jorgensen and co-workers studied free energy surfaces to explore the reaction mechanism of four of the designed Kemp elimination catalysts<sup>17</sup> by means of free-energy perturbation (FEP) methods within hybrid QM/MM potentials based on Monte Carlo (MC) simulations.<sup>71</sup> Some approximations were considered in the simulations, like keeping the protein backbone fixed, the use of a semiempirical Hamiltonian (PM3) to describe the QM region, or keeping the interatomic distance between the proton donor and acceptor atoms as constant. According to their results, the catalytic mechanisms of Kemp elimination catalysed by KE07, KE10(V131N), KE15 were found to be concerted with proton transfer more advanced in the TS than the breaking of the isoxazolyl N-O bond, while a step-wise mechanism preceded by the proton transfer was found for the KE16. The trend predicted from the computed free energy barriers, by comparison with the experimental rate constants, was not encouraging.<sup>71</sup> The application of this strategy to the study of the mechanism of KE59, KE70 and several mutants rendered similar conclusions.<sup>72</sup> The study of the effect of the mutation Glu50Asp in the 34E4 catalytic antibody, showed an increase in the free energy barrier of  $2.4 \text{ kcal}\cdot\text{mol}^{-1}$ ,<sup>73</sup> in good agreement with experimental observation. Again, the same methodology allowed justifying the promiscuity of the 4B2 antibody in catalysing the allylic isomerization and the Kemp elimination reactions.<sup>74</sup>

Different approaches based on density functional theory (DFT) methods, ranging from reduced models in gas phase to those including the full protein and solvent water molecules in ONIOM QM/MM MD simulations, were carried out by Houk and coworkers to rationalize the experimental data of several designed catalysts of Kemp elimination.<sup>75</sup> Computed activation barriers were in qualitative agreement with experiments, allowing distinguishing the inactive from the active catalysts but with a weak correlation with the energy barriers deduced from the experimentally determined rate constants.

Warshel and coworkers<sup>45,76,77</sup> have recently pointed out that the design of a new catalyst must be focused not only in placing the reacting fragments in proper places around the substrate but in optimizing the full environment preorganization. This preorganization accounts for the required electrostatic properties of the enzyme to lower the activation free energy of the reaction, a property that depends not only on residues of the active site. Using an Empirical Valence Bond (EVB) approach, these authors studied several Kemp eliminases (catalytic antibodies, albumines and *de novo* designs) remarking the difficulties of designing a correctly preorganized environment adapted for the small



changes in the electronic distribution observed during the formation of the TS. A larger preorganization of the environment is closely related to a smaller reorganization of the protein needed to attain the TS of the reaction. This reorganization implies an energy penalty that should be minimized to improve the catalysts design, an aspect that was emphasized also by Labas et al.<sup>78</sup>

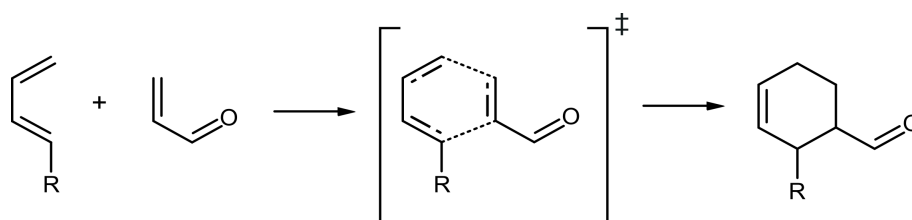
A different approach, based on studies of failed designs of new enzymes, has been applied for the development of more efficient catalysts. In particular, Mayo and co-workers used an iterative approach starting from an inactive protein scaffold, HG-1, that was designed to convert the xylan binding pocket of a thermostable xylanase into a Kemp eliminase.<sup>79</sup> Based on MD simulations and analysis of structures obtained from X-ray crystallography, the authors found that the inactivity could be due to bonded waters and high flexibility of residues in the active site. These findings guided the design of a more embedded binding pocket, which moved the active site deeper into the protein and resulted in an active Kemp eliminase, HG-2. MD simulations of this last structure led to an additional mutation that provided a better packing around the substrate by reducing the active site conformational heterogeneity observed in HG-2. The new designed protein, HG-3, exhibited activity comparable to the best Rosetta design.<sup>79</sup> This new catalyst was proved to be particularly amenable to direct evolution based on multiple rounds of mutagenesis and screening. The turnover number for proton transfer of the best variant obtained in the Hilvert laboratory after 17 rounds of mutagenesis and screening, HG3.17, was comparable to the efficiency of many natural enzymes.<sup>80</sup> The three factors that were suggested to be decisive in the improvement of the Kemp eliminase activity from HG-3 to HG3.17 were: 1) the extraordinary high shape complementarity between the binding pocket of the protein and the substrate (while two possible conformers of substrate were able to bind the pocket of HG3, only one orientation was detected in HG3.17); 2) the ligand alignment with Asp127, the catalytic base, was optimized by evolution resulting in an unusual short hydrogen bond; and 3) a new catalytic group, Gln50, stabilized the development of a negative charge in the TS.<sup>80,81</sup>

Recently, hybrid QM/MM MD simulations for the Kemp elimination in this new HG3.17 enzyme have provided a deeper insight into the origin of its catalytic efficiency.<sup>82</sup> In this case, we have demonstrated the presence of different conformations with significant different reactivity. The larger reactivity has been demonstrated to be related with a better electrostatic preorganization of the environment that creates a more

favourable electrostatic potential for the reaction to proceed. In HG3.17, efforts to improve the catalytic properties must be focused in possible mutations increasing the preorganization and decreasing the reorganization around the oxyanion hole. Mutations should be considered not only in the first shell of residues but in further shells since protein electrostatics is a long range property.

### 3. Diels Alder

Reactions involving the forming of two carbon-carbon bonds, such as the Diels-Alder reaction (see Scheme 2), are very powerful tools in organic synthesis and they have been identified in several metabolites. Nevertheless, the number of enzymes catalysing this cycloaddition reaction, if any, is very low.<sup>83,84</sup> In fact, a recent work has claimed for the first *in vitro* characterization of an enzyme committed to the catalysis of a [4+2] cycloaddition reaction.<sup>85</sup>



Scheme 2. Representation of the Diels Alder cycloaddition reaction

High regioselectivity (ortho and para isomers are favoured with respect to the meta), diastereoselectivity (endo conformers are favoured with respect to the exo) and enantioselectivity (the reaction creates up to four new stereocenters) are probably key features that are searched when designing new enzymes catalysing the addition of a diene with a dienophile.<sup>9,11</sup> The enhanced reaction rate measured in new designed Diels-Alderase has been proposed to be due to a lowering of the entropy of activation by binding the two substrate molecules, also known as “entropy trap”.<sup>86</sup> However, a reduction in the activation enthalpy has been also proposed as the origin of catalysis.<sup>87</sup>

#### 3.1 catalytic antibodies

The first designed catalytic antibody (1E9) showing activity as a Diels-Alderase was reported by Hilvert and co-workers.<sup>88</sup> Braisted and Schultz, one year later, designed another catalytic antibody (39A11) with a different hapten.<sup>89</sup> Series of site-directed

mutagenesis carried out by Romesberg and Schultz allowed obtaining a mutant that displays an increase of one order of magnitude in its catalytic activity.<sup>90</sup> Three-dimensional crystal structures of these antibodies complexed with their respective TSA have suggested an almost perfect shape complementarity between protein cavity and the substrates in the TS conformation.<sup>91,92</sup> A deep analysis of 39A11 structure and its mutant revealed the presence of not only a favourable packing but hydrogen bond interactions that could be responsible of the control of the relative geometries of the bound substrates and the electronic distribution in the dienophile.<sup>92</sup> A theoretical study on the germline catalytic antibody and its matured form, 39A11, based on hybrid QM/MM MD simulations, suggested a complex indirect effect through coupled movements of the backbone of the protein and the substrate in the germline, after mutation.<sup>93</sup> These results stress the importance of backbone flexibility in docking studies and in the design of new enzymes.

