

Publishers' page

Publishers' page

Publishers' page

Publishers' page

Dedication Page
(optional)

June 27, 2014 23:32

PSP Book - 9in x 6in

000-main

Contents

| | |
|---|----|
| 1. Computational Enzyme Engineering: Activity Screening using Quantum Chemistry. | 1 |
| 1.1 Abstract | 1 |
| 1.2 Motivation | 2 |
| 1.3 Introduction | 3 |
| 1.4 Methods | 4 |
| 1.4.1 Calculation Engines | 4 |
| 1.4.2 Molecular Modeling | 6 |
| 1.4.3 Software | 9 |
| 1.5 Applications | 10 |
| 1.5.1 Overview | 10 |
| 1.5.2 Engineering <i>Candida antarctica</i> Lipase B . | 10 |
| 1.5.3 Engineering <i>Bacillus circulans</i> Xylanase . | 17 |
| 1.6 Conclusions | 23 |

Chapter 1

Computational Enzyme Engineering: Activity Screening using Quantum Chemistry.

Martin R. Hediger^a

^a*ma.hed@bluewin.ch*

1.1 Abstract

In this article, the application of quantum chemical methods for screening of enzyme activity is discussed. Using a linear interpolation approach, where the enzyme structure is divided into discrete structural frames along the reaction coordinate of the rate determining step and the wave function derived semi-empirical energies of each frame are evaluated to provide an approximate reaction profile, the amide hydrolysis activity in *Candida antarctica* lipase B (CALB) and glycosidic bond hydrolysis activity in *Bacillus circulans* xylanase (BCX)) and derived mutants is screened. Application of this approach to a programmatically generated library of mutant structures allows ranking and identification of mutants with likely increased activity towards an artificial

Book Title

Editor Name

Copyright © Year Pan Stanford Publishing Pte. Ltd.

ISBN

www.panstanford.com

substrate. Experimental verification of the CALB results yield a prediction rate where 15 out of 22 mutants are correctly characterized in terms of both increased and decreased activity. The screening of BCX mutants allows identification of structural patterns among the most active mutants providing new hypothesis for rationalizing the catalytic enhancing determinants of the BCX active site. At the time of writing, the presented applications are the largest quantum chemical screens of enzyme catalysts and present a valuable tool for computational enzyme engineering.

1.2 Motivation

Let us imagine two companies a and b . Both companies use very similar technical equipment to carry out the biotechnological process $X \xrightarrow{\text{Biocatalyst}} Y$ using two different versions of a biocatalyst. Company a uses an enzyme with a rate constant $k_{\text{cat},a} = 1000\text{s}^{-1}$ while company b uses an enzyme with $k_{\text{cat},b} = 2000\text{s}^{-1}$. Letting all other things be equal, the process of company b will therefore only require half the time to produce one Mole of product compared to the time required for company a . Company b therefore can save energy because the reaction volume does not need to be heated for as long as company a needs to in order to produce the same amount of product. The need for efficient catalysts arises from such an outline and the commercial implications of these considerations are immediate.¹

Increasing the performance of enzymes however is still far from trivial and forms a growing body of research. What is clear though is that the development of such catalysts is costly, in terms of manpower, material and energy – if carried out in the laboratory. A number of companies have in fact formed around this demand: Novozymes (DK), Genzyme (US) or DSM (NL) to name but a few [Kirk *et al.* (2002); Beilen and Li (2002); Schmid *et al.* (2002); Meyer *et al.* (2013)].

Laboratory costs can however be saved to a large part if initial development is carried out *in silico*. The proof that computational

¹In this work, we use the terms *enzyme* and *bio-/catalyst* interchangeably.

results are as reliable as experimental results has been provided not too long ago[Clayssens *et al.* (2006)]. The foundation for successful application of computational enzyme engineering has thus been laid out and its success was acknowledged by awarding the Nobel price in Chemistry in 2013 to pioneers of the field.

1.3 Introduction

In this review, we aim to provide an introduction into the topic of modeling of enzyme catalysis for people from both within or from outside the field, with emphasis on aspects of activity screening and some emphasis on commercial applicability. We outline the latest achievements, methods in use and required developments to further establish computational enzyme engineering and screening also in commercial practice.

While the process engineering team in company *a* from above is working on improving reaction conditions, consider the enzyme engineer required to engineer the currently used catalyst to match (and possibly outperform) the activity of the catalyst used by company *b*. Following the hypothesis that the function of an enzyme depends on its structure, in order to change the function, some kind of change to the structure has to be introduced and this in general requires the introduction of one or more mutations of amino acids in the primary structure of the enzyme. However, it is almost impossible to predict, how a mutated variant of an enzyme will behave. Will it be possible to express it, will it even fold and if so, will it be more or less active relative to the wild type? Are its pH properties still the same and is it of similar temperature stability? Even the most reasonably suggested mutations (except for elimination of the catalytic residues) will hardly allow to predict *qualitatively* with confidence if the mutated enzyme will be more or less active. Therefore it appears as if the only reasonable strategy were to carry out systematic screening studies.² As was pointed out initially, such studies are expensive in the laboratory and therefore an alternative is to carry them out on the computer. The aim of

²By *screening* we mean the construction of a large library of mutants which is inspected and compared for a specific property like activity.

4 | Computational Enzyme Engineering: Activity Screening using Quantum Chemistry.

such initial computational screening is not necessarily to provide high-accuracy kinetic data but much rather to provide a reliable, qualitative activity estimate of a preferably large mutant library. It is important to note that while the initial computational screening might be of limited accuracy, its major value is the provision of a systematically developed library of mutants and qualitative results indicating which of these are likely to be most active. Based on the screening, promising candidates can then be pipelined into a more demanding computational method for further characterization.

Also outside a commercial environment, screening a library of candidates can be of use. As we show in the second application, it is possible to identify structural patterns in the more active mutants and such observations can lead to new hypothesis about structure/function relationships.

The article is organized as follows. First, a number of technical aspects are introduced in the methods section, then we will discuss application of the approach to the modeling of two enzymes: *Candida antarctica* lipase B (CALB) and *Bacillus circulans* xylanase (BCX).

1.4 Methods

1.4.1 Calculation Engines

This article focuses on the application of quantum chemistry for screening of enzyme mutants. So instead of attempting to incompletely review modern computational chemistry methods, which are explained in text books on the subject[Young (2004); Jensen (2007); Cramer (2013)], only a few selected facts relevant for the following are highlighted.

Most fundamentally, computational descriptions of a molecular system are done using either (parametric) *force fields* or *ab initio* electronic structure methods.³ Force field methods are useful in studying structural behavior of enzymes over a significant time period (approaching the millisecond scale) and can provide details

³ *Molecular mechanics*, *empirical* or *classical mechanics* are other expressions for force field methods.

about structural rearrangement of an enzyme. These simulations are usually referred to as *molecular dynamics* simulations.

Chemical reactivity however is accompanied by changes in the electronic structure of the system and which are described by quantum chemical methods. Quantum chemical methods comprise so called *semi-empirical*, *density functional* or *ab initio* methods ('semi-empirical' methods are different to the empirical mechanics methods of the previous paragraph).

