

COMPUTATIONAL ENZYME DESIGN: ADVANCES, HURDLES AND POSSIBLE WAYS FORWARD

Mats Linder^{a,*}

Abstract: This mini review addresses recent developments in computational enzyme design. Successful protocols as well as known issues and limitations are discussed from an energetic perspective. It will be argued that improved results can be obtained by including a dynamic treatment in the design protocol. Finally, a molecular dynamics-based approach for evaluating and refining computational designs is presented.

MINI REVIEW ARTICLE

I. Introduction

A key factor in the success of the human race is our inclination to exploit nature for our own ends. Even though it has been disastrous many times throughout history, we would be nowhere near the current height of our civilization the without this trait.

One frontier waiting to be conquered is the control over enzyme catalysis. Enzymes have intrigued us for as long as they have been known because of their formidable complexity and proficiency, stemming from a relatively low number of catalyzed chemical transformations.[1] It is conceived that if we can take control of these 'unit operations' and their structure-function relationships, we will be able to design environmentally benign catalysts, yet featuring superior efficiency and selectivity compared to those in use today.[2] Another, more fundamental aspect is that the successful design of an enzyme is a stringent demonstration that the underlying mechanics have been understood. Given the increasing interest for industrial biocatalysis and the need for highly specific solutions to synthetic problems, the expectations on artificial (or *de novo*) enzymes for the coming decades are high.[2]

Recent years have seen tremendous advancement in computer-aided (or computational) enzyme design. The field can be defined as the use of *in silico* methods to understand, model and enhance/construct enzyme catalysis. As with all computational disciplines, the limits for what can be done are continuously pushed forward, from early approaches with limited side-chain rotamer optimization,[3-8] to more sophisticated optimization algorithms available today.[9-14] In addition, highly interesting crowd-sourcing applications have been applied in recent years, such as Rosetta@Home[15] and FoldIt.[16,17]

A long-standing vision in computational protein design has been to *automate*, i.e. letting an algorithm make most critical decisions, much like directed evolution is an automated exploitation of natural selection. In the light of the recent successful designs using such protocols,[18-21] it is however sobering to note that each active

structure is accompanied by a multitude of false positives.[22,23] No reported *de novo* enzyme has significantly outperformed the numerous catalytic antibodies elicited in recent decades.[21,22,24] let alone wild-type analogs. Moreover, it has recently been reported that some *de novo* designs do not tolerate closer scrutiny[25] and that established protocols sometimes fail altogether.[26]

This mini review will focus on recent trends in improving *de novo* designs beyond the automation protocols, discussed within the framework of contemporary understanding of how enzymes work.[27-31] It does not cover all methodologies and results published recently; for this purpose, the reader is directed to other recent reviews[2, 23, 32-35] (also see ref. 2 and references therein).

2. Automated protein design

Due to the sheer size of the sequence and conformational space of proteins, no computational method is able to assess them exhaustively. For example, the number of unique configurations when only considering a typical active site is $\sim 10^{65}$. [11] Hence, computational design relies on the ability to i) understand where to focus the efforts and ii) develop methods that efficiently examines and selects among the 'right' variables. To this end, all automated design protocols so far start with an existing protein scaffold and redesigns its sequence around some introduced functionality while maintaining a rigid backbone (Figure 1a). We will refer to such approaches as 'static'. Protocols from the Mayo and Hellinga labs pioneered automated design, [3,4,36-38] and the first computational 'designer enzyme' appeared in the early 2000's.[39] It was followed by an array of (re)designed binding proteins and enzymes.[8,18,38,40] Although several reports from the Hellinga lab have been challenged[25] and in some cases retracted,[41] the numerous reports of *de novo* functions promised a bright future for computational enzyme design.

One of the most successful approaches has been to couple the structure prediction utilities in the Rosetta design package,[9,42] developed by Kuhlman, Baker and coworkers,[10,14,43] with an 'inside-out' active site-design initiated by a 'theozyme'[44] optimization. Three very impressive enzyme designs were published rapidly (the 'Rosetta enzymes'), catalyzing a retro-aldol reaction,[19] a Kemp elimination,[20] and most recently a Diels-Alder reaction.[21] The inside-out design method is a clean-cut example of the general philosophy described above. It optimizes a minimal model

^aApplied Physical Chemistry, KTH Royal Institute of Technology, Teknikringen 30, SE-100 44, Stockholm, Sweden

* Corresponding author. Tel.: +46 87908429
E-mail address: linder2@kth.se (Mats Linder)