A combination of experimental studies and quantum mechanical gas phase calculations, carried out by Gouverneur et al., was focused on the stereoselectivity of newly generated catalytic antibodies. In particular, the cycloaddition between trans-1-N-acylamino-1,3-butadiene and N,N-dimethylacrylamide was used to explore the control of the endo/exo and absolute configuration of the eight possible reaction products.<sup>94</sup> These substrates were in fact the ones used in the *de novo* design of new Diels-Alderase, as shown in next section. A new catalytic antibody, 13G5, was made to catalyse the disfavoured exo Diels-Alder transformation, yielding a single regioisomer in high enantiomeric excess.<sup>95</sup> This surprising absolute enantioselectivity of 13G5 was explained based on gas phase quantum mechanical calculations followed by docking and MD simulations.<sup>96</sup>

### 3.2. *de novo* design of new enzymes

The first *de novo* design of an enzyme catalysing the bimolecular Diels-Alder cycloaddition reaction, with high stereoselectivity and substrate specificity was performed by Siegel et al.<sup>18</sup> After studying the normal-electron-demand reaction by quantum mechanics in gas phase, frontier molecular orbital analysis dictated that narrowing the energy gap between the highest occupied molecular orbital (HOMO) of the diene with the lowest unoccupied molecular orbital (LUMO) of the dienophile should reduce the activation energy and thus it would increase the rate of the reaction. This was achieved by positioning a hydrogen bond acceptor to interact with the former

(raising the energy of the HOMO) and a hydrogen bond donor to interact with the later (thus lowering the energy of the LUMO). These electrostatic requirements were added to efforts to improve the binding of the substrates in optimal position in the active site of the designed enzyme. The design process starts from generating three dimensional models of minimal active sites (theozymes) that, in this case, consist in a glutamine or an asparagine to accept hydrogen bond from the N-H moiety of the diene carbamate, and a serine, threonine or tyrosine to donate a hydrogen bond to the carbonyl oxygen of the dienophile amide moiety. Afterwards, RosettaMatch<sup>21</sup> program was used to search stable protein scaffolds that allow the theozyme to be placed without making substantial steric clashes with the protein backbone. Then, RosettaDesign<sup>24,25</sup> was used to optimize the set of putative stable proteins. And finally, a reduced number of designs were selected for experimental validation proving that only two of them (DA\_20 and DA\_42) had catalytic activity. After rounds of directed evolution, further improvement of catalytic activity of DA\_20\_10 (100-fold over the original design) was obtained, with high stereoselectivity that favoured formation of only one of the stereoisomers (3R,4S). *De novo* design can be improved if backbone flexibility is considered. In this sense, the search for adequate conformations can be enhanced with creative strategies such as the one used by the Foldit multiplayer online game.<sup>97</sup> In this game, players compete and collaborate solving structure refinements problems where backbone flexibility is considered. This strategy was applied to remodel the backbone of DA\_20\_10 Diels-Alderase to enable additional interactions with the substrates, resulting in a design (CE6) with a modest increase of the enzymatic activity of about 20-fold in  $k_{cat}$ .<sup>98</sup> Rounds of directed evolution were used by Hilvert and coworkers to obtain a Diels-Alderase (CE20) 6-times more efficient than CE6, reshaping the active site pocket for a better substrate preorganization and TS stabilization.<sup>99</sup>

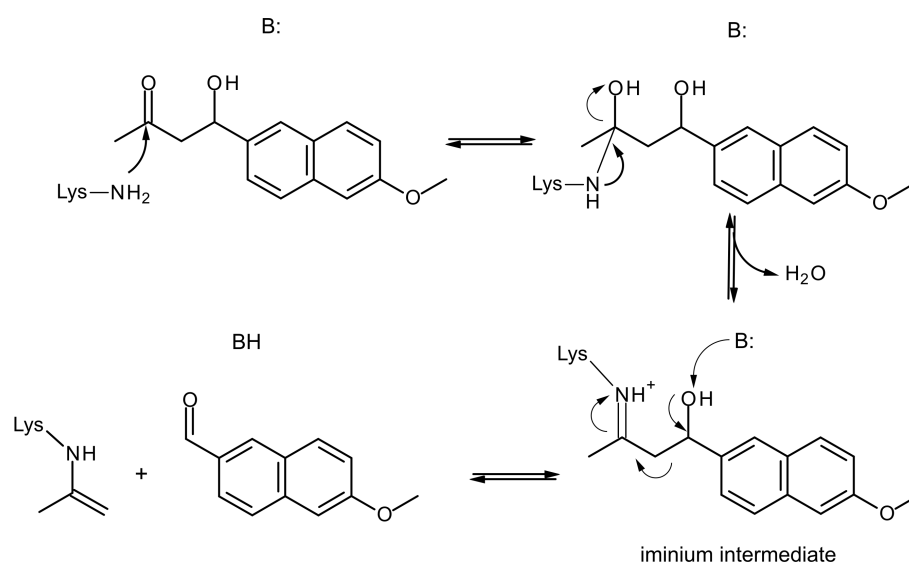
A different semirational combinatorial design protocol has been used by Brinck and coworkers to design new Diels-Alderases with significant rate accelerations.<sup>100,101,102</sup> The computational protocol employs molecular docking, MD and quantum chemical calculations at DFT level to find reasonable candidate protein and substrate combinations for which protein-substrate interactions were optimized. The estimation of the free energy of activation is based on two steps; the free energy difference going from the bound state to conformations named as the “near attack-conformation” (NAC),<sup>103</sup> and the energy required to go from NAC to the TS. The first term is computed from the MD simulations while the second term is derived from the DFT

calculations. In particular, efforts were focused in finding and optimizing a protein structure for a suitable oxyanion hole that would stabilize the LUMO of the dienophile,<sup>101</sup> or in developing a route based on *in situ* acid-base catalysed diene generation.<sup>102</sup>

Other strategies to obtain new Diels-Alderases based on the use of metalloenzymes will be presented in section 5.

#### 4. Retro-Aldol reaction

The Retro-Aldol reaction, in contrast to previous presented reactions, consists in a multi-step process that results in a C-C bond breaking. The reaction mechanism involves a nucleophilic lysine and the formation of an iminium intermediate (see Scheme 3).



Scheme 3. Representation of the Retro-Aldol reaction mechanism

To account for the multistep reaction mechanisms, the designed new enzyme has to present an active site simultaneously compatible with different TSs and intermediates. Thus, there must be a compromise between structural reorganization during the catalytic process and the preorganization of the active site for such multistep enzyme catalysed reaction. The enzyme must present some motions to reach some degree of active site reorganization to stabilize each TS. Hilvert, Houk and co-workers have stated that these motions may be divided into intrastep and interstep reorganization motions.<sup>104</sup>

#### 4.1 catalytic antibodies

The first catalytic antibodies that showed activity for the Aldol reaction, 38C2 and 33F12,<sup>105</sup> were generated to catalyse the carbon-carbon bond formation. These antibodies showed to also catalyse the Retro-Aldol reaction.<sup>106,107</sup> Later, new aldolase catalytic antibodies were designed with special attention to the enantio- and diastereoselectivity.<sup>108</sup> The mechanism for antibody catalysis of this reaction mimics that used by natural class I aldolase enzymes that utilize the  $\epsilon$ -amino group of an active site lysine residue to form a Schiff base with the substrate (see Scheme 3). X-ray crystallographic studies on 33F12 confirmed the presence of a lysine residue, deeply buried in a hydrophobic pocket at the base of the binding site, thereby accounting for its perturbed  $pK_a$ .<sup>109</sup> Further structural analysis by X-ray diffraction studies gave a direct observation of the enamine intermediate.<sup>110</sup> Characterization of the Aldol reaction mechanism and the proposal of a TSA was made by Arnó and Domingo based on quantum mechanical calculations within reduced models.<sup>111</sup>

#### 4.2. *de novo* design of new enzymes.

The first *de novo* protein design for the Retro-Aldol reaction was published in 2008 by Baker and co-workers.<sup>19</sup> In this case, four different theozymes were prepared all presenting a lysine as a Schiff base. The four active site motifs optimized with quantum mechanical methods differ in the general acid/base for deprotonation of the  $\beta$  alcohol: Lys/Asp dyad, Tyr, His/Asp dyad or a water molecule. The first proposal was analogous to that with D-2-deoxyribose-5-phosphate aldolase, while the other three motifs were inspired on the active site of catalytic antibodies, as described above. Using Rosetta program<sup>24,25</sup> for different active site geometries of the four different motifs that fitted into the binding pockets of 71 protein scaffolds, a total of 72 designs were selected for experimental characterization. The backbone geometries and side-chain orientation of two apo structures solved by X-ray diffraction studies showed an excellent agreement with those of the designed enzymes. 32 of them showed detectable retro-aldolase activity, although the rate enhancements were between two and three orders of magnitude lower than the observed in the catalytic antibodies.<sup>107,108</sup>

A computational study of these *de novo* designed enzymes, based on MD simulations, showed that the protein distorted the designed geometry of the substrate-enzyme intermediate thus suggesting that the substrate and the catalytic residues were poorly packed in the models. These results emphasize that the assumption of the rigid

backbone during the computational enzyme designs is one important limitation. Further studies on one of the most active new enzymes, corresponding to the last motif, revealed that the catalysis was largely due to hydrophobic substrate binding interactions and, to a lesser extent, to the shifting of the  $pK_a$  of the catalytic lysine.<sup>112</sup> Structural analysis of covalently bound mechanism based inhibitor showed to be similar but not identical to the design models.<sup>113</sup> Thus, while the substrate appears to be well oriented in the binding pocket through hydrophobic interactions with the naphthyl moiety, as predicted in the model, the relative orientation with the catalytic lysine is different. The bound water network appears to be more extensive, and the bound substrate analog exhibits more conformational heterogeneity than typical native enzyme-inhibitor complexes.

In an attempt to improve the efficiency of the new designed enzymes, Hilvert, Baker and co-workers optimized the protein via directed evolution,<sup>114,115</sup> from the previously computationally retro-aldolases that were designed from scratch.<sup>19</sup> The specific activity was boosted more than 4,400-fold by random mutagenesis and screening, affording catalytic efficiencies approaching those of natural enzymes. Surprisingly, dramatic changes were detected in the active site by comparison with the original structure. In particular, the reactive catalytic lysine and the ordered water molecule were unexpectedly abandoned and another lysine residue was found at another binding pocket created during the optimization process.<sup>115</sup> The drawback of the original computational design was the assumption of a rigid protein scaffold during the simulations.<sup>115</sup> More recently, Hilvert and co-workers have designed a new Retro-Aldolase by directed evolution from the previously computationally designed RA110.<sup>116</sup> The original activity, that accelerated substrate cleavage by a factor of 12000 over background, was increased more than 1000-folds. The X-ray structure of the new catalyst with an inhibitor showed, again, substantial deviations from the designed model. Finally, Baker and co-workers have investigated an alternative active site design in which the catalytic water molecule was replaced by the side chain of a glutamic acid. After site-directed evolution, the rate enhanced of the new enzyme, RA110.4-6, was still below the best values obtained for catalytic antibodies and the previous example.