Quantum Chemical Methods. Quantum chemical methods attempt at solving the Schrödinger equation for the molecule, the solution of this equation will yield the wave function of the system. However, depending on the specifics of the quantum chemical method, the solutions are, albeit likely similar, different. Now, since every quantum chemical method provides its own description of the wave function (or electron density in density functional methods) and all properties, such as energy or charge distribution, are *derived* from the wave function (or the electron density), it is essential to note that every method has *its* own chemistry depending to the wave function (or density) it produces. As a consequence, hydrogen bonding patterns of a structure can be different when optimized with two different quantum chemical methods. Therefore, it is good practice to verify the predictions of one quantum chemical method against, perhaps, higher-level methods or literature references. With regard to the modeling of enzymatic reactions, if a method can reproduce the generally accepted mechanism then there is a good chance that at least qualitatively reliable conclusions can be made using this method.

In the applications described below, the semi-empirical PM6 method[Stewart (2007, 2009)] was mainly used together with the MOZYME technology[Stewart (1996)]⁴. Initial test calculations on small active site models of two enzymes allowed to reproduce the generally accepted mechanisms and in a comparative study, PM6 is found to provide results approaching the accuracy of density functional methods[Schenker *et al.* (2011)]. The PM6 method is actively being developed further to take into account properties such as dispersion or solvation[Rezác *et al.* (2009);

⁴*Semi-empirical* essentially means that parts of the calculation are parametrized from the start which results in reduced computational demands.

Rezác and Hobza (2011)]. The MOZYME technology allows to localize the molecular orbitals obtained during the self-consistent field calculation which results in computational demands scaling linearly with system size. In essence, these two technologies in combination therefore allow to efficiently model enzymatic reactivity in a screening assay.

1.4.2 Molecular Modeling

A variety of molecular modeling approaches are available in the literature, all suitable for different applications. We briefly outline two of them and then introduce in more detail the approach used in the screening studies of the application section.

Cluster approach. Among the most popular approaches is the so called cluster approach, where a larger or smaller part of the enzyme active site is extracted into a separate model and then all calculations are carried out on this smaller model (consisting of usually 10-15 residues[Lind and Himo (2013)]). Structural integrity of the model is preserved by applying constraints on the truncation points and among the most prominent applications of this approach is the kinetic discrimination between possible competing mechanistic alternatives[Noodleman *et al.* (2004); Himo (2006); Siegbahn and Himo (2009)]. This technique, in principle, can be used in enzyme activity screening studies, however care has to be taken that no mutations are introduced into the model which can only be accommodated because of the reduced model size, *i.e.* because there appears to be space for a side chain where in reality there would not be any. An advantage of cluster models is that because of the reduced model size, higher level methods can be applied.

Full Enzyme Models. Using a full model of the enzyme structure has advantages but also raises a number of issues. Including the full enzyme structure in the calculation naturally increases the computational demands significantly. Thus when running calculations on a full enzyme structure, careful consideration of the trade-off between accuracy and efficiency becomes very important. One way of addressing this issue is by using so called hybrid methods which combine a quantum chemical and a classical

mechanics based calculation. The advantage of these methods is that they are capable of producing highly accurate results, the setup, especially definition of the boundary between the classical and quantum regions, of such calculations is however still not as straightforward as is required for large scale screening approaches. An alternative is to describe the complete system using the same, albeit lower level of theory (*i.e.* PM6) which reduces efforts to setup the calculation significantly. This is the approach of the applications described below.

As was outlined above, semi-empirical methods in combination with linear scaling technologies, which are readily available in a number of programs (GAMESS, MOPAC and others), are sufficiently efficient and allow for some degree of programmatic customization, which is also of great relevance for later data analysis.

Calculation of Reaction Barrier. Irrespective of the method used, calculating the activation energy requires to estimate the difference between the transition state and the most stable state of the system before the transition state on the reaction coordinate. This is a non-trivial problem and no generally applicable methods exist to do this, solving this problem will therefore always at one point require manual input from the modeler. Furthermore, characterization and identification of a transition state is a computationally very demanding operation. A reasonable approach is therefore to approximate the transition state by linearly interpolating the structure along the reaction coordinate of the rate determining step of the total reaction. If carried out carefully, this approach can provide very good approximations of the transition state.

Linear Interpolation. In both applications outlined below, the basic procedure to obtain a reaction barrier consists of preparing structures for the enzyme substrate complex (“ES”) and the first intermediate (CALB: “TI”⁵, BCX: “GE”⁶) and calculating the reaction barrier inbetween the ES_{CALB}-TI and ES_{BCX}-GE pairs, respectively. For these two pairs, the reaction barrier is calculated by preparing a set of 10 intermediate structures which approximate

⁵Tetrahedral intermediate

⁶Glycosyl enzyme

the structure of the system along the reaction coordinate, Fig. 1.1A.⁷ By evaluating the energies for each interpolation frame for

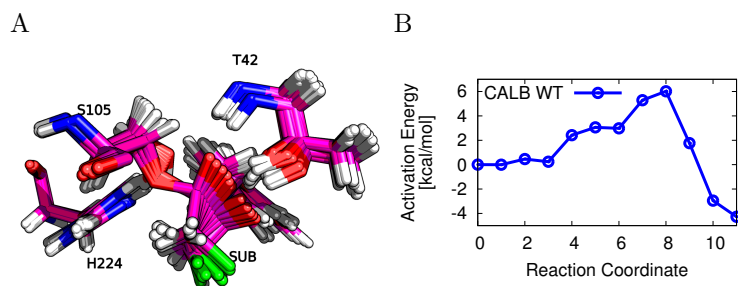


Figure 1.1 **A:** Illustration of linear interpolation. It is visible how the substrate and the nucleophilic oxygen of S105 approach each other. Also visible is the proton transfer to H224 in the late interpolation frames and how the oxyanion hole (T42) follows the increasingly negative carbonyl oxygen of the substrate. **B:** Evaluation of the system energy for each interpolation frame results in an approximate reaction barrier (calculation done with PM6).

the two reactions $ES_{\text{CALB}} \rightarrow \text{TI}$ or $ES_{\text{BCX}} \rightarrow \text{GE}$, an approximate potential energy surface of the reaction is obtained, Fig. 1.1B. The quality of the results of this approach are strongly dependent on how careful the modeling of each of these structures is carried out. Since the geometry of every interpolation frame is optimized individually, it is possible that adjacent interpolation frames optimize into significantly different local minima, this is a frequent reason for meaningless results. It is therefore advisable that the two end points of the interpolation are not too different in structure. This approach provides an estimate of k_{cat} and in the working hypothesis it is assumed that binding of the substrate by different mutants is similar, an observation which is also made in experiments[Ludwiczek *et al.* (2013)].