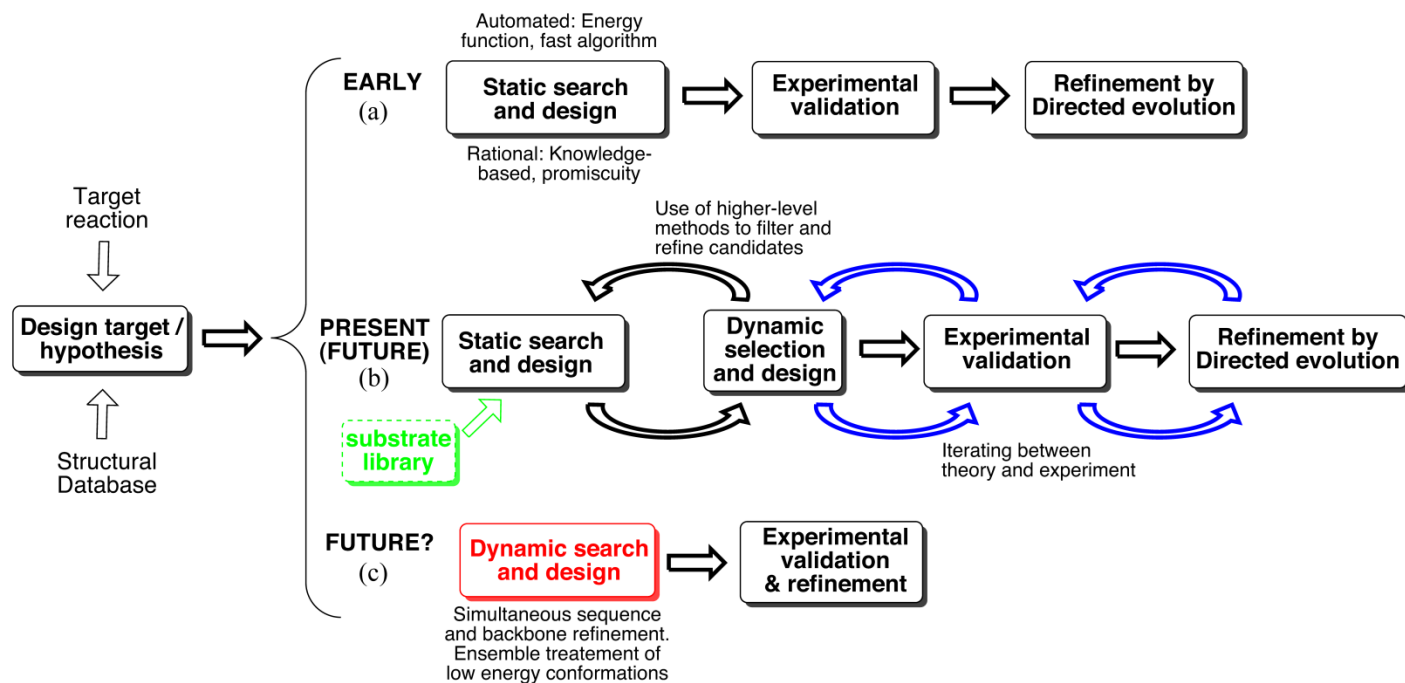


Figure 1. Evolution of enzyme design strategies. Nascent applications are indicated in blue and green and unlikely (or very distant) in red. (a) Approximate protocol based on automated and/or rational design followed by experimental validation and screening. Optionally, the designs can be further refined through directed evolution. (b) An iterative approach including molecular dynamics for selection and further design. (c) Tentative, automated computational design incorporating conformational flexibility and dynamics in the search algorithm.

involving the transition state (TS) and a few key interacting residues using quantum chemical (QC) methods, and the resulting theozyme is then fitted into a suitable active site, which is repacked with the Rosetta program. The idea relies on the finding that native proteins have nearly optimal sequences for their structure,[43] so that accommodating a new function can be done in a self-consistent way.

However, an enzyme needs to be exquisitely tuned in every part, not just the active site residues, to be proficient or even active.[45] As exemplified in Table I, enzymes straight out of the computer are typically quite weak catalysts. But fine-tuning distant residues is (so far) a bit to diffuse for computational methods,[45] and coupling with directed evolution has proven to be a more successful approach.[20,46] Nevertheless, such refined enzymes have at best had a factor $\sim 10^2$ added to their efficiencies (Table I).

Before proceeding with a discussion on how to improve prediction and refinement of catalytic power in computational design, we will now briefly review how enzyme catalysis is understood. The framework will then be used to quantify substrate binding and dynamic contributions to the catalytic effect.

3. The underlying theory

In the famous words of Linus Pauling,[47,48] enzymes work by "stabilization of the transition state" of the reaction (relative to the solvent). Transition state theory (TST) has acquired additional flavors since then,[49] but it is now generally agreed that the bulk of the catalytic effect can be attributed to the quasi-thermodynamic transition state (TS) stabilization.[28] This is good news for the computational enzyme designer, since the 'extra-thermodynamic'[28] terms (re-crossing, tunneling and non-equilibrium effects) are arguably too subtle to be incorporated in computational predictions (yet).

TS stabilization is understood as higher affinity for the TS of the enzyme with respect to the solvent, described by a 'dissociation constant' K_{TS} . [50] This hypothetical TS binding is illustrated in the pseudo-thermodynamic cycle in Figure 2 and is defined as

$$K_{TS} = K_M \frac{K_{\text{uncat}}^\ddagger}{K_{\text{cat}}^\ddagger} = K_M \frac{k_{\text{uncat}}}{k_{\text{cat}}} \quad (1)$$

with the associated dissociation energy

$$\Delta\Delta G_{TS} = -RT \ln K_{TS} = \Delta G_{\text{uncat}}^\ddagger - \Delta G_{\text{cat}}^\ddagger - \Delta G_{\text{bind}} \quad (2)$$

(We allow ourselves to approximate K_M with $\exp[\Delta G_{\text{bind}}/RT]$.) The catalytic proficiency of enzymes is thus the reciprocal of K_{TS} and can in native enzymes be as high as 10^{24} M^{-1} . [51]

Exactly how enzymes achieve their TS stabilization has been, and still is, widely debated.[31,52-54] That enzymes selectively bind their substrates is well understood[31] and reflected in small values of the Michaelis constant (K_M). Rate enhancement, defined as $k_{\text{cat}}/k_{\text{uncat}}$, requires a lower enzymatic reaction barrier ($\Delta G_{\text{cat}}^\ddagger < \Delta G_{\text{uncat}}^\ddagger$); if the difference is zero, K_{TS} simply equals K_M . One way or another, a proficient enzyme must therefore allow a reaction path for which $k_{\text{cat}}/k_{\text{uncat}} \gg 1$, equivalent to $\Delta\Delta G_{\text{cat}}^\ddagger = \Delta G_{\text{cat}}^\ddagger - \Delta G_{\text{uncat}}^\ddagger \ll 0$.

