Int. J. Mol. Sci. **2012**, *13*, 12428-12460; doi:10.3390/ijms131012428

OPEN ACCESS

International Journal of
Molecular Sciences
ISSN 1422-0067
www.mdpi.com/journal/ijms

Review

Computational Protein Engineering: Bridging the Gap between Rational Design and Laboratory Evolution

Alexandre Barrozo [1], Rok Borstnar [1,2], Gaël Marloie [1] and Shina Caroline Lynn Kamerlin [1,*]

[1] Department of Cell and Molecular Biology, Uppsala Biomedical Center (BMC),
Uppsala University, Box 596, S-751 24 Uppsala, Sweden;
E-Mails: alexandre.barrozo@icm.uu.se (A.B.); rok.borstnar@ki.si (R.B.);
gael.marloie@icm.uu.se (G.M.)

[2] Laboratory for Biocomputing and Bioinformatics, National Institute of Chemistry, Hajdrihova 19,
SI-1000 Ljubljana, Slovenia

* Author to whom correspondence should be addressed; E-Mail: kamerlin@icm.uu.se;
Tel.: +46-18-471-4423; Fax: +46-18-530-396.

Received: 20 August 2012; in revised form: 16 September 2012 / Accepted: 17 September 2012 / Published: 28 September 2012

Abstract: Enzymes are tremendously proficient catalysts, which can be used as extracellular catalysts for a whole host of processes, from chemical synthesis to the generation of novel biofuels. For them to be more amenable to the needs of biotechnology, however, it is often necessary to be able to manipulate their physico-chemical properties in an efficient and streamlined manner, and, ideally, to be able to train them to catalyze completely new reactions. Recent years have seen an explosion of interest in different approaches to achieve this, both in the laboratory, and *in silico*. There remains, however, a gap between current approaches to computational enzyme design, which have primarily focused on the early stages of the design process, and laboratory evolution, which is an extremely powerful tool for enzyme redesign, but will always be limited by the vastness of sequence space combined with the low frequency for desirable mutations. This review discusses different approaches towards computational enzyme design and demonstrates how combining newly developed screening approaches that can rapidly predict potential mutation “hotspots” with approaches that can quantitatively and reliably dissect the catalytic step can bridge the gap that currently exists between computational enzyme design and laboratory evolution studies.

Keywords: *de novo* enzyme design; enzyme redesign; protein engineering; directed evolution; computational enzymology

---

1. Introduction

Since Friedrich Wöhler first synthesized urea in 1828 [1], organic chemists have worked on synthesizing a breathtaking range of organic compounds that play a role in all aspects of modern life, from food preservation to dyes, pharmaceuticals and fuels. Due to their ubiquity, a great variety of artificial catalysts have been created to accelerate organic synthesis, from compounds using simple acid/base catalysis to organometallic catalysts using transition metals such as palladium [2], platinum [3], iridium [4], rhodium [5] or gold [6] as their metal centers. Particularly, the use of transition metals has permitted us to increase our ability to form carbon-carbon bonds [7–9], a process which forms the basic building block of almost all synthetic organic chemistry. However, such progress has not been without its challenges. That is, in addition to cost concerns, many industrial processes harnessing such catalysts occur under environmentally unfriendly conditions that go against the important current emphasis on "green chemistry", which encourages sustainable development [10].

In principle, enzymes should provide the perfect solution to these problems. As one example, the half-life for the hydrolysis of an amide bond is in the range of hundreds of years (the half-life for the hydrolysis of a glycine-glycine dipeptide in neutral solution at 25 °C is 350 years [11]). If the same dipeptide is transferred to 1 M HCl or NaOH solutions (at the same temperature), the half-life for the reaction drops to approximately 150 and 2 days, respectively [12]. While this is of course impressive, under enzymatic catalysis, the reaction rate is increased by a factor of 10^{12} [11,12] (and there are many examples of enzymes facilitating even more extreme rate enhancements, see e.g., [12,13]). Such proficiencies have never been matched in a man-made catalyst (although changing the reaction medium from water to alcohol has been demonstrated to produce enzyme-like rate enhancements in the case of dinuclear zinc catalysts of phosphate diester hydrolysis [14]). In addition to their extreme proficiencies, they are also biodegradable and reusable catalysts [15], making them ideal green reagents. Also, while most enzymes work within limited and tightly controlled temperature ranges, there do exist enzymes that can act anywhere within a temperature range from 0 to 100 °C (for the two extremes observed in extremophile bacteria) [16]. On top of this, enzymes have chiral active sites, making them able to discriminate between different stereoisomers and regioisomers, with at times quite high efficiency, making them ideal catalysts for enantio- and regioselective chiral chemistry, in order to generate isomerically pure pharmaceuticals and fine chemicals [17,18].

The challenge is that the same properties that make enzymes such proficient catalysts are directly orthogonal to the needs of biotechnology [19]. That is, enzymes have had millions of years in order to evolve to become the proficient catalysts that we see today, and they need to operate under tight *in vivo* regulation. As a result of this, they are picky about what substrates they are willing to accept (which can be limited by both electrostatic, and simple steric considerations), their efficiencies vary dramatically for different substrates [20], and, since they need to be turned off as well as on, often they can be inhibited by even their own product [21]. They are also often highly resistant to environments other than the one they have evolved for. In contrast, biotechnology needs catalysts that can work under harsh conditions if necessary, tolerate changes in environment, catalyze a broad range of substrates to generate maximal amounts of product, and, ideally, can catalyze completely new reactions for which no catalyst currently exists. Therefore, in order to fulfill these requirements, it is often not sufficient to take enzymes "as is", but rather, it is necessary to be able to modify their physico-chemical and functional properties in a fast, efficient and streamlined manner. In light of this, it is perhaps unsurprising that interest has exploded in engineering enzymes for biocatalysis, and, as of 2012, that this area is one of the fastest growing of the biotechnology sector [22].

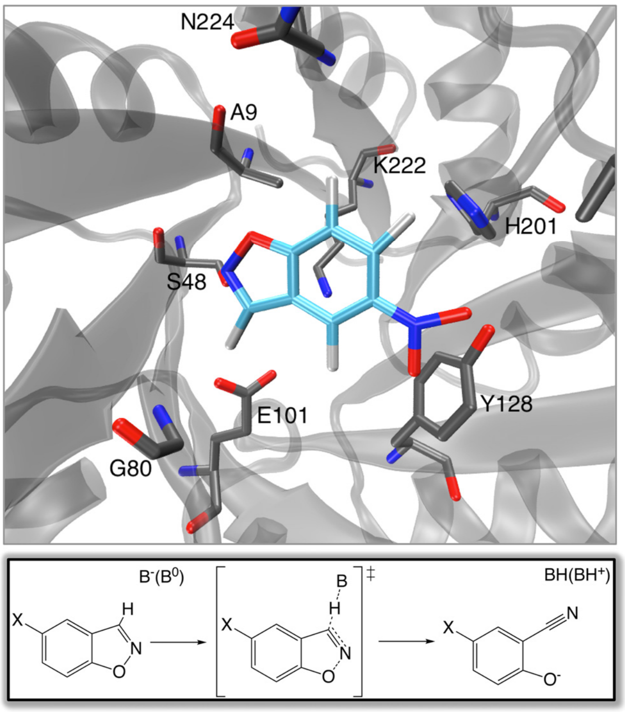
While it could be argued that "biocatalysis" dates back thousands of years to Sumerians and Egyptians using microorganisms to produce beer, the development of techniques such as error-prone PCR [23] and DNA shuffling [24] have essentially revolutionized biocatalysis [25], as they allow enzyme properties to be altered through random mutagenesis, in a "directed" fashion (directed evolution, DE), where the enzyme is iteratively refined until a property of interest (such as, for instance, a specific function or improved selectivity towards a given enantiomer) is observed. However, despite the tremendous power of this approach and the advances it has allowed in protein engineering, it is still subject to a number of limitations, the most pressing of which is the sheer vastness of the sequence space that needs sampling. That is, a simple enzyme comprised of 300 amino acids will have a sequence space of 20^{300} possible mutations (if one uses the naturally occurring building blocks alone), which is far out of the reach of even the best synthetic libraries [26,27]. This problem is further compounded by the fact that, in a random library, the frequency of a beneficial mutation occurring is only $\sim 10^{-3}$ [28,29], whereas that of a deleterious mutation is ≥ 0.33 [30]. Various approaches have been developed to deal with this issue, including approaches such as iterative saturation mutagenesis [31] and simultaneous multiple-site saturation mutagenesis [32], which focus on limiting the library sizes necessary to effectively obtain enhancements in activity, or even using ancestral libraries [26] (to name a few examples). However, clearly, as the rounds of directed evolution progress, most mutations quickly become unviable, and the library sizes that need managing become simply astronomical. In addition to this, while one of the strengths of using directed evolution is that, in principle, little-to-no knowledge of the actual catalytic mechanism being employed by the enzyme is required in order to be able to obtain substantial catalytic enhancements, this also poses a weakness, as without insight into why DE mutants are catalytic (or anticatalytic) it becomes challenging to further improve and refine the selection procedure for a given property.

As there are a large number of reviews addressing the advantages and challenges of laboratory-based evolution techniques for enzyme optimization [33,34], we will not be addressing this issue in the present work. However, computers have the potential to play a major role in guiding and directing experimental work, and increasing realization of this fact has led to an explosion in activity in this area. While computational enzyme design could still be classed as a nascent field, recent years have seen major advances in a broad spectrum of areas, from (computational) enzyme redesign, to the *de novo* design of enzymes from the basic building blocks upwards. This review will outline some recent developments in this field, introducing different "philosophies" towards computational enzyme design, and present suggestions for important future directions. Our focus will be on improving functional properties of enzymes rather than their physico-chemical properties such as thermostability and solubility, which are out of the scope of the present review. Specifically, we will illustrate the fact that, when combined with ever-increasing computational power, current computational approaches are rapidly reaching a stage where it is possible to rationally mimic the process of laboratory evolution *in silico*, bridging the gap that exists between the proficiencies of artificially designed catalysts and the outcome of laboratory evolution studies.

2. "*De Novo*" Enzyme Design

One of the "Holy Grails" of biotechnology would be the ability to rapidly design effective biological catalysts for new chemistry completely from scratch on demand. This goal was one step closer to realization in 2008, with the computational design of eight enzymes capable of performing base-catalyzed benzisoxazole ring opening (Kemp elimination) with rate accelerations of $\sim 10^5$, and multiple (≥ 7) turnovers [35], although this is similar to the rate in solution when the cost of bringing the reacting fragments into the "reacting cage" is taken into account (see [36,37]). Catalysis of this reaction, which does not occur in nature, is well characterized, and has been performed by systems ranging from catalytic antibodies [38,39], to "off-the-shelf" proteins such as bovine (and other) serum albumens [40,41], to small synthetic enzyme mimics made of organic molecules ("synzymes") [42]. The successful design of *de novo* enzymatic catalysts for this reaction was achieved using a multi-level approach [35]. A first port of call was the design of an idealized active site for benzisoxazole ring opening, which incorporated the necessary catalytic machinery for general base catalysis, as well as a number of other residues necessary for efficient chemistry [43] (see Figure 1). This was achieved by constructing an appropriate "theozyme", which optimizes relevant functional groups necessary for efficient chemistry around a proposed transition structure based on quantum mechanical calculations [44]. The next step was to match this theozyme to a suitable scaffold, which was done using the Rosetta Match hashing algorithm [45]. A search through a scaffold set containing >100,000 members, which included multiple protein folds such as β-propellers, jelly rolls, Rossman folds and lipocalins ultimately favored a TIM barrel (which is also a widespread fold in naturally occurring enzymes). Subsequent experimental characterization of this design demonstrated an r.m.s.d. deviation of <1 Å (full backbone plus active site sidechains) between the designed and crystallographic structures, with rate accelerations (k_{cat}/k_{uncat}) up to 10^5 as outlined above, and k_{cat}/K_M values up to 163 M^{-1} s^{-1}, making it impressive that it was possible to design an effective catalyst using this approach. Nevertheless, the resulting catalytic efficiencies still range far behind that of catalytic antibodies or serum albumens [41], although still presenting progress on elegant earlier work that used computational search algorithms to identify catalytically active enzyme-like active sites in protein scaffolds [46]. Initial *in vitro* evolution was able to modestly increase the catalytic activity of the designed enzymes to provide rate accelerations in the region of factors of 10^4 to 10^6 (*i.e.*, a k_{cat} value of 1.37 s^{-1}). This is a 75-fold increase over the initial designed variant used as a starting point in this work, which had a k_{cat} of 0.018 s^{-1} [35,47]). In addition, the k_{cat}/K_m value was improved to 2590 M^{-1} s^{-1} (compared to 12.2 M^{-1} s^{-1} in the initial design [35,47]), after the introduction of eight mutations over seven rounds of evolution, showing the evolvability of the designed systems. Specifically, these initial mutations were determined to act by tuning the electrostatic environment of the active site, as well as correcting a problem in the initial design, which introduced a lysine as a general acid for leaving group protonation, but placed it such that it's pK_a would be detrimentally shifted by "quenching" from the catalytic base. Subsequent combined experimental and theoretical work [48] on a different (and more active) designed enzyme, KE70 [35] (k_{cat} of 0.14 s^{-1} and k_{cat}/K_M of 126 M^{-1} s^{-1}), managed to further fine-tune the active site electrostatics, redesign the active site in order to improve substrate binding, as well as stabilize a catalytic His-Asp dyad in an optimal catalytic conformation. This, in combination, resulted in a >400-fold improvement in catalytic proficiency (for practicality, we distinguish here between catalytic enhancement (or efficiency) and proficiency, where by catalytic enhancement (or efficiency) we refer to the rate of the catalyzed *versus* the uncatalyzed reactions (*i.e.*, k_{cat}/k_{uncat}), whereas by proficiency we refer to the overall k_{cat}/K_M value.), with k_{cat}/K_M values of as high as 5×10^4 M^{-1} s^{-1} in the best variants [48], and a ~75-fold increase in k_{cat} compared to the original design. A more recent work [49] has also introduced consensus mutations into one of the best of the original designs [35] ($k_{cat}/K_M \sim 160$, k_{cat} was not measurable due to instability of the design), and screened for enzymatic activity with a number of substrates (5,7-dichloro, 6-chloro and 6-fluoro benzisoxazole), none of which showed activity above the detection limit. In all cases, however, after 16 rounds of mutation, not only were activities observed, but also, in the case of 5,7-dichloro substituted benzisoxazole, a k_{cat} of 21.2 s^{-1} and a k_{cat}/K_M of 573,090 M^{-1} s^{-1} was observed. Now while Kemp elimination presents something of a "test system" for artificial enzyme design, nevertheless, it provides a good proof-of-concept of the fact that it is possible to start from a minimal, idealized active site, and match this to a pre-existing scaffold to end up with an active (albeit inefficient) catalyst, that can be further improved by subsequent laboratory evolution.

