

Designing a New Diels–Alderase: A Combinatorial, Semirational Approach Including Dynamic Optimization.

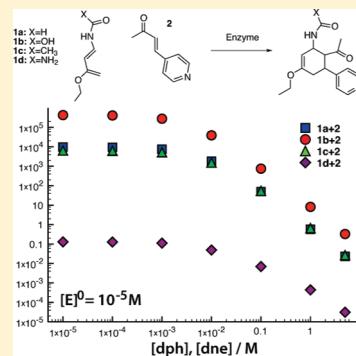
Mats Linder,[†] Adam Johannes Johansson,[†] Tjelvar S. G. Olsson,[‡] John Liebeschuetz,[‡] and Tore Brinck*,[†]

[†]Physical Chemistry, Royal Institute of Technology (KTH), Teknikringen 30, S-100 44 Stockholm, Sweden

[‡]Cambridge Crystallographic Data Centre, Cambridge, U.K.

Supporting Information

ABSTRACT: A computationally inexpensive design strategy involving ‘semirational’ screening for enzymatic catalysis is presented. The protocol is based on well-established computational methods and represents a holistic approach to the catalytic process. The model reaction studied here is the Diels–Alder, for which a successful computational design has recently been published (Siegel, J. B. et al. *Science* **2010**, *329*, 309–313). While it is a leap forward in the field of computational design, the focus on designing only a small fraction of the active site gives little control over dynamics. Our approach explicitly incorporates mutagenesis and the analysis of binding events and transition states, and a promising enzyme–substrate candidate is generated with relatively little effort. We estimate catalytic rate accelerations of up to 10^5 .



INTRODUCTION

Enzymatic catalysis of the Diels–Alder reaction, a [4 + 2] pericyclic reaction, has been an enigmatic and captivating problem ever since the emergence of modern enzyme engineering. Not only because enzyme-catalyzed Diels–Alder is extraordinarily rare in nature¹ but also because of the synthetic utility of creating two carbon–carbon bonds with up to 4 stereocentra in one step.² This makes the creation of an efficient and selective Diels–Alderase something of a holy grail in enzymatic catalysis.

Several approaches toward designing a Diels–Alderases have been employed. The first successful method was to develop catalytic antibodies around transition state-mimicking haptens.^{3–5} The best catalytic antibody 1E9⁴ is to date still the best-performing engineered Diels–Alder biocatalyst.^{3,6}

The concept of catalytic promiscuity⁷ has frequently been invoked as a means of rationally (re)designing enzymes to access new reactions. Two main approaches have been employed (sometimes in concert): rational design and directed evolution. Despite several fruitful efforts,^{8,9} the most successful attempts show results far from the impressive rate accelerations displayed by the most proficient enzymes. While rational design, utilizing a small number of point mutations, can be directly related to our understanding of the enzymatic mechanism, it often fails to encompass the subtleties of the overall function of the enzyme. Directed evolution has the power to improve existing mechanisms, but there is no obvious rationale for introducing new ones. Moreover, variants improved by directed evolution often have important mutations far from the active site and thus the improved performance becomes difficult to understand. In either case, a thorough understanding of all the subtle interactions leading to the efficient catalysis remains elusive.

Recently, the groups of David Baker, Kendall N. Houk, and Donald Hilvert have developed a design protocol in which a quantum chemically designed active site is incorporated into a protein scaffold using the Rosetta methodology.^{10,11} Expressing and assaying the best candidates, followed by several generations of directed evolution, has so far created designed enzymes for the Kemp elimination,¹² retro-alcohol reaction,¹³ and, most recently, a Diels–Alder reaction.¹⁴ The advantage of their protocol is that it efficiently screens the bulk of all possible mutants *in silico*, capturing putative catalysts according to some design template. Subsequently, directed evolution refines the computationally designed enzyme.

The results mentioned above show the potential of computational and experimental methods working together. However, some issues remain. It has been argued that the designed retro-alcohol enzyme¹³ does not actually stabilize the transition state (TS) to the extent they were designed for, nor do they display stereoselectivity associated with chiral active sites.^{15,16} A recent study similarly suggests that TS stabilization in Kemp eliminase¹² is much lower than that predicted in the design.¹⁷ Furthermore, the computationally designed catalytic machinery in the Diels–Alderase was predicted to have a TS stabilization ($\Delta\Delta G_{cat}^\ddagger$) of almost 5 kcal/mol,¹⁴ corresponding to a rate constant ratio (k_{cat}/k_{uncat}) of $\approx 5 \cdot 10^3$, but the experimental k_{cat}/k_{uncat} were at best $\approx 10^2$.

The above points show that further investigation of different design protocols is necessary in order for man-made enzymes to truly compete with the fine-tuned activity of wild types. One area

Received: April 19, 2011

Published: July 22, 2011

that current methodologies fail to address, in particular, is the dynamic nature of proteins.¹⁵

Given the growing wealth of structural information in the protein data bank (PDB),¹⁸ a knowledge-based approach to designing a Diels–Alderase seems increasingly viable. By combining molecular docking, molecular dynamics (MD), and quantum chemical (QC) methods, we have previously explored the effect of point mutations computationally.¹⁹ Although no experimental validation was carried out, our approach enabled the evaluation of thermodynamically relevant quantities and a prediction of the total rate enhancement as a function of concentrations. Computational design is emerging as an efficient method to handle the first phases of enzyme engineering, and our reported protocol provides a means to investigate the reaction pathway in detail. However, it is not a high-throughput method, and some screening method, such as RosettaMatch¹⁰ or a knowledge-based search, is required upstream to find a candidate enzyme in the first place.

Herein we present an alternative computational screening method which optimizes both substrates and enzyme concurrently. The design is based on varying substituents on a generic transition state scaffold in order to find matches with a particular enzyme. Using a combinatorial approach, protein variants with potentially high activities can be designed with relatively little effort. These variants can then be further optimized through rational design. We view this approach as an intermediate and necessary step toward the ultimate goal which is a methodology for the *in silico* design of a catalytically efficient Diels–Alderase designed to generate a predetermined product of commercial value. To achieve that goal we need first to develop computational methodologies that can most easily be validated by experiment. Our ‘semirational’ approach satisfies that criterion as it aims to alter the activity of already experimentally well studied enzyme targets with a limited number of mutational modifications.

The protocol outlined here builds on our previous work¹⁹ and employs molecular docking, molecular dynamics, and density functional theory (DFT). Our philosophy is to take a holistic approach to the design problem, where all major parameters are studied and optimized. As a reference, we investigate the Diels–Alderase design DA_20_00 from Siegel et al.¹⁴ using the same protocol. The aim of our research is to find reasonable candidate protein and substrate combinations for which the transition state protein–substrate interactions are optimized, and the calculated catalytic efficiency is expected to be high. These may then be candidates for experimental studies with a hope of finding catalytic efficiencies comparable to those observed in nature. If that is shown to be the case such artificial enzymes would provide excellent foundations for studies to broaden or change substrate specificity that might lead to commercially valuable enzymes.

This rest of this paper is organized as follows: In the ‘Theory and Methods’ section, the design strategy is first explained in general terms, followed by the particular methods used at each stage. The thermodynamics of our two-substrate enzyme-catalyzed reaction is then discussed along with methods used to calculate the thermodynamic parameters central to the design. In the ‘Results and Discussion’ section, results for the system used to test our concepts are presented and discussed in light of the methodology employed. The overall results are discussed and compared to the DA_20_00 system, which is also compared to experimental data.¹⁴ We conclude with an outlook and suggestions for improvement of the design strategy.

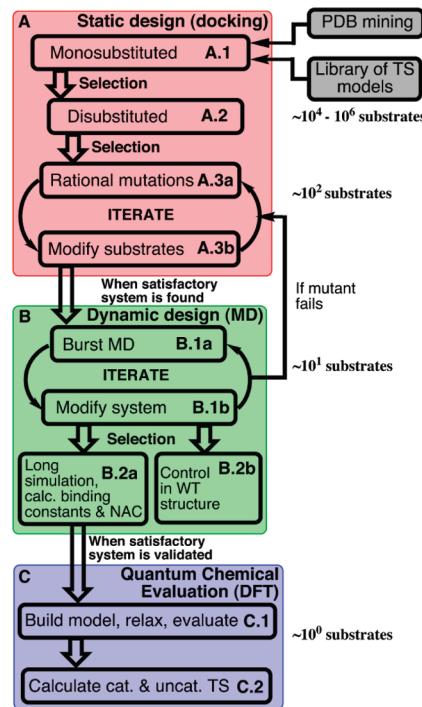


Figure 1. Flowchart representation of the semirational, combinatorial design protocol. Stages A to C are divided into two or more substeps (A.1, B.2, etc.) which are referred to in the text when needed. Numbers on the right-hand side indicate how many substrate structures are processed in each step.