The discourse regarding the origins of TS stabilization can, in simple terms, be said to be about whether they have completely pre-organized active sites that are evolved to match the TS geometry and electrostatics,[31,55] or if they, e.g. by ground state destabilization (or entropy trapping),[56] desolvation[57] or dynamic motion,[30] increase the fraction of pre-organized states in conformational space.[53,54] While it is reasonable to suggest that all contributions play a role,[58] it is still disputed which one dominates in a particular enzyme.[30,59-61]

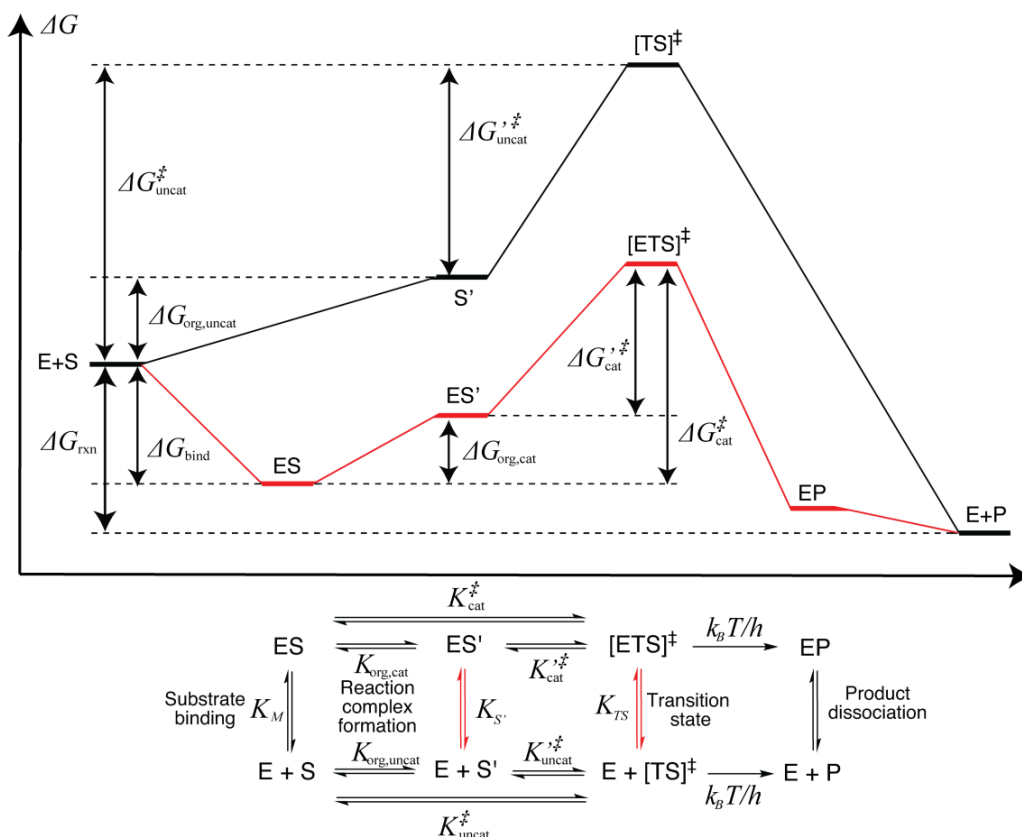


Figure 2. Schematic description of general enzyme catalysis (represented by one substrate and only the rate-determining step). The thermodynamic relationships in the bottom panel has been adapted from Wolfenden.[27] Fictitious equilibria are indicated in red.

Regardless of what explanation one prefers, it is instructive for the present discussion to define a pre-organized state (S' or ES') as any substrate conformation that is distorted with respect to the ground state (S or ES), so that it lies 'en route' to the TS on the reaction coordinate (Figure 2). Note that ES is a Boltzmann ensemble and contains all conformations defined to belong to ES' ; the crucial delimiter between the two ensembles can be seen as ES not necessarily being conformationally similar to the TS while ES' is. Thus, $\Delta G_{org} > 0$, and

$$\Delta G^\ddagger = \Delta G_{org} + \Delta G'^\ddagger \quad (3)$$

where ΔG^\ddagger is the free energy of activation from the pre-organized state. We can from Figure 2 define a 'pre-organization constant' $K_{S'}$ as

$$K_{S'} = K_M \frac{K_{org,uncat}}{K_{org,cat}} \quad (4)$$

and it follows that

$$K_{TS} = K_{S'} \exp \left[\frac{\Delta G_{cat}^\ddagger - \Delta G_{uncat}^\ddagger}{RT} \right] = K_{S'} K'_{TS} \quad (5)$$

This manipulation effectively divides the activation barrier in two parts, where ΔG_{org} can be assumed to contain mainly non-electrostatic and entropic contributions, whereas $\Delta G'^\ddagger$ mainly contains the charge-transfer associated with the chemical step.[62]

From an experimental perspective it may seem pointless to

consider such a 'micro-history'[28,63] of the reaction coordinate, since the only observables are K_M and k_{cat} , but it is a useful computational tool to characterize the enzyme. And, as will be argued below, to design better ones. The introduction of an intermediate state does not dismiss any of the various theories regarding the origins of enzyme catalysis. Indeed, a very proficient native enzyme can be regarded as having an $ES' \approx ES$ due to a pre-organized active site, in which case $\Delta G_{org,cat} \approx 0$ [60] and the extended model reduces to the usual one. That is, $\Delta G_{cat}^\ddagger \approx \Delta G'^\ddagger_{cat}$ and contains solely electrostatic contributions[31] while all pre-organization effects are captured in ΔG_{bind} .

However, in designer enzymes the match might not be so ideal, even with the appropriate catalytic groups in place. Such a situation will render a larger $\Delta G_{org,cat}$, and k_{cat}/k_{uncat} will be small even with seemingly ideal interactions in the TS. In this case, the enzyme might be improved by 'pulling down' ES' towards ES , thereby lowering the barrier, whereas in a structure with insufficient electrostatic pre-organization, the effort must focus on placement of the catalytic residues to lower $[ETS]^\ddagger$.

4. Where automated design fails

How do these mechanistic considerations influence computational design strategies? Since the most direct design approach is to optimize an active site around a computed TS model or perhaps TS hybrid, the designer enzyme should ideally be fairly good at binding the substrate(s) in their pre-organized conformations (ES') as well, with a consequently low $\Delta G_{org,cat}$ penalty. In other words, optimizing TS binding should in principle yield adequate pre-arrangement as well. It

has been pointed out that it is essentially the same strategy as employed when raising catalytic antibodies against haptens.[38]

As mentioned, such an approach essentially assumes that the design retains its structural integrity with respect to the scaffold, or at least that the region around the active site is rigid. Herein lies, the author argues, an important reason why computational design algorithms rarely exceed a rate enhancement of $\sim 10^6$ (Table 1),[22] and why the success rate is in fact very low.[21,26] The dynamic nature of enzymes includes both backbone and side-chain movements, and although it is a good approximation to assume that the backbone remains rigid with respect to ~ 10 introduced mutations, even a small distortion can severely affect the positioning of catalytic residues. A related problem is that backbone movements are slow on the molecular time scale, and they are therefore seldom captured in standard MD simulations.