Requirements. The most basic requirements are that a structure representing a bound intermediate along the reaction pathway is available. For the modeling it is advantageous to have a structure

⁷In the literature, the terms *interpolation*, *linear transit scan*, *reaction coordinate calculation* or *adiabatic mapping* are frequently used to indicate essentially the same kind of procedure.

with a bound intermediate which can then be modified into the substrate, while preserving its natural binding pose. This model is then used as a template for both stationary points adjacent to the transition state of the rate determining step in the total reaction. Using a single model as the template for both the intermediate of the reaction, as well as the preceding stationary point is additionally advantageous because the structural differences between these structures will be small and will less likely lead to non-conclusive reaction barriers in the interpolation. A second requirement is naturally to have a reasonable understanding of the reaction mechanism. In case the mechanism is not available from the literature, computational experiments on small active site models can provide a working hypothesis.

Preparation of Mutant Structures. The structure of the mutants can be prepared with any kind of molecular modeling software, however it is advantageous if the software can be controlled programmatically. When preparing molecular structures of the mutants, it is frequently observed that specific pairs of mutants can not be accommodated in the same model because the two new side chains would clash. In the simplest version of the approach, these variants are simply discarded from the screen, assuming that such mutants would not be expressed *in vivo*. Furthermore, side chain orientations can be oriented in various ways and can be defaulted using a local optimization procedure (as the one which is built into PYMOL, see below). Ionization states of side chains greatly complicate the screening and at the current development state, only default ionization states are considered.

1.4.3 Software

The largest part of the calculations in the applications below have been carried out with the MOPAC software[Stewart (1990)]. This software is designed to be convenient to use. It offers various semi-empirical calculation engines and linear scaling techniques and is optimized towards working with PDB formatted data structures. To prepare the molecular models, the protein visualization software PYMOL[Schrödinger, LLC (2010)] offers a huge range of functions and can be controlled programmatically, which is critical for

the preparation of a large, systematic set of mutants. It has to be noted, that molecular modeling and computational screening still consists of a lot of customization work, especially final data analysis usually requires development of software for extraction of relevant data from the output files. This might also be contributing to preventing quantum chemical methods from becoming more popular in industrial environments. Docking methods for example are implemented in software which is designed towards user friendliness and have established themselves in industry[Goodsell *et al.* (1996)]. Another project is GTKDynamo, combining the modeling features of PYMOL directly with quantum chemical calculation software, which looks very advanced but so far however has gained traction only with a small community[Bachega *et al.* (2013)].

1.5 Applications

1.5.1 Overview

From a technical point of view, a computational screening assay is characterized by its computational efficiency, *i.e.* how long it takes for the calculations to complete, its user friendliness, its accuracy and the modeling prerequisites. In the following, two applications of the presented screening method are introduced. It is shown how the method performs relative to these mentioned criteria.

1.5.2 Engineering *Candida antarctica* Lipase B

Relevance and Summary. Using CALB as a test case, we develop an efficient modeling protocol, verify the calculations with experimental results and show how quantum chemical methods can be used for enzyme activity screening[Hediger *et al.* (2012, 2013a)]. The reaction studied is the hydrolysis of *N*-benzyl-2-chloroacetamide.

CALB has been highly characterized over the years such that its structure, physical properties and expression systems are well documented[Uppenberg *et al.* (1994, 1995)]. The enzyme

is a highly established catalyst with applications in organic synthesis, formulation technology[Gayot *et al.* (2003)] and in kinetic resolution of racemic mixtures[Gotor-Fernández *et al.* (2006); Naik *et al.* (2010); Chaput *et al.* (2012)]. Furthermore, since the enzyme is known for its reactive promiscuity[Bornscheuer and Kazlauskas (2004); Svedendahl *et al.* (2008)], it provides an ideal development platform for applications for which existing biocatalysts are only of limited efficiency. One such application is the hydrolysis of amide bond containing lipophilic compounds. While in principle amide bonds can be cleaved by proteases or peptidases, these enzymes are found to be less active towards hydrophobic substrates such as amide containing lipids[Nakagawa *et al.* (2007)]. Therefore, engineering CALB activity towards CO – NH bond containing hydrophobic substrates would greatly expand its applicability.

Structure and Mechanism. Briefly, the active site in CALB consists of a catalytic triad (S105, D189 and H224) and an associated oxyanion hole. The mechanism is understood to follow the generally accepted mechanism of serine protease active sites[Hedstrom *et al.* (2002)], Fig. 1.2. The rate determining step

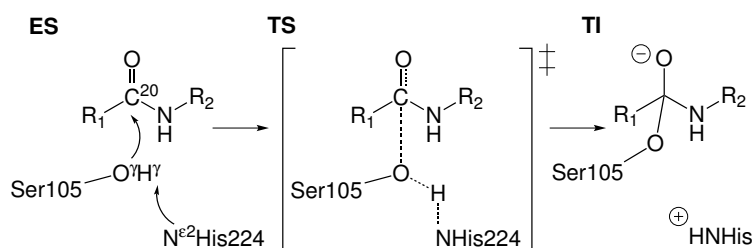


Figure 1.2 Reaction scheme for the formation of the tetrahedral intermediate in CALB. R₁: -CH₂-Cl, R₂: -CH₂-C₅H₆[Hediger *et al.* (2013a)].

is believed to be the nucleophilic attack by O^γ of S105 on the carbonyl carbon C²⁰ of the substrate.

It is noted that the CALB model used consists of not the complete structure, but rather a comparably large cluster model consisting of around 800 atoms (including hydrogens), where ordinarily cluster

models usually consist of up to 120 atoms. CALB has around 320 amino acids in the backbone, which would make the calculations of the full structure take exceedingly long time to complete, this will be discussed further in the next section.

Calculations, Experimental Verification and Calibration.

After establishing a working proof-of-concept of the screening approach using a limited set of three mutants[Hediger *et al.* (2012)], the method was scaled up to screen a set of 386 mutants of the CALB active site[Hediger *et al.* (2013a)]. This set consisted of single- to eight-fold mutants and of these, the activity of 22 mutants was measured experimentally.

The accuracy of the method was calibrated in terms of qualitative predictivity, *i.e.* how likely can the method predict if a mutant will have higher or lower activity than the wild type. For this calibration, the following points are to be noted. In the experiments, the activity relative to the WT, in principle, ranges from zero to infinity. In the calculations on the other hand, the activity of a mutant is determined by the difference between the mutant reaction barrier and the wild type barrier.⁸ However, even for a very active mutant, the reaction barrier is not expected to be zero and so in this approach, since lower barriers are related to higher activities, the amount by which a mutant can be more active than the wild type is in principle limited. Conversely, since the reaction barrier of a mutant can in principle be arbitrarily high, the amount by which a mutant could be less active than the wild type might be infinite. Establishing a direct quantitative relationship between differences in reaction barrier heights and relative experimental activities is therefore not very meaningful. Instead, it is more meaningful to only qualitatively compare activities, *i.e.* if a mutant has higher or lower activity than the WT (both experimentally and calculated), and use these results for determining the predictivity.