## **5. Artificial Metalloenzymes**

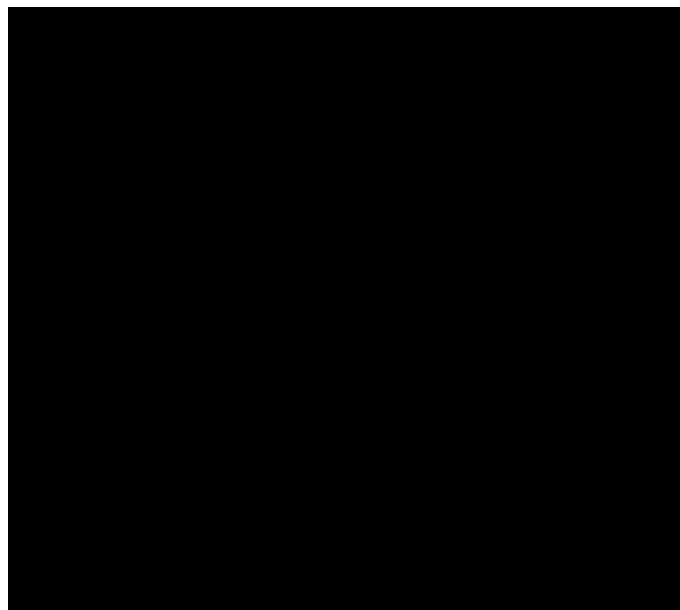
Metalloproteins, which represents almost half of the enzymes in nature, catalyse some of the most difficult biological transformations by complex reaction mechanisms.<sup>117,118</sup>

This is the origin of the extensive work devoted to the development of strategies for the design of artificial metalloenzymes as stressed in recent reviews.<sup>119,120,121,122,123,124,125,126</sup>

One of the main targets in this field is to merge the attractive properties of homogeneous catalysis and biocatalysis (see Figure 2). It is envisioned that artificial metalloenzymes can combine the best of both worlds; the broad catalytic scope, a hallmark of homogeneous catalysis, and the high activity and selectivity under mild conditions that characterize enzymatic catalysis. The main structural characteristic of metalloenzymes is the presence of a metal-ligand complex that is anchored to the biomolecule host and provides the first coordination sphere whereas the host provides the second one. Two different anchoring strategies can be used; noncovalent, which includes anchoring through supramolecular and dative interactions, and covalent anchoring, usually through a cysteine residue.<sup>117,118,119,120</sup> Nevertheless, non-native functionalities has also been expanded through the incorporation into the protein scaffold of metal-containing cofactors not found in nature.<sup>118</sup> In nature, proteins incorporate metal cofactors such as haem by means of non-covalent and covalent interactions.<sup>118</sup> The *de novo* design of an artificial metalloenzyme is also related with the modelling of the protein scaffold from first principles. This can imply the redesign of a critical loop to improve the key side-chain interactions established with the substrate.<sup>127</sup> Other efforts have been focused in the engineering of metal-binding sites containing cysteine residues in an existing three-helix bundle,<sup>128</sup> or the generation of combinatorial libraries of *de novo* proteins (named “artificial superfamilies”) containing, for instance, four-helix bundles and presenting catalytic activity for a variety of functions.<sup>129</sup> It is then not surprising that the border between the *de novo* protein design and the redesign based on existing proteins is not well defined.<sup>126,130</sup>

As exposed above, due to the large number of designed metalloenzymes, only selected examples will be highlighted in this review. In particular, we will first focus on metalloenzymes with catalytic activity in the Diels-Alder reaction. A section devoted to the design of proteins with zinc hydrolytic centers will be followed by examples of *heme no-heme* artificial enzymes, and we will finish with some examples of technologies based on the supramolecular anchoring approach.

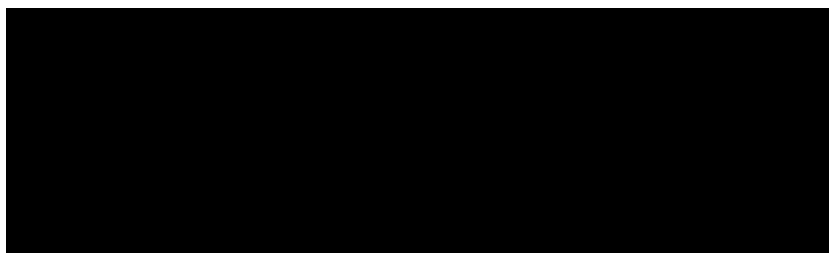




**Figure 2.** Schematic representation of an artificial metalloenzyme, where a metal ligand complex is posed in the cavity of an enzyme by means of different possible anchoring strategies.

### 5.1 Diels-Alder

A metalloenzyme that catalyses the Diels-Alder cycloaddition of azachalcone (see Scheme 4) have been created by designing a Cu(II)-specific binding site with coordinating amino acids at geometrically appropriate positions in a thermoestable host protein (tHisF).<sup>131,132</sup> The catalytic activity has been assumed to be due to the five-membered chelate structure made by the lone electron pairs at the carbonyl O atom and the pyridine N atom, activating the substrate by lowering the energy of the LUMO. Authors observed notable enantioselectivity, enhanced endo-selectivity, and higher reaction rate. The same asymmetric Diels-Alder reaction has been used to design an artificial metalloenzyme from a different protein scaffold, a sterol carrier protein type 2 like domain (SCP-2L), obtaining a promising enantioselectivity of 25% ee and good endo selectivity<sup>133</sup>



Scheme 4. Representation of the Diels-Alder cycloaddition between azachalcone and cyclopentadiene

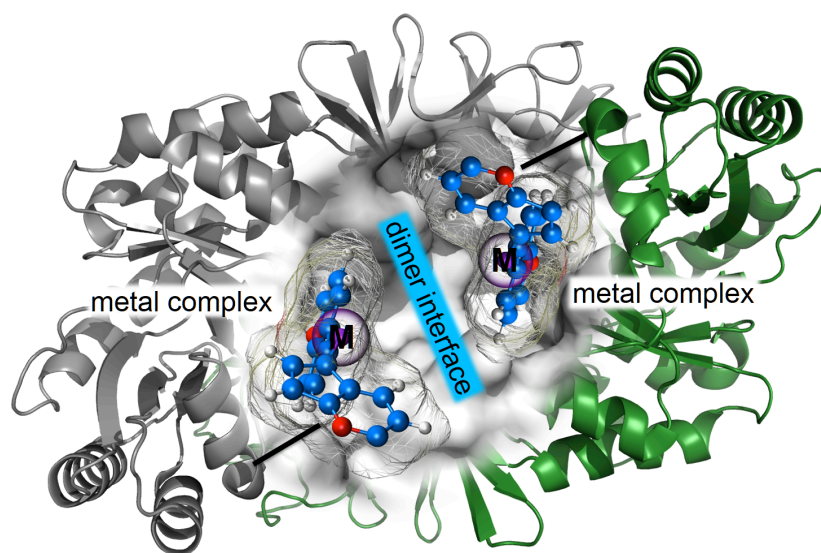


Figure 3. Schematic representation of of proposed artificial metalloenzymes based on Cu(II) complexes grafted on the dimer interface of a protein.

Roelfes group, using a different strategy, has also studied the Diels-Alder cycloaddition between azachalcone and cyclopentadiene as a benchmark reaction to design new metalloenzymes.<sup>134</sup> First, an active site is generated on the dimer interface of the Lactococcal multidrug resistance Regulator (LmrR) (see Figure 3). A copper(II) phenantroline complex was covalent anchored in the hydrophobic pocket of the protein using a cysteine conjugation strategy. The resulting artificial metalloenzyme showed excellent enantioselectivity (up to 97%) and diastereoselectivity (endo/exo 95:5).<sup>134</sup>

Similar to proteins, DNA can offer a defined chiral second coordination sphere for the construction of an artificial metalloenzyme, as introduced by Roelfes and Feringa for

the asymmetric Diels-Alder catalysis.<sup>135,136,137</sup> A DNA-based catalyst was designed using a copper complex of a non-chiral ligand that can bind to DNA (see Figure 4). The active  $\text{Cu}^{2+}$  center was brought into proximity of the DNA double helix, allowing transfer of chirality from DNA to the product of  $\text{Cu}^{2+}$ -catalyzed Diels-Alder reaction. Products were obtained with very high diastereoselectivities and enantioselectivities (83-98%).<sup>135,136</sup> As explained, the catalytic ensemble comprises an intercalator and a copper-chelating moiety. In the first generation, these two moieties were linked by a spacer, whereas in the second generation both sites are integrated into a single bipyridine moiety and thus the reactive copper center is brought into closer contact with the chiral environment. These findings underline the fact that the design of the ligand is an important factor in the development of this kind of artificial catalysts.<sup>117</sup> The second generation of catalysts are also capable of enhancing the rate of the reaction up to two orders of magnitude.<sup>137</sup>

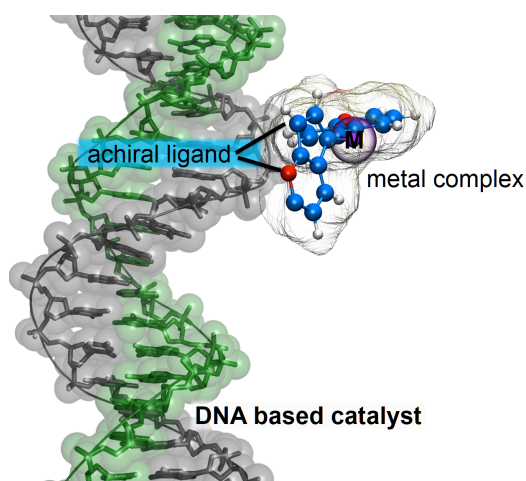


Figure 4. Schematic representation of of proposed artificial metalloenzymes based on modular assembly of a copper complex of a non-chiral ligand to a DNA scaffold.