Several studies have investigated the Rosetta enzymes,[46,64-67] and most of their conclusions point in the same direction: Unaccounted protein dynamics leads to a unsatisfying description of enzyme-substrate interactions and the ability to pre-organize into the target TS, which results in an over-optimistic prediction of activity. In other words, the static nature of the design process fails to predict the actual Boltzmann distribution of ES conformations. In addition, Warshel and coworkers have pointed out that the extensively studied Kemp Eliminases attain a significant part of their catalytic effect by bringing the substrate close to the catalytic base, whereas actual TS stabilization is poor and extremely difficult to achieve due to a small difference in charge distribution compared to the ground state.[66]

It should be noted that the automated computational protocols do a fairly good job at predicting the (albeit simple) reaction mechanism and stereoselectivity, but do not manage very well to separate actives from inactives.[19-21] Thus, selection is a major challenge for automated protein design.[22] The problem of picking out a few structures from an immense ensemble is in principle equivalent to what troubles other disciplines in computational chemistry: calculating small energy differences between large systems.

Table 1: Examples rate enhancements and specificities of computationally designed enzymes.^a

Name	Reaction	$\log \frac{k_{\text{cat}}}{k_{\text{uncat}}}$	$\log \frac{k_{\text{cat}}}{K_M}$	$-\log K_{TS}$	Ref.
P7D2	Ester hydrolysis	2.26	0.43	6.03	39
G ₄ -DF _{tet}	Phenol oxidation	$\approx 3^b$	1.42	6.08	18
RA61	Retro-Aldol	4.36	-0.15	8.04	19
KE07	Kemp Elimination	4.19	1.11	7.04	20
→KE07* ^c	Kemp Elimination	6.07	3.40	9.34	20
KE70	Kemp Elimination	5.08	2.10	8.04	46
→KE70* ^c	Kemp Elimination	6.63	4.75	10.7	46
HG-3 ^d	Kemp Elimination	5.77	2.63	8.56	79
DA_20_00	Diels-Alder ^e	0.61	-1.26	3.90	21
→DA_20_10 ^f	Diels-Alder ^e	1.93	0.73	5.90	21,83
→CE6	Diels-Alder ^e	1.96	1.94	7.11	83

^a'A' →' indicates the design has been developed from the closest design above.

^bOnly an approximate rate enhancement was reported by the authors. ^cThe 'R7 10/11G' variant of KE07,[20] and 'R6 4/8B' variant of KE70,[46] evolved by directed evolution and containing 8 and 14 mutations compared to their respective progenitor. ^dRefined in three generations using a combination of small-molecule placement,[11] MD and experimental techniques. ^eA bimolecular reaction, reported values therefore contain $k_{\text{cat}}(K_M1K_M2)$. ^fEvolved from DA_20_00 by rational design and contains 6 mutations with respect to the progenitor. ^gEvolved from DA_20_10 by exchanging an unstructured loop on the fringe of the active site to a helix-turn-helix motif that better encapsulates the substrates. The design was found by employing the community of FoldIt[16] players.

5. Searching for matching functionalities

In order to mitigate some of the problems associated with introducing a *de novo* mechanism in an existing scaffold, an alternative strategy could be to search for structures with functionalities matching the desired machinery already in place. We have been interested in re-designing enzymes containing an 'oxyanion hole' moiety[68] to catalyze the intermolecular Diels-Alder reaction, which is virtually unknown in nature[69-71] and therefore highly interesting for *de novo* design.[2] Our work was initially guided by a rational design framework (Figure 1a),[72] which can be said to rely on the concept of 'catalytic promiscuity'.[73] Catalytic promiscuity is in turn a manifest of evolutionary heritage,[23,74,75] where a wide array of reactions can be catalyzed by structurally and functionally similar enzymes.[23]

We recently adapted a 'semi-rational' approach,[76,77] which contains a biased search in the Protein Data Bank (PDB) for structures containing the target functionality (an oxyanion hole). The search is coupled with a search of a combinatorial substituent library for matching substrates, thereby increasing the chance of structural complementarity with the active site while keeping the number of mutations at a minimum (Figure 1b, green). After a pre-selection, the remaining structures are evaluated by molecular docking of TS analogs. The results then form the basis for further selection and rational mutations. Recently, Nosrati and Houk published a program ('SABER') designed to automatically search the PDB for 'pre-designs' having a set of predefined functionalities in place for promiscuous catalysis.[78] Their approach is consistent with ours, albeit in a more automated fashion. To the author's knowledge, only the original proof of principle study has been published to date. One can anticipate, however, future *de novo* designs employing the SABER methodology.

6. Including dynamics

Another key feature of our approach is that we rely on molecular dynamics (MD) simulations for evaluation, selection and quantification of different variants,[72,76] a practice that has recently begun appearing in several other studies (Figure 1b).[21,67,79] We refer to this as 'dynamic design'.

The virtue of dynamic design is that it relaxes a static design in response to the functionalities introduced, and in addition provides both qualitative and quantitative tools to distinguish between systems with just a few atoms discrepancy (even though it does not warrant a perfect description of reality). It was shown by Kiss *et al.* that MD-derived metrics could be used to significantly improve computational predictions of active designs[67] of *de novo* Kemp Eliminases.[20] The groups of Mayo, Houk and Hilvert then went on to use MD in an iterative protocol[11] to refine an inactive design for the same reaction to an active one (HG-3 in Table 1).[79]

Despite the use of both QC and MD methods in several dynamic protocols, few attempts have been made so far to use them for quantitative predictions. Warshel and coworkers recently used their 'empirical valence bond' (EVB) method to correlate computed ΔG_{cat} values with experiment, and argued that this approach can be used for screening[80] and refinement purposes.[45] It is the author's opinion that reliable energetic predictions of the designed reaction coordinate is the most robust way to improve the predictive power of computational enzyme design.