Figure 1. Model of the active site of a representative *de novo* catalyst of the Kemp elimination reaction [43] (shown in its generalized form in the inset below). Note the position of E101, which was introduced to act as a general base, relative to that of the substrate. However, it is important to point out that the design was executed relative to the transition state and not the native substrate, and, therefore, the contacts in this figure are not truly idealized.



A number of other examples of *de novo* enzyme design have been observed in the literature recently. For instance, the approach used in [35] was expanded in order to design an active site that can accommodate multiple TS and reaction intermediates during the course of multistep reaction pathways. Specifically, a catalyst was designed *de novo* that was capable of facilitating the cleavage of carbon-carbon bonds [50] (a reaction which is fundamental to organic chemistry) with multiple turnovers and modest rate accelerations of up to four orders of magnitude (modest again in comparison to accelerations routinely obtained by natural systems) in the designed variants. An overview of the retro-aldol reaction catalyzed by this enzyme, as well as a schematic of a representative designed active site is shown in Figure 2. As with previous work [35], an initial theozyme was constructed, which harnessed two lysines close to each other in order to reduce the pK_a of the nucleophilic lysine, as well as an aspartate which allows for general base catalysis. This was then coupled to an exhaustive search of 181,555 distinct protein scaffolds that could accommodate this theozyme, which resulted ultimately in 72 designs that were selected for experimental characterization. Of these, 32 showed detectable activity [50] (k_{cat} values up to 9.3×10^{-3} s^{-1}, k_{cat}/K_M values up to 0.74 M^{-1} s^{-1}). Once again, this is not a highly proficient catalyst, however, as a proof-of-concept study of the possibility for designing catalysts from scratch for complex chemical reactions, this is elegant work. Subsequent analysis of the designed systems [51] suggested that the observed catalytic enhancements come from a number of sources, most importantly the effect of the enzyme environment on lowering the pK_a of the catalytic lysine, as well as hydrophobic binding interactions in the enzyme active site. It was therefore suggested that improvements in binding interactions and the placement of catalytic groups could improve design efforts, a suggestion compounded by more recent studies into the interaction of these enzymes with covalent enzyme-substrate analog complexes [52]. Once again, site directed mutagenesis and laboratory evolution could be applied to the designed retroaldolases [53], extending on the catalytic motifs used in the original work [50], allowing for increases of up to 0.71 min^{-1} for k_{cat}, and 490 M^{-1} s^{-1} for k_{cat}/K_M.

A third recent example of *de novo* enzyme design is the computational design of a biological catalyst for a stereoselective bimolecular Diels Alder reaction [54]. The overall reaction mechanism and a representation of the designed active site are shown in Figure 3. This system is significant, as there is no natural counterpart capable of catalyzing the Diels-Alder reaction. The fact that it can catalyze the formation of two carbon-carbon bonds to stereoselectively yield four new stereogenic centers is of great importance to organic chemistry, although now the catalytic proficiency is even poorer for this challenging reaction, with k_{cat} values of at most 2.13 h^{-1}, presenting a modest rate acceleration of only 12-fold compared to the uncatalyzed reaction at 298 K [54]. However, to achieve this modest acceleration, the designed variant uses a clever chemical trick. Specifically, the Diels-Alder reaction is under orbital control, with the reaction rate depending on the gap between the highest occupied molecular orbital (HOMO) of the diene, and the lowest unoccupied molecular orbital (LUMO) of the dienophile. Therefore, in the designed variant, two hydrogen bonds were introduced to increase the HOMO of the diene and decrease the LUMO of the dienophile, respectively, and these between them, if correctly positioned, were expected to provide ~4.7 kcal/mol transition state stabilization (based on quantum chemical calculations of the theozyme [54]), although, as mentioned above, the actual observed catalytic enhancement was much smaller.

Figure 2. Model of the active site of a representative *de novo* designed retroaldolase [50], with a generalized version of the aldol reaction pathway using a lysine nucleophile and acid-base catalysis shown in the inset. Note that the figure depicts the enzyme in complex with the native substrate, and *not* with the transition state, and therefore the contacts are not truly idealized.

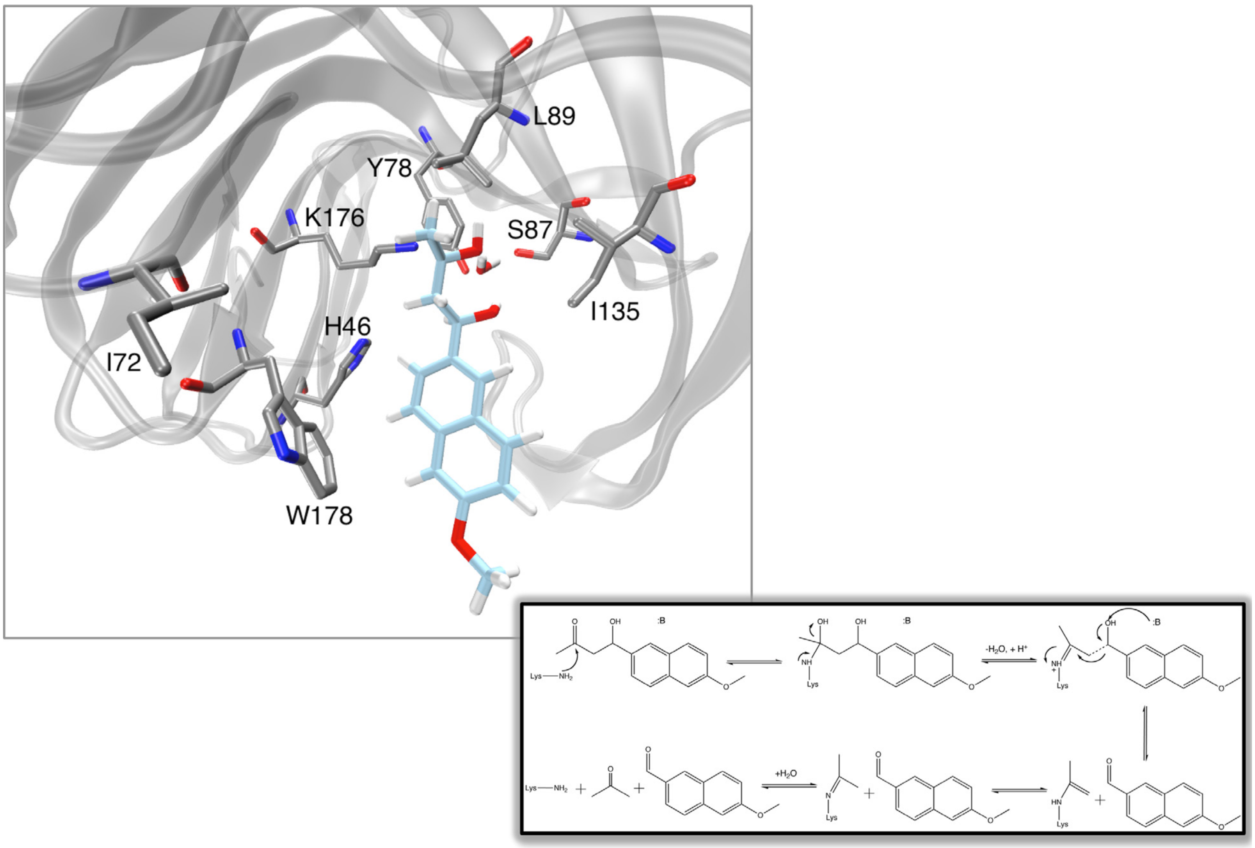
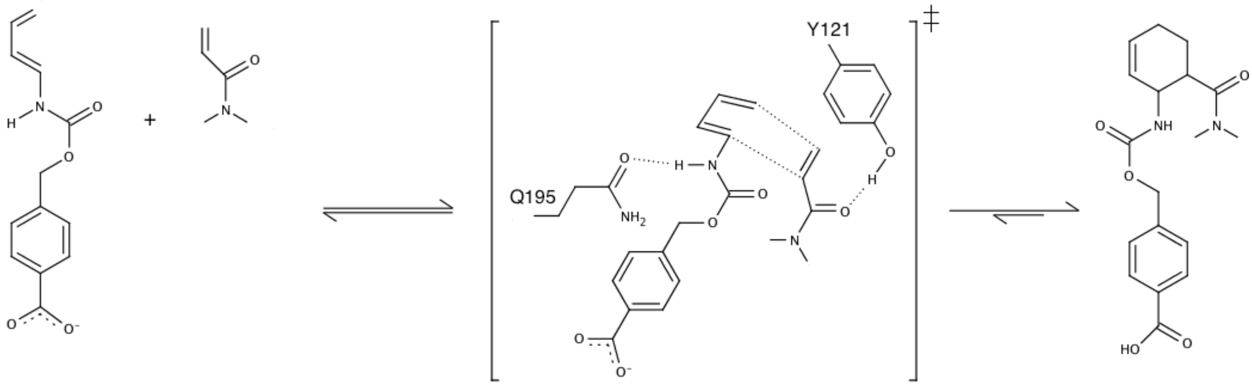


Figure 3. The Diels-alder reaction, showing a pericyclic (4 + 2) cycloaddition to form a chiral cyclohexene ring. Also illustrated here are the residues selected to act as donor and acceptor in the designed Diels alderase discussed in the main text [54].



We have presented here a few examples of successful *de novo* enzyme design, which have focused on using the generalized protocol illustrated in Figure 4. This is a field that has grown rapidly in interest, and space considerations prevent us from discussing all recent work in similar detail, and there have been other recent reviews that address this issue (e.g., [54], as one example). However, some works that we would like to mention here include a recent work by Faiella and coworkers, in which not only the active site, but also the scaffold was designed from first principles [55]. This resulted in an artificial di-iron oxo-protein, with phenol oxidase activity and, depending on substrate, k_{cat} values of up to 13.5 min^{-1}, and k_{cat}/K_M values of up to 6315 M^{-1} min^{-1} (for the best substrates). Another recent example comes in the form of a computationally designed homodimeric zinc-mediated protein interface [56], which was originally designed in order to study protein-protein interactions, but has recently been demonstrated to be a promiscuous catalyst capable of at least two functionalities [57] (Figure 5), namely the hydrolysis of *p*-nitrophenyl acetate and phosphate, with rate accelerations of 10^5 and 10^4-fold, as well as k_{cat}/K_M values of 630 and 14 M^{-1} s^{-1}, respectively. This is noteworthy in light of the fact that this system was not designed with catalytic ability in mind, but in comparison to many designed enzymes it shows surprising potential as a catalyst, opening other interesting avenues in protein engineering using simple model systems. Clearly, the works presented here demonstrate that we are moving closer to being able to design effective enzymes from scratch. However, there are still a number of significant issues. As will be discussed in the Sections 4 and 5, any rational design strategy requires in-depth knowledge of an enzyme's catalytic mechanism, making it very difficult to predict how an enzyme will behave *a priori*, as was illustrated in the example of the pK_a quenching problem in the designed Kemp eliminase (see [35,47]). Additionally, current design strategies such as those highlighted above cannot in and of themselves distinguish between active and inactive constructs, and therefore subsequent testing is left to experiment. The third issue is that, at present, the efficiencies of *de novo* designed systems are far too low to be useful in commercial settings and can only be improved to come closer to that of natural enzymes following subsequent laboratory evolution. While this has led to significant improvements in activity, this will ultimately always be limited to the same challenges all directed evolution studies face, namely an inability to explore the entire sequence space and library sizes. Therefore, for such designs to ultimately be effective, approaches will be needed to bridge the gap between *in silico* design and laboratory evolution. Therefore, Section 3 will address advances in and insights obtained from rational enzyme design, and Section 4 will address machine-learning approaches and novel screening approaches that can be used to greatly aid the design process.

Figure 4. A generalized version of a currently popular protocol for *de novo* enzyme design, adapted and abridged from [50]. Note that while molecular dynamics (MD) or quantum mechanical/molecular mechanical (QM/MM) characterization was not explicitly included in [50], it has been successfully used to aid the design process.

