

Publishers' page

Publishers' page

Publishers' page

Publishers' page

Dedication Page
(optional)

Contents

1. Computational Enzyme Engineering: Activity Screening using Quantum Chemistry.	1
1.1 Abstract	1
1.2 Motivation	2
1.3 Introduction	4
1.4 Methods	7
1.4.1 Calculation Engines	7
1.4.2 Molecular Modeling	8
1.4.3 Modeling Software	11
1.5 Applications	13
1.5.1 Overview	13
1.5.2 Engineering <i>Candida antarctica</i> Lipase B .	13
1.5.3 Engineering <i>Bacillus circulans</i> Xylanase .	19
1.6 Conclusions	26

Chapter 1

Computational Enzyme Engineering: Activity Screening using Quantum Chemistry.

Martin R. Hediger^a

^a *Zürich*
m—@—.ch

1.1 Abstract

- QM Screening of CALB and BCX
- Screening 317 single and 111 double mutants with av. 29 h per barrier and 1 CPU per interpolation point
- Correct mechanisms using PM6
- Barriers in agreement (calc. 18.5, exp. 17.0 kcal/mol)
- 15/22 correctly predicted (using many approximations and not full model)
- No bias, more active double mutants identified
- largely automated
- For CALB lowest barriers close to WT, BCX lowest barriers lower than WT

Book Title

Editor Name

Copyright © Year Pan Stanford Publishing Pte. Ltd.

ISBN

www.panstanford.com

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 $k_a = 1000s^{-1}$ while company b uses an enzyme with $k_b = 2000s^{-1}$. 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 required to keep the reaction volume at a certain temperature. 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[Meyer *et al.* (2013); Kirk *et al.* (2002); Beilen and Li (2002); Schmid *et al.* (2002)].

Laboratory costs can however be saved to a large part if the development is carried out *in silico*. The proof that computational 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. It is therefore natural to ask: Why hasn't this become more important yet? Why are there not more job openings at pharmaceutical and biotechnology companies looking for computational enzyme engineers? The field, after all, did receive the Nobel price in Chemistry in 2013. As a side note, computational modeling of aero- and fluid dynamics, semiconductor properties, pharmacokinetics/pharmacodynamics or civil engineering is well beyond the point of having to justify itself as of strategic relevance. More important questions asked and answered in this review:

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

- Is it possible to screen for mutant activity using quantum chemical methods?
- Which quantum chemical model can reproduce the accepted mechanism?
- What are the time requirements depending on the model size?
- How should the program be configured?
- What modeling procedure/sequence is needed? Which tools are useful?

By giving an overview and some details in this article, I hope to, at least in part, answer the above questions.

[ARE THE QUESTIONS ANSWERED?]

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, as is well known in the community, it is almost impossible to predict, how a mutated variant of an enzyme will behave[REFS?, Bornscheuer?]. Will it at all be possible to express it and will it even fold and if so, will it be more or less active and specific 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 not 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 such initial computational screening is not 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

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

value is the provision of a systematically developed library of mutants. Based on the screening, promising candidates can then be pipelined into a more demanding computational method for further characterization.

[Physical property of interest:]

Ideally the physical property of interest guides the selection of the appropriate computational method and not the other way around, therefore a brief overview on computational methods is provided. The method has to be adapted to the physical property of interest

- Thermal and structural integrity
- pH dependent activity[Ludwiczek *et al.* (2013)]
- Active site hydration
- Binding, *c.f.* drug design where the target protein is constant
- **Binding effects not considered**
- Catalysis: *C.f.* biofuel production or synthesis applications where the substrate is mostly non-variable.

[Different modeling objectives:]

- 1) One Enzyme – Multiple Substrates: These tasks are usually addressed by docking approaches and studies involving millions of substrates have been published[Zhou and Caflisch (2010)].
- 2) Multiple Enzymes – One Substrate: The approaches to this task, as it appears, are much less established. Why is that? A number of factors can be of influence among which are that modeling of different variants of an enzyme will also mean to differentiate between multiple (possibly coupled) protonation states which is non-trivial. Furthermore, modeling of reactivity always involves modeling of a transition state, which is non-trivial and for which there is no standard fail safe approach.

[Application context:]

- Academic: screening can assist in generation of new hypothesis about structure/function relationships
- Commercial: commercial application is restricted by time-to-delivery of predictions, accuracy/reliability, screening spectrum and convenience of use. Notefully, in a commercial context, accuracy can be of less value than reliable qualitative prediction of relative activities.

[Review structure and audience:]

This review is written for a general audience. For specialized

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

details, the reader is referred to the primary literature. Since molecular modeling of enzymes has been reviewed extensively over the recent years, we do not attempt to provide all fundamentals of the subject. Instead, we focus on screening aspects of computational enzyme engineering. The intention is to provide an accessible introduction into the field of quantum chemical modeling of enzymatic reactions, highlight advantages of using quantum chemistry and also point out situations where special care needs to be taken.

By 2014, quantum chemical methods have been in use to study enzymatic reactions for around three decades. What is new however is that quantum chemical methods can now be used also in *in silico* screening assays of enzymatic activity and this is also part of the focus of this review.