We took an early interest in probing the possibility of quantitative evaluation using MD and QC. One reason for this being that the docking protocol used in the early stages of design is

inherently static, like most other protocols employed thus far, and does not provide a particularly precise basis for selection of catalytic activity. Our studies have led to a dynamic protocol employing standard MD and cluster QC calculations,[76, 77] which for a preliminary benchmark test gives encouraging results.

To quantify the quality of a particular design, one needs to determine K_{TS} (eq. 1), which requires explicit predictions of the substrate binding and activation free energies. An estimate of K_M can be obtained from the MD trajectory while the rate constants in principle require *ab initio* calculations. As seen from Figure 2, $\Delta G^{\ddagger}_{cat}$ must be calculated with respect to ES, but this state is not correctly sampled at the QC level if not conformationally very restricted. Although QM/MM may be one solution to this problem, we have utilized that ES' can be fairly well defined within both MD and QC frameworks due to its conformational constraints, and use eq. 3 to estimate $\Delta G^{\ddagger}_{cat}$. The pre-organized complex can be defined in several ways, and we have used the concept of 'near attack conformers' (NACs) championed by Bruice[81, 53] to determine $\Delta G_{org,cat}$. [60]

$$\Delta G_{org,cat} \approx \Delta G_{NAC} = RT \ln \frac{N_{NAC}}{N_{total}} \quad (6)$$

where N_{NAC} denotes the number of instances in the ensemble obeying the criteria defined for a NAC. Hence, we compute $\Delta G^{\ddagger}_{cat}$ using a cluster model of the active site, and obtain $\Delta G_{org,cat}$ as a statistical average from MD. The uncatalyzed states are determined in a traditional way from QC.[72,76,77]

The NAC concept is debated[28,54,60,61] and suffers from its inherently arbitrary definition,[28] but we argue that it is a useful tool to optimize substrate binding in enzyme catalysis. In the notation of Figure 2, $NAC \equiv ES'$ (or S'), and as pointed out above, the ideal design goal is to equate this state with ES. Computational design typically produces multiple candidates with essentially identical catalytic residues, and in a QC evaluation of the TS stabilization it is convenient to use the same (or similar) models to save CPU time. They can instead be ranked by augmenting a common $\Delta G^{\ddagger}_{cat}$ with $\Delta G_{org,cat}$ obtained from MD. (This is essentially what is done by Houk and coworkers,[46,67] although without estimating energetics.)

In addition, this treatment is useful for quantifying the origins of catalysis (or lack thereof!). If ES' and S' are defined equivalently, the electrostatic portion of the TS can be determined by $\Delta G^{\ddagger}_{cat} - \Delta G^{\ddagger}_{uncat}$ (i.e. K'_{TS} in eq. 5; this comparison is similar to the 'caged' reference state used by Frushicheva *et al.*[45,66]), which can be taken as a measure of the quality of the designed catalytic machinery.

As a test of the predictive power of this concept, the kinetic constants of four active Diels-Alderases published by the Baker group were estimated in ref. 77. We used a 'LIE+ γ SASA' method to calculate binding constants from MD, and employed the theozyme reported by Siegel *et al.* as our cluster model. Despite its crudity, the approach predicted K_{TS} within one order of magnitude and ranked three of four designs correctly; the largest error was a factor 60 overestimation for CE6 ($\approx 2.4 \text{ kcal}\cdot\text{mol}^{-1}$). Predicted in the same way, two variants designed using our semi-rational protocol have estimated rate enhancements of $5.9\cdot 10^4$ and $4.1\cdot 10^6 \text{ M}$, and efficiencies of $8.6\cdot 10^4$ and $9.5\cdot 10^4 \text{ s}^{-1}\text{M}^{-1}$, respectively.[76,77] We have not yet been able to express these designs for experimental validation, but speculatively, our simulations suggest that the main difference between these variants and those published by the Baker group is the propensity to form NACs (or ES'). In our designs, we explicitly designed for alignment of the (ternary) ES complex to resemble the tentative TS, thereby lowering the $\Delta G_{org,cat}$ penalty. We recorded N_{NAC}/N_{total} values of ≤ 0.10 for the Baker variants, versus 0.25–0.50

for our best designs.

Interestingly, the computed $\Delta G^{\ddagger}_{cat}$ are not so different in these examples. In other words, we conclude that K'_{TS} is rather similar, whereas K_S is smaller in our computational designs. This observation suggests that the moderate rate enhancements of the Diels-Alderases given in Table I can be improved by dedicated improvement of substrate pre-arrangement. In addition, the treatment reveals the difficulty in attaining a large (electrostatic) TS stabilization. We have that $\Delta G^{\ddagger}_{uncat} - \Delta G^{\ddagger}_{cat} \leq 3 \text{ kcal/mol}$ in our design studies. This can possibly be attributed to lack of electrostatic reorganization in the TS, similar to the case of Kemp-eliminases.[66]

Again, note that partitioning K_{TS} as in eq. 5 is not always justified, and must be done with care depending on what one seeks to analyze. We have used it as a tool to improve designs suffering from poor substrate pre-arrangement and quantitatively compare systems with similar active sites. Furthermore, the QC model needs to be large enough to capture all essential electrostatics, and the theozyme used for evaluating the Baker Diels-Alderases is, strictly speaking, too small.[45] Nevertheless, we obtained remarkable correlation with experiment. In our design works, we use cluster models of the active site for the QC calculations, which are typically size-converged at 150-200 atoms.[82]

7. Outlook

Ideally, one would wish for a completely dynamic design approach, where everything is treated in a time-resolved fashion (Figure 1c). For example, one would be able to measure the direct response to a point mutation or change in conformation of individual residues in 'real time'. Some methodologies have begun incorporating backbone flexibility,[13,14,32] and the EVB approach suggested by Warshel and coworkers[45,66,80] in principle maps the whole reaction coordinate. However, completely dynamic treatments will need time to mature, if ever feasible. A common and resilient problem is of course the coarseness of force field approaches, introducing dependencies on different parameterizations and failing to treat more subtle aspects of chemical interactions.