To cluster the results into increased/decreased activities it is therefore necessary to decide above/below which activity a mutant is considered being part of either the higher-activity or lower-activity clusters. Considering the relatively narrow spread of

⁸If a mutant has a lower reaction barrier than the WT, it is understood to have higher activity.

experimental activities⁹, mutants with ≥ 1.2 times the wild type activity are considered improving while mutants with < 0.8 times the wild type activity are considered degrading (mutants inbetween are considered as neither improving nor degrading)[Hediger *et al.* (2013a)]. The computational results are clustered in a similar way. A cutoff value for the calculated reaction barriers is introduced *below* which the mutant is considered improving and *above* which the mutant is considered degrading. It is found that at a specific cutoff value (12.5 kcal/mol), the method qualitatively predicts activity of 15 out of 22 mutants correctly, Fig. 1.3. Noteworthy, mutants with both higher as well as lower activity

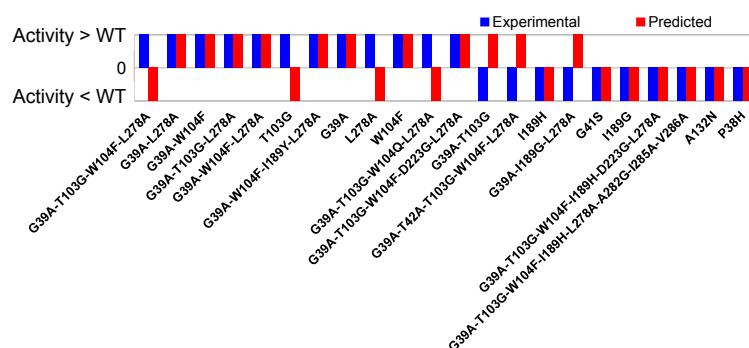


Figure 1.3 Agreement between predicted and experimental qualitative activity relative to wild type[Hediger *et al.* (2013a)].

are correctly predicted. Importantly, the advantage of clustering the mutants in this way is not apparent maximum agreement with the experimental data in the first place, but much rather to provide a lower limit for the predictivity under the given set of approximations. It is important to keep in mind that it will be unlikely to have calibration data available in a computational enzyme activity assay, since ideally it is performed *before* wet-lab experiments are started. Therefore, such an approach is further justified since at the initial outset of a screening study, the focus is mostly on narrowing down possible candidates from a large

⁹The experimental improvement factors range from 0 to roughly 11 times wild type activity.

library, *i.e.* the focus is on obtaining qualitative conclusions rather than quantitative results [Agresti *et al.* (2010)]. Following such a screening analysis, out of the mutants predicted to have higher activity, a set consisting of as many mutants as can be further characterized is selected and suggested for further study, either computationally or experimentally.

Screening Results and Discussion of Mutants. Finding mutants with reaction barriers significantly lower than the wild type is difficult. In the set of 386 screened mutants, we identified only three mutants with a barrier lower than the wild type (a double, a triple and a four-fold mutant) while a large proportion of the mutants has barriers in the range from 12 to 14 kcal/mol. We find that a large number of mutants need to be screened to identify the best few candidates for further study. This conclusion is likely even more true in studies where the enzyme is either intrinsically of only very little activity towards a specific substrate (“it does not work”).

Further analysis of the screening results was focused on two residues: W104 and I189. These are selected because mutations of these positions can provide more space for the substrate which is hypothesized to lead to increased activity, Fig. 1.4. As seen, W104 and I189 are in direct proximity to the substrate. From the single mutant data, it is found that the mutants W104Q or W104Y have barriers just around the cutoff value of 12.5 kcal/mol. However, when in combination with other mutations, such as G39A-W104F-A141Q-I189A (8.3 kcal/mol) or G39A-T103G-W104Y-A141N (9.8 kcal/mol), these mutations are among the most active ones found. It is noted that it would be very difficult to arrive at these combinations by pure rational considerations of the active site.

Analysis of mutations of I189, Fig. 1.5, provides additional interesting insight into the non-linear effects of various mutations. Five different mutations of this residue are screened but to illustrate the point, the focus is on I189G. As a single mutant, I189G is found to be completely unproductive (18.9 kcal/mol, far right red triangle in the row “Mutation Order = 1” of Fig. 1.5). In combination with another mutation (row 2, the other mutation is not explicitly labeled in the diagram), the barrier is reduced in one case to around 17.5 kcal/mol and in G39A-

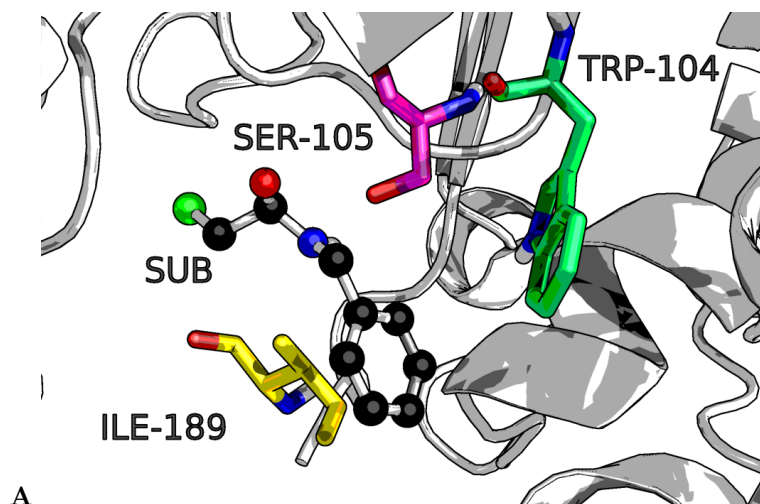


Figure 1.4 CALB active site, some residues omitted for clarity. SUB labels the substrate.

A141Q-I189G-L278A, the I189G mutation is among the top three mutants found (6.3 kcal/mol, row 4). Conversely, the mutant G39A-A141Q-L278A (not shown) has a moderately low barrier (10.9 kcal/mol). These results emphasize the fact that the activity modulating effect of specific point mutations can behave strongly non-linearly and non-additive and should be kept in mind when considering which mutations to include in further studies. Put differently, these findings illustrate that it is difficult to predict the activity of combination mutants based on single mutant data alone. Furthermore, since single mutant data is apparently not sufficient to predict combination mutant activity, screening a wide library of mutants should form an essential initial part of any enzyme engineering project. It thus becomes clear that computational activity screening can provide a major contribution to cost and material use reduction of the total enzyme engineering efforts.