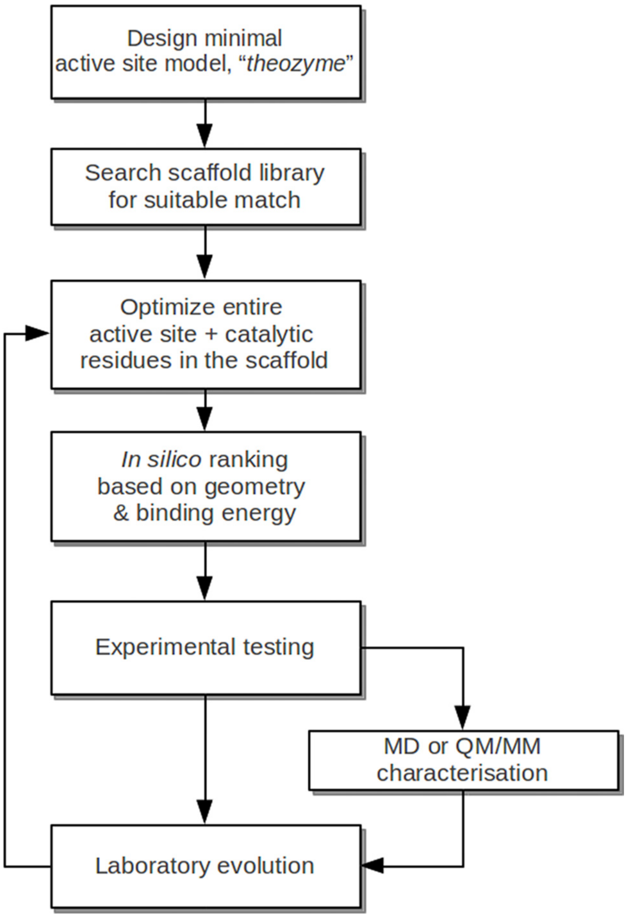
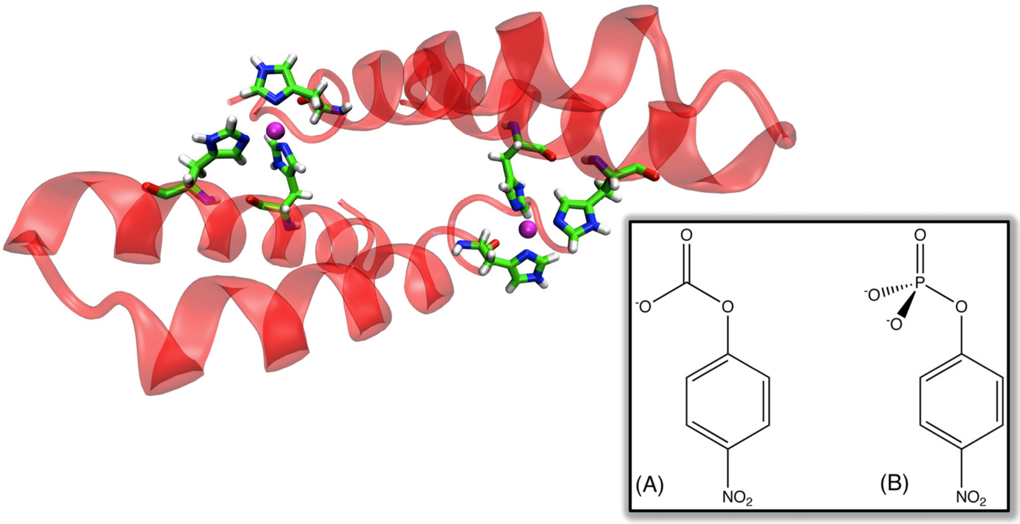

Design minimal active site model, "*theozyme*"
Search scaffold library for suitable match
Optimize entire active site + catalytic residues in the scaffold
In silico ranking based on geometry & binding energy
Experimental testing
MD or QM/MM characterisation
Laboratory evolution

Figure 5. Structural overview of a computationally designed zinc-mediated protein interface [57], that (unintentionally) is capable of promiscuously catalyzing both (**A**) *p*-nitrophenyl acetate (k_{cat}/K_M = 630 M^{-1} s^{-1}) and (**B**) *p*-nitrophenyl phosphate hydrolysis (k_{cat}/K_M = 14 M^{-1} s^{-1}).



3. Computational Enzyme Redesign

As was seen in Section 2, recent years have seen tremendous advances in our ability to design enzymatic catalysts for novel chemistry "from scratch". However, what about the enzymes that Nature has already provided? A wide variety of chemical reactions are necessary to facilitate life, and, as a result, catalysts have evolved for a range of reactions from group transfers, isomerization and hydrolysis reactions to oxidation/reduction reactions. Therefore, many templates already exist that can, theoretically, be modified for desired chemistry (as an example, compare the artificially designed retroaldolase [50] presented in Section 2 with natural aldolases such as DERA from *E-coli*, which has been evolved successfully for industrial biocatalytic applications e.g., [58]). In view of this, it is perhaps unsurprising that there has been interest in not just experimental [59,60] but also computational [61,62] protein redesign for several decades. Despite many promising studies, (rational) computational protein redesign of functional properties is not without its challenges, as it requires a reliable 3-D structure of the system of interest, as well as in-depth insight into the catalytic mechanisms, which can be changed by mutations. Additionally, in cases where high-quality crystal structures of the enzyme *do* exist, the fleeting nature of transition states means that the best one can hope for are complexes with transition state analogues [63]. While these can carry useful structural information in terms of potential contacts, they can also be deceptive in terms of the information they provide as to the actual chemistry [64,65] and need to be treated with care. As a result of this, the enhancements obtained using rational design approaches have been modest, particularly in comparison to the proficiencies of naturally occurring enzymes [66]. However, as with *de novo* enzyme design, increasing computer power has greatly expanded the range of possibilities with respect to protein redesign, both in terms of approaches to qualitatively and quantitatively probe and modify activity, as well as in terms of computational approaches to rapidly identify potential mutations that can affect enzyme function and stability. In this section, we will illustrate and discuss some noteworthy recent applications of computational enzyme design, whereas Section 3 will focus on screening approaches that can bring us closer to *in silico* directed evolution becoming a reality.

Returning to the Kemp eliminases presented in Section 2, there have been a number of recent studies that have addressed the problem of computationally predicting and improving the catalytic activity of the designed enzymes (and closely related systems). Initial mixed quantum mechanical/molecular mechanical (QM/MM) and free-energy perturbation (FEP) calculations [67] that aimed to explore the energetics and mechanism of the reaction catalyzed by a number of the designed constructs as well as the background reactivity in water predicted activity in three of the examined constructs. In the fourth case, the authors obtained significant product trapping, however, uncertainty in the protonation states of a number of active site residues made results on this system inconclusive. In all enzyme-catalyzed cases, the authors obtained concerted mechanisms proceeding through a single transition state in which the proton transfer was quite advanced in comparison to the breaking of the isoxazolyl N–O bond (for the reaction mechanism, see Figure 1). A similar mechanism was obtained when modeling the background reaction in aqueous solution using a hydroxide base. This led the authors to suggest that, since the enzymes do not alter the mechanism compared to solution, design studies should focus on further increasing the basicity of the catalytic base as well as optimizing the positioning of catalytic residues in the active site [67]. It should be noted, however, that this work provided poor quantitative accuracy, with large discrepancies between calculated and observed catalytic effects. Subsequent work has used a range of computational approaches [68], including quantum mechanical cluster models comprising a smaller subset of catalytic residues, QM/MM calculations that take into account the full system, as well as MD simulations that explore structural changes to explore the origins of activity in designed systems. As the authors acknowledge, the fact that the cluster and QM/MM models do not account for protein reorganization as well as changes in solvent accessibility along the reaction coordinate made these approaches insufficient to distinguish between active and inactive designs. However, performing molecular dynamics simulations allowed for exploration of structural features in cases where the activity or inactivity was known, leading to the identification of "design flaws" causing inactive enzymes [68].

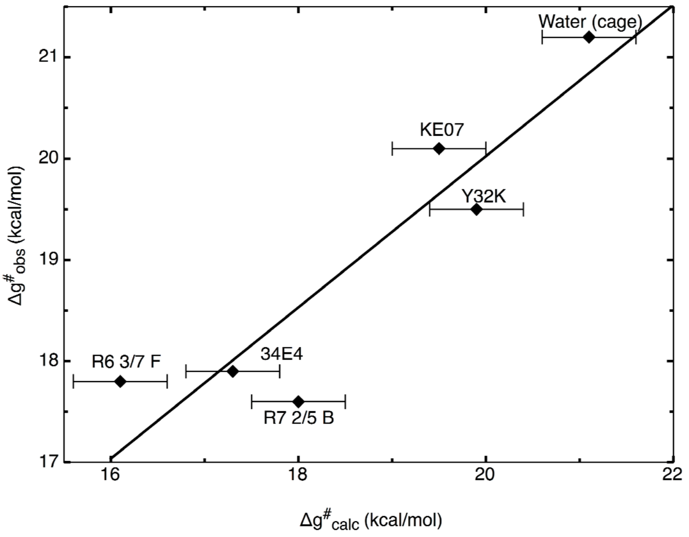
While such simulations are interesting and informative in the context of providing valuable structural insight, for effective enzyme redesign, it would be useful (and, indeed, most likely critical) to also have quantitative insight into the precise molecular basis for the observed catalytic (or anticatalytic) effect of different mutations, as well as, ideally, also, information about the quantitative contribution of different residues to activity and stability. To the best of our knowledge, most currently available quantum chemical approaches do not provide sufficient breakdowns of different contributions in order to be able to probe this issue, with the exception of a few notable examples such as the activation strain model [69], which provides a breakdown of the total energy intro contributions from electrostatic interactions, Pauli repulsion, and orbital interaction energies which could be used to dissect interactions between the reacting atoms (although even such elegant approaches become challenging in cases where one aims to explore the enzyme contributions, where using the linear response approximation with major sampling is to obtain reliable results, as demonstrated in [70]). Additionally, a second problem is computational cost: while, in principle, *ab initio* quantum mechanical approaches provide better precision, the cost involved in obtaining this precision makes it simply intractable to test numerous large systems, multiple mutations, different potential mechanisms and substrate binding modes, creating a major bottleneck in the design process.

Here, in our opinion, the empirical valence bond (EVB) approach of Warshel and coworkers [71–73] provides the perfect solution to these problems. The EVB approach is a semi-empirical QM/MM approach, based, as the name suggests, on the valence bond theory, which describes chemical reactivity by mixing resonance states corresponding to classical valence-bond structures for different possible reacting states and obtaining the free energy for moving between these using free-energy perturbation/umbrella sampling (FEP/US) [71,74]. It is important to point out that, despite the concerns that may arise from the fact that the EVB approach sacrifices precision due to the fact that it is semi-empirical, all quantum chemical approaches use approximations to varying extents, and, for instance, the inherent errors in commonly used DFT approaches can be quite significant (see discussion in e.g., [75] and references cited therein), and the reliability of the results are heavily dependent on the precise functional and basis set used. However, this becomes partially irrelevant, since the EVB accuracy is mainly in determining relative free energies (with respect to a well-defined reference state), or catalytic effects. Therefore, while it is in principle fully feasible to increase the precision of the EVB using approaches such as the paradynamics approach [76], unless one has the computational resources to perform quantum chemical calculations at a very high level of theory, any inherent error introduced through using empirical force fields is not greater than that existing in most commonly used DFT approaches (once the force field has been rigorously and carefully parameterized), particularly as an important core point of the EVB approach is careful calibration to not just *ab initio* but also experimental data available on the reference reaction in water. The fact that precision should not be a concern to the reader can be illustrated in the repeated success of the EVB approach at reproducing catalytic effects in wild-type and mutant enzymes with very high levels of quantitative accuracy, and without the need for adjustable parameters [13,36,37,77]. The strength of the EVB approach as a tool for computational enzyme redesign lies in three key features: (1) it's relative speed, in that while it is still computationally demanding to perform the extensive sampling required to obtain physically meaningful convergent free energies, it is much faster than doing the same at a higher level of theory without compromising precision; (2) the tremendous amount of chemical information it carries as well as the energy-gap reaction coordinate, which allows for a reliable description of bond-breaking and bond-making events, while the energy-gap coordinate reduces the time required to obtain convergent results compared to other currently popular approaches, as recently shown in extensive benchmarks by Fuxreiter and coworkers [78]; and (3) the fact that the EVB approach is based on calibration to the correct reference reaction in solution, such that one is directly examining catalytic effects, and that it provides a detailed breakdown of such effects, allowing one to pinpoint the precise molecular basis for the observed catalytic effect. In combination, these approaches make the EVB ideal for not just understanding enzyme catalysis, but also as a tool for rapid predictive protein redesign, due to its ability to predict mutational effects with high accuracy.

To illustrate the power of the EVB approach in protein redesign, we refer the reader to detailed EVB studies [36,37] of catalysis of the Kemp elimination shown in Figure 1. The first issue that we would like to highlight here again is the aforementioned one of precision. Namely, in these works, the authors first examined the molecular basis for catalysis of the Kemp elimination reaction in a wide range of systems, from catalytic antibodies through to serum albumens, and also computationally designed Kemp eliminases (KE) and designed variants. As can be seen from Figure 6, the authors were able to reproduce the catalytic effect of all known systems with very high accuracy (maximum difference between calculated and experimental activation barriers being within ~1 kcal/mol in approximately 20 tested systems [36,37] with standard deviations of maximum 1 kcal/mol over up to 20 trajectories).

Having established the reliability of the computational procedure, the authors then provided a number of significant insights into the reason for the observed catalytic effect of the studied mutants, as well as the challenges involved with obtaining proficient catalysis of this system. Specifically, it was observed that, in contrast to natural systems, which achieve efficient catalysis through stabilizing the transition state of the reaction, the mutant KEs are achieving enhanced catalytic activity through *destabilizing* the ground state of the reaction [36,37]. In principle one could argue that exactly how the catalysis is being achieved is not that significant as long as one can improve the efficiency of the enzyme. However, not only is this not observed in native enzymes [13], but also, the designed and evolved KEs are also not taking optimal advantage of this effect. This is important in light of suggestions that one should target the catalytic base [67], since, as discussed in [37], even if it were possible to create strong desolvation for the base in the ground state, this would result in a very large pK_a that would not help at physiological pH as the base would simply be protonated by the bulk. The second challenge that ties in with this is that the charge change between the ground state and transition state for this reaction is very small, making it extremely difficult to exploit active site polarity in order to better catalyze this reaction [36] (and, as was observed by [36,37], as a result, mutations that tend to effect activity also tend to be farther from the active site). This could also explain to some extent the current problems with trying to significantly improve this system, whether computationally or experimentally.