THEORY AND METHODS

Design Strategy. The strategy outlined here can be described as consisting of 3 stages (A–C), each with several internal steps. They are illustrated in the flowchart given in Figure 1, and a general description is given in this section. Details of the methods used in this project are given in the ‘Methods and Software’ subsection. Note that an inherent property of our approach is its flexibility and that moving upstream may be necessary at any one point if one reaches a dead end. This is illustrated by the iteration cycles in Figure 1.

Stage A: Screening and Static Design. When a design project is initiated, the PDB is mined for relevant structures. The design philosophy is to search for enzymes possessing a given property that could be of use for catalyzing the reaction of interest. This can reduce the number of mutations needed to create an active variant. Ideally, one can retain the whole catalytic machinery and only make careful structural modifications. However, for a reaction such as Diels–Alder, this is hardly possible. It is thus a good example of a general case in which parts of a native catalytic machinery needs to be removed or altered to create the desired activity.

The structural data in the PDB can be screened using an appropriate method of choice, until a sufficiently low number of candidates can be selected. We stress that visual inspection of the active sites is an important factor to weigh in the final evaluation of which structures to choose for further investigation. The selected enzymes are subject to ‘virtual screening’ of a library consisting of one or more transition state (TS) scaffolds with a number of substituents, using molecular docking. Docking of a TS model, as opposed to docking the two substrates, serves two

purposes. First, current docking methodologies make it impractical to dock more than one substrate at a time. Currently multicomponent docking has to be carried out in a sequential fashion which makes it more difficult to sample conformational space efficiently. Second, since the catalyst is supposed to stabilize the TS *more* than the bound substrates, it makes intuitive sense to try and drive the selection process toward interactions that are optimized for the TS model.

The library of TS models can of course be made in various ways. We have used the approach of attaching a set of substituents to quantum chemically optimized TS scaffolds. For this reason, we refer to our set of TS models as a ‘Combinatorial library’. As shown in Figure 1, two steps of screening are taken in our particular design (steps A.1 and A.2), first using a monosubstituted subset of the combinatorial library and then a filtered disubstituted version. The main reason for this is to save computational time and retain some control of the process, as discussed in more detail below.

After the initial steps of screening, rational design can be performed on both the enzyme and TS model (step A.3). Whereas the main objective of Steps A.1 and A.2 was to find a reasonably good geometric fit between the active site and the reacting substrates, in step A.3 the main objective is to alter the function of the enzyme. Step A.3 is also the stage where new catalytic interactions can be introduced. This is the most time-consuming part of the static design; in our experience it takes 5–10 iterations before a system is ready for the dynamic stage (B in Figure 1). Note that we do not intend for there to be a specific protocol for modifying the system. In our attempts of designing new Diels–Alderases, mutations and substrate modifications have been performed as needed, following no specific order. The key to converging toward one or a small set of candidates is to have adequate tools for measuring success, which can be challenging when using molecular docking (see the Supporting Information for more details).

Stage B: Dynamic Design and Evaluation. Taking the design system into the dynamic realm is a first test to validate the design. Even if long simulations are normally required to validate a successful design, we utilize the fact that false predictions from the static stage are typically detected within a few nanoseconds. Hence, in the first step at this stage (B.1), we convert the lead substrate pair(s) from static TS models to individual molecules and produce ‘burst’ MD trajectories. These are then analyzed to see if the predicted poses and interactions are retained to a satisfactory extent. If not, one can choose to go back to the static stage and remodify the system. However, since the dynamic behavior of the substrates in the active site is unfeasible to design for using molecular docking, alterations can be introduced directly in step B.1, as will be described below. In principle one can modify both enzyme and substrates during step B.1 m although we in this work focus more on substrate modification after the static stage.

When a system has passed the burst MD test, it is subject to longer simulations to produce trajectories of several nanoseconds (step B.2a). From these simulations, we calculate key thermodynamic parameters as described in the ‘Thermodynamics’ section. Key interactions are statistically analyzed using the ptraj tool within AMBER. In addition, the lead substrates are combined with the unmodified wild-type enzyme, and corresponding simulations are carried out to measure how productive the introduced mutations are (step B.2b).

Stage C: Quantum Chemical Evaluation. Optimizing the transition state of the reaction in an active site model is necessary to calculate the central rate constant (see eq 1). It is equally important to evaluate whether the enzyme–substrate interactions designed for are present in the transition state, since this is the one point on the reaction coordinate that needs to be stabilized the most for efficient catalysis. The result will always depend somewhat on the model, which is built by taking the core of the active site from the enzyme structure.

Depending on the computational resources at hand, models consisting of 100 to ~200 atoms should be feasible. Note that herein lies an important difference between our approach and the one taken by the Baker group.^{12–14} We build the quantum chemical model based on an existing active site with introduced catalytic moieties, while they ‘fit’ an enzyme to a predefined catalytic machinery.

METHODS AND SOFTWARE

PDB Mining. We have earlier focused on oxyanion holes as a primary catalytic functionality,^{19,20} so the present study was limited to a set of such proteins, studied by Simón and Goodman.²¹ This data set contains 310 structures. Using the protein–ligand search tools in Relibase+,²² the data set was reduced by removing duplicates and filtering out the most high-resolution entries. The CavBase tool^{23,24} was then used to visually inspect and classify active sites.

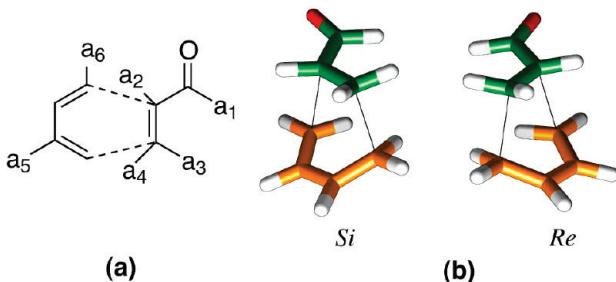
Among several interesting structures, the active site of PDB entry 1EB9²⁵ caught our attention. It is a W128A mutant of *Manihot esculenta* (cassava) hydroxynitrile lyase (HNL). W128 covers the entrance to the active site, and the authors sought to investigate its role in substrate-specificity. To their surprise, the mutant crystal structure bound two 4-hydroxybenzaldehyde molecules in a sandwich-type conformation, much resembling the geometry of a Diels–Alder transition state.²⁵

The HNL mutant W128A is thus a promising structure on which to test our combinatorial strategy and will be the system used to illustrate our concepts throughout the rest of this paper.

Substituent Library. A combinatorial library was generated by adding substituents to a B3LYP/6-31G(d)-optimized TS scaffold for the reaction of butadiene and acrolein, as shown in Scheme 1a. The normally unfavored²⁶ *s-trans* geometry was selected since our previous investigations have shown that oxyanion holes are often buried with bulky residues around,¹⁹ and a *s-trans* double bond on the dienophile is more exposed than a *s-cis* bond. Since the stereochemical outcome depends on the face of attack, both the *Re* and *Si* TSs were considered (Scheme 1b). Six attachment points were defined, four on the dienophile and two on the diene, since activating (or deactivating) groups on the diene are most significant in positions *ortho* and *para* to the dienophile electron-withdrawing group.²⁰

A library of 40 common substituents with varied properties was generated using the MOE suit.²⁷ A detailed presentation of this selection is given in the Supporting Information (SI). Although 40 is a small number for building up a virtual library, it is possible to generate $>8 \cdot 10^9$ substrates with this set (2×40^6), which leads computationally demanding screening and contains a large portion of unreasonable structures. Instead, we chose to limit substitution to a maximum of two substituents, one on the dienophile (a_1 – a_4) and one on the diene (a_5 – a_6). In this way, the full combinatorial set is limited to 25600 entries. These were generated using the QSAR Combinatorial tool in MOE.²⁸

Scheme 1. (a) Schematic Representation of the TS Scaffold with Attachment Points a_1 - a_6 ^a



^a During the first step of screening (A.1 in Figure 1), only a_1 - a_4 were considered, while in step A.2 we considered combinations with one group on one of a_1 - a_4 and one on either a_5 or a_6 . (b) The *Re*-TS and *Si*-TS scaffolds, optimized at the B3LYP/6-31G(d) level of theory.

Note, however, that the substituent library can be made arbitrarily large, based on the required amount of flexibility and the computational resources at hand.

Docking Protocol (Stage A). The docking was conducted with GOLD Suite 5.0²⁹ in two steps (A.1 and A.2). First a set of 320 monosubstituted substrates were docked (attachment points a_1 - a_4), and a selection of disubstituted TSs was made based on the results and processed in the same way. The results were analyzed based on the anticipated interaction with the enzyme's oxyanion hole and the presence of other potentially catalytic interactions as well as the scoring function. In all cases, an individual rescoring function was applied to filter out bad poses.

During the rational design step (A.3) substrates were modified using standardized functional groups. Mutations were introduced by replacing whole side chains and giving the new groups some rotational freedom. A more extensive description of the docking procedure is provided in the Supporting Information.

Molecular Dynamics (Stage B). MD simulations were undertaken using the AMBER 10³⁰ package and the ff03 force field,³¹ and otherwise according to a previously described protocol.¹⁹ The general Amber force field (GAFF)³² was used for all substrates, and initial ligand positions were taken from suitable docking poses. Periodic boundary conditions were used in all simulations. A truncated octahedral box of TIP3P³³ water, with an 8 Å shell outside the enzyme, was used for solvation.