## 5.2 Zinc Hydrolytic centers

In the preceding section we have presented two different strategies for the design of Diels-Alder metalloenzymes, based on the complexation of a  $\text{Cu}(\text{II})$  metal ion, or a metal complex. In this section we focus in the case metalloenzyme designs based in the complexation of just metal ions such as  $\text{Zn}(\text{II})$  or  $\text{Fe}(\text{II})$ .

As stated in a recent review by Alexandrova and co-workers,<sup>138</sup> the selection of the catalytic metal in the process of metalloenzymes evolution can be due to factors other

than optimizing catalysis alone, such as metal availability or toxicity. Enzymes evolved under these restrictions are good enough for vital purposes but they do not necessarily reach the best catalytic performance. Consequently, these limitation provides an excellent opportunity to improve the catalytic properties of these enzymes in the laboratory substituting the metal for another one. In this sense, computational predictions based on DFT calculations have shown that  $\beta$ -lactamases could be improved via replacement of zinc by nickel.<sup>139</sup> It must be taken into account that one of the main limitations in the computational study of metalloenzymes is the description of metal ions and its coordination shell. Typically, classical force fields and semiempirical methods, and also some DFT functional or *ab initio* treatments, can lead to erroneous descriptions of the coordination index of the metal and in particular its change during the reaction process.

Active site redesign changing metal ions, has been found to be a promising strategy to improve not only the primary catalytic reaction but also secondary processes in promiscuous enzymes. As an example, Leitgeb and Nidetzky enhanced the secondary hydrolase activity of a promiscuous oxygenase  $\beta$ -Diketone-cleaving enzyme by changing  $\text{Fe}^{2+}$  to  $\text{Zn}^{2+}$  in the active site.<sup>140</sup>

Another strategy to design metalloproteins was reported by the Baker group as an extension of their *de novo* enzyme design methodology to exploit the reactivity of metal ions in existing enzyme active sites.<sup>141</sup> In particular, the method was based in the redesign of existing suitable native crystallized proteins containing mononuclear zinc centers with at least one free coordination site. TS models of organophosphate hydrolase reaction were modelled into the zinc sites and additional hydrogen-bond interactions were identified using their RosettaMatch program. After maturation using directed evolution, a  $10^7$ -fold increase in the catalytic activity was obtained. Warshel and co-workers have suggested that designs on mononuclear zinc metalloenzyme can be improved by optimizing the preorganization of the environment, which is the key factor in enzyme catalysis.<sup>142</sup>

Recently, Pecoraro and co-workers have designed efficient hydrolytic zinc sites to create an artificial hydrolase, with catalytic activity that compares well with natural carbonic anhydrases.<sup>143,144,145</sup> X-ray crystallography showed that the designed protein contained two different metal ions; Zn(II), which is important for catalysis, and a Hg(II) ion, which provides structural stability. In these hydrolytic zinc metalloenzymes, it was

demonstrated the influence of location of the active site formed by the metal and three histidine residues in a three-stranded coiled-coil oligopeptide.

Kuhlman and co-workers serendipitously discovered that a computationally designed Zn(II) site at the interface of a homodimer (MID1-Zn) could effectively catalyze carboxyester and phosphoester hydrolysis.<sup>146,147</sup> A Rosetta-based approach was used for the rational design of a protein monomer to form a zinc-mediated, symmetric homodimer. The computationally designed zinc-mediated protein showed a good overall agreement with the X-ray structure,<sup>146</sup> and suggested that the clefts formed by the protein-protein interactions were well suited for creating enzyme active sites.<sup>147</sup>

### 5.3 heme no-heme artificial enzymes

This section deals with catalysis by bimetallic centers, where one of the metals forms part of a heme complex and the other is directly complexed by enzymatic residues. Example of key reactions that are catalysed by heme no-heme artificial enzymes are the reduction of NO to NO<sub>2</sub> and the reduction of O<sub>2</sub> to H<sub>2</sub>O. Nitric oxide reductase (NOR), is a metalloenzyme that participate in the nitrification pathway and is critical for life. A first example of the success of designing a NOR artificial enzyme is the conversion from a myoglobin by introducing a non-heme Fe<sub>B</sub> site in the distal heme pocket. One glutamate residue and two histidines were required in this Fe<sub>B</sub> active site for both iron binding and NOR activity.<sup>148,149</sup> The introduction of a second glutamate residue in a myoglobin protein scaffold, that appears to interact with a water molecule forming a hydrogen bonding network, was demonstrated to increase 100-fold the NOR activity. It was suggested the importance of this hydrogen bonding network in facilitating the proton delivery for NOR reactivity.<sup>150</sup> Replacing the Fe metal in the no-heme center by other cations shows that Cu dependent NOR activity was significantly higher than that of Fe in the same metal binding site.<sup>151</sup> Molecular dynamics simulations on a mutant with an additional distal histidine (L29H) and a distal tyrosine (F43Y) in the heme pocket, to mimic the active site of cytochrome cd1 nitrite reductase from *Ps. aeruginosa*, revealed that the new protein contains the necessary structural features of the later. This strategy provides rationality for protein design and can guide further protein engineering.<sup>152</sup>

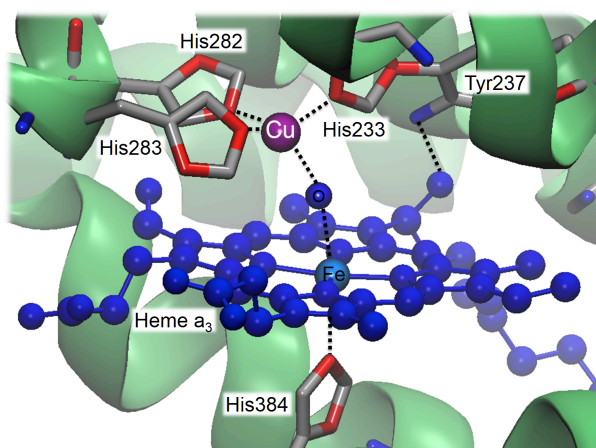


Figure 5. Crystal structure of the binuclear center of cytochrome ba3 oxidase (PDB code 1XME).

In a similar way, myoglobins have been also engineering into a heme copper oxidase (HCO) that performs the reduction of oxygen to water, by introducing a histidine residue near the heme site to create a Cu<sub>B</sub> site. Thus, the corresponding active site consists of a heterobinuclear center containing a heme Fe and a histidine ligated Cu (Cu<sub>B</sub>) as well as a post-translationally installed crosslink between a tyrosine residue and a close lying histidine (see Figure 5). The designed functional metalloenzyme showed a HCO activity with over one thousand turnovers.<sup>153,154</sup> Genetic incorporation of an unnatural aminoacid, which mimics the cross-linked Tyr-His ligand, allowed stressing the effect of Tyr and His with and without cross-linking (obtained after a selected mutation that avoided its formation). The new designed model achieved higher activity for O<sub>2</sub>-reduction, with an eightfold increase in selectivity and a threefold increase in catalytic turnover.<sup>154</sup>

#### 5.4 Supramolecular Anchoring Approach

The concept of introducing a homogeneous catalyst within a protein environment to create artificial metalloenzymes was first proposed by Wilson and Whitesides in the early 1970s.<sup>155</sup> Examples of biomolecular scaffolds capable of embedding a catalytically active transition metal complex include avidin, streptavidin, bovine serum albumin and myoglobin.<sup>134</sup> The extraordinary affinity of the molecule biotin for either avidin or streptavidin makes this case as one of the most employed systems. Moreover,

these protein scaffolds present a binding pocket that is large enough to bind the organometallic catalyst and still leave space for the substrate.<sup>117,134</sup> Biotin is covalently linked to a catalyst precursor, a metal complex, ensuring that the metal moiety is quantitatively incorporated within the host protein upon stoichiometric addition of (strep)avidin (see Figure 6).<sup>156</sup> Another important advantage of these hybrid supramolecular catalysts is that, since the biotinylated catalyst precursor and the host protein are developed separately of one another, they can be optimized separately through synthetic and molecular biology approaches.<sup>126</sup>

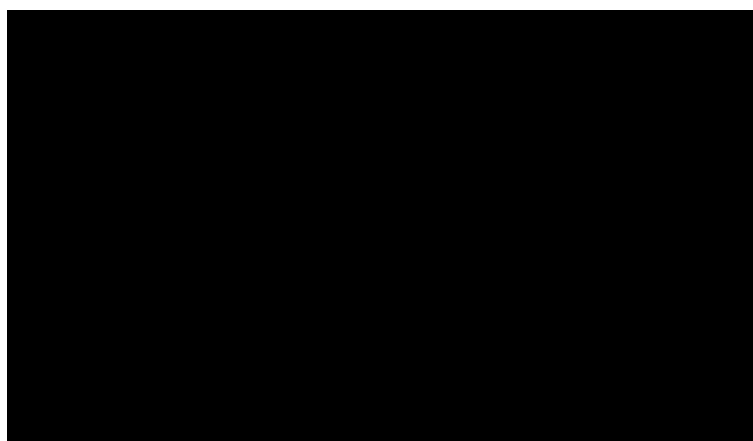


Figure 6. Schematic representation of artificial metalloenzymes based on the biotin-avidin technology.