Conclusions and Outlook. In the study presented in the next section, the method is improved with respect to a number of points. Among these is the fact that due to reasons of exceeding computational effort, not the complete enzyme structure was used in the calculations, but only amino acids within 8Å of the substrate

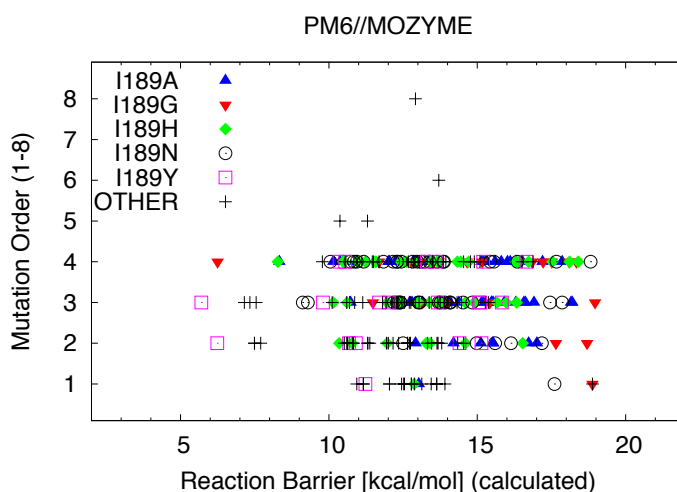


Figure 1.5 Screening of position 189. Each data point corresponds to a mutant with calculated reaction barrier on the x -axis and number of further mutations indicated by the value of the y -axis. The data point symbols indicate either of the point mutations I189A/G/H/N or Y. Data points/Mutants labeled as “Other” do not contain a mutation of I189. *E.g.* the far left purple square on row 3 indicates a mutant containing I189Y and two other mutations which are not explicitly indicated in the graph.

have been included in the model. The backbone sequence is therefore interrupted, introducing ambiguity in which amino acids to include in the model and in how to orient hydrogen atoms used to complete the valences at the interruption sites and possibly allowing for non-realistic structural rearrangements during the geometry optimizations of the structure. Secondly, the mutated positions were selected, in part, based on prior knowledge of the enzyme, instead of systematically screening all active site positions, however a working proof-of-concept is provided by the data. Lastly, it remained to be shown that the presented approach could also be adapted to mechanistically different reactions of other enzymes. These points are addressed in the study of the following section.

1.5.3 Engineering *Bacillus circulans* Xylanase

Relevance and Summary. In order to address the points summarized at the end of the previous section, activity of the xylanase of *Bacillus circulans* (BCX) towards an artificial substrate (*ortho*-nitrophenol- β -xylobioside), Fig. 1.6, is engineered. Understanding

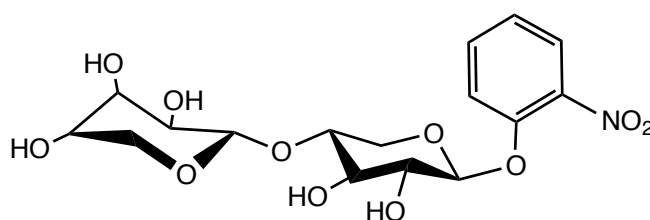


Figure 1.6 *ortho*-nitrophenol- β -xylobioside substrate used in BCX study consisting of two xylose units and ONP.

mechanisms of glycoside hydrolases can also be of major relevance to related fields such as biofuel production[Ragauskas *et al.* (2006); Yeoman *et al.* (2010); Zhang (2011); Gao *et al.* (2013)].

Structure and Mechanism. Due to its use in paper production[Bajpai (1999); Buchert *et al.* (1994)], BCX is also well characterized, a number of PDB structures are available and because of its interesting mechanistical features, it is also the subject of ongoing basic experimental research[Ludwiczek *et al.* (2013)]. In this study, the mechanism was taken from the literature as reported previously[Joshi *et al.* (2000, 2001)].

In a first step, the wild type (WT) reference reaction barrier is established. This is a crucial step because all subsequent conclusions about mutants are depending on how good the WT reaction barrier is estimated. As was done for CALB, only the rate determining step is modeled, Fig. 1.7. Noteworthy, in the formation of the covalently linked glycoside-enzyme complex, the substrate isomerises from a chair to a boat conformation. As outlined in the methods section, the reaction barrier is mapped out by fixing the reaction coordinate (x_1 in Fig. 1.7) at discrete, decrementing values while optimizing the rest of the protein structure. Since the reaction consists of two concerted events (nucleophilic attack

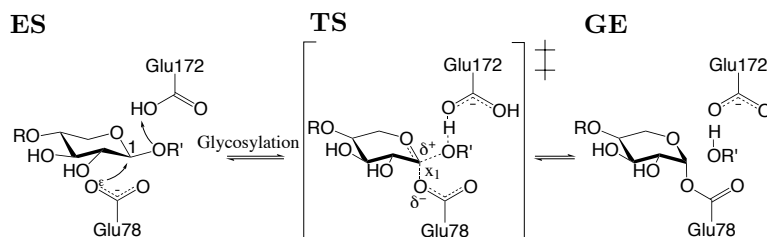


Figure 1.7 Rate determining step in conventional glycosylation. x_1 : constrained reaction coordinate; R : xylose; R' : *ortho*-nitrophenol (ONP). C_1 indicating nucleophilic carbon of first xylose unit[Hediger *et al.* (2013b)].

by E78 and proton transfer from E172 to ONP), in principle a two-dimensional potential energy surface could be mapped out to determine the reaction barrier. However, since proton transfer is associated with a rearrangement of the electronic structure, this reaction coordinate is left unconstrained and is handled solely by the quantum chemical method, *i.e.* the proton is free to transfer from or remain on E172. Forcing the quantum chemical model to do something it would not, if left unconstrained, most likely results in an increased reaction barrier and therefore it is advisable to keep the number of constraints as low as possible.

A PDB structure comprising a covalently linked inhibitor is available (1BVV) and is used as a starting point for the modeling procedure because it gives a very clear indication of how the substrate is located in the active site. As shown in Fig. 2 in Hediger *et al.* (2013b), two possible modeling pathways were worked out. In the first one, the ES complex is prepared from the GE structure *before* the GE structure is optimized (see footnote 6). Alternatively, the GE structure is optimized first and then used as a template for the ES structure. Within this alternative approach, the preparation of the ES structure then just requires minor modifications of substrate conformation before the structure is optimized.

Because in this second approach the ES structure very closely resembles the optimized GE structure, it is possible to concentrate the optimization of the ES structure only on the active site, while the remote part of the enzyme remains fixed at the optimized

geometry of the GE structure. This results in very large gains in computational efficiency, which depend on the layer of residues being optimized around the active site, Fig. 1.8A.¹⁰ Fixing, *i.e.* not reoptimizing an already optimized part of the enzyme however also results in slightly increased energy of the ES structure relative to the fully optimized GE structure and so the reaction barriers will be decreased, Fig. 1.8B. This increase (of ES energy relative to GE) results from multiple minor structural strain forming at the interface between the fixed and optimized layers of the structure.

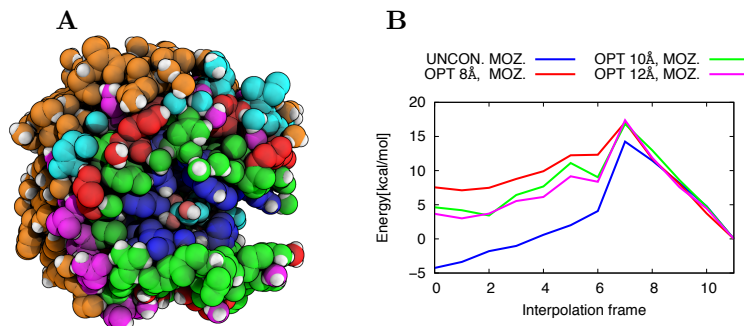


Figure 1.8 **A:** Illustration of optimization layers of BCX with bound ONP substrate. Color coded residue layers have at least one atom within the indicated distance to any atom of the substrate: *green* 8 Å, *red* 10 Å, *purple* 12 Å, *cyan* 14 Å. *Brown* residues remain at their optimized GE geometry, *dark blue* are residues which are screened for active mutations. The substrate is visible in the center of the dark blue residues. **B:** Calculated reaction barriers for different layers of optimized residues (based on Hediger *et al.*(2013b)). “MOZ.”: Energy evaluated using PM6 and the MOZYME linear scaling method.