Systems selected from the burst simulations (B.1) were subject to 4 ns of unconstrained simulations. For each substrate pair, simulations containing only the diene, only the dienophile, and both were performed. Apart from evaluating the consistency of the interactions when one or both substrates are bound, we used these three systems to independently calculate the two binding constants and the NAC-forming penalty to the free energy of activation (see the 'Thermodynamics' section below). In addition, the conformational stability of the lead mutant was tested by a 10 ns simulation of the uncomplexed enzyme compared to the corresponding simulation of the parent 1EB9 structure.

Quantum Chemical Calculations (Stage C). When an active site model had been selected, it was truncated to limit the number of atoms to the active parts of each residue while still retaining some flexibility of the groups. The cavity shape was retained by fixing the terminal carbon and one hydrogen on each truncated residue. All optimizations were performed using density functional theory (DFT) at the B3LYP/6-31+G(d) level of theory,³⁴

using Jaguar 7.6.³⁵ Activation energies were calculated relative to a reactant complex ('NAC', see below) and compared most to the favorable uncatalyzed channel. Thermodynamic corrections were added from frequency analyses. Single-point energies were then obtained with Gaussian09³⁶ at the M062X/6-311+ +G(d,p) level³⁷ with a SMD-PCM³⁸ solvent correction, using $\epsilon = 4$ for the active site model, and $\epsilon = 78.39$ for the uncatalyzed reaction. We selected the M062X functional since it has been shown to perform well for Diels–Alder reactions, in line with spin-component scaled MP2.³⁹

Graphical Representation. All molecular representations were created using the UCSF Chimera package.^{40,41} Diagrams were created using Plot.⁴²

Thermodynamics. Here follows a brief discussion of the thermodynamic parameters that we wanted to optimize during the design procedure. A more detailed discussion is given elsewhere.¹⁹ The uncatalyzed Diels–Alder reaction is of second order, while an enzyme-catalyzed pathway can be said to follow extended Michaelis–Menten kinetics.⁴³ The relative catalytic rate is then given by eq 1^{3,44,45}

$$\frac{v_{cat}}{v_{uncat}} = \frac{k_{cat}}{k_{uncat}} \frac{[E^0]}{K_{M1}K_{M2} + K_{M1}[dne] + K_{M2}[dph] + [dne][dph]} \quad (1)$$

Here, k_{cat} and k_{uncat} are the rate constants for the enzyme-catalyzed and uncatalyzed reactions, respectively. K_{M1} and K_{M2} are the Michaelis constants for the first and second substrates, in this case the dienophile and diene. Concentrations are given as square-bracketed variables; $[E^0]$ is the total enzyme concentration, while $[dph]$ and $[dne]$ represent dienophile and diene, respectively.

In order for a quantitative prediction of catalytic activity to be made, K_{M1} and K_{M2} , k_{uncat} and k_{cat} need to be calculated. K_{M1} and K_{M2} are the Michaelis constants and can, to first order, be approximated with $\exp(\Delta G_{bind}/RT)$. The K_M constant expressed in the rate coefficients in Michaelis–Menten kinetics is

$$K_M = \frac{[E][S]}{[ES]} = \frac{k_{-1} + k_{cat}}{k_1} \quad (2)$$

and since k_{cat} is normally several orders of magnitude smaller than k_{-1} the approximation holds. It is readily validated once k_{cat} has been determined.

Binding energies are calculated from MD simulations using the Linear Interaction Energy (LIE) method.⁴⁶ In this method the free energy of binding is estimated from the differences in electrostatic and nonpolar (van der Waals) interaction energies between the protein-bound and solvated ligands, using a parameterized relationship

$$\begin{aligned} \Delta G_{bind} = & \alpha [\langle V_{1-s}^{elec} \rangle_{bound} - \langle V_{1-s}^{elec} \rangle_{free}] \\ & + \beta [\langle V_{1-s}^{vdw} \rangle_{bound} - \langle V_{1-s}^{vdw} \rangle_{free}] + \gamma \end{aligned} \quad (3)$$

For uncharged molecules, α and β have been determined to 0.43 and 0.18, respectively, while the parameter γ has been shown to depend on the hydrophobicity of the binding pocket.⁴⁷ We have found for a variety of hydrolytic enzymes and ligands that accurate binding energies can be obtained by estimating γ from the nonpolar solvation energy contribution (ΔG_{SA}) to ligand binding as calculated in the MM-GBSA routine of AMBER. So far, the ΔG_{SA} -corrected LIE energies have given binding

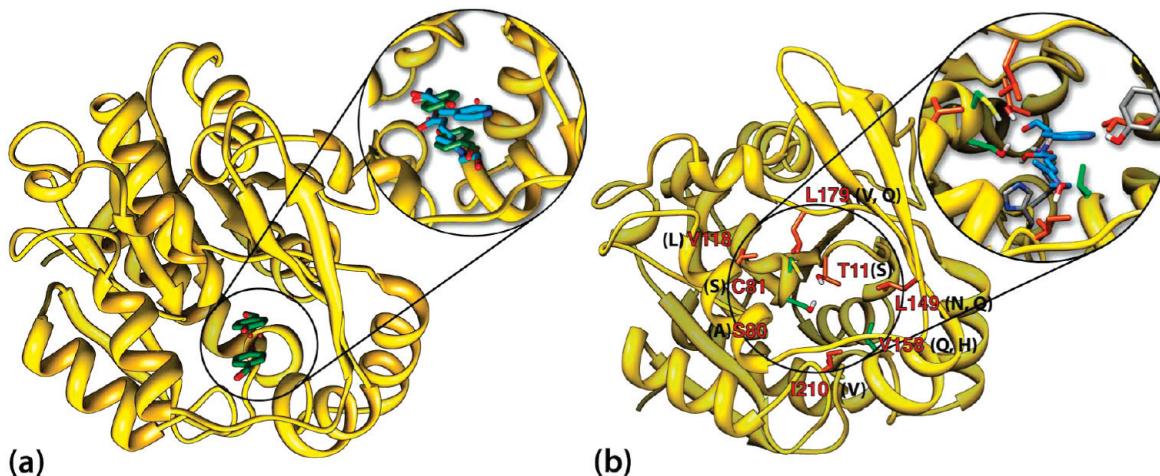


Figure 2. (a) Crystal structure of the W128A mutant of *M. esculenta*, PDB entry 1EB9. Two 4-hydroxybenzaldehyde molecules are seen stacked in the active site, together with a water molecule. The inset shows a closeup of the active site with an overlay of the two 4-hydroxybenzaldehydes and an early combinatorial design (*cf.* A.3 in Scheme 2). (b) Amino acids mutated in this study. Residues are labeled red, and investigated mutants are given in black letters. Mutated amino acids that appear in the lead design HNL-3 are colored green, and the remaining ones are orange. The inset shows the amino acids in relation to an early substrate design.

constants deviating less than a factor 10¹ from experimental data.⁴⁸

The free energy of activation for the bond forming step ($\Delta G_{cat}^{\ddagger}$) has to be calculated by some quantum chemical method, in our case DFT. The reaction coordinate can thus be divided into one classical and one quantum mechanical regime, as illustrated in Figure 3. To join these regimes, we use the fact that ‘near-attack conformers’ (NACs), described by Bruice and co-workers,^{49,50} can be represented in both. A NAC can be viewed as a point on the potential energy surface (PES) where the atoms involved in the bond forming process are aligned and within van der Waals distance. Toro-Lobbe and co-workers have shown for a number of reactions that most electronic changes take place in a narrow transition region on the PES,⁵¹ which justifies the division into different regimes. The NAC energy (ΔG_{NAC}) can be seen as an estimate of how well the interactions that bind the substrates to the active site stabilize the transition state. Thus for an enzyme that is preorganized to stabilize the transition state, the substrates are expected to bind in a configuration resembling the transition state and ΔG_{NAC} should be small.