Computational simulations of protein-ligand docking and experimental studies based on electron paramagnetic resonance spectroscopy, chemical modifications and mutagenesis studies, suggest that a vanadyl ion is located within the biotin-binding pocket and interacts only via the second coordination sphere contacts with strep(avidin).<sup>157</sup> Many reactions have been catalysed following this synthetic strategy.<sup>156,158</sup> Interestingly, a theoretical study based on MD simulations has shown significant flexibility of the biotinylated ligand in artificial metalloenzymes based on streptavidin.<sup>159</sup> The creation of a bifunctional artificial metalloenzyme with a reduced flexibility was generated by introducing a base (an aspartate or a glutamate residue) into the streptavidin in concert with a docked biotinylated rhodium(III) complex to enable asymmetric carbon-hydrogen activation.<sup>160</sup> A similar strategy has been used to improve the catalytic performance in the transfer hydrogenation of imines by an *in silico* design of an

artificial metalloenzyme. The designed structure, confirmed by X-ray crystallography, contained a suitable positioned histidine residue that firmly anchor, via a dative bond, a biotinylated rhodium piano stool complex within streptavidin.<sup>161</sup>

An integrated computational approach, combining protein-ligand docking, quantum mechanical, and QM/MM calculations, has been used to provide a deep insight into the mechanistic behaviour of the first artificial metalloenzyme generated by supramolecular anchoring approach able to reduce enantioselectively cyclic imines.<sup>162</sup> The identification of the TS geometries leading to (R)- and (S)-reduction products in the presence of the full hybrid system allowed the authors to identify the distinctive pattern of interactions between the substrate and the binding site, responsible of the observed results.

## 6 Future perspectives

The design of new biocatalysts is an ambitious target that requires a detailed knowledge of the reaction to be catalyzed, and how natural enzymes achieve their catalytic efficiency in mild environmental conditions. In this regard, the combination of experimental and computational techniques has been revealed as a very reasonable strategy. As mentioned by Hilvert and co-workers, precision is essential for development of new biocatalysts based on the combination of *de novo* computational design with laboratory evolution techniques.<sup>80</sup> New and more sophisticated engineering protocols are required to approach to the rate constants usually observed in native enzyme catalysed processes.

The success in the design of new enzymes with new functions has potential applications in different chemical and biochemical industries. This review has been focused on the advances obtained in the field of the *de novo* design during the last years using examples of increasing complexity. Kemp, Diels-Alder and retro-aldol, three of the most popular benchmarks reactions in the field designing new enzymes have been used to illustrate recent achievements. The first one has been chosen due to its simplicity and the fact that no natural enzyme catalyses this reaction. Diels-Alder is a bimolecular reaction that can produce many different isomers. The control of the products specificity is a feature that, in fact, can be more important than enhancing its rate constant. Finally, the design of new biocatalysts for the retro-aldol reaction represents an additional difficulty since the reaction takes place in several chemical steps. Thus, the new enzyme must properly represent the features of several TSs and, at least, one intermediate. Conformational changes of the protein to form the Michaelis complex (preorganization)



has to take place along the reaction to stabilize different intermediates of the different steps. This requires a coupling between the reaction progress and the conformational changes of the protein to get a proper preorganization for each chemical step. Finally, a section has been devoted to the particularities of the design of metal-containing enzymes, a promising field due to the combination of the best features of homogeneous catalysis and enzyme catalysis; the broad catalytic scope and the high activity and selectivity under mild conditions, respectively. Nevertheless, since one or more metal ions, or metal complexes, are involved in the process, the difficulties of designing this kind of biocatalysts dramatically increase.

From the pioneer work of Mayo and co-workers in the *de novo* design of an esterase,<sup>14</sup> several groups have focused the attention into the design of catalysts for the three mentioned reactions by means of similar strategies, still with a limited success. Different improvements have been incorporated during the past decade, such as partially considering the flexibility of the backbone of the template protein or the use of non-canonical amino acids. Nevertheless, the rate enhancement reached by these new biocatalysts is usually far from those of natural enzymes. As mentioned by Baker,<sup>37</sup> and stressed in the introduction section of this review, the *de novo* based design strategies can fail due to an imperfect theozyme, which does not represent the real TS of the reaction, a distortion of the designed active site into a given protein scaffold, or due to the effect of the long range electrostatic interactions and/or protein dynamics that can be incompatible with catalysis. Nevertheless, long range electrostatic interactions, their knowledge and simulation, can be used as a guide to the design of the new enzyme, instead of considering them as a limitation. Mutations far from the active site can be proposed to improve this kind of interactions, and the flexibility of the protein must be considered in any step of the design protocol. It must be kept in mind that while efforts have been focused exclusively in the chemical step of the reaction to be catalyzed, it will be interesting to extend these methodologies to analyze the influence of mutations on steps like substrate binding or product release. This could improve the substrate affinity and/or the catalytic turnover, respectively. Increasing the preorganization of the active site would likely improve the catalytic efficiency. In addition, a more complete design strategy should consider not only the TS stabilization but the differential stabilization of the TS with respect to the reactants. Thus, further improvements on the *de novo* design of new enzymes will be probably based in better localization and characterization of the TS and reactants by means of high level computational methods.

The use of flexible molecular models treated by QM/MM multiscale methods can be the bedrock of future successful studies. A proper sampling of the protein conformational landscape, performed by Molecular Dynamics simulations, could give also information related with the capability of a designed enzyme to catalyze a multistep-process. In these more complex targets, the active site of the enzyme has to be prepared not only to stabilize the TS but to accommodate and stabilize the system for different steps. In this refinement, the improved packing should be considered as not only a mechanical embedding but also an electrostatic one. The role of not only residues belonging to the first shell of the active site should be taken into account.

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 from scratch. This is the final goal in rational design of proteins.

### **Acknowledgements**

This work was supported by the Spanish Ministerio de Economía y Competitividad for project CTQ2012-36253-C03, Universitat Jaume I (project P1·1B2014-26), Generalitat Valenciana (PROMETEOII/2014/022 and ACOMP/2014/277 projects), Polish National Center for Science (NCN) (grant 2011/02/A/ ST4/00246, 2012–2017), the Polish Ministry of Science and Higher Education ("Iuventus Plus" program project no. 0478/IP3/2015/73, 2015-2016) and the USA National Institute of Health (ref. NIH R01 GM065368). Authors acknowledge computational resources from the Servei d'Informàtica of Universitat de València on the 'Tirant' supercomputer and the Servei d'Informàtica of Universitat Jaume I.