Calculations and Experimental Verification. Based on this second approach, *i.e.* using the optimized GE structure as the template for the ES structure, and when not applying any constraints (apart from the reaction coordinate x_1), the reaction barrier is found to be 18.5 kcal/mol which is in very close agreement to the experimentally reported value of 17.0 kcal/mol obtained

¹⁰The structural constraints are parameters to the calculation program and are prepared programmatically using a PYMOL script. The way the parameters are submitted to the calculation may vary depending on the software used for calculations.

from transition state theory[Joshi *et al.* (2000)]. This strongly supports that the outlined modeling procedure and the quantum chemical method provide a meaningful model of the enzyme kinetics. As pointed out above, when applying constraints on parts of the structure which are located away from the active site, this can decrease the reaction barrier because the ES structure is increased in energy relative to the GE reference point (which is set to zero). However, it is reasonable to assume that this relative increase will be the same for all studied structures and so cancels out once mutants are compared relative to each other. Overall, application of constraints to remote parts of the enzyme is necessary to reach the desired calculation efficiency of one reaction barrier per day when running each interpolation step in parallel. Application of constraints has been widely adopted by the community[Himo (2006); Siegbahn and Himo (2009); Liao and Thiel (2012); Lonsdale *et al.* (2013)]. The time required for optimizing the structures with or without constraints are found to differ by a factor of six (Fig. 4B in Hediger *et al.* (2013b)).

Screening Results and Discussion of Mutants. By a similar approach as the one outline for the modeling of the WT reaction, it is found that different ways of modeling the set of mutants results in significantly different levels of efficiency, both in terms of required manual efforts as well as in computational time requirements. We refer to the original publication for details[Hediger *et al.* (2013b)] but point out that when, similarly as for the WT, the ES structures of the mutants are based on templates of optimized mutant GE structures, again large gains in computational efficiency are possible (see Fig. 7 in Hediger *et al.* (2013b)).

Using PYMOL[Schrödinger, LLC (2010)], models of all possible single mutants within a given distance of the substrate are prepared (excluding the catalytically active E78 and E172).¹¹ It is noted that in the current implementation of the presented approach, all mutated side chains are in their default ionization state at physiological pH. Furthermore, in a number of cases it is not possible to introduce a large side chain in a spatially restricted environment, such cases are discarded. Lastly, instead of using

¹¹The PYMOL script used for the preparation of the mutants is available at <https://github.com/mzhKU/Enzyme-Screening/blob/master/vsc-bcx.py>.

a side chain rotamer library, each mutated side chain is locally optimized (using an empirical builtin PYMOL function) in the environment of the enzyme before being submitted to the quantum calculation. This approach allowed to screen a set of 317 single mutants, where on average 29 hours are required to complete the calculations of a full reaction barrier using one CPU for each interpolation point.¹² From this set of single mutants, for every position i, j, \dots in the active site, the mutation resulting in the mutant with the lowest barrier X_i, Y_j, \dots is identified. Then, from this set of single mutants all possible double mutants (X_i, Y_j) are formed. Analysis of the reaction barriers of the obtained candidates reveals that the set of double mutants on average has lower activation barriers than the set of single mutants. Furthermore, the lowest barriers are found among the double mutants, Fig. 1.9.

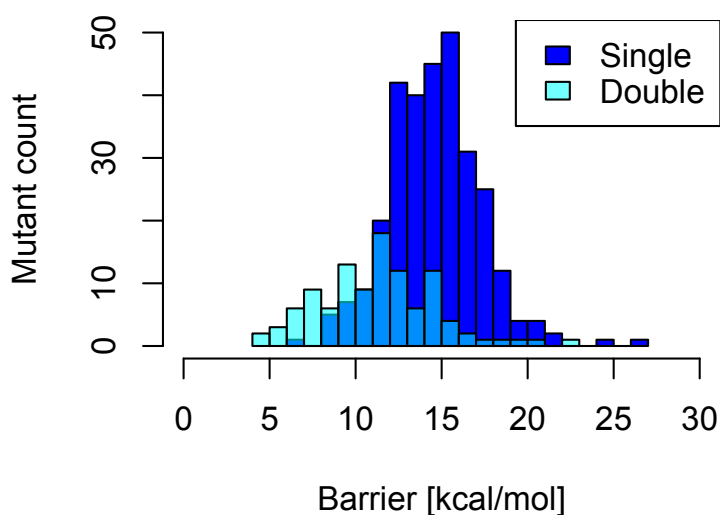


Figure 1.9 Barrier distribution of single and double BCX mutants[Hediger *et al.* (2013b)]. The plot is a histogram of reaction barriers, i.e. how many mutants are identified for a given reaction barrier.

Analysis of the set of mutants indicates that position 127 has

¹²All obtained barriers are provided at <http://www.scribd.com/doc/133445214/Supp-Mat-Paper-4>.

strong influence on the activity (9 of 20 best single mutants carry a mutation at that position). Inspection of the structural arrangement of the Q127W, W9D/E and N35E mutations points to a number of possible explanations of the observed lower barriers. One observation is highlighted in Fig. 1.10. As discussed above, the mechanism is postulated to involve

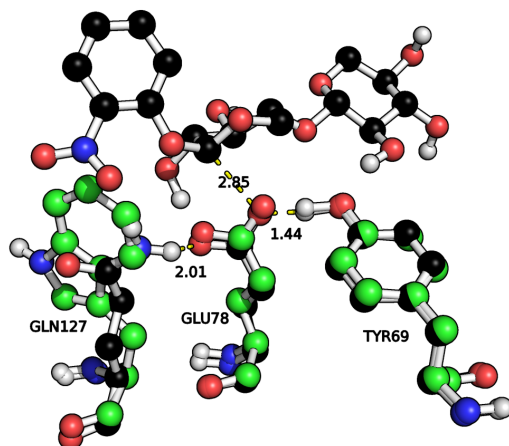


Figure 1.10 Overlay of Q127W mutant (green spheres) and wild type enzyme substrate complex. The Q127W mutation preserves structural integrity but prevents hydrogen bond formation to the nucleophilic E78. The substrate is in black spheres in the upper part of the figure. Distances in Å[Hediger *et al.* (2013b)].

a catalytically active, negatively charged E78 residue as the nucleophile. If Q127 stabilizes the negative charge by hydrogen bonding, the nucleophilic character of E78 is reduced. Replacing the potential hydrogen bond donor Q127 with a non-hydrogen bonding substitute that can preserve overall structural integrity with a side chain of similar size, such as Q127W, could result in increased nucleophilicity of E78, Fig. 1.10. Furthermore, as illustrated in Fig. 1.7, the transition state involves a partial positive charge on C₁ of the substrate. Negative charges on side chains (such as in W9D/E) could act as stabilizers of this negative charge

and increase catalytic activity in doing so, *i.e.* by stabilization of the transition state. Similarly, when ionized, N35E could act in stabilizing C₁ or when in neutral protonation state, it could also be acting as a hydrogen bond donor to the catalytically active E172. This rationalization of N35 mutations would be in agreement with a recently proposed *reverse protonation* mechanism[Joshi *et al.* (2000)], which still is the subject of ongoing research.