For Diels–Alder, the NAC is a sandwich-like overlap between diene and dienophile. Using it as a reference point, the free energy of activation ($\Delta G_{cat}^{\ddagger}$) is given by eq 4 (*cf.* Figure 3)

$$\Delta G_{cat}^{\ddagger} = \Delta G_{NAC} + \Delta G_{NAC}^{\ddagger} \quad (4)$$

ΔG_{NAC} and $\Delta G_{NAC}^{\ddagger}$ are the free energy differences going from the bound state to NAC and from NAC to TS, respectively. If the enzyme–substrate complex is adequately sampled, ΔG_{NAC} can be calculated from an ensemble average

$$\Delta G_{NAC} = -RT\ln P_{NAC} \geq 0 \quad (5)$$

where P_{NAC} is the fraction of sampling points belonging to the NAC ensemble. It should be noted that the uncertainty increases with poor sampling; no meaningful value of ΔG_{NAC} can be calculated from simulation of nanosecond time scale if $P_{NAC} \ll 1\%$. In addition, we recognize that the uncertainty associated with all three methods employed to calculate the energies

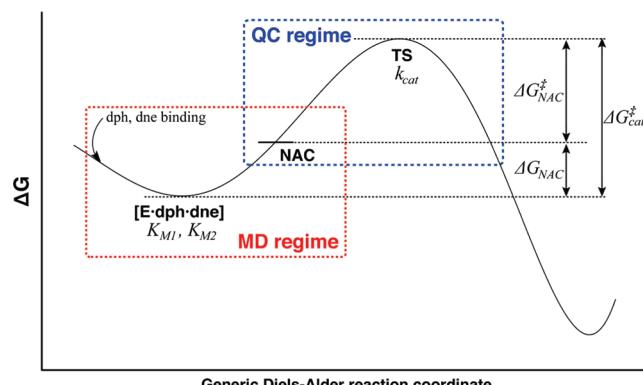


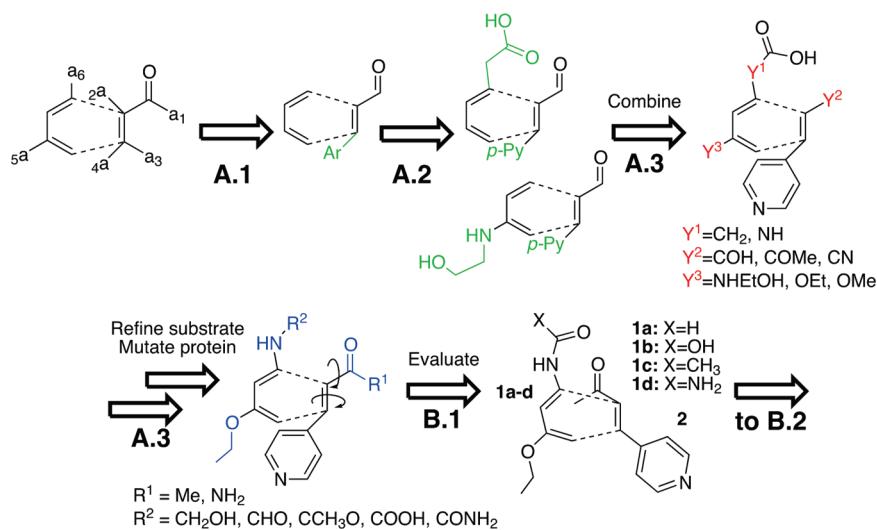
Figure 3. A schematic representation of an enzyme-catalyzed Diels–Alder pathway, showing the classical (MD) and quantum chemical (QC) regimes, and the overlap between them. The NAC can be seen as a point in the intercrossing region.

entering eq 1 can potentially add up to a large error, but we do think that our approach is well suited for comparing different substrates in similar enzyme environments.

RESULTS AND DISCUSSION

Static Design (A). A detailed account of the molecular docking performed is given in the Supporting Information, but a few key points are noted here. The stepwise evolution of the substrate design is also illustrated in Scheme 2. After the first docking round (A.1), it was clear that the enlarged cavity of 1EB9 could host a Diels–Alder TS, with the carbonyl oxygen bound to the oxyanion hole. We found that most high-scoring poses had aromatic substituents in a_4 -position. Therefore, TS models having such substituents at this position were selected from the disubstituted library and docked to screen for additional models with favorable interactions. This selection meant docking ~ 3000 structures rather than 26000 in step A.2.

Step A.2 resulted in highlighting two distinct substituents in positions a_5 and a_6 , as shown in Scheme 2 and Figure S2, both

Scheme 2. Stepwise Evolution of Substrate Design^a

having a *p*-pyridyl group in a₄-position H-bonded to Y133. The a₅ substituent interacted with the F125 backbone oxygen, while the a₆ substituent formed a H-bonding network with H236 and K237. It was thus rational to try the effect of using both substituents, and this was found to conserve the interactions (see Figure S2 and Figure 2). The *Re* conformation was preferred almost exclusively over *Si*.

Apart from combining two individual TS models from step A.2, the rational design step (A.3) consisted of several iterations. Inspired by Siegel et al., an activating —NH— group was introduced in diene C1-position. The a₅-substituent was slightly modified to become smaller and less reactive, and the position of the carbonyl group of the dienophile relative to the diene was varied.

In total, mutations of T11, S80, C81, V118, L149, L158, L179, and I210 were attempted in this work (see Figure 2b for an overview). Most explored mutations were conservative and tested to reshape the active site cavity to provide more space in otherwise strained regions. L158 is an exception, which was mutated to a Asn, Gln, and His to create a potential catalytic contact with the —NH— substituent introduced on the diene. From the ligand and residue positions, it was deduced that a catalytic effect similar to that in the Siegel et al. system could be obtained, and the docking results showed promising concurrence for the L158Q variant.

The catalytic residues were modified by removing the nucleophilic residue S80 (obsolete in the design reaction) and replacing it with Ala and replacing C81 with Ser. The former modification is common practice when dealing with hydrolytic enzymes, especially if one wants to utilize the oxyanion hole without forming any covalent complexes with the enzyme. The latter was attempted since it gave better access to the T11 OH group without losing any H-bonding capability.

While most mutations attempted did not lead to any detectable improvements of docking scores or poses, L158Q, S80A, and C81S seemed to modify the enzyme in a productive direction. We therefore created the (W128A)/S80A/C81A/L158Q mutant, henceforth referred to as HNL-3, and were able to reproduce the poses produced by uncombined mutants.

The lead TS models at the end of the static design stage [1+2] are shown in Scheme 2, with some flexibility introduced in the form of different binding motifs in the a₁ and a₆ positions.

Dynamics Design and Evaluation (B). The initial burst simulations readily showed that the putative *p*Py-Y133 interaction did not need to be conserved, mainly due to two reasons. Backbone movement of the protein and dienophile tilting toward a more hydrophobic region (mainly consisting of L121, L122, and L179), resulting in Y133 and the pyridyl moving further away from each other. After checking different conformations of the dienophile using new bursts, we decided to leave the *p*-Py group unaltered on the dienophile although the lack of a specific interaction could potentially allow quite unconstrained variations. We deduced that a bulky residue helps align the molecule in a productive orientation. With the Y133 interaction dropped, the MD bursts favored *endotrans* conformation rather than *exocis*, which was predicted by docking. The *endotrans* conformation granted more access to the oxyanion hole for the dienophile and aligned the diene's butadiene core closer to the dienophile double bond. It was therefore used in subsequent modeling.

Both R₁ groups (CH₃ and NH₂) were introduced in the docking stage (A.3) and performed similarly. From the MD bursts, we chose —CH₃ as it constrained the motion of the carbonyl group while not interfering with the binding interactions of the diene. Four substrate pairs were thus selected for further study in HNL-3, dienes 1a–d and dienophile 2 (shown in Scheme 2). The selected systems were submitted to the long simulation step (B.2a in Figure 1), and the uncomplexed HNL-3 was compared to the parent 1EB9 structure (referred to as HNL-W128A) to investigate the conformational stability. The same approach was used on the DA_20_00 system, designed by Siegel et al.¹⁴ to validate our approach and to study in detail the behavior of a system with documented success.

Overlays of some snapshots from the simulations of uncomplexed HNL-3 and HNL-W128A are shown in Figure 4. Both simulations are well-behaved with no large structural rearrangements occurring. The backbone rmsd values toward the end of each simulation were in the region of 2.10 for HNL-3 and 2.00 for

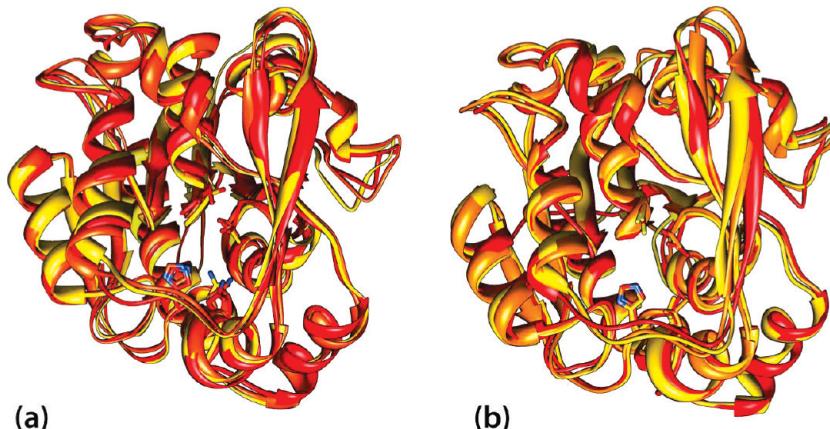


Figure 4. Ribbon representations of the (W128A)/S80A/C81S/L158Q mutant HNL-3 (a) and the ‘wild-type’ 1EB9 structure HNL-W128A (b). The overlaid structures show the proteins at 0 (yellow), 4 (orange), and 8 (red) ns. T11, H236, S/C80, and V/Q158 are drawn in stick representations.