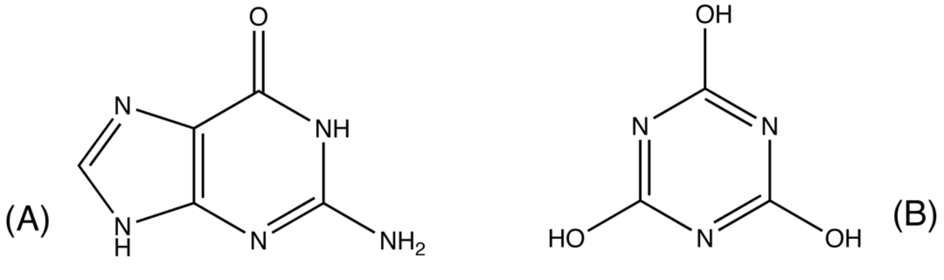
Figure 6. Correlation between calculated ($\Delta g^{\neq}_{\mathrm{calc}}$) and observed ($\Delta g^{\neq}_{\mathrm{exp}}$) activation barriers for the catalysis of the Kemp elimination reaction (illustrated in Figure 1) for a range of catalysts, including two catalytic antibodies (34E4 and Y32K), and a number of designed proteins in wild-type and mutant form (KE07, R6 3/7F and R7 2/5B). This plot is based on data presented in Table 1 of [37]. Calculations were performed using the empirical valence bond (EVB) approach, and, in all cases, both the difference between the calculated and experimentally observed values, as well as the standard deviation in the calculated values (over 20 trajectories) was <1 kcal/mol.



To address this issue, a number of screening approaches were presented [36,37] that could be used to screen for potential mutation hotspots (which could then be tested by EVB), as well as to predict stability change upon mutation, however, we will leave the discussion of these to Section 4, which deals with currently available approaches for performing *in silico* directed evolution. Here, we would just like to comment on a recent work [79] that used iterative MD and experimental analysis to explore previously inactive artificial KEs and used insights from MD and structural analysis to improve activity. We agree with the authors of this work [79] that an iterative procedure is important, as well as the fact that understanding inactive designs can provide insights that can potentially guide future design effort by avoiding repetition of "mistakes" that were made during the laboratory evolution (or initial computational design). However, this brings us back to the question at the end of Section 2: Considering the wide variety of templates Nature already provides, what is the best starting point? Here, we believe that the most effective approach for redesign, regardless of the starting point, is one which can provide quantitative as well as qualitative insight (such as that of [36,37]), which allows for the identification of factors that can be rapidly modified, and which can then be combined with an iterative approach such as that illustrated in [79] for rapid redesign.

Having presented current computational enzyme redesign philosophies in great detail, we would like to only mention here some more noteworthy works to provide the interested reader with an overview of recent activity in the field. An area that has gained significant interest in recent years is enzyme specificity, as a result of increasing awareness [80–83] of the fact that, contrary to the classical image of enzyme catalysis, many (if not most) enzymes are catalytically promiscuous, catalyzing one or more chemically distinct reactions in addition to their native reaction. Significant effort has been invested into experimental redesign of enzyme specificity [84,85]. Along with this, inserting or modifying activity has also been the topic of recent computational studies. Earlier examples of this include Park *et al.*'s work [86], which introduced β-lactamase activity into a glyoxylase scaffold, while completely destroying the original activity. More recently, Korendovych and coworkers were able to computationally redesign calmodulin, which is a regulatory calcium binding protein, into an allosterically controlled Kemp eliminase that is activated upon Ca^{2+} binding by means of a single mutation [87]. Another recent work [88] used computational loop remodeling to control interactions between active site residues and the bound substrate and was able to alter the specificity of a human guanine deaminase to make it 100-fold more active with a structurally similar but chemically distinct substrate ammelide (Figure 7), and 2×10^4-fold less active for the native guanine substrate, compared to the wild-type enzyme (resulting in a net specificity change of 2.5×10^6-fold). A third significant recent work [89] focused on computationally redesigning a zinc-containing mouse deaminase to catalyze organophosphate hydrolysis (specifically the R(P) isomer of a coumarinyl analogue of the nerve agent cyclosarin), which, while giving only a modest initial phosphatase activity (k_{cat}/K_M of 4 M^{-1} s^{-1}), compared to the wild-type, which showed no detectable deaminase activity. However, this could be improved to a k_{cat}/K_M of ~10^4 M^{-1} s^{-1} after directed evolution. Therefore, rational enzyme redesign was again successfully used to provide a starting point for subsequent evolution to generate an enzyme with an activity comparable to that of an moderately efficient naturally occurring enzyme [20].

Figure 7. Structural comparison between (**A**) guanine (*i.e.*, the native substrate of human guanine deaminase), and (**B**) ammelide.



Finally, to conclude this section, we would like to briefly mention the increased interest in the challenging problem of designing efficient catalysts for chiral chemistry. Being able to control enzyme enantio- or regioselectivity is a problem of significant interest to the pharmaceutical industry, due to the role of chirality in drug efficacy, as well as potentially playing a major role in the industrial production of fine chemicals. However, computational redesign of stereo- or regioselecitvity poses a significant computational problem, due to the very small differences that govern selectivity as well as the tight dance between steric and electronic effects. There have been several computational attempts to address the origin of enzyme selectivity, however, these have often focused either on just ground states [90] or limited small models of the active site where the results can be very dependent on the precise starting structure used [91]. A more promising recent work [75] addressed the selectivity of the soluble epoxide hydrolase (sEH), which catalyzes epoxide ring opening to yield the corresponding vicinal diols. Through a combination of low-level QM/MM umbrella sampling molecular dynamics (MD) simulations and high-level *ab initio* calculations, the authors were able to rationalize the experimentally observed selectivity of the enzyme for phenyl *vs*. benzylic attack, identifying factors responsible for governing this selectivity. This work highlights the importance of proper sampling and examining multiple conformations, however, the obtained results only provided reasonable agreement with experimental rate constants when using a very high level of theory. Here, it should be noted that Frushicheva and Warshel recently examined the selectivity of wild-type and mutant *Candida antarctica* lipase using the EVB approach [92], and, as with earlier work on the Kemp eliminases [36,37], were able to obtain very high quantitative agreement with both the wild-type enzyme, as well as correctly reproducing (and rationalizing) the observed changes in selectivity upon mutation. It is worth highlighting here that this is a particularly challenging system that faces problems such as different degrees of water penetration, and, therefore, this work demonstrates the crucial importance of extensive sampling in order to obtain convergent results and to be able to reliably rationalize enantioselectivity [92].

4. *In Silico* "Directed Evolution": Approaches to Reduce the Sequence Space in Laboratory Evolution Studies

Despite extensive improvements to rational enzyme design using computational approaches, the arguably most powerful tool in enzyme (re)design for biocatalytic purposes remains simply random (or semi-random) laboratory evolution [25,93]. Despite its power, the limitations in searching sequence space as well as the complexity of the enzymes' catalytic actions remain inherent bottlenecks in all directed evolution studies. As a result of this, several experimental strategies are being developed to handle these problems in order to allow for high accuracy "low-throughput" screening (an issue discussed in detail in e.g., [94]). The question then arises of the extent to which computational approaches can contribute to this approach, guiding and rationalizing the ongoing experimental work. In light of the importance of this problem, it should not be surprising that there are also several computational strategies being developed in order to limit the search space in directed evolution studies as well as to perform screening of mutations *in silico*, allowing for a form of computational directed evolution. Here, we will address a number of these that can be used for improving enzymatic activity (many more of which have been reviewed elsewhere, see e.g., [95–98]). We aim to demonstrate that we are reaching a stage in the field where, rather than being limited to the early stages of artificial enzyme design or small-scale qualitative modifications, computational approaches can start to be used as a bridge across the gap that currently exists between rational design and laboratory evolution.

Early computational approaches towards reducing sequence space in DE studies included applying mean-field theory to explore protein fitness landscapes as well as the structural tolerance of individual residues [99], suggesting that mutations that are beneficial to activity and stability tend to occur at amino acid positions that are tolerant to substitutions. This was then matched by comparison to experimental directed evolution on subtilisin E and the T4 lysozyme, demonstrating that during experimental DE, favorable mutations were accumulated at the position predicted by this approach. A related approach is the SIRCH algorithm [100], which uses a second-order mean-field approach to identify residue-residue clashes from rotamer libraries, using atomistic representations to calculate rotamer-backbone, romater-intrinsic and rotamer-rotamer conformational energies. This can be used for truncating libraries in experimental directed evolution. More recent approaches for targeting directed evolution studies by exploring sequence function space include the protein sequence activity relationship algorithm (ProSAR) [101], currently owned by CODEXIS Inc™. This approach uses Kaufmann's NK model [102] to describe the fitness landscape. Here, in this case N is the number of variable positions in the protein, and K is the degree of coupling between the variable positions. ProSAR then exploits a genetic algorithm to search the NK landscape, based on an iterative procedure in which a subset of protein sequences are generated from the full library of possible sequences. These are then ordered in terms of decreasing fitness, the top protein is retained as the backbone for the design work, and the list of residues at each position in the backbone is expanded by addition of subsequent protein sequences in decreasing order of fitness. This allows for the identification of residues potentially contributing most to fitness, and then parents are "mated" and the procedure is repeated iteratively until the desired property is observed (for more details, see [101]). A clear advantage of this approach in an industrial context is the limited knowledge of detailed chemistry that is required to achieve noticeable improvements in efficiency, and an example of a practical application of this approach was illustrated for instance in the 4000-fold improvement in the volumetric productivity of a cyanation process necessary for the synthesis of a cholesterol-lowering drug, atorvastatin (Lipitor), by a bacterial dehalogenase [103]. A few last examples include statistical coupling analysis (SCA), which is a bioinformatics procedure that uses sequence information to determine networks of energetically coupled co-evolving residues in proteins [104] and has already been applied to the design of allosteric communication in proteins [105] as well as the design of artificial sequences capable of folding to target structures [106]. Finally, while the SCA approach is powerful, an alternative approach based on evaluating energy-based allosteric matrices, which has already been successfully applied to explore the effect of mutations on transition state energetics and fidelity in DNA polymerases [107], is likely to be a powerful tool for not only the rational design of TSAs and drug design, but also for computational enzyme design.

Another issue that has gathered increasing attention recently is that of protein promiscuity, which has been hypothesized to play an important role in guiding enzyme evolution [80,81,83]. This is of potential importance to enzyme design [83], as reverse evolution to a promiscuous "progenitor" enzyme capable of catalyzing multiple reactions provides an attractive starting point for the insertion of novel functionality. Therefore, in addition to general interest in resurrecting ancestral proteins [108,109], there is also interest in mapping and redirecting enzyme evolutionary trajectories to modify enzyme functionality [110]. Finally, apart from standard bioinformatics-based approaches that can be used to do this, such as those harnessed in the aforementioned works, recent years have also seen a tendency to borrow from the social or economics sciences, using approaches such as Pareto efficiency or Pareto optimality [111] as an optimization procedure in protein design [112,113] for cases where there are multiple objects present, and an efficient and fast optimization procedure is needed.

A key common feature of the approaches described above is the fact that they rely on limited to no knowledge of the molecular details for the actual chemical step being catalyzed by the systems of interest. As the examples brought here highlight, there are cases where such approaches can rapidly provide improvements in enzyme activity, and are particularly useful in, for instance, industrial settings where a modest improvement at minimal cost is often more desirable than approaches that can provide better improvements but require far more input in terms of both manpower hours and chemical or computational expense. To conclude this section, we would like to highlight a number of recently developed screening approaches that directly target the chemical step, as well as the effect of mutations on subsequent folding stability, and highlight the fact that the gap between "black-box" approaches and those that require more insight into the actual system are rapidly decreasing allowing for effective targeting of actual chemistry and, therefore, potentially much higher improvements in catalytic efficiency than current approaches allow. As a starting point, considering the multitude of computational approaches that are currently available, it is important to focus on what is exactly required of a computational approach in order to be able to do effective computational enzyme design. As one of the main challenges with directed evolution studies is focusing the search space in an effective way, and the fact that it is hard to deduce the molecular basis for the beneficial or detrimental effects observed with different mutations, one would require from a computational approach that it can reliably reproduce catalytic effects in both wild-type and mutant enzymes (see also discussion in e.g., [114]). In such a way, it is possible to then use as a scoring function the expected effect of mutations on the catalytic effect in order to directly rank different constructs. A second issue, however, is the cost-to-benefit ratio of such an approach. Recent years have seen tremendous improvements in approaches for experimental directed evolution (see e.g., [94]), both increasing the library sizes that can be handled and also finding ways to focus the search space. Therefore, it is already possible to obtain quite impressive improvements on enzymatic activity using experimental approaches alone, and for a computational approach to be advantageous, it needs to provide an insight that the experiment cannot provide by itself, and to do so in a way that is faster than one can currently perform with experimental evolution. As discussed in Section 3, and illustrated by the data shown in Figure 6, we believe that one of the most effective ways of currently doing this is the EVB approach, as it not only allows you to rapidly explore mutational effects without the need for any further parameters adjustment once the reference state has been correctly calibrated, but also, it provides detailed breakdowns of the different contributors to this effect, allowing one to dissect the molecular basis for catalysis. In principle, one could do brute-force EVB to test different mutations (driven by chemical insight into the native reaction). However, considering the fact that it is necessary to do extensive sampling over multiple trajectories in order to get physically meaningful convergent results, computational cost will still become a limiting factor. Therefore, it can be useful to have a toolkit with which to do rapid initial screening of mutation hotspots prior to subsequent more detailed testing of the most promising constructs using EVB (and ultimately experiment). A number of these have been discussed and compared in detail in [114,115]. Specifically, these works compared the efficacy of different computational approaches for modeling the enzyme-catalyzed rearrangement of chorismate to prephenate (Figure 8), as well as for probing the effect of mutations on the observed catalytic activity of the different mono-, di- and trimeric forms of chorismate mutase (CM). This particular system was chosen as it has been important both for design studies [116], as well as for probing the validity of transition state analogues [64,116]. Specifically, the authors compared using the EVB directly to screen for mutations with focusing only on the TS electrostatic energy, using in this case the semimacroscopic version of the protein dipole Langevin dipole method, in its linear response approximation version (PDLD/S-LRA, for methodological details see [117]). The authors also explored the effectiveness of using the changes in EVB reorganization energies upon mutation as a screening tool. Here, it was found that, clearly, the EVB was the most effective tool for quantitative screening, due to its ability to reliably reproduce mutational effects in a quantitative manner (see e.g., Figures 6 and 9). In comparison, examining reorganization energies was found to be useful but suffered from convergence problems, and also, the PDLD/S-LRA (or the group contribution approach which dissects the contribution of different residues to the activation barrier or to TS binding energies) could be useful for identifying residues that can contribute to electrostatic stabilization of the TS, however, as it does not include the important effect of protein reorganization, it is mainly helpful as an initial screening approach to identify residues that could then be examined further by EVB.