Conclusions and Outlook. In overcoming the points raised at the end of the previous section, we note that in this application the method is extended such that the reaction barriers of all active site single mutants are screened. Furthermore, the complete BCX structure is used as a model, rather than just a large cluster model as was done for CALB. Contrary to the results from the CALB study, we identified numerous mutants with barriers which are considerably lower than the WT barrier. This fact might however also be attributed to significantly promoted catalysis by a better fitting of the substrate in the active site compared to the substrate fitting in the CALB study. A further important finding is that by applying constraints to the outer, non-active site part of the enzyme, the computational demands for structural optimizations can be reduced to around two days, this makes the method sufficiently fast for industrial applications. It is noted however, that most likely application of similarly layered constraints in the CALB study would have permitted to use the complete structure in the optimizations without leading to too large computational time requirements.

1.6 Conclusions

It is concluded that using linear scaling methods and a number of customized software packages, it is possible to carry out large scale quantum chemistry based enzyme activity screening studies. An important finding is that single mutant data is not enough to predict combination mutant activities because the activity modulating effects of the point mutations behave non-linearly when in combination with other mutations. In detail this means that two mutations which appear activating as single mutants are not

24 | *Computational Enzyme Engineering: Activity Screening using Quantum Chemistry.*

necessarily activating when in combination in a double mutant. Also, a deactivating single mutation can become an activating mutation when in combination with other mutations (I189G in the CALB study). On average, double mutants, as seen in the BCX study, have lower reaction barriers than single mutants and a systematic screening of the active site is required to identify the mutants with highest apparent activity. Lastly, using such a screening approach allows not only to identify mutants for further characterization by higher level methods or experimental verification, but also to formulate hypothesis on how the various mutations actually contribute to increased catalysis.

As an outlook on the computational enzyme engineering field, we note that some improvements and developments are anticipated. Among these is to consider later reaction steps such as isomerizations or product release steps. Also, it would be of great value to the field if the software would evolve to a state where it can be routinely used with much less manual and customization efforts to setup, carry out and analyse the calculations.

Bibliography

- Agresti, J. J., Antipov, E., Abate, A. R., Ahn, K., Rowat, A. C., Baret, J.-C., Marquez, M., Klibanov, A. M., Griffiths, A. D. and Weitz, D. A. (2010). Ultrahigh-throughput screening in drop-based microfluidics for directed evolution, *Proceedings of the National Academy of Sciences* **107**, 9, pp. 4004–4009.
- Bachega, J. F. R., Timmers, L. F. S., Assirati, L., Bachega, L. R., Field, M. J. and Wymore, T. (2013). Gtcdynamo: A pymol plug-in for qc/mm hybrid potential simulations, *Journal of computational chemistry* **34**, 25, pp. 2190–2196.
- Bajpai, P. (1999). Application of enzymes in the pulp and paper industry, *Biotechnology Progress* **15**, 2, pp. 147–157.
- Beilen, J. and Li, Z. (2002). Enzyme technology: an overview, *Current Opinion in Biotechnology* **13**, 4, pp. 338–344.
- Bornscheuer, U. T. and Kazlauskas, R. J. (2004). Catalytic promiscuity in biocatalysis: using old enzymes to form new bonds and follow new pathways, *Angewandte Chemie International Edition* **43**, 45, pp. 6032–6040.
- Buchert, J., Tenkanen, M., Kantelinen, A. and Viikari, L. (1994). Application of xylanases in the pulp and paper industry, *Bioresource technology* **50**, 1, pp. 65–72.
- Chaput, L., Sanejouand, Y.-H., Balloumi, A., Tran, V. and Graber, M. (2012). Contribution of both catalytic constant and michaelis constant to calb enantioselectivity: Use of fep calculations for prediction studies, *Journal of Molecular Catalysis B: Enzymatic* **76**, pp. 29–36.
- Claeyssens, F., Harvey, J., Manby, F., Mata, R., Mulholland, A., Ranaghan, K., Schütz, M., Thiel, S., Thiel, W. and Werner, H. (2006). High-Accuracy Computation of Reaction Barriers in Enzymes, *Angewandte Chemie* **118**, 41, pp. 7010–7013.
- Cramer, C. J. (2013). *Essentials of computational chemistry: theories and models* (John Wiley & Sons).
- Gao, D., Chundawat, S. P., Sethi, A., Balan, V., Gnanakaran, S. and Dale, B. E. (2013). Increased enzyme binding to substrate is not necessary for more efficient cellulose hydrolysis, *Proceedings of the*

26 | Bibliography

- National Academy of Sciences* **110**, 27, pp. 10922–10927.
- Gayot, S., Santarelli, X. and Coulon, D. (2003). Modification of flavonoid using lipase in non-conventional media: effect of the water content, *Journal of biotechnology* **101**, 1, pp. 29–36.
- Goodsell, D., Morris, G. and Olson, A. (1996). Automated docking of flexible ligands: applications of autodock, *Journal of Molecular Recognition* **9**, 1, pp. 1–5.
- Gotor-Fernández, V., Busto, E. and Gotor, V. (2006). Candida antarctica lipase b: an ideal biocatalyst for the preparation of nitrogenated organic compounds, *Advanced synthesis & catalysis* **348**, 7-8, pp. 797–812.
- Hediger, M. R., De Vico, L., Rannes, J. B., Jäckel, C., Besenmatter, W., Svendsen, A. and Jensen, J. H. (2013a). In silico screening of 393 mutants facilitates enzyme engineering of amidase activity in calb, *PeerJ* **1**, p. e145.
- Hediger, M. R., De Vico, L., Svendsen, A., Besenmatter, W. and Jensen, J. H. (2012). A computational methodology to screen activities of enzyme variants, *PLoS ONE* **7**, 12, p. e49849, doi:10.1371/journal.pone.0049849, URL <http://dx.doi.org/10.1371/journal.pone.0049849>.
- Hediger, M. R., Steinmann, C., De Vico, L. and Jensen, J. H. (2013b). A computational method for the systematic screening of reaction barriers in enzymes: searching for bacillus circulans xylanase mutants with greater activity towards a synthetic substrate, *PeerJ* **1**, p. e111.
- Hedstrom, L. *et al.* (2002). Serine protease mechanism and specificity, *Chemical reviews* **102**, 12, pp. 4501–4524.
- Himo, F. (2006). Quantum chemical modeling of enzyme active sites and reaction mechanisms, *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)* **116**, 1, pp. 232–240.
- Jensen, F. (2007). *Introduction to computational chemistry* (John Wiley & Sons).
- Joshi, M., Sidhu, G., Nielsen, J., Brayer, G., Withers, S. and McIntosh, L. (2001). Dissecting the electrostatic interactions and pH-dependent activity of a family 11 glycosidase, *Biochemistry* **40**, 34, pp. 10115–10139.
- Joshi, M., Sidhu, G., Pot, I., Brayer, G., Withers, S. and McIntosh, L. (2000). Hydrogen bonding and catalysis: a novel explanation for how a single amino acid substitution can change the ph optimum of a glycosidase1, *Journal of molecular biology* **299**, 1, pp. 255–279.
- Kirk, O., Borchert, T. V. and Fuglsang, C. C. (2002). Industrial enzyme applications, *Current opinion in biotechnology* **13**, 4, pp. 345–351.