## References

- <sup>1</sup> A. Warshel, P. K. Sharma, M. Kato, Y. Xiang, H. Liu, M. H. M. Olsson. *Chem. Rev.* 106 (2006) 3210-3235
- <sup>2</sup> S. J. Benkovic, S. Hammes-Schiffer. *Science* 301 (2003) 1196-1202
- <sup>3</sup> M. Garcia-Viloca, J. Gao, M. Karplus, D. G. Truhlar. *Science* 303 (2004) 186-195
- <sup>4</sup> S. Martí, M. Roca, J. Andrés, V. Moliner, E. Silla, I. Tuñón. *J. Bertrán Chem. Soc. Rev.* 33 (2004), 98-107
- <sup>5</sup> K. Steiner, H. Schwab, *Comput. Struc. Biotech. J.* 2 (2012) e2012209010
- <sup>6</sup> M. N. Carbone, F. H. Arnold. *Curr. Opin. Struct. Biol.* 17 (2007) 454-459.
- <sup>7</sup> S. Martí, J. Andrés, V. Moliner, E. Silla, I. Tuñón, J. Bertran. *Chem. Soc. Rev.* 37 (2008) 2634-2643.
- <sup>8</sup> K. L. Morley, R. J. Kazlauskas. *Trends Biotechnol.* 23 (2005) 231-237.
- <sup>9</sup> G. Kiss, N. Celebi-Olcum, R. Moretti, D. Baker, K. N. Houk. *Angew. Chem. Int. Ed.* 52 (2013) 5700-5725.
- <sup>10</sup> E. Keinan (Ed.). *Catalytic Antibodies*, Wiley-VCH, Verlag Weinheim, 2005.
- <sup>11</sup> D. Hilvert. *Annu. Rev. Biochem.* 82 (2013) 447-470.
- <sup>12</sup> M. D. Toscano, K. J. Woycechowsky, D. Hilvert. *Angew. Chem. Int. Ed.* 46 (2007) 3212-3236.
- <sup>13</sup> M. S. Humble, P. Berglund. *Eur. J. Org. Chem.* (2011) 3391-3401.
- <sup>14</sup> D. N. Bolon, S. L. Mayo. *Proc. Natl. Acad. Sci. USA* 98 (2001) 14274-14279.
- <sup>15</sup> B. I. Dahiyat, S. L. Mayo. *Science* 278 (1997) 82-87.
- <sup>16</sup> J. K. Lassila, H. K. Privett, B. D. Allen, S. L. Mayo. *Proc. Natl. Acad. Sci. USA* 103 (2006) 16710-16715.
- <sup>17</sup> 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, D. Baker. *Nature* 453 (2008) 190-95.
- <sup>18</sup> J. B. Siegel, A. Zanghellini, H. M. Lovick, G. Kiss, A. R. Lambert, J. L. St. Clair, J. L. Gallaher, D. Hilvert, M. H. Gelb, B. L. Stoddard, K. N. Houk, F. E. Michael, D. Baker. *Science* 329 (2010) 309-313.
- <sup>19</sup> 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 III, D. Hilvert, K. N. Houk, B. L. Stoddard, D. Baker. *Science* 319 (2008) 1387-1391.
- <sup>20</sup> D. J. Tantillo, J. Chen, K. N. Houk. *Curr. Opin. Chem. Biol.* 2 (1988) 743-750.
- <sup>21</sup> A. Zanghellini, L. Jiang, A. M. Wollacott, G. Cheng, J. Meiler, E. A. Althoff, D. Röthlisberger, D. Baker. *Protein Science*, 15 (2006) 2785-2794.
- <sup>22</sup> J. Meiler, D. Baker. *Proteins: Structure, Function and Bioinformatics*, 65 (2006) 538-548.
- <sup>23</sup> C. E. Tinberg, S. D. Khare, J. Dou, L. Doyle, J. W. Nelson, A. Schena, W. Jankowski, C. G. Kalodimos, K. Johnsson, B. L. Stoddard, D. Baker. *Nature* 501 (2013) 212-216.
- <sup>24</sup> A. Leaver-Fay et al. in *Methods in Enzymology*, volume 487, pp. 545-574. Elsevier 2011.
- <sup>25</sup> F. Richter, A. Leaver-Fay, S. D. Khare, S. Bjelic, D. Baker. *PLoS ONE*, 6 (2011) e19230.
- <sup>26</sup> D. J. Mandell, T. Kortemme. *Curr. Opin. Biotech.* 20 (2009) 420-428.
- <sup>27</sup> P-S. Huang, Y-E A. Ban, F. Richter, I. Andre, R. Vernon, W. R. Schief, D. Baker. *PLoS ONE* 6 (2011) e24109.
- <sup>28</sup> P. D. Renfrew, E. J. Choi, R. Bonneau, B. Kuhlman. *PLoS ONE*, 7 (2012) e32637.

- 
- <sup>29</sup> J. H. Mills, S. D. Khare, J. M. Bolduc, F. Forouhar, V. K. Mulligan, S. Lew, J. Seetharaman, L. Tong, B. L. Stoddard, D. Baker. *J. Am. Chem. Soc.* 135 (2013) 13393-13399.
- <sup>30</sup> I. Drienovská, A. Rioz-Martínez, A. Draksharapu and G. Roelfes. *Chem. Sci.* 6 (2015) 770-776.
- <sup>31</sup> F. Khatib, S. Cooper, M. D. Tyka, K. Xu, I. Makedon, Z. Popovic, D. Baker, Foldit Players. *Proc. Natl. Acad. Sci. USA.* 108 (2011) 18949-18953.
- <sup>32</sup> H. Fazelinia, P. C. Cirino, C. D. Maranas. *Protein Science*, 18 (2008) 180-195.
- <sup>33</sup> C. Malisi, O. Kohlbacher, B. Höcker. *Proteins* 77 (2009) 74-83.
- <sup>34</sup> Y. Lei, W. Luo, Y. Zhu. *Protein Science*, 20 (2011) 1566-1575
- <sup>35</sup> X. Huang, K. Han, Y. Zhu. *Protein Science*, 22 (2013) 929-941.
- <sup>36</sup> G. R. Nosrati, K. N. Houk. *Protein Science* 21 (2012) 697-706.
- <sup>37</sup> D. Baker. *Protein Science*, 19 (2010) 1817-1819.
- <sup>38</sup> I. Samish, C. M. MacDermaid, J. M. Perez-Aguilar. *J. G. Saven, Annu. Rev. Phys. Chem.* 62 (2011) 129-149.
- <sup>39</sup> U. T. Bornscheuer, G. W. Huisman, R. J. Kazlauskas, S. Lutz, J. C. Moore, K. Robins. *Nature* 485 (2012) 185-194.
- <sup>40</sup> A. Barrozo, R. Borstnar, G. Marloie, S. C. L. Kamerlin, *Int. J. Mol. Sci.* 13 (2012) 12428-12460.
- <sup>41</sup> H. Kries, R. Blomberg, D. Hilvert. *Curr. Opin. Chem. Biol.* 17 (2013) 221-228.
- <sup>42</sup> S. D. Khare, S. J. Fleishman. *FEBS Letters*, 587 (2013) 1147-1154.
- <sup>43</sup> H. J. Wijma, D. B. Janssen. *FEBS Journal* 280 (2013) 2948-2960.
- <sup>44</sup> J. Damborsky, J. Brezovsky. *Curr. Opin. Chem. Biol.* 19 (2014) 8-16.
- <sup>45</sup> M. P. Frushicheva, M. J. L. Mills, P. Schopf, M. K. Singh, R. B. Prasad, A. Warshel. *Curr. Opin. Chem. Biol.* 21 (2014) 56-62.
- <sup>46</sup> M.L. Casey, D.S. Kemp, K. G. Paul, D. D. Cox. *J. Org. Chem.* 38 (1973) 2294-2301.
- <sup>47</sup> D.S. Kemp, M.L. Casey. *J. Am. Chem. Soc.* 95 (1973) 6670-6680.
- <sup>48</sup> D.S. Kemp. *Nature.* 373 (1995) 196-197.
- <sup>49</sup> S.N. Thorn, R.G. Daniels, M.T.M. Auditor, D. Hilvert. *Nature* 373 (1995) 228-230.
- <sup>50</sup> A. Genre-Grandpierre, C. Tellier, M-J. Loirat, D. Blanchard, D. R. W. Hodgson, F. Hollfelder, A. J. Kirby. *Bioorg. Med. Chem. Lett.* 7 (1997) 2497-2502.
- <sup>51</sup> R. Müller, E. W. Debler, M. Steinmann, F. P. Seebeck, I. A. Wilson, D. Hilvert. *J. Am. Chem. Soc.* 129 (2007) 460-461.
- <sup>52</sup> L.C. James, D. S. Tawfik. *Protein Sci.* 10 (2001) 2600-2607.
- <sup>53</sup> E. W. Debler, S. Ito, F.P. Seebeck, A. Heine, D. Hilvert, I.A. Wilson. *Proc. Nat. Acad. Sci. USA.* 102 (2005) 4984-4989.
- <sup>54</sup> F.P. Seebeck, D. Hilvert. *J. Am. Chem. Soc.* 127 (2005) 1307-1312.
- <sup>55</sup> E. W. Debler, R. Müller, D. Hilvert, I.A. Wilson. *J. Biol. Chem.* 283 (2008) 16554-16560.
- <sup>56</sup> E. W. Debler, R. Müller, D. Hilvert, I.A. Wilson. *Proc. Nat. Acad. Sci. USA.* 106 (2009) 18539-18544.
- <sup>57</sup> A.J. Kirby. *Acc. Chem. Res.* 30 (1997) 290-296.
- <sup>58</sup> J. Na, K.N. Houk, D. Hilvert. *J. Am. Chem. Soc.* 118 (1996) 6462-6471.
- <sup>59</sup> K. Kikuchi, S.N. Thorn, D. Hilvert. *J. Am. Chem. Soc.* 118 (1996) 8184-8185.
- <sup>60</sup> F. Hollfelder, A.J. Kirby, D.S. Tawfik. *Nature* 383 (1996) 60-63.
- <sup>61</sup> F. Hollfelder, A.J. Kirby, D.S. Tawfik, K. Kikuchi, D. Hilvert. *J. Am. Chem. Soc.* 122 (2000) 1022-1029.
- <sup>62</sup> Y. Hu, K.N. Houk, K. Kikuchi, K. Hotta, D. Hilvert. *J. Am. Chem. Soc.* 126 (2004) 8197-8205.