Figure 8. Overview of the mechanism for the rearrangement of chorismate to prephenate.

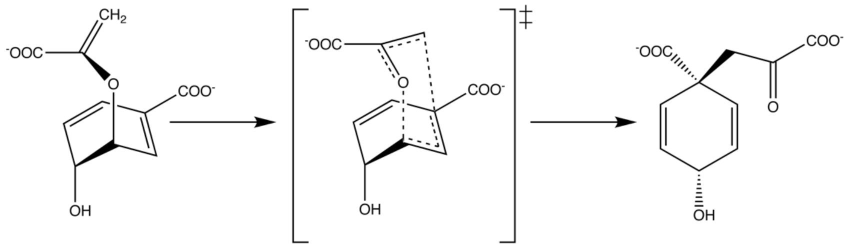
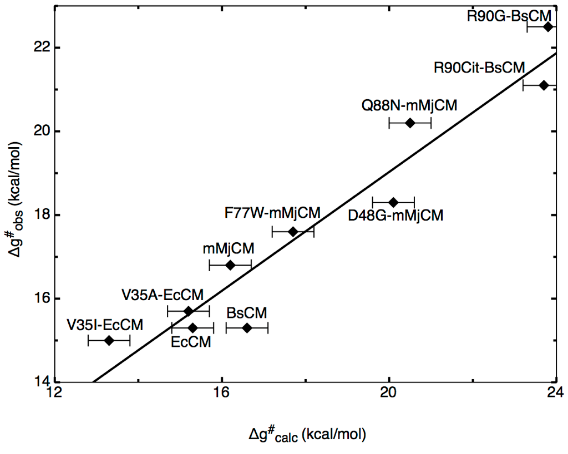


Figure 9. Correlation between calculated ($\Delta g^{\neq}_{calc}$) and observed ($\Delta g^{\neq}_{exp}$) activation barriers for the rearrangement of chorismate to prephanate catalyzed by the wild-type mono- (MjCM), di- (EcCM) and trimeric (BsCM) forms of chorismate mutase. The corresponding data for a range of mutants are also shown here. Note that R90Cit denotes a mutation to a non-standard amino acid (citrulline). Based on data presented in Table 2 of [114]. Calculations were performed using the empirical valence bond approach. In all cases, the calculated value is within ~2 kcal/mol of the experimental value.

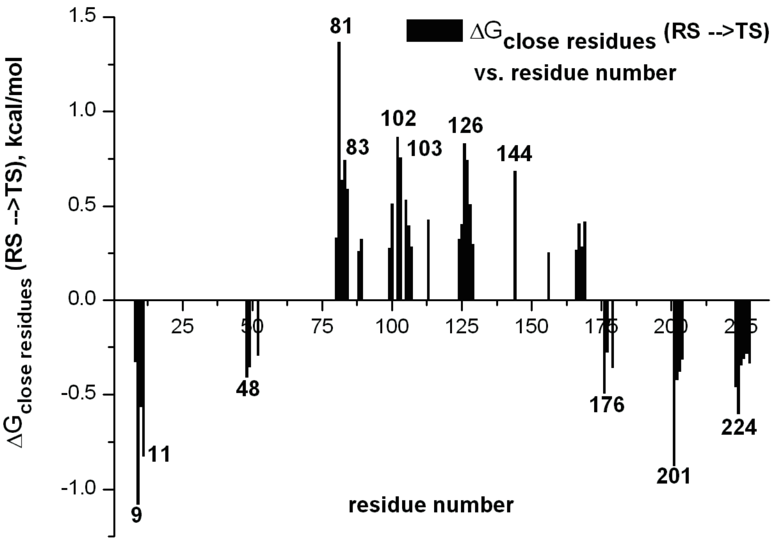


The authors have since evolved and expanded on their screening approaches, highlighting their power as applied to the cases of the aforementioned Kemp elimination reaction [36,37] and also to an enantioselective lipase from *Candida Antarctica* [92] (CAL A). Here, the authors' starting point is a simple and effective rapid screening approach, which bases itself on using the PDLD/S-LRA approach and estimating the electrostatic contribution of different residues ("groups") to the activation barrier using the formulation:

$$\Delta\Delta g^{\ddagger}_{elec} \cong 332\sum_{ij} \bar{q}_j \Delta Q_i / r_{ij}\bar{\varepsilon}_{ij} \tag{1}$$

Here, $\bar{q}_j$ refers to the effective charge (or dipole in the case of polar residues) of the qth residue, ΔQ_{i} refers to the change in substrate residual charges upon moving from the ground to the transition state, and ε_{ij} represents the dielectric constant for a given interaction. The effect of mutating each residue can then be explored by artificially assigning a charge of 1.0 to all residues in the protein, and then identifying the charge change that will lead to the most negative $\Delta\Delta g^{\ddagger}_{\mathrm{elec}}$ (*i.e.*, the greatest stabilization upon mutation). An example of this in the case of a designed Kemp eliminase [36] is shown in Figure 10.

Figure 10. Predicting mutation hotpots for a designed Kemp eliminase. The y-axis denotes the interactions (in kcal/mol) between the protein residues with the substrate as its charge changes upon moving from the reactant state (RS) to the transition state (TS). Large negative contributions suggest optimal sites for mutations likely to enhance the catalytic effect. This figure was originally presented in [36].



This then leads to the formulation:

$$(\Delta\bar{q}_j)_{opt} \cong -\alpha\partial\Delta\Delta g^{\ddagger}_{elec} / \partial q_j = -\alpha\sum_{i} \Delta Q_i / r_{ij}\varepsilon_{ij} \tag{2}$$

where the optimal values of $\Delta\bar{q}_j$ are proportional to the electrostatic group contributions in the case when all the protein groups are positive charges, and α is a proportionality constant. Such an approach is particularly useful in cases where the effect is predominantly electrostatic, and also for challenging cases where the charge change upon moving from ground to product state is sufficiently small to makes it hard to directly exploit the polarity of the active site residues (see [36]). However, it can be insufficient in cases such as enzyme enantioselectivity, where the observed effect is due to not only electrostatic but also steric effects, by combining the linear response approximation (LRA) approach [118], which can provide a reliable estimate of the free energy associated with changing the electrostatic potential of the system from one potential to another, with the steric (nonelectrostatic) component of Åqvist's linear interaction energy (LIE) approach [119,120], which introduces an empirical parameter (β) that scales the van der Waals component of the protein-ligand interaction. Such a combined LRA/β approach has shown significant promise in the test case of CAL A [118]. A final critical issue is, of course, the effect of such mutations on protein stability. That is, stability should minimally not be significantly impaired, and ideally retained or even optimized. This can be done in an effective way using a focused dielectric constant [121], which approximates the folding energy using the expression:

$$\Delta G_{fold} \cong 332\sum_{ij} q_i q_j / r_{ij}\varepsilon_{focus} \tag{3}$$

Here, $\varepsilon_{\text{focus}}$ denotes the optimal dielectric constant for a given interaction, and q_{i} and q_{j} are the charges of the relevant ionized residues at a given pH [121]. For other recent examples of predicting protein stability upon mutation using computational approaches we refer the reader to [122–126], amongst others. As discussed in e.g., [37], combining approaches to optimize TS stabilization with protein stability constraints can provide a highly efficient approach to predict optimal mutations of particularly distant (from the active site) ionized residues, and, as can be seen, while this is an area still in its infancy, significant progress is being made at a rapid pace, and we believe that approaches such as these will be the most effective currently available in the move towards *in silico* directed evolution.

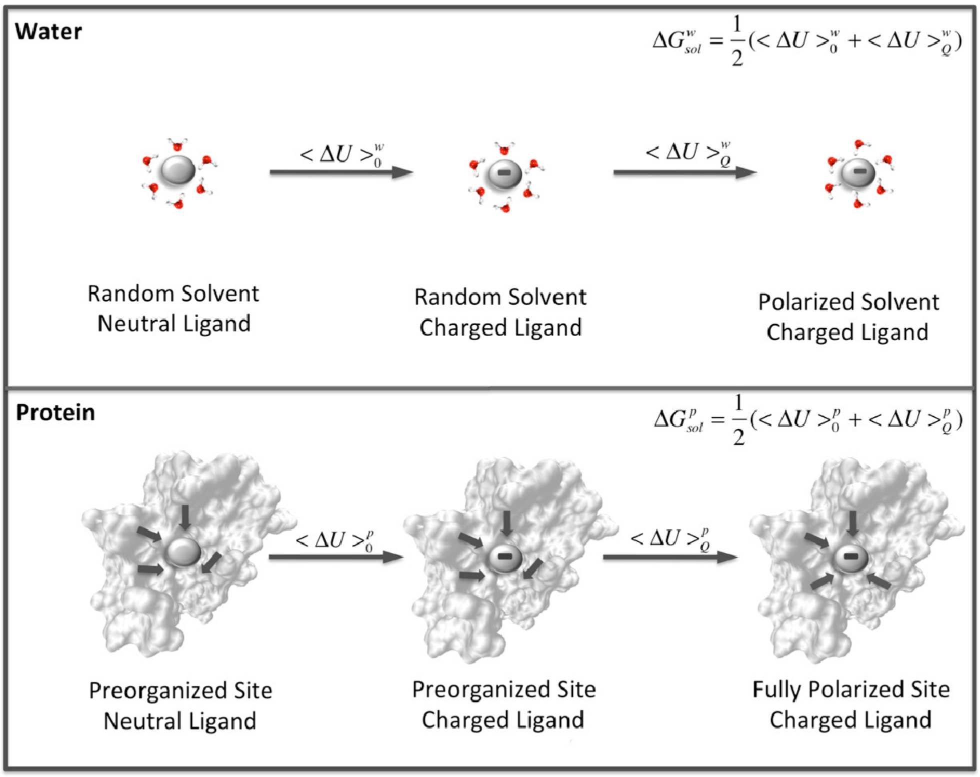
5. How do Enzymes Actually Work, or Why is Computational Enzyme Design so Difficult?

Computers are constantly increasing in power, and, in theory, computational approaches should play a major role in the design of novel catalysts. However, as illustrated in Sections 2 and 3, despite impressive advances in computational enzyme design (both *de novo* and by redesigning existing systems), the actual rate enhancements obtained by the designed systems are quite poor, and clearly far from that of naturally occurring systems. The question then becomes why this should be the case? This has been a topic of significant discussion [36,37,66,116], and suggestions as to the source of this problem have included, at the most qualitative level, simply that either the active site construct is incorrect (based on the argument that enzymes require idealized active sites), that the idealized active site is not realized in practice, and that the designed active site is not supported by the scaffold it is placed into [66].

While valid points, such arguments reflect the focus that has been placed on achieving shape complementarity in many of the enzyme design studies discussed in this work. Here, we would like to argue that the actual situation is more complex than this. That is, enzymes are chemical catalysts, and therefore, in order to be able to engage in effective enzyme (re)design, it is important to have detailed insight into how enzymes actually work at the molecular level. Indeed, being able to design effective artificial enzymes is perhaps one of the best proofs that we have finally solved this problem. However, the question of how enzymes actually work is one that has eluded enzymologists for over a century, and multiple hypothesis have been put forward to try to rationalize the tremendous catalytic proficiencies of enzymes, including desolvation effects, strain, and entropy loss during the reaction coordinate (see discussion in e.g., [13,127,128]). The problems with these (and related) hypotheses have been discussed in detail elsewhere [13,70,129,130]. However, we would like to focus on a hypothesis that has gained a lot of popularity in recent years, namely the potential importance of dynamical effects to enzyme catalysis [131–135]. One of the problems with this hypothesis is that it has not been clearly defined: That is, clearly, once atoms are above a few degrees K they move, and enzymes are dynamic entities. Additionally, many enzymes utilize conformational modification as an important part of their catalytic cycles. However, whether such dynamics can actually contribute to the *chemical* step of enzyme catalysis is a different issue, and one that is not possible to test experimentally, as there is no experiment that can actually check this. Therefore, arguments in favor of the dynamical hypothesis have been mainly based on *indirect* observations such as similar timescales for the conformational and chemical steps [132,133]. We explored this issue in great detail [128] and demonstrated that, based on our simulations, this would strongly suggest that the chemical step has no memory of any preceding conformational transition [136]. Additionally, even anticatalytic mutations that apparently provide direct evidence for the importance of dynamics in enzyme catalysis [134] were demonstrated to be electrostatic in origin [77]. Other workers than us have also expressed concern over this issue [137–139]. We bring this up here due to the increasing popularity of invoking dynamical effects as being important for artificial enzyme design (see e.g., [140]). Clearly, conformational changes can be very important for enzyme function, and by modulating such changes one also modulates the activity of the enzyme. However, as one is only likely to impair the catalytic activity of the enzyme through this, and as the enzyme appears to have no memory of the conformational change once the chemical step has started, it is unclear how such effects are likely to actually aid enzyme design.