Table 1. LIE Binding Energies from MD Simulations

	ΔG_{bind}^a	$K_M^{b,c}$
HNL-3, 1a	-2.7	9.7
HNL-3, 1b	-3.4	3.1
HNL-3, 1c	-2.4	17
HNL-3, 1d	+0.1	1210
HNL-3, 2	-3.0	6.2
HNL-W128A, 1b	0.3	1620
HNL-W128A, 2	-1.3	119
DA_20_00, diene ^d	-0.5	411 (146)
DA_20_00, dienophile ^d	-2.7	9.7 (3.5)

^a Energies in kcal/mol. ^b Binding constants in 10^{-3} M^{-1} . ^c Using eq 2. ^d Experimental K_M values shown in parentheses.¹⁴

HNL-W128A. The core of the active site was particularly rigid and did not change significantly despite the three mutations in HNL-3. This rigidity explains why interactions with the substrates predicted by docking within the core of the active site were essentially preserved after the burst simulations, whereas predicted interactions in more flexible regions (Y133) were lost.

From simulations of the separate substrates, the binding free energies were estimated according to eq 3, and the results are given in Table 1. As can be seen, K_M values are on the millimolar scale, comparable to DA_20_00 and a significant improvement from HNL-W128A. The latter still exhibits good binding, consistent with the fact that substantial substrate optimization was undertaken prior to enzyme mutation. The correlation between experimental K_M values of of DA_20_00 by Siegel et al.,¹⁸ and the calculated counterparts presented here, is a reassuring validation for the methodology employed herein.

Together with DFT activation energies, the NAC penalty determines the rate of bond formation in the protein complex (Figure 3 and eq 4). From simulations of both substrates, the occurrence of NACs was calculated by monitoring the incipient C–C distances between diene and dienophile. Defining a NAC as a sampling point where both carbons are within 4.0 Å from each other (a somewhat more loose definition¹⁵ than originally suggested by Bruice⁴⁹) we obtain results for substrates **1a** through **1d** and DA_20_00 given in Table 2.

From Table 1, it is seen that the R₂ substituent with the lowest ΔG_{NAC} was the original COOH group, (i.e., **1b**), which forms a

Table 2. NAC Statistics

	P_{NAC}	$\Delta G_{NAC}^{a,b}$
HNL-3, 1a+ 2	0.226	0.9
HNL-3, 1b+ 2	0.612	0.3
HNL-3, 1c+ 2	0.035	2.0
HNL-3, 1d+ 2	0.005	3.2
HNL-W128A, 1b+ 2	0.360	0.6
DA_20_00	0.0002	>4.5

^a Energies in kcal/mol. ^b Using eq 5.

H-bonding network with H236 and K237. These interactions seem to strongly influence the ability to form NACs, since both ligands are held close to each other by the enzyme. The situation is worse by 1–2 orders of magnitude **1c** and **1d**, resulting in a 2–3 kcal/mol higher ΔG_{NAC} , while **1a** also performs well (Table 2). The statistic distribution of key interactions in the MD trajectories give further insight into the underlying reasons for these results. The measured distributions are summarized in Figure 5. Panel (a) shows **1b** and **2** within the active site together with key residues and distances.

It is clear from Figure 5 that the only substrate pair retaining all desired interactions is **1b+ 2** (panel (c)). Binding to Q158 and H236 holds the diene in place underneath the dienophile, which has one strong interaction to T11. S81 prefers to bind to T11, making its OH group more acidic and a more apt H-bond donor. The CC1 and CC2 distributions shown in the inset illustrate how tightly packed the Diels–Alder motif is in the active site. The on average shorter CC1 distance is in accordance with the generally asynchronous geometry of polar Diels–Alder reactions.^{20,52} The mean CC1 distance is just below the sum of two carbon van der Waals (vdW) radii, with a significant portion being closer. CC2 is centered around 4 Å and also frequently decreases below the vdW distance.

The main difference between **1b** and the other dienes is the lack of an interaction with H236 in the latter, which in turn affects positioning of both substrates. This is best illustrated by the C–C distributions, where the majority lie outside the 4.0 Å limit for at least one of the distances. (In panels (d) and (e), we even see that the more important CC1 distance is longer, which is unfavorable from a catalytic perspective.) The relatively high number of NACs produced by **1a+ 2** can be explained by the

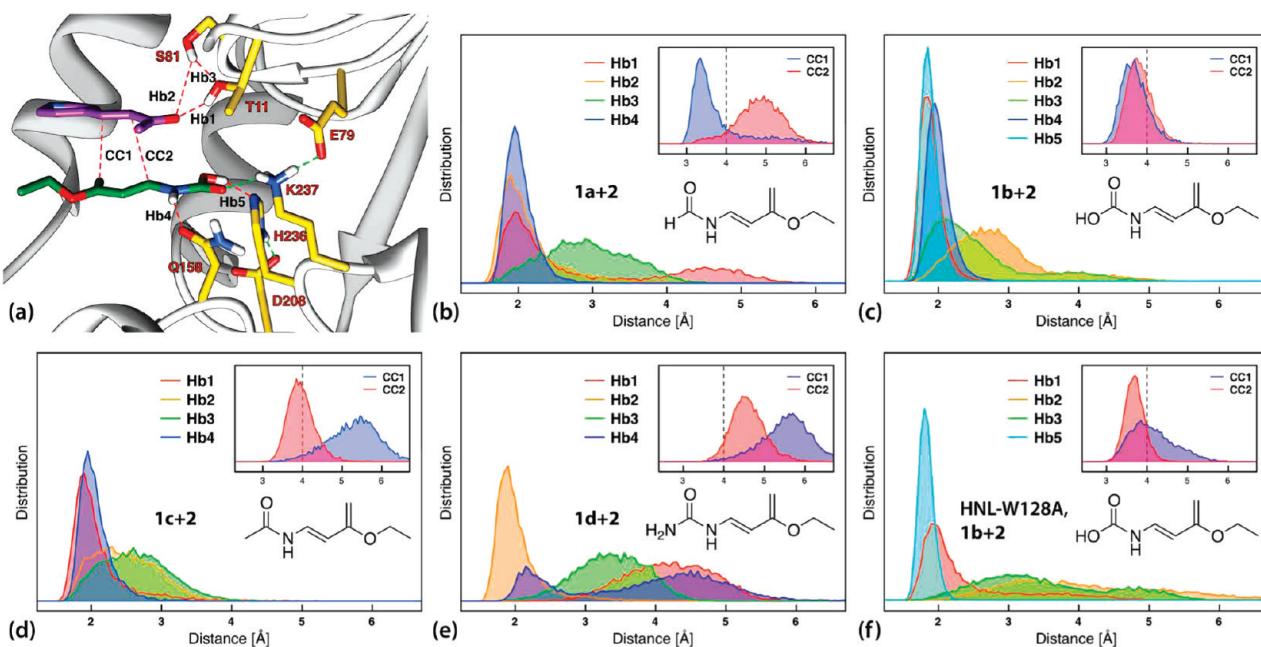


Figure 5. (a) Snapshot of **1b** and **2** in the active site of HNL-3. Bonds depicted in the diagrams are shown as red dashed lines. Other influential H-bonds are shown in green. H-bond statistics are shown for (b) **1a+2**, (c) **1b+2**, (d) **1c+2**, (e) **1d+2**, and (f) **1b+2** in HNL-W128A. Distribution of the two incipient C–C distances are shown in the insets.

smaller size of the aldehyde group. While it does not have complementary interactions with the deep end of the pocket, it is not as repelled as the bulkier keto and amide groups of **1c** and **1d**.

Interestingly, there seems to be a correlation between P_{NAC} and the binding constant of the diene (K_M). Looking at the one-ligand trajectories (used to calculate K_M), we observe that the diene has a preference for the oxyanion hole (T11 and S81) instead of the H236–K237 cavity when alone in the active site. Only **1b** remains in the same conformation as when both substrates are present. It seems that tighter, more specific substrate binding not only favors NAC formation but also pushes down the K_M value.

Panel (f) in Figure 5 shows the distribution for **1b+2** in HNL-W128A. There is less specific binding between enzyme and substrates, leading to higher K_M values (Table 1), but P_{NAC} is relatively high (Table 2). This can be attributed to the strong interaction between **1b** and H236, i.e. a substrate property. Thus, the ability to form NACs in our system can be attributed more to the choice of substrate than the modification on the enzyme. The latter on the other hand has a significant influence on the binding energy.

Quantum Chemical Evaluation (C). From docking and simulations, we decided to build our DFT model from the catalytically important residues T11, E79, A80, S81, Q158, D208, H236, K237, and Q239. E79, D208, and Q239 do not bind directly to the substrates but interact strongly with K237 and H236. Our choice of model is similar to a recent study of wild-type *Hevea brasiliensis* HNL.⁵³ Note that the rest of the cavity is built up of mainly hydrophobic residues, important in substrate binding and alignment, but arguably not as crucial between the NAC point and TS (cf. Figure 3). The model is shown in Figure 6. As can be seen, all residues were truncated to retain only important parts of the side chains, except for the A80–C81 peptide that interacts with the T11 side chain. A detailed representation of the model with indicated frozen atoms is shown in Figure S3.