- 
- <sup>63</sup> F. Hollfelder, A.J. Kirby, D.S. Tawfik. *J. Am. Chem. Soc.* 119 (1997) 9578-9579.
- <sup>64</sup> I. V. Korendovych, D.W. Kulp, Y. Wu, H. Cheng, H. Roder, W.F. DeGrado. *Proc. Nat. Acad. Sci. USA.* 108 (2011) 6823-6827.
- <sup>65</sup> O.V. Moroz, Y.S. Moroz, Y. Wu, A.B. Olsen, H. Cheng, K.L. Mack, J. M. McLaughlin, E. A. Raymond, K. Zhezherya, H. Roder, I.V. Korendovych. *Angew. Chem. Int. Ed.* 52 (2013) 6246-6249.
- <sup>66</sup> M. Merski, B. K. Shoichet. *Proc. Nat. Acad. Sci. USA.* 109 (2012) 16179-16183.
- <sup>67</sup> V. Nanda. *Nat. Chem. Biol.* 4 (2008) 273-275.
- <sup>68</sup> O. Khersonsky, D. Röthlisberger, O. Dym, S. Albeck, C.J. Jackson, D. Baker, D.S. Tawfik. *J. Mol. Biol.* 396 (2010) 1025-1042.
- <sup>69</sup> O. Khersonsky, D. Röthlisberger, A.M. Wollacott, P. Murphy, O. Dym, S. Albeck, G. Kiss, K.N. Houk, D. Baker, D.S. Tawfik. *J. Mol. Biol.* 407 (2011) 391-412.
- <sup>70</sup> O. Khersonsky, G. Kiss, D. Röthlisberger, O. Dym, S. Albeck, K.N. Houk, D. Baker, D.S. Tawfik. *Proc. Nat. Acad. Sci. USA.* 109 (2012) 10358-10363.
- <sup>71</sup> A.N. Alexandrova, D. Röthlisberger, D. Baker, W.L. Jorgensen. *J. Am. Chem. Soc.* 130 (2008) 15907-15915.
- <sup>72</sup> M. Sparta, A. N. Alexandrova. *Mol. Simul.* 37 (2011) 557-571
- <sup>73</sup> A.N. Alexandrova, W.L. Jorgensen. *J. Phys. Chem. B.* 113 (2009) 497-504.
- <sup>74</sup> O. Acevedo. *J. Phys. Chem. B.* 113 (2009) 15372-15381.
- <sup>75</sup> G. Kiss, D. Röthlisberger, D. Baker, K.N. Houk. *Prot. Sci.* 19 (2010) 1760-1773.
- <sup>76</sup> M.P. Frushicheva, J. Cao, Z.T. Chu, A. Warshel. *Proc. Nat. Acad. Sci. USA.* 107 (2010) 16869-16874.
- <sup>77</sup> M.P. Frushicheva, J. Cao, A. Warshel. *Biochemistry* 50 (2011) 3849-3858.
- <sup>78</sup> A. Labas, E. Szabo, L. Mones, M. Fuxreiter. *Biochim. Biophys. Acta* 1834 (2013) 908-917.
- <sup>79</sup> H.K. Privett, G. Kiss, T.M. Lee, R. Blomberg, R.A. Chica, L.M. Thomas, D. Hilvert, K.N. Houk, S.L. Mayo. *Proc. Nat. Acad. Sci. USA.* 109 (2012) 3790-3795.
- <sup>80</sup> R. Blomberg, H. Kries, D.M. Pinkas, P.R.E. Mittl, M.G. Grütter, H.K. Privett, S.L. Mayo, D. Hilvert. *Nature* 503 (2013) 418-421.
- <sup>81</sup> M. Höhne, U.T. Bornscheuer. *Angew. Chem. Int. Ed.* 53 (2014) 1200-1202.
- <sup>82</sup> K. Swiderek, I. Tuñón, V. Moliner, J. Bertran. 2014. *ACS Catalysis*. In Press.
- <sup>83</sup> S. Laschat. *Angew. Chem. Int. Ed.* 35 (1996) 289-291.
- <sup>84</sup> W.L. Kelly. *Org. Biomol. Chem.* 6 (2008) 4483-4493
- <sup>85</sup> Hak Joong Kim, Mark W. Ruzsyczky, Sei-hyun Choi, Yung-nan Liu, Hung-wen Liu. *Nature* 473 (2011) 109-112.
- <sup>86</sup> M.I. Page, W.P. Jencks. *Proc. Natl. Acad. Sci. USA.* 68 (1971) 1678-1683.
- <sup>87</sup> B. Golinelli-Pimpaneau. *Curr. Op. Struct. Biol.* 10 (2000) 697-708.
- <sup>88</sup> D. Hilvert, K.W. Hill, K.D. Nared, M.T.M. Auditor. *J. Am. Chem. Soc.* 111 (1989) 9261-9262.
- <sup>89</sup> A.C. Braisted, P.G. Schultz. *J. Am. Chem. Soc.* 112 (1990) 7430-7431.
- <sup>90</sup> F.E. Romesberg, P.G. Schultz. *Bioorg. Med. Chem. Lett.* 9 (1999) 1741-1744.
- <sup>91</sup> J. Xu, Q. Deng, J. Chen, K.N. Houk, J. Bartek, D. Hilvert, I.A. Wilson. *Science* 286 (1999) 2345-2348.
- <sup>92</sup> F.E. Romesberg, B. Spiller, P.G. Schultz, R.C. Stevens. *Science* 279 (1998) 1929-1933.
- <sup>93</sup> S. Marti, J. Andres, V. Moliner, E. Silla, I. Tuñón, J. Bertran. *Chem. Eur. J.* 14 (2008) 596-602.
- <sup>94</sup> V.E. Gouverneur, K.N. Houk, B. Pascual-Teresa, B. Beno, K.D. Janda, R.A. Lerner. *Science* 262 (1993) 204-208.

- 
- <sup>95</sup> A. Heine, E. A. Stura, J. T. Yli-Kauhaluoma, C. Gao, Q. Deng, B.R. Beno, K.N. Houk, K.D. Janda. *I.A. Wilson. Science* 279 (1998) 1934-1940.
- <sup>96</sup> C.E. Cannizzaro, J.A. Ashley, K.D. Janda, K.N. Houk. *J. Am. Chem. Soc.* 125 (2003) 2489-2506.
- <sup>97</sup> S. Cooper, F. Khatib, A. Treuille, J. Barbero, J. Lee, M. Beenen, A. Leaver-Fay, D. Baker, Z. Popovic, F. Players. *Nature Letters*, 466, (2010) 756-760.
- <sup>98</sup> C. B. Eiben, J.B. Siegel, J. B. Bale, S. Cooper, F. Khatib, B. W. Shen, F. Players, B. L. Stoddard, Z. Popovic, D. Baker. *Nature Biotechnology* 30 (2012) 190-192.
- <sup>99</sup> N. Preiswerk, T. Beck, J. D. Schulz, P. Milovnik, C. Mayer, J. B. Siegel, D. Baker, D. Hilvert. *Proc. Natl. Acad. Sci. USA* 111 (2014) 8013-8018.
- <sup>100</sup> M. Linder, A.J. Johansson, T.S.G. Olsson, J. Liebeschuetz, T. Brinck. *J. Chem. Inf. Model.* 51 (2011) 1906-1917.
- <sup>101</sup> M. Linder, A.J. Johansson, T.S.G. Olsson, J. Liebeschuetz, T. Brinck. *J. Comput. Aided Mol. Des.* 26 (2012) 1079-1095.
- <sup>102</sup> M. Linder, A.J. Johansson, B. Manta, P. Olsson, T. Brinck. *Chem. Comm.* 48 (2012) 5665-5667.
- <sup>103</sup> F.C. Lightstone, T.C. Bruice. *J. Am. Chem. Soc.* 118 (1996) 2595-2605.
- <sup>104</sup> A.J.T. Smith, R. Müller, M.D. Toscano, P. Kast, H.W. Hellings, D. Hilvert, K.N. Houk. *J. Am. Chem. Soc.* 130 (2008) 15361-15373.
- <sup>105</sup> J. Wagner, R.A. Lerner, C.F. Barbas III. *Science* 270 (1995) 1797-1800.
- <sup>106</sup> T. Hoffmann, G. Zhong, B. List, D. Shabat, J. Anderson, S. Gramatikova, R.A. Lerner, C.F. Barbas III. *J. Am. Chem. Soc.* 120 (1998) 2768-2779.
- <sup>107</sup> B. List, C.F. Barbas III, R.A. Lerner. *Proc. Natl. Acad. Sci. USA.* 95 (1998) 15351-15355.
- <sup>108</sup> G. Zhong, R.A. Lerner, C.F. Barbas III. *Angew. Chem. Int. Ed.* 38 (1999) 3738-3741.
- <sup>109</sup> C.F. Barbas III, A. Heine, G. Zhong, T. Hoffmann, S. Gramatikova, R. Björnstedt, B. List, J. Anderson, E. A. Stura, I.A. Wilson, R.A. Lerner. *Science* 278 (1997) 2085-2092.
- <sup>110</sup> X. Zhu, F. Tanaka, R.A. Lerner, C.F. Barbas III, I.A. Wilson. *J. Am. Chem. Soc.* 131 (2009) 18206-18207.
- <sup>111</sup> M. Arnó, L.R. Domingo. *Org. Biomol. Chem.* 1 (2003) 637-643
- <sup>112</sup> J.K. Lassila, D. Baker, D. Herschlag. *Proc. Natl. Acad. Sci. USA* 107 (2010) 4937-4942.
- <sup>113</sup> L. Wang, E.A. Althoff, J. Bolduc, L. Jiang, J. Moody, J.K. Lassila, L. Giger, D. Hilvert, B. Stoddard, D. Baker. *J. Mol. Biol.* 415 (2012) 615-625.
- <sup>114</sup> E.A. Althoff, L. Wang, L.K. Jiang, L. Giger, J.K. Lassila, Z. Wang, M. Smith, S. Hari, P. Kast, D. Herschlag, D. Hilvert, D. Baker. *Prot. Sci.* 21 (2012) 717-726.
- <sup>115</sup> L. Giger, S. Caner, R. Obexer, P. Kast, P. Kast, D. Baker, N. Ban, D. Hilvert. *Nat. Chem. Biol.* 9 (2013) 494-498.
- <sup>116</sup> R. Obexer, S. Studer, L. Giger, D. M. Pinkas, M. G. Grütter, D. Baker, D. Hilvert. *ChemCatChem.* 6 (2014) 1043-1050.
- <sup>117</sup> J. Steinreiber, T.R. Ward. *Coord. Chem. Rev.* 252 (2008) 751-766.
- <sup>118</sup> Y. Lu, N. Yeung, N. Sieracki, N. M. Marshall. *Nature* 460 (2009) 855-862.
- <sup>119</sup> T. Heinisch, T. R. Ward. *Curr. Op. Chem. Biol.* 14 (2010) 184-199.
- <sup>120</sup> F. Rosati, G. Roelfes. *ChemCatChem.* 2 (2010) 916-927.
- <sup>121</sup> J. G. Saven. *Curr. Op. Chem. Biol.* 15 (2011) 452-457.
- <sup>122</sup> M.R. Ringenberg, T. R. Ward. *Chem. Comm.* 47 (2011) 8470-8476.
- <sup>123</sup> I.D. Petrik, J. Liu, Y. Lu. *Curr. Op. Chem. Biol.* 19 (2014) 67-75.