While it is unlikely that dynamical effects play an important role in the actual rate acceleration that enzymes have evolved to optimize, a number of other factors clearly are to different degrees. These include well-characterized chemical effects such as acid-base catalysis and covalent catalysis, both of which can account in relevant cases for some of the observed catalytic effect. The major contributor, however, as has been observed by countless simulation studies [13], is the electrostatic preorganization of the active site [141] (Figure 11). Specifically, the reorganization penalty associated with reorienting randomly oriented water dipoles as the environment becomes polarized by changing charge on the substrate is much higher than the corresponding effect in an enzyme active site, where the relevant dipoles and charges are already optimally organized. The importance of electrostatic complementarity has been particularly highlighted in the case of ketosteroid isomerase [65,142,143], which has been one of the "classical" systems to challenge the electrostatic preorganization idea [144–146], and where it was demonstrated that almost the entire origin of the catalytic effect is electrostatic [65]. Now the fact that an increasing number of enzymes are being demonstrated to be "catalytically promiscuous" [83], catalyzing multiple, chemically distinct transition states in addition to their native transition state, with proficiencies that can at times almost compete with that for their native reaction [147–150], would superficially appear to be at odds with the idea of electrostatic preorganization (and the argument above that enzymes have "idealized" active sites). However, a recent computational work (to the best of our knowledge) that has comparatively explored both structural and electrostatic features driving catalysis of multiple substrates in a quantitative fashion in an extremely promiscuous arylsulfatase [151,152] has demonstrated that the enzyme is able to identify and distinguish between different, chemically distinct transition states, but that the promiscuity is simply driven by the ability of the promiscuous substrates to exploit the pre-existing electrostatic preorganization of the active site for the native substrate. Similar qualitative observations were made for a related system based on experimental work [153]. Therefore, the promiscuity appears to even in such a case be electrostatically-based [152], suggesting a role for chemistry-driven protein evolution. It is also worth noting that in this study, the two best substrates examined (out of four) proceeded through the most compact and expansive transition states, respectively. Also, out of a number of promiscuous substrates, the one that the enzyme preferentially catalyzed was the largest and bulkiest of these (in agreement with experimental data [147]), illustrating that the enzyme can exploit its active site plasticity to adapt to the substrate if the electrostatics match. Therefore, it could be argued that, rather than focusing on shape complementarity, it would be much more effective to focus on electrostatic complementarity, and use rational chemistry-driven protein redesign in order to obtain effective catalysts (see also discussion in [36,37]).

Figure 11. Schematic overview of the preorganization effect. This figure first considers the charging of a substrate in an environment that has not been polarized by the substrate and then illustrates the effect of the polarization of the solvent by the field of the substrate, which is substantially larger in (**A**) water than in (**B**) protein. This figure was originally presented in [65].


Water
$\Delta G^w_{sol} = \frac{1}{2}(<\Delta U>^w_0 + <\Delta U>^w_Q)$
$<\Delta U>^w_0$
$<\Delta U>^w_Q$
Random Solvent
Neutral Ligand
Random Solvent
Charged Ligand
Polarized Solvent
Charged Ligand
Protein
$\Delta G^p_{sol} = \frac{1}{2}(<\Delta U>^p_0 + <\Delta U>^p_Q)$
$<\Delta U>^p_0$
$<\Delta U>^p_Q$
Preorganized Site
Neutral Ligand
Preorganized Site
Charged Ligand
Fully Polarized Site
Charged Ligand

6. Conclusions and Future Perspectives

Artificial enzymes have the potential to play a major role in sustainable development, providing green reusable catalysts for processes encompassing all aspects of life, from generating new therapeutics to the food industry to their use as detergents. Therefore, interest in using enzymes as artificial catalysts has exploded, and biocatalysis is one of the most rapidly growing current fields (see e.g., [15,34,154,155]). However, design of artificial enzymes requires an intimate understanding of how enzymes work, and to date, the precise molecular details of enzyme catalysis still remain controversial and to some extent elusive, although electrostatics clearly plays a central and dominant role [13]. Therefore, any progress in artificial enzyme design will be aided by parallel progress in our understanding of enzymology, and the ability to effectively design artificial enzymes will in turn be the best proof that we have finally understood how enzymes work.

Despite these challenges, recent years have seen significant progress in both *de novo* computational enzyme design, from minimal active site models, as well as computational protein redesign, based on existing templates, and constant increases in computational power are expected to continue to accelerate such advances. Nevertheless, impressive as such studies have been, as has been illustrated by the examples brought up in this review that there is still a very large gap between the proficiencies of natural and designed enzymes. Many arguments have been put forward to rationalize this [66,116], some of which have focused on the importance of having an idealized active site and the problems with realizing such a situation. However, as was outlined in Section 5, many promiscuous enzymes do not have idealized active sites, recognizing multiple chemically distinct transition states (e.g., [147–150,153]), with often quite high efficiency. Here, in the only example known to us where such promiscuity has been *quantitatively* dissected using computational approaches to compare the catalysis of multiple substrates, it appeared that the main driving force for the promiscuity was simply the electrostatic preorganization of the active site, which, while it is optimized for the native substrate, can also be flexible enough to accommodate multiple, chemically distinct substrates [152]. If it can be extended to other systems, such electrostatic flexibility, whether naturally occurring or engineered, can clearly be exploited for enzyme design. This then raises a number of questions. The first is that of what the best starting point actually is for enzyme design. That is, while clearly the ability to engineer novel catalysts completely *de novo* is impressive in and of itself, Nature already provides a vast range of templates for many different types of chemistry, and naturally occurring enzymes are evolving all the time. Therefore, we believe that there is a strong argument in favor of starting from a naturally occurring system that demonstrates evolvability and manipulating its activity, rather than performing completely *de novo* enzyme design. However, regardless of the starting point, the next issue becomes that of how to effectively improve such systems. As was illustrated in Section 2, even in cases where the initial catalytic activities of *de novo* designed systems were relatively poor, subsequent rounds of experimental evolution can improve on such poor activities. Despite its power, however, directed evolution studies will inherently be limited by the vastness of the sequence space combined with low frequencies of desirable mutations even in targeted libraries [26]. In light of this, we are therefore at a very exciting time in the field. As illustrated in Section 4, recent years have seen substantial advances in approaches to reduce the search space needed in directed evolution studies, but also, even more importantly (in light of the fact that enzymes are chemical catalysts), in approaches that can effectively and reliably screen for mutation hotspots and predict the effect of mutations *in silico*, even prior to any experimental testing. Such approaches can directly target the chemical step, dissecting the contribution of different residues to catalysis, and quantifying the effect of modification of different residues both directly in the active site and even quite far from it. Critically, such approaches (such as, e.g., those discussed in [36,37,114,115,118]) allow for extensive sampling and take into account the reorganization of the protein environment, which is often missing in many design studies. This means that they can be used to study complex problems such as that of enzyme selectivity [118], where a very subtle balance between steric and electrostatic effects can completely determine the preference of an enzyme for one form of the substrate, or position of attack, over another. Clearly there are still challenges involved in such chemistry-driven approaches, which include the problems of dealing with systematic searches of the effect of multiple mutations simultaneously. Now this problem can, for example, be alleviated by use of a coarse-grained model that can be extended to examine multiple mutants at once, while then also extrapolating back to an all-atom model, as discussed in [118]). More critical, however, is the risk that incorporated mutations could cause large-scale conformational rearrangements, or even collapse, that are not captured by the computational approach (although many approaches for predicting protein stability upon mutation are currently being developed, with varying degrees of success [121–126]). In the long run, once such approaches can be coupled with approaches that can reliably predict the effect of mutations on enzyme physico-chemical properties such as thermal stability and solubility, this would then provide an ideal starting point for experimental testing. Such an iterative approach, in which the experimental evolution is guided by rational *in silico* evolution (which is in turn guided by detailed knowledge of the molecular machinery for catalysis), with theory being used to guide and rationalize experiment and experiment being used to test, validate and refine theory, will ultimately allow for a far more efficient design strategy. This, in addition to accelerating the design process overall by allowing a significant part of it to be performed computationally, also accelerates the design process by allowing for the construction of focused "smart" libraries for experimental evolution, significantly reducing the size of the sequence space that needs sampling. When this is combined with parallel advances in *de novo* enzyme design and structure-based protein redesign, this will provide a much-needed bridge, which closes the gap that currently exists between computational enzyme design and laboratory evolution.

Acknowledgments

The authors would like to thank the Swedish Research Council (Vetenskapsrådet, 2010-5026) and the Carl Tryggers Foundation for Scientific Research (CTS 11:226) for research funding, as well as the Slovenian Research agency for providing a visiting studentship to Rok Borstnar (J1-2014). The European Research Council has provided financial support under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement n° 306474. Finally, the authors thank Arieh Warshel for helpful discussions and Maria Frushicheva for providing the original high-resolution version of Figure 10.

References

1. Wöhler, F. Ueber künstliche bildung des harnstoffs. *Pogg. Ann.* **1828**, *88*, 253–256.
2. Hickman, A.J.; Sanford, M.S. High-valent organometallic copper and palladium in catalysis. *Nature* **2012**, *484*, 177–185.
3. Marinetti, A.; Julien, H.; Voituriez, A. Enantioselective, transition metal catalyzed cycloisomerizations. *Chem. Soc. Rev.* **2012**, *41*, 4884–4908.
4. Tosatti, P.; Nelson, A.; Marsden, S.P. Recent advances and applications of iridium-catalysed asymmetric allylic substitution. *Org. Biomol. Chem.* **2012**, *10*, 3147–3163.

5. Candeias, N.R.; Afonso, C.A.; Gois, P.M. Making expensive dirhodium(II) catalysts cheaper: Rh(II) recycling methods. *Org. Biomol. Chem.* **2012**, *10*, 3357–3378.
6. Yan, Z.Y.; Xiao, Y.; Zhang, L. Gold-catalyzed one-step construction of 2,3-dihydro-1*H*-pyrrolizines with an electron-withdrawing group in the 5-position: A formal synthesis of 7-methoxymitosene. *Angew. Chem. Int. Ed.* **2012**, *51*, 8624–8627.
7. Heck, R.F.; Nolley, J.P. Palladium-catalyzed vinylic hydrogen substitution reactions with aryl, benzyl, and styryl halides. *J. Org. Chem.* **1972**, *37*, 2320–2322.
8. King, A.O.; Okukado, N.; Negishi, E. Highly general stereo-, regio- and chemoselective synthesis of terminal and internal conjugated enynes by Pd-catalyzed reactions of alkynylzinc reagents with alkenyl halides. *J. Chem. Soc. Chem. Commun.* **1977**, 683–684.
9. Miyaura, N.; Yamada, K.; Suzuki, A. A new stereospecific cross-coupling by the palladium-catalyzed reaction of 1-alkenylboranes with 1-alkenyl or 1-alkynyl halides. *Tetrahedron Lett.* **1979**, *20*, 3437–3440.
10. Green Chemistry, United States Environmental Protection Agency. Available online: http://www.epa.gov/greenchemistry/ (accessed on 20 July 2012).
11. Radzicka, A.; Wolfenden, R. Rates of uncatalyzed peptide bond hydrolysis in neutral solution and the transition state affinities of proteases. *J. Am. Chem. Soc.* **1996**, *118*, 6105–6109.
12. Wolfenden, R.; Snider, M.J. The depth of chemical time and the power of enzymes as catalysts. *Accounts Chem. Res.* **2001**, *34*, 938–945.
13. Warshel, A.; Sharma, P.K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M.H.M. Electrostatic basis for enzyme catalysis. *Chem. Rev.* **2006**, *106*, 3210–3235.
14. Brown, R.S.; Lu, Z.-L.; Liu, T.; Tsang, W.Y.; Edwards, D.R.; Neverov, A.A. Dinuclear Zn(II) catalysts as biomimics of RNA and DNA phosphoryl transfer enzymes: Changing the medium from water to aclohol provides enzyme-like rate enhancements. *J. Phys. Org. Chem.* **2009**, *23*, 1–15.
15. Schmid, A.; Dordick, J.S.; Hauer, B.; Kiener, A.; Wubbolts, M.; Witholt, B. Industrial biocatalysis today and tomorrow. *Nature* **2001**, *409*, 258–268.
16. García-Arellano, H.; Valderrama, B.; Saab-Rincón, G.; Vazquez-Duhalt, R. High temperature biocatalysis by chemically modified cytochrome C. *Bioconjug. Chem.* **2002**, *13*, 1336–1344.
17. Ramesh, N.P. Synthesis of chiral pharmaceutical intermediates by biocatalysis. *Coord. Chem. Rev.* **2008**, *252*, 659–701.
18. Sukamaran, J.; Hanefeld, U. Enantioselective C–C bond synthesis catalysed by enzymes. *Chem. Soc. Rev.* **2005**, *34*, 530–542.
19. Wiessman, K. Rational or random? *Chem. World* **2004**, *1*. Available online: http://www.rsc.org/chemistryworld/Issues/2004/July/rational.asp (accessed on 28 September 2012).
20. Bar-Even, A.; Noor, E.; Savir, Y.; Libermeister, W.; Davidi, D.; Tawfik, D.S.; Milo, R. The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* **2011**, *50*, 4402–4410.
21. Frieden, E.; Walter, C. Prevalence and significance of the product inhibition of enzymes. *Nature* **1963**, *198*, 834–837.
22. Sanderson, K. Chemistry: Enzyme expertise. *Nature* **2011**, *471*, 397–398.
23. McCullum, E.O.; Williams, B.A.; Zhang, J.; Chaput, J.C. Random mutagenesis by erorr-prone PCR. *Methods Mol. Biol.* **2010**, *634*, 103–109.