More likely is consolidation of the trend for which publications from the last couple of years[21,46,67,78,79,83] provide mounting evidence: development of composite approaches that utilize rational, computational and experimental tools iteratively. Iteration between models and experiment has recently been demonstrated (Figure 1b, blue),[79] and additional protocols are likely to emerge. An intriguing example of the beneficial use of human resources is an enhanced Diels-Alderase design developed with the help of FoldIt players.[83] They collectively improved substrate binding of DA_20_10 by redesigning a loop surrounding the active site. The new design CE6 showed a 20-fold increased efficiency (see Table I).

It has been argued that K^{1}_{TS} values larger than 10^{11} M^{-1} ($\Delta\Delta G_{TS} \approx 15 \text{ kcal}\cdot\text{mol}^{-1}$) are virtually impossible for enzymes exhibiting non-covalent mechanisms.[29] Non-covalent mechanisms have been defined as those involving covalent enzyme-substrate intermediates, general acid/base catalysis, metal coordination or low-barrier hydrogen bonds. It follows from Table I that all designs leave room for significant improvement. To accomplish this, enzyme designers must be prepared to envision more complex reactions.

Another pertinent aspect is that enzymes earn much of their proficiency from catalyzing reactions that are astoundingly slow in water,[27,52] and for which catalysis stabilize large charge reorganizations in the TS.[84] Several of the reactions discussed herein are comparatively fast in solution, and thus limit the maximum

rate enhancement. To significantly improve Diels–Alder catalysis, for example, it is perhaps necessary to both find a slower background reaction and re-route the catalytic mechanism. We recently presented an acid/base-mediated mechanism that utilized the catalytic machinery of ketosteroid isomerase.[85] This preliminary study indicated a dramatic rate enhancement of a reaction that is very slow in solution, provided that the substrates could bind to the active site and form pre-arranged conformations.

Enzyme design is evidently a complex, non-linear process and requires more than an ever-so-elegant algorithm. More advanced, diverse and cheap design tools, both computational and experimental, become available every year. The literature discussed in this review testifies that if a systematic application of the entire toolbox is conducted, dramatically improved results will definitely ensue.

Acknowledgements

This work has been supported by the Cambridge Crystallographic Data Centre (CCDC). The author acknowledges Prof. Tore Brinck for his support and encouragement.

Citation

Linder M (2012) Computational Enzyme Design: Advances, hurdles and possible ways forward. *Computational and Structural Biotechnology Journal*. 2 (3): e201209009.
doi: <http://dx.doi.org/10.5936/csbj.201209009>

References

- Silverman RB (2002) *The Organic Chemistry of Enzyme-Catalyzed Reactions*. Academic Press, London
- Bornscheuer UT, Huisman GW, Kazlauskas RJ, Lutz S, Moore JC, Robins K (2012) Engineering the third wave of biocatalysis. *Nature* 485:185–194
- Hellinga HW, Richards FM (1991) Construction of new ligand binding sites in proteins of known structure: I. computer-aided modeling of sites with pre-defined geometry. *J Mol Biol* 222:763–785
- Dahiyat BI, Mayo SL (1996) Protein design automation. *Protein Sci* 5:895–903
- Dahiyat BI, Mayo SL (1997) De novo protein design: Fully automated sequence selection. *Science* 278:82–87
- Voigt CA, Mayo SL, Arnold FH, Wang ZG (2001) Computational method to reduce the search space for directed protein evolution. *Proc Natl Acad Sci* 98:3778–3783
- Looger LL, Hellinga HW (2001) Generalized dead-end elimination algorithms make large-scale protein side-chain structure prediction tractable: implications for protein design and structural genomics. *J Mol Biol* 307:429–445
- Looger LL, Dwyer MA, Smith JJ, Hellinga HW (2003) Computational design of receptor and sensor proteins with novel functions. *Nature* 423:185–190
- Rohl CA, Strauss CE, Misura KM, Baker D (2004) Protein structure prediction using Rosetta. In: Brand L, Johnson ML (eds) *Numerical Computer Methods, Part D, Methods Enzymol* 383:66–93
- Zanghellini A, Jiang L, Wollacott AM, Cheng G, Meiler J, Althoff EA, Röthlisberger D, Baker D (2006) New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci* 15:2785–2794
- Lassila JK, Privett HK, Allen BD, Mayo SL (2006) Combinatorial methods for small-molecule placement in computational enzyme design. *Proc Natl Acad Sci* 103:16710–16715
- Allen BD, Nisthal A, Mayo SL (2010) Experimental library screening demonstrates the successful application of computational protein design to large structural ensembles. *Proc Natl Acad Sci* 107:19838–19843
- Leaver-Fay A, Jacak R, Stranges PB, Kuhlman B (2011) A generic program for multistate protein design. *PLoS ONE* 6:e20,937
- Huang PS, Ban YEA, Richter F, Andre I, Vernon R, Schief WR, Baker D (2011) Rosetta re-model: A generalized framework for flexible backbone protein design. *PLoS ONE* 6:e24,109
- rosetta@home. <http://boinc.bakerlab.org/rosetta/>
- Cooper S, Khatib F, Treuille A, Barbero J, Lee J, Beenen M, Leaver-Fay A, Baker D, Popovic Z, players F (2010) Predicting protein structures with a multiplayer online game. *Nature* 466:756–760
- Khatib F, DiMaio F, Cooper S, Kazmierczyk M, Gilski M, Krzywdka S, Zabranska H, Pichova I, Thompson J, Popovic Z, Jaskolski M, Baker D (2011) Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nat Struct Mol Biol* 18:1175–1177
- Kaplan J, DeGrado WF (2004) De novo design of catalytic proteins. *Proc Natl Acad Sci* 101:11566–11570
- Jiang L, Althoff EA, Clemente FR, Doyle L, Röthlisberger D, Zanghellini A, Gallaher JL, Betker JL, Tanaka F, Barbas CF, Hilvert D, Houk KN, Stoddard BL, Baker D (2008) De Novo Computational Design of Retro-Aldol Enzymes. *Science* 319:1387–1391
- Röthlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, Gallaher JL, Althoff EA, Zanghellini A, Dym O, Albeck S, Houk KN, Tawfik DS, Baker D (2008) Kemp elimination catalysts by computational enzyme design. *Nature* 453:190–195
- Siegel JB, Zanghellini A, Lovick HM, Kiss G, Lambert AR, StClair JL, Gallaher JL, Hilvert D, Gelb MH, Stoddard BL, Houk KN, Michael FE, Baker D (2010) Computational Design of an Enzyme Catalyst for a Stereoselective Bimolecular Diels–Alder Reaction. *Science* 329:309–313
- Baker D (2010) An exciting but challenging road ahead for computational enzyme design. *Protein Sci* 19:1817–1819
- Gerlt JA, Babbitt PC (2009) Enzyme (re)design: lessons from natural evolution and computation. *Curr Opin Chem Biol* 13:10–18
- Hilvert D (2000) Critical analysis of antibody catalysis. *Annu Rev Biochem* 69:751–793
- Schreier B, Stumpp C, Wiesner S, Höcker B (2009) Computational design of ligand binding is not a solved problem. *Proc Natl Acad Sci* 106:18491–18496
- Morin A, Kaufmann KW, Fortenberry C, Harp JM, Mizoue LS, Meiler J (2011) Computational design of an endo-1,4- β -xylanase ligand binding site. *Protein Eng Des Selec* 24:503–516
- Wolfenden R, Snider MJ (2001) The depth of chemical time and the power of enzymes as catalysts. *AccChem Res* 34:938–945
- Garcia-Viloca M, Gao J, Karplus M, Truhlar DG (2004) How enzymes work: Analysis by modern rate theory and computer simulations. *Science* 303:186–195
- Zhang X, Houk KN (2005) Why enzymes are proficient catalysts: beyond the Pauling paradigm. *Acc Chem Res* 38:379–385
- Hammes GG, Benkovic SJ, Hammes-Schiffer S (2011) Flexibility, diversity, and cooperativity: Pillars of enzyme catalysis. *Biochemistry* 50:10,422–10,430
- Warshel A (1998) Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. *J Biol Chem* 273:27,035–27,038
- Lassila JK (2010) Conformational diversity and computational