Three points on the reaction coordinate were considered; a preattack complex (relaxed NAC), TS, and product. The optimized geometries of **1b+2** in the active site are shown in Figure 6(a), (b), and (c), respectively. All TS and product structures show close resemblance (see Table 1 in the SI), and the catalytic machinery essentially works as anticipated.¹⁴ ΔG values for the active site model and uncatalyzed reactions are given in Table 3, and we note that, upon addition of the NAC term according to eq 4, the **1b** TS is lower by 1–4 kcal/mol than the others. As seen in Figure 6, two proton transfers take place upon forming the reaction complex (and subsequent TS and product) with **1b**. The basic H236 deprotonates the COOH group of **1b** and is in turn deprotonated by D208. For the nonacidic substrates, no such proton transfers occurs (see Figure S4). This result raises the question as to when deprotonation takes place and how it affects substrate binding (see below). However, the carboxylic acid is not essential to obtain substantial TS stabilization in the active site, since the main TS stabilization comes from H-bonding to T11 and Q158. This is seen by the similar $\Delta G_{\text{NAC}}^{\ddagger}$ values in Table 3.

The uncatalyzed reactions have activation barriers of 20–22 kcal/mol, a large part of which comes from the entropic cost of bringing the reactants together. To get a sense for how large the TS stabilization is, we compare $\Delta G_{\text{NAC}}^{\ddagger}$ to $\Delta G_{\text{NAC,uncat}}^{\ddagger}$. As seen from the corresponding columns in Table 3, the ‘absolute’ TS stabilization is small and amounts to at most –2.2 kcal/mol (**1c+2**). This result correlates with previous findings by ourselves^{19,20} and others,^{15,16} stating that the stabilization of the TS is a relatively small portion of the net catalytic effect. The fact that $\Delta G_{\text{NAC}}^{\ddagger}$ for **1d** is larger than the corresponding uncatalyzed energy is a consequence of interactions in the active site that hinders the TS formation. These interactions are already present in the NAC, as reflected by the high ΔG_{NAC} , but they have a larger effect in the TS.

Comparison with DA_20_00. The DA_20_00 crystal structure has been published (PDB code 3I1C)¹⁴ and could thus be

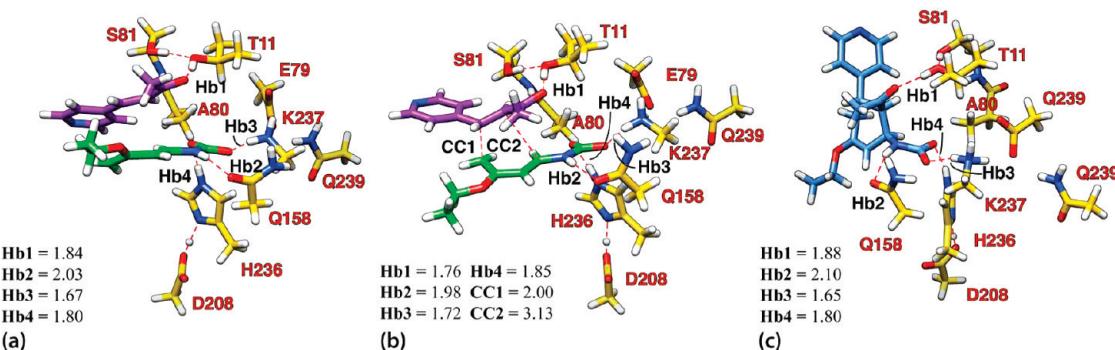


Figure 6. DFT active site model with **1b+2** optimized geometries. Preattack complex (relaxed from NAC), TS, and product are shown in panels (a), (b), and (c), respectively. All distances are in Angstroms. Images and coordinates of the TSs and distance data for all substrates are available in the Supporting Information.

Table 3. M062X/6-311++G(d,p) Free Energies of the Active Site Model^a

	$\varepsilon = 4$			$\varepsilon = 78.39$			
	$\Delta G_{NAC}^{\ddagger}$	$\Delta G_{cat}^{\ddagger}$ ^b	$\Delta G_{uncat}^{\ddagger}$ ^c	$\Delta G_{NAC,uncat}^{\ddagger}$ ^d	$\Delta\Delta G_{cat}^{\ddagger}$ ^e	k_{cat}/k_{uncat}	$\Delta G_{rxn}^{\ddagger}$
1a+2	14.6	15.5	22.0	16.4	-6.5	$5.8 \cdot 10^4$	-18.3
1b+2	13.4	13.7	21.8	15.4	-8.1	$8.2 \cdot 10^5$	-12.8
1c+2	13.0	15.0	21.6	15.2	-6.6	$6.7 \cdot 10^4$	-15.5
1d+2	14.5	17.7	20.4	13.4	-2.7	$9.7 \cdot 10^1$	-10.3

^a Energies in kcal/mol. Optimized geometry and thermodynamic corrections are from the B3LYP/6-311+G(d) level. ^b Calculated using eq 4 with values from Table 2. ^c Relative to separated reactants. ^d Free energies of activation relative to a NAC-like reactant complex in solution. Comparing these to $\Delta G_{NAC}^{\ddagger}$ gives an estimate of the TS-stabilizing effect of the active site. ^e $\Delta\Delta G_{cat}^{\ddagger} = \Delta G_{cat}^{\ddagger} - \Delta G_{uncat}^{\ddagger}$.

studied using the same docking and MD protocol as used in the design methodology reported in this paper. Docking generated high scores slightly above the lead substrates in HNL-3 but did not spontaneously bind to the putative H-bond donors and acceptors (Y121 and Q195). Only when constraints were imposed could the poses postulated by Siegel et al. be reproduced. Subsequent MD simulations of the diene, dienophile, and reaction complex inside DA_20_00, followed by LIE analysis gave binding energies that correlate well with experimental K_M values for the diene and dienophile, as seen in Table 1. However, the enzyme–substrate interactions were less specific, which could be the reason for the reported low k_{cat}/k_{uncat} .¹⁴ We note that the computed ‘theozyme’⁵⁴ geometry, where Y121 and Q195 are oriented to simultaneously bind to the dienophile, is not well-represented in the designed enzymes during the MD simulations, neither in the uncomplexed form nor when substrates are present. The substrates are therefore not frequently oriented in a conformation similar to the optimized theozyme (described in the Supporting Information of ref 14) and move around more freely in the active site. We think that by considering the dynamics of the design system at an early stage, such issues can be addressed, which would allow more efficient designs emerge.

Our approach can be viewed as the inverse of Siegel et al.: Instead of grafting a predesigned active site into an existing protein fold, we let the target enzyme determine the direction of the design. The fewer mutations with this approach have the advantage that the likelihood of conformational changes

that alter the predicted substrate-enzyme interactions become reduced.

Comparing our design with DA_20_00, we conclude that energies of both substrate binding and computed TS stabilization¹⁴ in the quantum chemical models are in the same ballpark. The only large difference we observe is the time substrates spend in the ‘correct’ conformation for the anticipated catalytic step. It is difficult to see how this important property can be addressed without considering the dynamics of the system.

Summary and Overall Performance. The procedure outlined herein has allowed us to perform a semirational design of a HNL enzyme for catalyzing the Diels–Alder reaction, using only a few specific mutations. Several substrates have been designed and varied alongside the enzyme in order to find good matches. Using a series of computational methods has allowed us to study the most important events along the reaction coordinate; binding of the substrates, NAC formation, and formation of the TS followed by the product.

The calculated binding and rate parameters can be inserted into eq 1 to yield v_{cat}/v_{uncat} as a function of substrate and enzyme concentrations ([dne], [dph], and [E⁰], respectively). Keeping [E⁰] fixed at a typical 10^{-5} mol and setting [dne] = [dph], we can plot the catalytic performance as a function of concentrations, as shown in Figure 7. It is seen that optimum catalysis is reached at 10^{-3} M or lower, with v_{cat}/v_{uncat} values ranging from ~ 0.1 (**1d**) to $\sim 10^5$ (**1b**). We note that the DA_20_00 enzyme performs in parity with **1d+2** for the same substrate concentration, showing that the catalytic success of this design depends heavily on keeping enzyme concentrations high (see the Supporting Information of ref 14).