24. Stemmer, W.P. Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* **1994**, *370*, 389–391.
25. Turner, N.J. Directed evolution drives the next generation of biocatalysts. *Nat. Chem. Biol.* **2009**, *5*, 567–573.
26. Alcolombri, U.; Elias, M.; Tawfik, D.S. Directed evolution of sulfotransferases and paraoxonases by ancestral libraries. *J. Mol. Biol.* **2011**, *411*, 837–853.
27. Lutz, S. Beyond directed evolution: Semi-rational protein engineering and design. *Curr. Opin. Biotechnol.* **2010**, *21*, 734–743.
28. Bershtein, S.; Tawfik, D.S. Ohno's model revisited: Measuring the frequency of potentially adaptive mutations under various mutational drifs. *Mol. Biol. Evol.* **2008**, *25*, 2311–2318.
29. Soskine, M.; Tawfik, D.S. Mutational effects and the evolution of new protein functions. *Nat. Rev. Genet.* **2010**, *11*, 572–582.
30. Wickens, M.; Cox, M.M. Critical reviews in biochemistry and molecular biology. Introduction. *Crit. Rev. Biochem. Mol. Biol.* **2009**, *44*, 2.
31. Reetz, M.T.; Carballeira, J.D. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. Protoc.* **2007**, *2*, 891–903.
32. Helge, J.; Bornscheuer, U.T. Natural diversity to guide focused directed evolution. *ChemBioChem* **2010**, *11*, 1861–1866.
33. Reetz, M.T. Laboratory evolution of stereoselective enzymes: A prolific source of catalysts for asymmetric reactions. *Angew. Chem. Int. Ed.* **2011**, *50*, 138–174.
34. Jäckel, C.; Hilvert, D. Biocatalysis by evolution. *Curr. Opin. Biotechnol.* **2010**, *21*, 753–759.
35. Röthlisberger, D.; Kheronsky, O.; Wollacott, A.M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J.L.; Althoff, E.A.; Zanghellini, A.; Dym, O.; *et al*. Kemp elimination catalysts by computational enzyme design. *Nature* **2008**, *435*, 190–195.
36. Frushicheva, M.P.; Cao, J.; Chu, Z.T.; Warshel, A. Exploring challenges in rational enzyme design by simulating the catalysis in artificial Kemp eliminase. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 16869–16874.
37. Frushicheva, M.P.; Cao, J.; Warshel, A. Challenges and advances in validating enzyme design proposals: The case of Kemp eliminase catalysis. *Biochemistry* **2011**, *50*, 3849–3858.
38. Thorn, S.N.; Daniels, R.G.; Auditor, M.-T.M.; Hilvert, D. Large rate accelerations in antibody catalysis by strategic use of haptenic charge. *Nature* **1995**, *373*, 228–230.
39. Debler, E.W.; Ito, S.; Seebeck, F.; Heine, A.; Hilvert, D.; Wilson, I.A. Structural origins of efficient proton abstraction from carbon by a catalytic antibody. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 4984–4989.
40. Hollfelder, F.; Kirby, A.J.; Tawfik, D.S. Off-the-shelf proteins that rival tailor-made antibodies as catalysts. *Nature* **1996**, *383*, 60–63.
41. Hollfelder, F.; Kirby, A.J.; Tawfik, D.S.; Kikuchi, K.; Hilvert, D. Characterization of proton-transfer catalysis by serum albumins. *J. Am. Chem. Soc.* **2000**, *122*, 1022–1029.
42. Hollfelder, F.; Kirby, A.J.; Tawfik, D.S. Efficient catalysis of proton transfer by synzymes. *J. Am. Chem. Soc.* **1997**, *119*, 9578–9579.

43. Na, J.; Houk, K.N.; Hilvert, D. Transition state of the base-promoted ring-opening of isoxazoles. Theoretical prediction of catalytic functionalities and design of haptens for antibody production. *J. Am. Chem. Soc.* **1996**, *118*, 6462–6471.
44. Tantillo, D.J.; Chen, J.; Houk, K.N. Theozymes and compuzymes: Theoretical models for biological catalysis. *Curr. Opin. Chem. Biol.* **1988**, *2*, 743–750.
45. Zanghellini, A.; Jiang, L.; Wollacott, A.M.; Cheng, G.; Meiler, J.; Althoff, E.A.; Röthlisberger, D.; Baker, D. New algorithms and an *in silico* benchmark for computational enzyme design. *Protein Sci.* **2006**, *15*, 2785–2794.
46. Bolon, D.N.; Mayo, S.L. Enzyme-like proteins by computational design. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 14274–14279.
47. Khersonsky, O.; Röthlisberger, D.; Dym, O.; Albeck, S.; Jackson, C.J.; Baker, D.; Tawfik, D.S. Evolutionary optimization of computationally designed enzymes: Kemp eliminases of the KE07 series. *J. Mol. Biol.* **2010**, *396*, 1025–1042.
48. Khersonsky, O.; Röthlisberger, D.; Wollacott, A.M.; Murphy, P.; Dym, O.; Albeck, S.; Kiss, G.; Houk, K.N.; Baker, D.; Tawfik, D.S. Optimization of the *in-silico*-designed kemp eliminase KE70 by computational design and directed evolution. *J. Mol. Biol.* **2010**, *407*, 391–412.
49. Kheronsky, O.; Kiss, G.; Röthlisberger, D.; Dym, O.; Albeck, S.; Houk, K.N.; Baker, D.; Tawfik, D.S. Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 10358–10363.
50. Jiang, L.; Althoff, E.A.; Clemente, F.R.; Doyle, L.; Roethlisberger, D.; Zanghellini, A.; Gallaher, J.L.; Betker, J.L.; Tanaka, F.; Barbas, I.; *et al. De novo* computational design of retro-aldol enzymes. *Science* **2008**, *319*, 1387–1391.
51. Lassila, J.K.; Baker, D.; Herschlag, D. Origins of catalysis by computationally designed retroaldolase enzymes. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 4937–4942.
52. Wang, L.; Althoff, E.A.; Bolduc, J.; Jiang, L.; Moody, J.; Lassila, J.K.; Giger, L.; Hilvert, D.; Stoddard, B.L.; Baker, D. Structural analyses of covalent enzyme-substrate analog complexes reveals strengths and limitations of de novo enzyme design. *J. Mol. Biol.* **2012**, *415*, 615–625.
53. Althoff, E.A.; Wang, L.; Jiang, L.; Giger, L.; Lassila, J.K.; Wang, Z.; Smith, M.; Hari, S.; Kast, P.; Herschlag, D.; *et al*. Robust design and optimization of retroaldol enzymes. *Protein Sci.* **2012**, *21*, 717–726.
54. Siegel, J.B.; Zanghellini, A.; Lovick, H.M.; Kiss, G.; Lambert, A.R.; St. Clair, J.L.; Hilvert, D.; Gelb, M.H.; Stoddard, B.L.; Houk, K.N.; *et al.* Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science* **2010**, *329*, 309–313.
55. Faiella, M.; Andreozzi, C.; de Rosales, R.T.M.; Pavone, V.; Maglio, O.; Nastri, F.; deGrado, W.F.; Lombardi, A. An artificial di-iron oxo-protein with phenol oxidase activity. *Nat. Chem. Biol.* **2009**, *5*, 882–884.
56. Der, B.S.; Machius, M.; Miley, M.J.; Mills, J.; Szyperski, T.; Kuhlman, B. Metal-mediated affinity and orientation specificity in a computationally designed protein homodimer. *J. Am. Chem. Soc.* **2011**, *134*, 375–385.

57. Der, B.S.; Edwards, D.R.; Kuhlman, B. Catalysis by a de novo zinc-mediated protein interface: Implications for natural enzyme evolution and rational enzyme engineering. *Biochemistry* **2012**, *51*, 3933–3940.
58. Jennewein, S.; Schürmann, M.; Wolberg, M.; Hilker, I.; Luiten, R.; Wubbolts, M.; Mink, D. Directed evolution of an industrial biocatalyst: 2-deoxy-D-ribose 5-phosphate aldolase. *Biotechnol. J.* **2006**, *1*, 537–548.
59. Gutte, B.; Daumigen, M.; Wittschieber, E. Design, synthesis and characterisation of a 34-residue polypeptide that interacts with nucleic acids. *Nature* **1979**, *281*, 650–655.
60. Russel, A.J.; Fersht, A.R. Rational modification of enzyme catalysis by engineering surface charge. *Nature* **1987**, *328*, 496–500.
61. Hellinga, H.W.; Richards, F.M. Construction of new ligand binding sites in proteins of known structure. 1. Computer-aided modeling of sites with pre-defined geometry. *J. Mol. Biol.* **1991**, *222*, 763–785.
62. Hellinga, H.W.; Caradonna, J.P.; Richard, F.M. Construction of new ligand binding sites in proteins of known structure II. Grafting of a buried transition metal binding site into Escherichia coli thioredoxin. *J. Mol. Biol.* **1991**, *222*, 787–803.
63. Schramm, V.L. Enzymatic transition states, transition-state analogs, dynamics, thermodynamics, and lifetimes. *Annu. Rev. Biochem.* **2011**, *80*, 703–732.
64. Barbany, M.; Gutierrez-de-Teran, H.; Sanz, F.; Villà-Freixa, J.; Warshel, A. On the generation of catalytic antibodies by transition state analogues. *ChemBioChem* **2003**, *4*, 277–285.
65. Kamerlin, S.C.L.; Sharma, P.K.; Chu, Z.T.; Warshel, A. Ketosteroid isomerase provides further support to the idea that enzymes work by electrostatic preorganization. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 4075–4080.
66. Baker, D. An exciting but challenging road ahead for computational enzyme design. *Protein Sci.* **2010**, *19*, 1817–1819.
67. Alexandrova, A.N.; Röthlisberger, D.; Baker, D.; Jorgensen, W.L. Catalytic mechanism and performance of computationally designed enzymes for Kemp elimination. *J. Am. Chem. Soc.* **2008**, *130*, 15907–15915.
68. Kiss, G.; Röthlisberger, D.; Baker, D.; Houk, K.N. Evaluation and ranking of enzyme designs. *Protein Sci.* **2010**, *19*, 1760–1773.
69. van Zeist, W.-J.; Bickelhaupt, F.M. The activation strain model of chemical reactivity. *Org. Biomol. Chem.* **2010**, *8*, 3118–3127.
70. Sharma, P.K.; Chu, Z.T.; Olsson, M.H.M.; Warshel, A. A new paradigm for electrostatic catalysis of radical reactions in vitamin B_{12} enzymes. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 9661–9666.
71. Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; John Wiley & Sons: New York, NY, USA, 1991.
72. Kamerlin, S.C.L.; Warshel, A. The EVB as a quantitative tool for formulating simulations and analyzing biological and chemical reactions. *Faraday Discuss.* **2010**, *145*, 71–106.
73. Kamerlin, S.C.L.; Warshel, A. The empirical valence bond model: Theory and applications, *WiRE: Comp. Mol. Sci.* **2011**, *1*, 30–45.

74. Hwang, J.-K.; King, G.; Creighton, S.; Warshel, A. Simulation of free energy relationships and dynamics of S_N2 reactions in aqueous solution. *J. Am. Chem. Soc.* **1988**, *110*, 5297–5311.
75. Lonsdale, R.; Hoyle, S.; Grey, D.T.; Ridder, L.; Mulholland, A.J. Determinants of reactivity and selectivity in soluble epoxide hydrolase from quantum mechanics/molecular mechanics modeling. *Biochemistry* **2012**, *51*, 1774–1786.
76. Plotnikov, N.; Kamerlin, S.C.L.; Warshel, A. Paradynamics: An effective and reliable coarse-grained model for *ab initio* QM/MM free energy calculations and related tasks. *J. Phys. Chem. B.* **2011**, *115*, 7950–7962.
77. Adamczyk, A.J.; Cao, J.; Kamerlin, S.C.L.; Warshel, A. The catalytic power of dihydrofolate reductase and other enzymes arises from electrostatic preorganization, not conformational motions. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 14115–14120.
78. Mones, L.; Kulhánek, P.; Simon, I.; Laio, A.; Fuxreiter, M. The energy gap as a universal reaction coordinate for the study of chemical reactions. *J. Phys. Chem. B* **2009**, *113*, 7867–7873.
79. Privett, H.K.; Kiss, G.; Lee, T.M.; Blomberg, R.; Chica, R.A.; Thomas, L.M.; Hilvert, D.; Houk, K.N.; Mayo, S.L. Iterative approach to computational enzyme design. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 3790–3795.
80. Jensen, R.A. Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.* **1976**, *30*, 409–425.
81. O'Brien, P.J.; Herschlag, D. Catalytic promiscuity and the evolution of new enzymatic activities. *Chem. Biol.* **1999**, *6*, R91–R105.
82. Aharoni, A.; Gaidukov, L.; Khersonsky, O.; McQ Gould, S.; Roodveldt, C.; Tawfik, D.S. The “evolvability” of promiscuous protein functions. *Nat. Genet.* **2005**, *37*, 73–76.
83. Khersonsky, O.; Tawfik, D.S. Enzyme promiscuity: A mechanistic and evolutionary perspective. *Ann. Rev. Biochem.* **2010**, *79*, 471–505.
84. Afriat-Jurnou, L.; Jackson, C.J.; Tawfik, D.S. Reconstructing a missing link in the evolution of a recently diverged phosphotriesterase by active-site loop remodelling. *Biochemistry* **2012**, *51*, 6047–6055.
85. Elias, M.; Tawfik, D.S. Divergence and convergence in enzyme evolution: Parallel evolution of paraoxonases from Quorum-quenching lactonases. *J. Biol. Chem.* **2012**, *287*, 11–20.
86. Park, H.-S.; Nam, S.-H.; Lee, J.K.; Noon, C.N.; Mannervik, B.; Benkovic, S.J.; Kim, H.-S. Design and evolution of new catalytic activity with an existing protein scaffold. *Science* **2006**, *311*, 535–538.
87. Korendovych, I.V.; Kulp, D.W.; Wu, Y.; Cheng, H.; Roder, H.; deGrado, W.F. Design of a switchable eliminase. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 6823–6827.
88. Murphy, P.M.; Bolduc, J.M.; Gallaher, J.L.; Stoddard, B.L.; Baker, D. Alteration of enzyme specificity by computational loop remodeling and design. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 9215–9220.
89. Khare, S.D.; Kipnis, Y.; Greisen, P., Jr.; Takeuchi, R.; Ashani, Y.; Goldsmith, M.; Song, Y.; Gallaher, J.L.; Silman, I.; Leader, H.; *et al.* Computational redesign of a mononuclear zinc mettaloenzyme for organophosphate hydrolysis. *Nat. Chem. Biol.* **2012**, *8*, 294–300.
90. Schiøtt, B.; Bruice, T.C. Reaction mechanism of soluble epoxide hydrolase: Insights from molecular dynamics simulations. *J. Am. Chem. Soc.* **2002**, *124*, 14558–14570.