- enzyme design. *Curr Opin Chem Biol* 14:676–682
33. Lutz S (2010) Beyond directed evolution: semi-rational protein engineering and design. *Curr Opin Biotechnol* 21:734–743
 34. Saven JG (2011) Computational protein design: engineering molecular diversity, nonnatural enzymes, nonbiological cofactor complexes, and membrane proteins. *Curr Opin Chem Biol* 15:452–457
 35. Pantazes RJ, Grisewood MJ, Maranas CD (2011) Recent advances in computational protein design. *Curr Opin Struct Biol* 21:467–472
 36. Dahiyat BI, Benjamin Gordon D, Mayo SL (1997) Automated design of the surface positions of protein helices. *Protein Sci* 6:1333–1337
 37. Benson DE, Wisz MS, Hellinga HW (1998) The development of new biotechnologies using metalloprotein design. *Curr Opin Biotechnol* 9:370–376
 38. Bolon DN, Voigt CA, Mayo SL (2002) De novo design of biocatalysts. *Curr Opin Chem Biol* 6:125–129
 39. Bolon DN, Mayo SL (2001) Enzyme-like proteins by computational design. *Proc Natl Acad Sci* 98:14,274–14,279
 40. Dwyer MA, Looger LL, Hellinga HW (2004) Computational design of a biologically active enzyme. *Science* 304:1967–1971
 41. Hayden EC (2009) Key protein-design papers challenged. *Nature* 461:859
 42. Richter F, Leaver-Fay A, Khare SD, Bjelic S, Baker D (2011) De novo enzyme design using rosetta3. *PLoS ONE* 6:e19,230
 43. Kuhlman B, Baker D (2000) Native protein sequences are close to optimal for their structures. *Proc Natl Acad Sci* 97:10383–10388
 44. Tantillo DJ, Jiangang C, Houk KN (1998) Theozymes and compuzymes: theoretical models for biological catalysis. *Curr Opin Chem Biol* 2:743–750
 45. Frushicheva MP, Cao J, Warshel A (2011) Challenges and advances in validating enzyme design proposals: The case of kemp eliminase catalysis. *Biochemistry* 50:3849–3858
 46. Khersonsky O, Röthlisberger D, Wollacott AM, Murphy P, Dym O, Albeck S, Kiss G, Houk K, Baker D, Tawfik DS (2011) Optimization of the in-silico-designed kemp eliminase ke70 by computational design and directed evolution. *J Mol Biol* 407:391–412
 47. Pauling L (1946) Molecular architecture and biological reactions. *Chem Eng News* 24:1375–1377
 48. Pauling L (1948) The nature of forces between large molecules of biological interest. *Nature* 161:707–709
 49. Truhlar DG, Garrett BC, Klippenstein SJ (1996) Current status of transition-state theory. *J Phys Chem* 100:12771–12800
 50. Schowen RL (1978) In: Gandour RD, Schowen RL (eds) *Transition States of Biochemical Processes*, Plenum, New York, p 77
 51. Radzicka A, Wolfenden R (1995) A proficient enzyme. *Science* 267:90–93
 52. Cannon WR, Benkovic SJ (1998) Solvation, reorganization energy, and biological catalysis. *J Biol Chem* 273:26257–26260
 53. Bruice TC (2002) A view at the millennium: the efficiency of enzymatic catalysis. *Acc Chem Res* 35:139–148
 54. Benkovic SJ, Hammes-Schiffer S (2003) A perspective on enzyme catalysis. *Science* 301:1196–1202
 55. Warshel A (1978) Energetics of enzyme catalysis. *Proc Natl Acad Sci* 75:5250–5254
 56. Page MI, Jencks WP (1971) Entropic contributions to rate accelerations in enzymic and intramolecular reactions and the chelate effect. *Proc Natl Acad Sci* 68:1678–1683
 57. Dewar MJ, Storch DM (1985) Alternative view of enzyme reactions. *Proc Natl Acad Sci* 82:2225–2229
 58. Kraut DA, Sigala PA, Pybus B, Liu CW, Ringe D, Petsko GA, Herschlag D (2006) Testing electrostatic complementarity in enzyme catalysis: Hydrogen bonding in the ketosteroid isomerase oxyanion hole. *PLoS Biol* 4:0501–0519
 59. Adamczyk AJ, Cao J, Kamerlin SCL, Warshel A (2011) Catalysis by dihydrofolate reductase and other enzymes arises from electrostatic preorganization, not conformational motions. *Proc Natl Acad Sci* 108:14115–14120
 60. Hur S, Bruice TC (2003) The near attack conformation approach to the study of the chorismate to prephenate reaction. *Proc Natl Acad Sci* 100:12015–12020
 61. Štrajbl M, Shurki A, Kato M, Warshel A (2003) Apparent nac effect in chorismate mutase reflects electrostatic transition state stabilization. *J Am Chem Soc* 125:10228–10237
 62. Toro-Labbe A, Gutierrez-Oliva S, Murray J, Politzer P (2007) A new perspective on chemical and physical processes: the reaction force. *Mol Phys* 105:2619–2625
 63. Schowen RL (2003) How an enzyme surmounts the activation energy barrier. *Proc Natl Acad Sci* 100:11931–11932
 64. Ruscio JZ, Kohn JE, Ball KA, Head-Gordon T (2009) The influence of protein dynamics on the success of computational enzyme design. *J Am Chem Soc* 131:14111–14115
 65. Lassila JK, Baker D, Herschlag D (2010) Origins of catalysis by computationally designed retroaldolase enzymes. *Proc Natl Acad Sci* 107:4937–4942
 66. Frushicheva MP, Cao J, Chu ZT, Warshel A (2010) Exploring challenges in rational enzyme design by simulating the catalysis in artificial kemp eliminase. *Proc Natl Acad Sci* 107:16869–16874
 67. Kiss G, Röthlisberger D, Baker D, Houk KN (2010) Evaluation and ranking of enzyme designs. *Protein Sci* 19:1760–1773
 68. Simón L, Goodman JM (2010) Enzyme catalysis by hydrogen bonds: The balance between transition state binding and substrate binding in oxyanion holes. *J Org Chem* 75:1831–1840
 69. Pohnert G (2001) Diels-Alderase. *ChemBioChem* 2:873–875
 70. Kelly WL (2008) Intramolecular cyclizations of polyketide biosynthesis: mining for a "Diels-Alderase"? *Org Biomol Chem* 6:4483–4493
 71. Guimaraes C, Udier-Blagovic M, Jorgensen W (2005) Macrophomate synthase: QM/MM simulations address the Diels-Alder versus Michael-aldol reaction mechanism. *J Am Chem Soc* 127:3577–3588
 72. Linder M, Hermansson A, Liebeschuetz J, Brinck T (2011) Computational design of a lipase for catalysis of the Diels-Alder reaction. *J Mol Model* 17:833–849
 73. Hult K, Berglund P (2007) Enzyme promiscuity: mechanism and applications. *Trends Biotechnol* 25:231–238
 74. Gerlt JA, Babbitt PC (1998) Mechanistically diverse enzyme superfamilies: the importance of chemistry in the evolution of catalysis. *Curr Opin Chem Biol* 2:607–612
 75. O'Brien PJ, Herschlag D (1999) Catalytic promiscuity and the evolution of new enzymatic activities. *Chem Biol* 6:R91–R105
 76. Linder M, Johansson AJ, Olsson TSG, Liebeschuetz J, Brinck T (2011) Designing a new Diels-Alderase: A combinatorial, semirational approach including dynamic optimization. *J Chem Inf Model* 51:1906–1917
 77. Linder M, Johansson AJ, Olsson TSG, Liebeschuetz J, Brinck T (2012) Computational design of a Diels-Alderase from a thermophilic esterase: the importance of dynamics. *J Comput-Aided Mol Des* 26:1079–1095
 78. Nosrati GR, Houk KN (2012) Saber: A computational method for identifying active sites for new reactions. *Protein Sci* 21:697–706
 79. Privett HK, Kiss G, Lee TM, Blomberg R, Chica RA, Thomas LM, Hilvert D, Houk KN, Mayo SL (2012) Iterative approach to