Applications. We will discuss application of the approach to the modeling of two enzymes: CALB and BCX.

1.4 Methods

1.4.1 Calculation Engines

The focus of this article is on the application of quantum chemistry for the screening of enzyme mutants. So instead of attempting to incompletely review modern computational chemistry methods, only a few selected facts relevant for the following are highlighted. A number of text books on the subject have been published[Young (2004); Jensen (2007); Cramer (2013)].

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 about structural rearrangement of an enzyme.

Chemical reactivity however is accompanied by changes in the electronic structure of the system which require quantum chemical methods to be described. For the sake of the argument, semi-empirical, density functional and *ab initio* methods are understood to form the set of quantum chemical methods.

Quantum Chemical Methods. Since quantum chemical methods each provide their own description of the wave function (or electron density in density functional methods) and all properties, such as energy, 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 to at least ensure that the obtained conclusions are consistent with the method against which results are compared. 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

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

conclusions can be made using this method.

For the discussion of the applications, the semi-empirical PM6 method[Stewart (2007, 2009)] together with the MOZYME technology[Stewart (1996)] are most relevant. In a comparative study, PM6 is found to provide results approaching density functional accuracy[Schenker *et al.* (2011)]. Furthermore, calculations on small active site models of the two enzymes allow to reproduce the generally accepted mechanisms. The MOZYME technology provides a method of localizing the molecular orbitals obtained during the self-consistent field calculation which results in computational demands scaling linearly with system size. Additionally, the implementation of the PM6 method is actively being developed further to take into account properties such as dispersion or solvation[Rezác *et al.* (2009); Rezác and Hobza (2011)]. In essence, these two technologies in combination therefore provide the accuracy and efficiency required for modeling of enzymatic reactivity in a screening assay.

1.4.2 Molecular Modeling

A variety of molecular modeling approaches has been presented, 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 are the so called cluster models, 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. 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.

Full Enzyme Models. Using a full model of the enzyme structure prevents such ambiguity, but raises a number of issues as well. 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. Calculations of full enzyme models also require to consider automated assignment of ionization states to amino acids. While a user can easily assign ionization states to a few dozen amino acids, such an approach becomes likely error prone in structures consisting of hundreds of amino acids. 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 computationally sufficiently efficient and allow for some degree of programmatic customization, which is of great relevance for later data analysis.

We have used such full enzyme models in two studies where new functionality has been introduced into an enzyme active site by quantum chemical methods[Hediger *et al.* (2012, 2013a,b)].

In both studies, the basic procedure consists of preparing structures for the enzyme substrate complex (“ES”) and the first intermediate (CALB: “TI”⁴, BCX: “GE”⁵) and calculating the reaction barrier between these two pairs of structures, respectively. The reaction barrier is calculated by interpolating between the ES_{CALB} and TI or ES_{BCX} GE structures, *i.e.* by preparing a set of 10 intermediate structures which approximate the structure of the system along the reaction coordinate.⁶ By evaluating the energies for each interpolation frame for the two reactions ES_{CALB} → TI or ES_{BCX} → GE, it becomes possible to obtain an approximate potential energy surface of the reaction, Fig. 1.1. The quality of the results of this approach are strongly dependent on how careful the modeling of each of these structures is carried out. A frequent reason for meaningless results is significant structural

⁴Tetrahedral intermediate

⁵Glycosyl enzyme

⁶In the literature, the terms *interpolation*, *linear transit scan*, *reaction coordinate calculation* or *adiabatic mapping* are frequently used to indicate the same kind of procedure.

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

rearrangement during the geometry optimization, requiring to interpolate between very different conformations of the two stationary points on each side of the reaction barrier.

Finally, the activities of different variants of an enzyme are ordered by comparing the activation energy required for the rate determining step of the overall reaction – low activation energies correspond to high anticipated activity of the variant. Implicitly this assumes high substrate concentration such that k_{cat} becomes rate limiting, a reasonable assumption under industrial conditions.

[Requirements.] The most basic requirements are that a structure representing a bound intermediate along the reaction pathway is available. Alternatively, a homology model could be used. Furthermore, ideally the mechanism for the rate determining step is established in the literature. In case no generally accepted mechanism is reported, a preceeding study could analyse kinetically competing mechanism alternatives.

- The structures of the enzyme in the relevant state of the reaction need to be available, ie. an enzyme-substrate complex (ES) and a transition state (TS) analog (usually an inhibited species)
- If the mechanism is not established, the reaction can usually follow various pathways – these would need to be checked in order to find a best guess at the rate determining step
- Identification of the transition state is difficult and depends intimately on the structure, optimizing two structures differing by 0.1 Å in a particular bond length can result in identification or not identification of the TS
- **Focus on catalysis, assume binding is similar over mutant range[Ludwiczek *et al.* (2013)]- K_{M} s are similar for mutants**

In the presented approach, large structural rearrangements can not be studied.