- Liao, R.-Z. and Thiel, W. (2012). Comparison of qm-only and qm/mm models for the mechanism of tungsten-dependent acetylene hydratase, *Journal of Chemical Theory and Computation* **8**, 10, pp. 3793–3803.
- Lind, M. E. and Himo, F. (2013). Quantum chemistry as a tool in asymmetric biocatalysis: Limonene epoxide hydrolase test case, *Angewandte Chemie* **125**, 17, pp. 4661–4665.
- Lonsdale, R., Houghton, K. T., Zurek, J., Bathelt, C. M., Foloppe, N., de Groot, M. J., Harvey, J. N. and Mulholland, A. J. (2013). Quantum mechanics/molecular mechanics modeling of regioselectivity of drug metabolism in cytochrome p450 2c9, *Journal of the American Chemical Society* **135**, 21, pp. 8001–8015.
- Ludwiczek, M. L., D'Angelo, I., Yalloway, G. N., Brockerman, J. A., Okon, M., Nielsen, J. E., Strynadka, N. C., Withers, S. G. and McIntosh, L. P. (2013). Strategies for modulating the pH-dependent activity of a family 11 glycoside hydrolase, *Biochemistry* **52**, 18, pp. 3138–3156.
- Meyer, H.-P., Eichhorn, E., Hanlon, S., Lütz, S., Schürmann, M., Wohlgemuth, R. and Coppolecchia, R. (2013). The use of enzymes in organic synthesis and the life sciences: perspectives from the swiss industrial biocatalysis consortium (sibc), *Catalysis Science & Technology* **3**, 1, pp. 29–40.
- Naik, S., Basu, A., Saikia, R., Madan, B., Paul, P., Chatterjee, R., Brask, J. and Svendsen, A. (2010). Lipases for use in industrial biocatalysis: Specificity of selected structural groups of lipases, *Journal of Molecular Catalysis B: Enzymatic* **65**, 1-4, pp. 18–23.
- Nakagawa, Y., Hasegawa, A., Hiratake, J. and Sakata, K. (2007). Engineering of *Pseudomonas aeruginosa* lipase by directed evolution for enhanced amidase activity: mechanistic implication for amide hydrolysis by serine hydrolases, *Protein Engineering Design and Selection* **20**, 7, pp. 339–346.
- Noodleman, L., Lovell, T., Han, W., Li, J. and Himo, F. (2004). Quantum chemical studies of intermediates and reaction pathways in selected enzymes and catalytic synthetic systems, *Chemical reviews* **104**, 2, pp. 459–508.
- Ragauskas, A. J., Williams, C. K., Davison, B. H., Britovsek, G., Cairney, J., Eckert, C. A., Frederick, W. J., Hallett, J. P., Leak, D. J., Liotta, C. L. *et al.* (2006). The path forward for biofuels and biomaterials, *science* **311**, 5760, pp. 484–489.
- Rezác, J., Fanfrlik, J., Salahub, D. and Hobza, P. (2009). Semiempirical quantum chemical pm6 method augmented by dispersion and h-bonding correction terms reliably describes various types of noncovalent complexes, *Journal of Chemical Theory and Computation* **5**, 7, pp. 1749–1760.

28 | Bibliography

- Rezáč, J. and Hobza, P. (2011). A halogen-bonding correction for the semiempirical pm6 method, *Chemical Physics Letters* **506**, 4, pp. 286–289.
- Schenker, S., Schneider, C., Tsogoeva, S. and Clark, T. (2011). Assessment of popular dft and semiempirical molecular orbital techniques for calculating relative transition state energies and kinetic product distributions in enantioselective organocatalytic reactions, *Journal of Chemical Theory and Computation* .
- Schmid, A., Hollmann, F., Park, J. B. and Bühler, B. (2002). The use of enzymes in the chemical industry in europe, *Current Opinion in Biotechnology* **13**, 4, pp. 359–366.
- Schrödinger, LLC (2010).
- Siegbahn, P. E. and Himo, F. (2009). Recent developments of the quantum chemical cluster approach for modeling enzyme reactions, *JBIC Journal of Biological Inorganic Chemistry* **14**, 5, pp. 643–651.
- Stewart, J. (1990). Mopac: a semiempirical molecular orbital program, *Journal of Computer-Aided Molecular Design* **4**, 1, pp. 1–103.
- Stewart, J. (1996). Application of localized molecular orbitals to the solution of semiempirical self-consistent field equations, *International Journal of Quantum Chemistry* **58**, 2, pp. 133–146.
- Stewart, J. (2007). Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements, *Journal of Molecular Modeling* **13**, 12, pp. 1173–1213.
- Stewart, J. (2009). Application of the PM6 method to modeling proteins, *Journal of molecular modeling* **15**, 7, pp. 765–805.
- Svedendahl, M., Carlqvist, P., Branneby, C., Allnr, O., Frise, A., Hult, K., Berglund, P. and Brinck, T. (2008). Direct epoxidation in candida antarctica lipase b studied by experiment and theory, *ChemBioChem* **9**, 15, pp. 2443–2451, doi:10.1002/cbic.200800318, URL <http://dx.doi.org/10.1002/cbic.200800318>.
- Uppenberg, J., Hansen, M. T., Patkar, S. and Jones, T. A. (1994). The sequence, crystal structure determination and refinement of two crystal forms of lipase b from candida antarctica, *Structure* **2**, 4, pp. 293–308.
- Uppenberg, J., Oehrner, N., Norin, M., Hult, K., Kleywegt, G., Patkar, S., Waagen, V., Anthonsen, T. and Jones, T. (1995). Crystallographic and molecular-modeling studies of lipase B from Candida antarctica reveal a stereospecificity pocket for secondary alcohols, *Biochemistry* **34**, 51, pp. 16838–16851.
- Yeoman, C. J., Han, Y., Dodd, D., Schroeder, C. M., Mackie, R. I. and Cann, I. K. (2010). Thermostable enzymes as biocatalysts in the biofuel industry, *Advances in applied microbiology* **70**, pp. 1–55.

- Young, D. (2004). *Computational chemistry: a practical guide for applying techniques to real world problems* (John Wiley & Sons).
- Zhang, Y.-H. P. (2011). What is vital (and not vital) to advance economically-competitive biofuels production, *Process Biochemistry* **46**, 11, pp. 2091–2110.