A word of caution is required for diene **1b**. Although it stands out as the most promising diene, with a shape complimentarity being a result from the rational screening, the stability of this carbamic acid derivative can be questioned. Moreover, as the basic H236 deprotonates **1b** in the DFT calculations, we performed complementary MD simulations with the proton transferred. They show much less stable behavior in the active site, with surprisingly little attraction between the protonated histidine and substrate carboxylate. This indicates that **1b** could be an unsuitable substrate because of the acid moiety. Nevertheless, the high, although possibly superficial performance of **1b+2** demonstrates the potential of carefully designing each step in the catalytic process. Moreover, dienes **1a** and **1c** are close behind in overall performance, as seen in Figure 7. We have far from exhausted all design possibilities toward this HNL-based

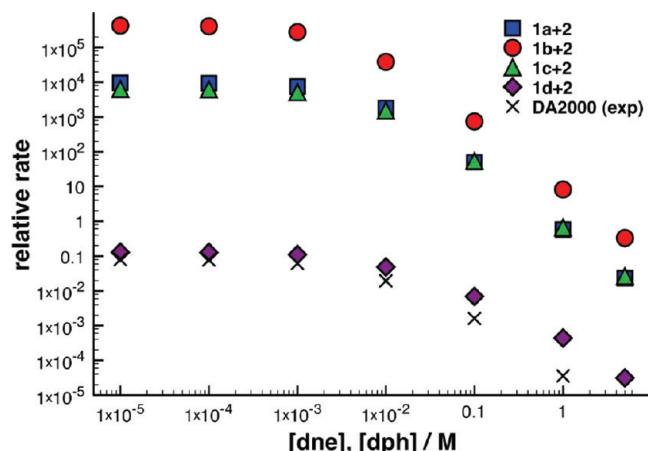


Figure 7. Theoretical v_{cat}/v_{uncat} plotted against substrate concentrations. DA_20_00 is included using experimental values of $K_{M,dne}$, $K_{M,dph}$ and k_{cat}/k_{uncat} from Siegel et al.¹⁴

Diels–Alderase, and this apparent flexibility toward the substrates implies that further improvements are possible.

How good is the semirational design approach, overall? It is difficult to say since we have not yet performed any experimental evaluation, but the good correlation with experimental K_M values in the DA_20_00 system is indeed promising. The reliability of the NAC term depends heavily on the sampling, and since it can be defined in several ways, it is probably the single most uncertain parameter. However, due to the logarithmic dependence in eq 5 quite small variations in ΔG_{NAC} will result from large deviations in P_{NAC} . The mean average deviation using M062X in DFT is ~ 1 kcal/mol,^{37,39} but smaller errors can be expected when comparing similar systems, e.g. the catalyzed and uncatalyzed TS. Most uncertainty can thus be said to be embedded in the MD simulations and the parameters calculated from them.

We do not expect that any verified activity of this design will be as high as that predicted, because of the large number of factors not considered during computational design, it may be lower by a few orders of magnitude. Most importantly, neither the long-term stability effects of mutations nor diffusion of substrates in and out of the cavity have been considered in this work, as they are hard to study using computational methods.

Nevertheless, we feel that the design protocol in this work demonstrates several valuable points to consider in enzyme design. First, the early stage application of molecular dynamics gives a more realistic picture of how the substrate and enzyme interact. This information can be used to modify both the substrates and the enzyme, which will be explored to a greater extent in a forthcoming study. Second, the overall computational cost is quite low. All calculations were performed on modern commodity computers. The total processor time was in the order of 10^4 hours, which is manageable for any normal-sized lab with modern computational resources.

It is worth emphasizing how quickly the combinatorial approach (steps A.1–A.2 in Figure 1) led to substrate candidates for further computational design. Albeit large, the data sets generated by molecular docking were in our experience quickly evaluated. Much more time needs to be invested in the rational steps (essentially steps A.3 through B.2). Evaluating the burst simulations (B.1) and modifying the design thereafter are the most time-consuming steps for the researcher. We note that although the combinatorial library is not optimized and could

be extended, the rational design can build upon relatively simple precursors once the general binding mode of the TS scaffold is established. It is possible that the same result would have been reached using a much smaller library (see the Supporting Information for an extended discussion).

CONCLUSIONS AND OUTLOOK

The design approach presented here shows great potential as it has properties of screening and detailed design alike. We have been able to reach a high degree of shape complementarity, leading to extremely high formation of ‘near-attack conformers’ compared to previous designs.^{9,19} Understanding and designing enzyme catalysis is a complex problem which requires a range of methods, and with this work, we have shown that a combination of several standard, well-established techniques gives detailed information and control of the design process. We have improved the chance of success by not designing to a specific pair of reagents, instead selecting reagent pairs most likely to be optimum for a given set of mutations. Although this process does not directly lead to the design of enzymes of commercial value it does suggest relatively straightforward experiments using modifications of existing enzymes, that could be used to validate the computational methodology presented here. This we feel is a necessary step toward the realization of full de novo Diels–Alderase design.

The DA_20_00 design was used as a means to validate the approach, with good results. We found that while binding its substrates strongly, the relatively poor shape complementarity and high flexibility of catalytic residues lead to unspecific binding and low NAC production.

The construction of the first artificial Diels–Alderase by Siegel et al. is a major milestone. The results obtained here hold promise that Diels–Alderases with much improved catalytic activity can be designed using rational means.

ASSOCIATED CONTENT

S Supporting Information. A description of the docking procedure and evaluation, additional figures, and .xyz molecular coordinates of optimized transition states in the enzyme models. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*E-mail: tore@physchem.kth.se.

ACKNOWLEDGMENT

This work has been supported by the Swedish Research Council (VR) and Cambridge Crystallographic Data Centre (CCDC). We would also like to thank the reviewers for their encouraging comments and constructive feedback.

REFERENCES

- (1) Pohnert, G. Diels-Alderases. *ChemBioChem* **2001**, *2*, 873–875.
- (2) Houk, K. N. Frontier molecular orbital theory of cycloaddition reactions. *Acc. Chem. Res.* **1975**, *8*, 361–369.
- (3) Kim, S. P.; Leach, A. G.; Houk, K. N. The Origins of Noncovalent Catalysis of Intermolecular Diels–Alder Reactions by Cyclodextrins,