91. Hopmann, K.H.; Himo, F. Insights into the reaction mechanism of soluble epoxide hydrolase from theoretical active site mutants. *J. Phys. Chem. B* **2006**, *110*, 21299–21310.
92. Fruschicheva, M.P.; Warshel, A. Towards quantitative computer-aided studies of enzymatic enantioselectivity: The case of *Candida antarctica* lipase A. *ChemBioChem* **2012**, *23*, 215–223.
93. Jäckel, C.; Kast, P.; Hilvert, D. Protein design by directed evolution. *Annu. Rev. Biophys.* **2008**, *37*, 153–173.
94. Goldsmith, M.; Tawfik, D.S. Directed enzyme evolution: Beyond the low-hanging fruit. *Curr. Opin. Struct. Biol.* **2012**, *22*, 406–412.
95. Voigt, C.A.; Mayo, S.L.; Arnold, F.H.; Wang, Z.-G. Computationally focusing the directed evolution of proteins. *J. Cell. Biochem. Supp.* **2001**, *37*, 58–63.
96. Gustafsson, C.; Govindarajan, S.; Minshull, J. Putting engineering back into protein engineering: Bioinformatics approaches to catalyst design. *Curr. Opin. Biotechnol.* **2003**, *14*, 366–370.
97. Romero, P.A.; Arnold, F.H. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell. Biol.* **2009**, *10*, 866–876.
98. Fox, R.J.; Huisman, G.W. Enzyme optimization. Moving from blind evolution to statistical exploration of sequence-function space. *Trends Biotechnol.* **2008**, *26*, 132–138.
99. Voigt, C.A.; Mayo, S.L.; Arnold, F.H.; Wang, Z.-G. Computational method to reduce the search space for directed protein evolution. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 3778–3783.
100. Moore, G.L.; Maranas, C.D. Identifying residue-residue clashes in protein hybrids by using a second-order mean field approach. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 5091–5096.
101. Fox, R.; Roy, A.; Govindarajan, S.; Minshull, J.; Gustafsson, C.; Jones, J.T.; Emig, R. Optimizing the search algorithm for protein engineering by directed evolution. *Prot. Eng.* **2003**, *16*, 589–597.
102. Kaufmann, S.A.; Weingerber, E.D. The NK model of rugged fitness landscapes and its application to maturation of the immune response. *J. Theor. Biol.* **1989**, *141*, 211–245.
103. Fox, R.J.; Davis, S.C.; Mundorff, E.C.; Newman, L.M.; Gavrilovic, V.; Ma, S.K.; Chung, L.M.; Ching, C.; Tam, S.; Muley, S.; *et al.* Improving catalytic function by ProSAR-driven enzyme evolution. *Nat. Biotechnol.* **2007**, *25*, 338–344.
104. Lockless, S.W.; Raganathan, R. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* **1999**, *286*, 295–299.
105. Lee, J.; Natarajan, M.; Nashine, V.C.; Socolich, M.; Vo, T.; Russ, W.P.; Benkovic, S.J.; Rangathan, R. Surface sites for engineering allosteric control in proteins. *Science* **2008**, *322*, 438–442.
106. Socolich, M.; Lockless, S.W.; Russ, W.P.; Lee, H.; Gardner, K.H.; Ranganathan, R. Evolutionary information for specifying a protein fold. *Nature* **2005**, *437*, 512–518.
107. Xiang, Y.; Oelschlaeger, P.; Florian, J.; Goodman, M.F.; Warshel, A. Simulating the effect of DNA polymerase mutations on transition-state energetics and fidelity: Evaluating amino acid group contribution and allosteric coupling for ionized residues in human pol β. *Biochemistry* **2006**, *45*, 7036–7048.
108. Thornton, J.W.; Bridgham, J.T. *Using Ancestral Gene Resurrection to Understand the Evolution of Protein Function*; Oxford University Press: Oxford, UK, 2006.

109. Bridgham, J.T.; Carroll, S.M.; Thornton, J.W. Evolution of hormone-receptor complexicity by molecular exploitation. *Science* **2006**, *312*, 97–101.
110. Huang, R.; Hippauf, F.; Rohrbeck, D.; Haustein, M.; Wenke, K.; Feike, J.; Sorrelle, N.; Piechulla, B.; Barkman, T.J. Enzyme functional evolution through improved catalysis of ancestrally nonpreferred substrates. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 2966–2971.
111. Fudenberg, D.; Tirole, J. *Game Theory*; MIT Press: Cambridge, MA, USA, 1983.
112. Garcia-Seisdedos, H.; Ibarra-Molero, B.; Sanchez-Ruiz, J.M. Probing the mutational interplay between primary and promiscuous protein functions: A computational-experimental approach. *PLoS Comput. Biol.* **2012**, *8*, e1002558.
113. Suarez, M.; Tortosa, P.; Carrera, J.; Jaramillo, A. Pareto optimization in computational protein design with multiple objectives. *J. Comp. Chem.* **2008**, *29*, 2704–2711.
114. Roca, M.; Vardi-Kilshtain, A.; Warshel, A. Toward accurate screening in computer-aided enzyme design. *Biochemistry* **2009**, *48*, 3046–3056.
115. Vardi-Kilshtain, A.; Roca, M.; Warshel, A. The empirical valence bond as an effective strategy for computer-aided enzyme design. *Biotechnol. J.* **2009**, *4*, 495–500.
116. Toscano, M.D.; Woycechowsky, K.J.; Hilvert, D. Minimalist active-site redesign: Teaching old enzymes new tricks. *Angew. Chem. Int. Ed.* **2007**, *46*, 4468–4470.
117. Sham, Y.Y.; Chu, Z.T.; Tao, H.; Warshel, A. Examining methods for calculations of binding free energies: LRA, LIE, PDLD-LRA, and PDLD/S-LRA calculations of ligands binding to an HIV protease. *Proteins: Struct. Funct. Genet.* **2000**, *39*, 393–407.
118. Lee, F.S.; Chu, Z.T.; Bolger, M.B.; Warshel, A. Calculations of antibody antigen interactions: Microscopic and semimicroscopic evaluation of the free-energies of binding of phosphorylcholine analogs to Mcpc603. *Prot. Eng.* **1992**, *5*, 215–228.
119. Almlof, M.; Carlsson, J.; Åqvist, J. Improving the accuracy of the linear interaction energy method for solvation free energies. *J. Chem. Theory Comput.* **2007**, *3*, 2162–2175.
120. Hansson, T.; Marelius, J.; Åqvist, J. Ligand-binding affinity prediction by linear interaction energy methods. *J. Comput-Aided Mol. Des.* **1998**, *12*, 27–35.
121. Vicatos, S.; Roca, M.; Warshel, A. Effective approach for calculations of absolute stability of proteins using focused dielectric constants. *Proteins* **2009**, *77*, 670–684.
122. Wickstrom, L.; Gallicchio, E.; Levy, R.M. The linear interaction energy method for the prediction of protein stability changes upon mutation. *Proteins* **2012**, *80*, 111–125.
123. Topham, C.M.; Srinivasan, N.; Blundell, T.L. Prediction of the stability of protein mutants based on structural environment-dependent amino acid substitution and propensity tables. *Protein Eng.* **1997**, *10*, 7–21.
124. Parhiban, V.; Gromiha, M.M.; Schomburg, D. CUPSAT: Prediction of protein stability upon point mutations. *Nucleic Acids Res.* **2006**, *34*, W239–W242.
125. Potapov, V.; Cohen, M.; Schreiber, G. Assessing computational methods for predicting protein stability upon mutation: Good on average but not in the details. *Protein Eng.* **2009**, *22*, 553–560.
126. Gromiha, M.M.; Huang, L.T. Machine learning algorithms for predicting protein folding rates and stability of mutant proteins: Comparison with statistical methods. *Curr. Protein Pept. Sci.* **2011**, *12*, 490–502.

127. Kamerlin, S.C.L.; Mavrí, J.; Warshel, A. Examining the case for the effect of barrier compression on tunneling, vibrationally enhanced catalysis, catalytic entropy and related issues. *FEBS Lett.* **2010**, *584*, 2759–2766.
128. Kamerlin, S.C.L.; Warshel, A. At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis? *Proteins: Struct. Func. Bioinformat.* **2010**, *78*, 1339–1375.
129. Warshel, A.; Åqvist, J.; Creighton, S. Enzymes work by solvation substitution rather than by desolvation. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 5820–5824.
130. Štrajbl, M.; Shurki, A.; Kato, M.; Warshel, A. Apparent NAC effect in chorismate mutase reflects electrostatic transition state stabilization. *J. Am. Chem. Soc.* **2003**, *125*, 10228–10237.
131. Henzler-Wildman, K.; Kern, D. Dynamic personalities of proteins. *Nature* **2007**, *450*, 964–972.
132. Henzler-Wildman, K.A.; Lei, M.; Thai, V.; Kerns, S.J.; Karplus, M.; Kern, D. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **2007**, *450*, 913–916.
133. Henzler-Wildman, K.A.; Thai, V.; Lei, M.; Ott, M.; Wolf-Watz, M.; Fenn, T.; Pozharski, E.; Wilson, M.A.; Petsko, G.A.; Karplus, M.; *et al.* Intrinsic motions along an enzymatic reaction trajectory. *Nature* **2007**, *450*, 838–844.
134. Bhabha, G.; Lee, J.Y.; Ekiert, D.; Gam, J.; Wilson, I.A.; Dyson, H.J.; Benkovic, S.J.; Wright, P.E. A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. *Science* **2011**, *332*, 234–238.
135. Villali, J.; Kern, D. Choreographing an enzyme's dance. *Curr. Opin. Chem. Biol.* **2010**, *14*, 636–643.
136. Pisliakov, A.V.; Cao, J.; Kamerlin, S.C.L.; Warshel, A. Enzyme millisecond conformational dynamics do not catalyze the chemical step. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 17359–17364.
137. Glowacki, D.R.; Harvey, J.N.; Mulholland, A.J. Taking Ockham's razor to enzyme dynamics and catalysis. *Nat. Chem.* **2012**, *4*, 169–176.
138. Glowacki, D.R.; Harvey, J.N.; Mulholland, A.J. Protein dynamics and enzyme catalysis: The ghost in the machine. *Biochem. Soc. Trans.* **2012**, *1*, 515–521.
139. Moliner, V. "Eppur si muove" (Yet it moves). *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 15013–15014.
140. Lassila, J.K. Conformational diversity and computational enzyme design. *Curr. Opin. Chem. Biol.* **2010**, *14*, 676–682.
141. Warshel, A. Energetics of enzyme catalysis. *Proc. Natl. Acad. Sci. USA* **1978**, *75*, 5250–5254.
142. Feierberg, I.; Åqvist, J. The catalytic power of ketosteroid isomerase investigated by computer simulation. *Biochemistry* **2002**, *41*, 15728–15735.
143. Warshel, A.; Sharma, P.K.; Chu, Z.T.; Aqvist, J. Electrostatic contributions to binding of transition state analogues can be very different from the corresponding contributions to catalysis: Phenolates binding to the oxyanion hole of ketosteroid isomerase. *Biochemistry* **2007**, *46*, 1466–1476.
144. Kraut, D.A.; Sigala, P.A.; Pybus, B.; Liu, C.W.; Ringe, D.; Petsko, G.A.; Herschlag, D. Testing electrostatic complementarity in enzyme catalysis: Hydrogen bonding in the ketosteroid isomerase oxyanion hole. *PLoS Biol.* **2006**, *4*, 0501–0519.
145. Schwans, J.P.; Kraut, D.A.; Herschlag, D. Determining the catalytic role of remote substrate binding interactons in ketosteroid isomerase. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 14271–14275.

146. Sigala, P.A.; Kraut, D.A.; Caaveiro, J.M.; Pybus, B.; Ruben, E.A.; Ringe, D.; Petsko, G.A.; Herschlag, D. Testing geometrical discrimination within an enzyme active site: Constrained hydrogen bonding in the ketosteroid isomerase oxyanion hole. *J. Am. Chem. Soc.* **2008**, *130*, 13696–13708.
147. Babtie, A.C.; Bandyopadhyay, S.; Olguin, L.F.; Hollfelder, F. Efficient catalytic promiscuity for chemically distinct reactions. *Angew. Chem.* **2009**, *48*, 3692–3694.
148. Jonas, S.; van Loo, B.; Hyvönen, M.; Hollfelder, F. A new member of the alkaline phosphatase superfamily with a formylglycine nucleophile: Structural and kinetic characterisation of a phosphate monoester hydrolyase/phosphodiesterase from Rhizobium leguminosarum. *J. Mol. Biol.* **2008**, *384*, 120–136.
149. Olguin, L.F.; Askew, S.E.; O'Donoghue, A.C.; Hollfelder, F. Efficient catalytic promiscuity of an enzyme superfamily: An arylsulfatase shows a rate acceleration of 10^{13} for phosphate monoester hydrolysis. *J. Am. Chem. Soc.* **2008**, *130*, 16547–16555.
150. van Loo, B.; Jonas, S.; Babtie, A.C.; Benjdia, A.; Berteau, O.; Hyvönen, M.; Hollfelder, F. An efficient, multiply promiscuous hydrolase in the alkaline phosphatase superfamily. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 2740–2745.
151. Luo, J.; van Loo, B.; Kamerlin, S.C.L. Examining the promiscuous phosphomonoesterase activity of *Pseudomonas aeruginosa* arylsulfatase: Comparison to analogous phosphatases. *Proteins: Struct. Func. Bioinformat.* **2012**, *80*, 1211–1226.
152. Luo, J.; van Loo, B.; Kamerlin, S.C.L. Catalytic promiscuity in pseudomonas aeruginosa arylsulfatase as an example of chemistry-driven protein evolution. *FEBS Lett.* **2012**, *11*, 1622–1630.
153. Ben-David, M.; Elias, M.; Filippi, J.J.; Duñach, E.; silman, I.; Sussman, J.L.; Tawfik, D.S. Catalytic versatility and backups in enzyme active sites: The case of serum paraoxonase 1. *J. Mol. Biol.* **2012**, *418*, 181–196.
154. Bornscheuer, U.T.; Kazlauskas, R.J. Catalytic promiscuity in biocatalysis: Using old enzymes to form new bonds and follow new pathways. *Angew. Chem. Int. Ed.* **2004**, *43*, 6032–6040.
155. Dalby, P.A. Engineering enzymes for biocatalysis. *Recent Pat. Biotechnol.* **2007**, *1*, 1–9.

© 2012 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).