- 
- <sup>124</sup> M. Dürrenberger, T. R. Ward. *Curr. Op. Chem. Biol.* 19 (2014) 99-106.
- <sup>125</sup> J. Bos, G. Roelfes. *Curr. Op. Chem. Biol.* 19 (2014) 135-143.
- <sup>126</sup> F. Yu, V.M. Cangelosi, M. L. Zastrow, M. Tegoni, J. S. Plegaria, A. G. Tebo, C. S. Mocny, L. Ruckthong, H. Qayyum, V. L. Pecoraro. *Chem. Rev.* 114 (2014) 3495-3578.
- <sup>127</sup> P.M. Murphy, J.M. Bolduc, J.L. Gallaher, B. L. Stoddard, D. Baker. *Proc. Nat. Acad. Sci. USA.* 106 (2009) 9215-9220.
- <sup>128</sup> S. Chakraborty, J. Y. Kravitz, P. W. Thulstrup, L. Hemmingsen, W.F. DeGrado, V.L. Pecoraro. *Angew. Chem. Int. Ed.* 50 (2011) 2049-2053.
- <sup>129</sup> S.C. Patel, L.H. Bradley, S.P. Jinadasa, M.H. Hecht. *Prot. Sci.* 18 (2009) 1388-1400
- <sup>130</sup> E. Farinas, L. Regan. *Prot. Sci.* 7 (1998) 1939-1946.
- <sup>131</sup> J. Podtetenieff, A. Taglieber, E. Bill, E. J. Reijerse, M.T. Reetz. *Angew. Chem. Int. Ed.* 49 (2010) 5151-5155
- <sup>132</sup> M.T. Reetz. *Chem. Rec.* 12 (2012) 391-406
- <sup>133</sup> P. J. Deuss, G. Popa, A. M. Z. Slawin, W. Laan, P.C. J. Kamer Chem Cat Chem. 5 (2013) 1184-1191.
- <sup>134</sup> J. Bos, F. Fusetti, A.J.M. Driessen, G. Roelfes. *Angew. Chem. Int. Ed.* 51 (2012) 7472-7475.
- <sup>135</sup> G. Roelfes, A. J. Boersma, B. L. Feringa. *Chem. Comm.* (2006) 635-637.
- <sup>136</sup> A. J. Boersma, B. L. Feringa, G. Roelfes. *Org. Lett.* 9 (2007) 3647-3650.
- <sup>137</sup> F. Rosati, A. J. Boersma, J.E.Klijjn, A. Meetsma, B. L. Feringa, G. Roelfes. *Chem. Eur. J.* 15 (2009) 9596-9605.
- <sup>138</sup> Crystal E. Valdez, Quentin A. Smith, Michael R. Nechay, and Anastassia N. Alexandrova. *Acc. Chem. Res.* 2014, 47, 3110–3117.
- <sup>139</sup> Crystal E. Valdez and Anastassia N. Alexandrova. *J. Phys. Chem. B.* 2012, 116, 10649–10656.
- <sup>140</sup> S. Leitgeb, B. Nidetzky. *ChemBioChem.* 11 (2010) 502-505.
- <sup>141</sup> S.D. Khare, Y. Kipnis, P Jr Greisen, R. Takeuchi, Y. Ashani, M. Goldsmith, Y. Song, J. L. Gallaher, I. Silman, H. Leader, J. L. Sussman, B. L. Stoddard, D. S Tawfik, D. Baker. *Nat. Chem. Biol.* 8 (2012) 294-300.
- <sup>142</sup> M.K. Singh, Z.T. Chu, A. Warshel. *J. Phys. Chem. B* 118 (2014) 12146–12152.
- <sup>143</sup> M.L. Zastrow, A.F.A. Peacock, J. A. Stuckey, V.L. Pecoraro. *Nat. Chem.* 4 (2012) 118-123.
- <sup>144</sup> M.L. Zastrow, V.L. Pecoraro. *J. Am. Chem. Soc.* 135 (2013) 5895-5903.
- <sup>145</sup> M.L. Zastrow, V.L. Pecoraro. *Biochemistry* 53 (2013) 957-978.
- <sup>146</sup> B.S. Der, M. Machius, M.J. Miley, J.L. Mills, T. Szyperski, B. Kuhlman. *J. Am. Chem. Soc.* 134 (2012) 375-385.
- <sup>147</sup> B.S. Der, D.R. Edwards, B. Kuhlman. *Biochemistry* 51 (2012) 3933-3940.
- <sup>148</sup> N. Yeung, Y-W. Lin, Y-G. Gao, X. Zhao, B. S. Russell, L. Lei, K. D. Miner, H. Robinson, Y. Lu. *Nature* 462 (2009) 1079-1082.
- <sup>149</sup> T. Hino, Y. Matsumoto, S. Nagano, H. Sugimoto, Y. Fukumori, T. Murata, S. Iwata, Y. Shiro. *Science* 330 (2010) 1666-1670.
- <sup>150</sup> Y-W. Lin, N. Yeung, Y-G. Gao, K.D. Miner, S. Tian, H. Robinson, Y. Lu. *Proc. Nat. Acad. Sci. USA.* 107 (2010) 8581-8586.
- <sup>151</sup> Y-W. Lin, N. Yeung, Y-G. Gao, K.D. Miner, L. Lei, H. Robinson, Y. Lu. *J. Am. Chem. Soc.* 132 (2010) 9970-9972.
- <sup>152</sup> Y-W. Lin, C-M. Nie, L-F. Liao. *J. Mol. Model.* 18 (2012) 4409-4415.
- <sup>153</sup> K. D. Miner, A. Mukherjee, Y-G. Gao, E.L. Null, I.D. Petrik, X. Zhao, N. Yeung, H. Robinson, Y. Yu. *Angew. Chem. Int. Ed.* 51 (2012) 5589-5592.

- 
- <sup>154</sup> X. Liu, Y. Yu, C. Hu, W. Zhang, Y. Lu, J. Wang. *Angew. Chem. Int. Ed.* 51 (2012) 4312-4316.
- <sup>155</sup> M.E. Wilson, G.M. Whitesides, *J. Am. Chem. Soc.* 100 (1978) 306-307.
- <sup>156</sup> T.R. Ward. *Acc. Chem. Res.* 44 (2011) 47-57.
- <sup>157</sup> A. Pordea, M. Creus, J. Panek, C. Duboc, D. Mathis, M. Novic, T. R. Ward. *J. Am. Chem. Soc.* 130 (2008) 8085-8088.
- <sup>158</sup> A. Pordea, D. Mathis, T.R. Ward. *J. Organometallic Chem.* 694 (2009) 930-936.
- <sup>159</sup> J.J. Panek, T.R. Ward, A. Jezierska-Mazzarello, M. Novic. *J. Comput. Aided Mol. Des.* 24 (2010) 719-732.
- <sup>160</sup> T. K. Hyster, L. Knörr, T. R. Ward, T. Rovis. *Science* 338 (2012) 500-503.
- <sup>161</sup> J. M. Zimbron, T. Heinisch, M. Schmid, D. Hamels, E.S. Nogueira, T. Schirmer, T. R. Ward. *J. Am. Chem. Soc.* 135 (2013) 5384-5388.
- <sup>162</sup> V. Muñoz Robles, P. Vidossich, A. Lledós, T.R. Ward, J-D. Maréchal. *ACS Catal.* 4 (2014) 833-842.