- Self-Assembling Capsules, Antibodies, and RNases. *J. Org. Chem.* **2002**, *67*, 4250–4260.
- (4) Hilvert, D.; Hill, K. W.; Nared, K. D.; Auditor, M. T. M. Antibody catalysis of the Diels-Alder reaction. *J. Am. Chem. Soc.* **1989**, *111*, 9261–9262.
- (5) (a) Gouverneur, V. E.; Houk, K. N.; de Pascual-Teresa, B.; Beno, B.; Janda, K. D.; Lerner, R. A. Control of the exo and endo pathways of the Diels-Alder reaction by antibody catalysis. *Science* **1993**, *262*, 204–208. (b) Blake, J. F.; Lim, D.; Jorgensen, W. L. Enhanced Hydrogen Bonding of Water to Diels-Alder Transition States. Ab Initio Evidence. *J. Org. Chem.* **1994**, *59*, 803–805. (c) Yli-Kauhaluoma, J. T.; Ashley, J. A.; Lo, C.-H.; Tucker, L.; Wolfe, M. M.; Janda, K. D. Anti-Metallocene Antibodies: A New Approach to Enantioselective Catalysis of the Diels-Alder Reaction. *J. Am. Chem. Soc.* **1995**, *117*, 7041–7047.
- (6) Chen, J.; Deng, Q.; Wang, R.; Houk, K.; Hilvert, D. Shape Complementarity, Binding-Site Dynamics, and Transition State Stabilization: A Theoretical Study of Diels–Alder Catalysis by Antibody 1E9. *ChemBioChem* **2000**, *1*, 255–261.
- (7) (a) O'Brien, P. J.; Herschlag, D. Catalytic promiscuity and the evolution of new enzymatic activities. *Chem. Biol.* **1999**, *6*, R91–R105. (b) Hult, K.; Berglund, P. Enzyme promiscuity: mechanism and applications. *Trends Biotechnol.* **2007**, *25*, 231–238.
- (8) (a) Branneby, C.; Carlqvist, P.; Magnusson, A.; Hult, K.; Brinck, T.; Berglund, P. Carbon-Carbon Bonds by Hydrolytic Enzymes. *J. Am. Chem. Soc.* **2003**, *125*, 874–875. (b) Carlqvist, P.; Svedendahl, M.; Branneby, C.; Hult, K.; Brinck, T.; Berglund, P. Exploring the Active-Site of a Rationally Redesigned Lipase for Catalysis of Michael-Type Additions. *ChemBioChem* **2004**, *6*, 331–336.
- (9) Svedendahl, M.; Carlqvist, P.; Branneby, C.; Allnér, O.; Frise, A.; Hult, K.; Berglund, P.; Brinck, T. Direct Epoxidation in *Candida antarctica* Lipase B Studied by Experiment and Theory. *ChemBioChem* **2008**, *3*, 2443–2451.
- (10) 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.
- (11) Kuhlman, B.; Baker, D. Native protein sequences are close to optimal for their structures. *Proc. Natl. Acad. Sci.* **2000**, *97*, 10383–10388.
- (12) Röthlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D. Kemp elimination catalysts by computational enzyme design. *Nature* **2008**, *453*, 190–195.
- (13) Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Röthlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas, C. F.; Hilvert, D.; Houk, K. N.; Stoddard, B. L.; Baker, D. De Novo Computational Design of Retro-Aldol Enzymes. *Science* **2008**, *319*, 1387–1391.
- (14) Siegel, J. B.; Zanghellini, A.; Lovick, H. M.; Kiss, G.; Lambert, A. R.; Clair, J. L.; St.; Gallaher, J. L.; Hilvert, D.; Gelb, M. H.; Stoddard, B. L.; Houk, K. N.; Michael, F. E.; Baker, D. Computational Design of an Enzyme Catalyst for a Stereoselective Bimolecular Diels-Alder Reaction. *Science* **2010**, *329*, 309–313.
- (15) Ruscio, J. Z.; Kohn, J. E.; Ball, K. A.; Head-Gordon, T. The Influence of Protein Dynamics on the Success of Computational Enzyme Design. *J. Am. Chem. Soc.* **2009**, *131*, 14111–14115.
- (16) Lassila, J. K.; Baker, D.; Herschlag, D. Origins of catalysis by computationally designed retroaldolase enzymes. *Proc. Natl. Acad. Sci.* **2010**, *107*, 4937–4942.
- (17) 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.* **2010**, *107*, 16869–16874.
- (18) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids. Res.* **2000**, *28*, 235–242.
- (19) Linder, M.; Hermansson, A.; Liebeschuetz, J.; Brinck, T. Computational design of a lipase for catalysis of the Diels-Alder reaction. *J. Mol. Model.* **2011**, *17*, 833–849.
- (20) Linder, M.; Brinck, T. Synergistic activation of the Diels-Alder reaction by an organic catalyst and substituents: a computational study. *Org. Biomol. Chem.* **2009**, *7*, 1304–1311.
- (21) Simón, L.; Goodman, J. M. Enzyme Catalysis by Hydrogen Bonds: The Balance between Transition State Binding and Substrate Binding in Oxyanion Holes. *J. Org. Chem.* **2010**, *75*, 1831–1840.
- (22) (a) Hendlich, M. Databases for Protein-Ligand Complexes. *Acta Crystallogr.* **1998**, *D54*, 1178–1182. (b) Hendlich, M.; Bergner, A.; Günther, J.; Klebe, G. Relibase: Design and Development of a Database for Comprehensive Analysis of Protein-Ligand Interactions. *J. Mol. Biol.* **2003**, *326*, 607–620.
- (23) Schmitt, S.; Hendlich, M.; Klebe, G. From Structure to Function: A New Approach to Detect Functional Similarity among Proteins Independent from Sequence and Fold Homology. *Angew. Chem., Int. Ed. Engl.* **2001**, *40*, 3141–3144.
- (24) Schmitt, S.; Kuhn, D.; Klebe, G. A New Method to Detect Related Function Among Proteins Independent of Sequence and Fold Homology. *J. Mol. Biol.* **2002**, *323*, 387–406.
- (25) Lauble, H.; Miehlich, B.; Förster, S.; Kobler, C.; Wajant, H.; Effenberger, F. Structure determinants of substrate specificity of hydroxynitrile lyase from Manihot esculenta. *Protein Sci.* **2002**, *11*, 65–71.
- (26) Kong, S.; Evanseck, J. Density Functional Theory Study of Aqueous-Phase Rate Acceleration and Endo/Exo Selectivity of the Butadiene and Acrolein Diels-Alder Reaction. *J. Am. Chem. Soc.* **2000**, *122*, 10418–10427.
- (27) *Molecular Operating Environment*, v. 2009.10, 2009. <http://www.chemcomp.com/software.htm> (accessed September 15, 2010).
- (28) Labute, P.; Nilar, S.; Williams, C. A probabilistic approach to high throughput drug discovery. *Comb. Chem. High Throughput Screening* **2002**, *5*, 135–145.
- (29) (a) GOLD 5.0, 2010. http://www.ccdc.cam.ac.uk/-products/life_sciences/gold/ (accessed 07–14, 2011). (b) Jones, G.; Willett, P.; Glen, R. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.* **1995**, *245*, 43–53. (c) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267*, 727–748. (d) Nissink, J. W. M.; Murray, C.; Hartshorn, M.; Verdonk, M. L.; Cole, J. C.; Taylor, R. A new test set for validating predictions of protein-ligand interaction. *Proteins* **2002**, *49*, 457–471. (e) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved Protein-Ligand Docking Using GOLD. *Proteins* **2003**, *52*, 609–623.
- (30) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A.; AMBER 10, University of California, San Francisco, CA, 2008.
- (31) Wang, J. M.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21*, 1049–1074.
- (32) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A. Development and testing of a general Amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (33) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (34) (a) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100. (b) Becke, A. D. A new mixing of Hartree-Fock and local density-functional theories. *J. Chem. Phys.* **1993**, *98*, 1372–1377. (c) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789. (d) Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-

- Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971**, *54*, 724–728.
- (35) *Jaguar v. 7.6*, 2009. www.schrodinger.com (accessed July 14, 2011).
- (36) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazeyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, Revision C.02*, Gaussian, Inc.: Wallingford, CT, 2004.
- (37) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215–241.
- (38) (a) Barone, V.; Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A* **1998**, *102*, 1995–2001. (b) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comput. Chem.* **2003**, *24*, 669–681. (c) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal solvation model based on solute electron density and a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.
- (39) Pieniazek, S.; Clemente, F.; Houk, K. Sources of Error in DFT Computations of C-C Bond Formation Thermochemistries: $\pi \rightarrow \sigma$ Transformations and Error Cancellation by DFT Methods. *Angew. Chem., Int. Ed.* **2008**, *47*, 7746–7746.
- (40) UCSF Chimera, 2004. <http://www.cgl.ucsf.edu/chimera> (accessed 07–14, 2011).
- (41) Pettersen, E.; Goddard, T.; Huang, C.; Couch, G.; Greenblatt, D.; Meng, E.; Ferrin, T. UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.
- (42) Wesemann, M. *Plot, v. 0.997*, 2007. <http://plot.micw.eu/> (accessed 07–14, 2011).
- (43) Michaelis, L.; Menten, M. L. Die Kinetik der Invertinwirkung. *Biochem. Z.* **1913**, *49*, 333.
- (44) Cleland, W. W. I Steady State Kinetics. *Enzymes* **1970**, *2*, 1–65.
- (45) Dalziel, K. I Kinetics and Mechanism of Nicotinamide-Nucleotide-Linked Dehydrogenases. *Enzymes* **1975**, *11*, 1–60.
- (46) (a) Åqvist, J.; Medina, C.; Samuelsson, J. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* **1994**, *7*, 385–391. (b) Marelius, J.; Hansson, T.; Åqvist, J. Calculation of ligand binding free energies from molecular dynamics simulations. *Int. J. Quantum Chem.* **1998**, *69*, 77–88.
- (47) Almlöf, M.; Brandsdal, B. O.; Åqvist, J. Binding affinity prediction with different force fields: Examination of the linear interaction energy method. *J. Comput. Chem.* **2004**, *25*, 1242–1254.
- (48) Linder, M.; Hermansson, A.; Brinck, T. 2010; unpublished results.
- (49) Lightstone, F. C.; Bruice, T. C. Ground State Conformations and Entropic and Enthalpic Factors in the Efficiency of Intramolecular and Enzymatic Reactions. 1. Cyclic Anhydride Formation by Substituted Glutarates, Succinate, and 3,6-Endoxo- Δ^4 -tetrahydrophthalate Mono-phenyl Esters. *J. Am. Chem. Soc.* **1996**, *118*, 2595–2605.
- (50) Bruice, T. C.; Lightstone, F. C. Ground State and Transition State Contributions to the Rates of Intramolecular and Enzymatic Reactions. *Acc. Chem. Res.* **1999**, *32*, 127–136.
- (51) (a) Toro-Labbe, A.; Gutierrez-Oliva, S.; Murray, J.; Politzer, P. A new perspective on chemical and physical processes: the reaction force. *Mol. Phys.* **2007**, *105*, 2619–2625. (b) Labet, V.; Morell, C.; Grand, A.; Toro-Labbe, A. Theoretical Study of Cytosine Deamination from the Perspective of the Reaction Force Analysis. *J. Phys. Chem. A* **2008**, *112*, 11487–11494. (c) Toro-Labbe, A.; Gutierrez-Oliva, S.; Murray, J.; Politzer, P. The reaction force and the transition region of a reaction. *J. Mol. Model.* **2009**, *15*, 707–710.
- (52) Domingo, L. R.; Saéz, J. A. Understanding the mechanism of polar Diels-Alder reactions. *Org. Biomol. Chem.* **2009**, *7*, 3576–3583.
- (53) Cui, F.-C.; Pan, X.-L.; Liu, J.-Y. Catalytic Mechanism of Hydroxynitrile Lyase from Hevea brasiliensis: A Theoretical Investigation. *J. Phys. Chem. B* **2010**, *114*, 9622–9628.
- (54) Zhang, X.; DeChancie, J.; Gunaydin, H.; Chowdry, A. B.; Clemente, F. R.; Smith, A. J. T.; Handel, T. M.; Houk, K. N. Quantum Mechanical Design of Enzyme Active Sites. *J. Org. Chem.* **2008**, *73*, 889–899.