- computational enzyme design. Proc Natl Acad Sci 109:3790–3795
80. Roca M, Vardi-Kilshtain A, Warshel A (2009) Toward accurate screening in computer-aided enzyme design. Biochemistry 48:3046–3056
81. Bruice TC, Lightstone FC (1999) Ground state and transition state contributions to the rates of intramolecular and enzymatic reactions. Acc Chem Res 32:127–136

Competing Interests:

The authors have declared that no competing interests exist.

Received: 31 July 2012

Received in revised form: 30 September 2012

Accepted: 12 October 2012



© 2012 Linder.

Licensee: Computational and Structural Biotechnology Journal.

This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are properly cited.

82. Siegbahn PEM, Himo F (2009) Recent developments of the quantum chemical cluster approach for modeling enzyme reactions. J Biol Inorg Chem 14:643–651
83. Eiben CB, Siegel JB, Bale JB, Cooper S, Khatib F, Shen BW, Players F, Stoddard BL, Popovic Z, Baker D (2012) Increased diels-alderase activity through backbone remodeling guided by foldit players. Nat Biotechnol 30:190–192
84. Warshel A, Sharma PK, Kato M, Xiang Y, Liu H, Olsson MHM (2006) Electrostatic basis for enzyme catalysis. Chemical Reviews 106:3210–3235
85. Linder M, Johansson AJ, Manta B, Olsson P, Brinck T (2012) Envisioning an enzymatic Diels-Alder reaction by in situ acidbase catalyzed diene generation. Chem Comm 48:5665–5667

What is the advantage to you of publishing in *Computational and Structural Biotechnology Journal (CSBJ)* ?

- ✚ Easy 5 step online submission system & online manuscript tracking
- ✚ Fastest turnaround time with thorough peer review
- ✚ Inclusion in scholarly databases
- ✚ Low Article Processing Charges
- ✚ Author Copyright
- ✚ Open access, available to anyone in the world to download for free

WWW.CSBJ.ORG