[Aspects of model size and modeling.] It is frequently observed that specific pairs of mutants can not be accommodated together because of the size of the side chains. In order to facilitate

the approach, these variants are discarded from the screen. Furthermore, side chain orientations can be oriented in various ways and are defaulted using some approach (local, empirical function based optimization in PyMOL). At the least, this removes ambiguity in the modeling step. Ionization states of side chains greatly complicate the screening and the best approach most likely is to consider different ionization states only at a later stage of the assay.

[Interpolation/Linear transit/Adiabatic mapping.]

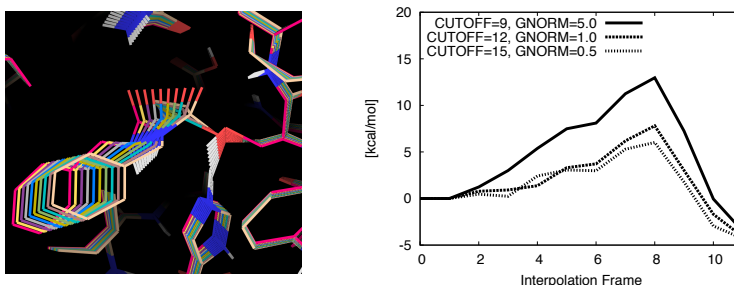


Figure 1.1 Intermediate structures along reaction coordinate of CALB reaction.

1.4.3 Modeling Software

Explain available software: GAMESS, MOPAC

Various engines are implemented in different calculation software. Methods on paper are of limited use if they can not be applied, therefore also available modeling software is a very important factor.

Modeling software is available, PyMOL, Schrödinger, Chem?? (Walter Thiel, commercially) but everything is still very much custom work (especially the data analysis, programs output format is like from last century), this is probably the major reason preventing the use of QM as standard method in industry. Compare docking and MD methods have software which is designed towards user friendliness and has become established in industry. GTKDynamo is an exceptionally good example for what it should

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

look like, unfortunately only small community.

Many studies claim that “*future developments are directed at further automating the presented approach*”. Obviously these remarks are directed at making the method industrially applicable by avoiding the requirement to carry out complicated setup and configuration steps[Rathore *et al.* (2013)]. In reality, such future developments, to the great unfortune of the whole field, are rarely realized – speculation on the reasons is beyond the scope of this article. Nevertheless, almost no publication fails to mention the potential applications in drug design. Computational considerations are only rarely taken into account but greatly help at putting the whole approach into perspective[Buch *et al.* (2011)].

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

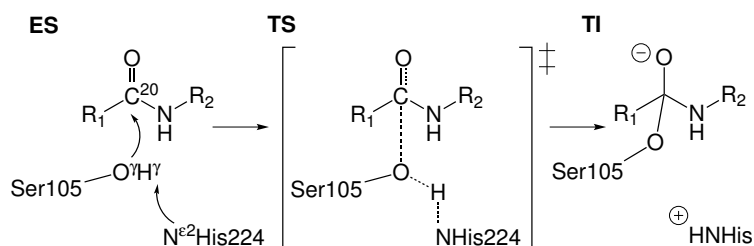


Figure 1.2 Reaction scheme for the formation of the tetrahedral intermediate in CALB. In this work, R_1 : $-\text{CH}_2\text{-Cl}$, R_2 : $-\text{CH}_2\text{-C}_5\text{H}_6$ [Hediger *et al.* (2013a)].

is believed to be the nucleophilic attack by O^γ of S105 on the carbonyl carbon C^{20} of the substrate.

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 issues need to be addressed. 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 barrier⁷. 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

⁷If a mutant has a lower reaction barrier than the WT, it is understood to have higher activity.

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 experimental activities⁸, 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 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

⁸The experimental improvement factors range from 0 to roughly 11 times wild type activity.

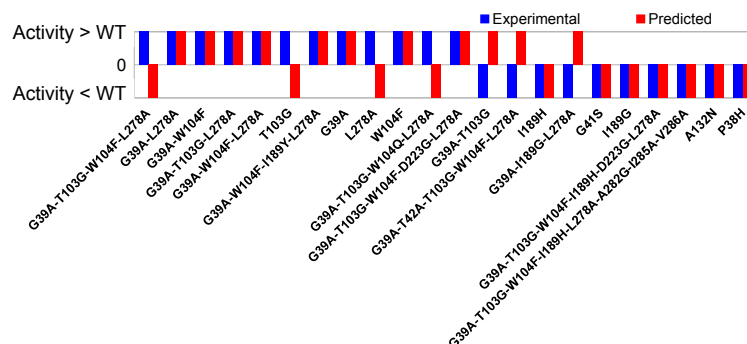


Figure 1.3 Agreement between predicted and experimental qualitative activity relative to wild type[Hediger *et al.* (2013a)].

large library, *i.e.* the focus is on obtaining qualitative conclusions rather than quantitative results[Agresti *et al.* (2010)]– this is the true value of computational screening. 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. Identification of mutants with reaction barriers significantly lower than the wild type is found to be 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”) or where the enzyme already has reached very high activity towards a specific substrate[Agresti *et al.* (2010)].

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

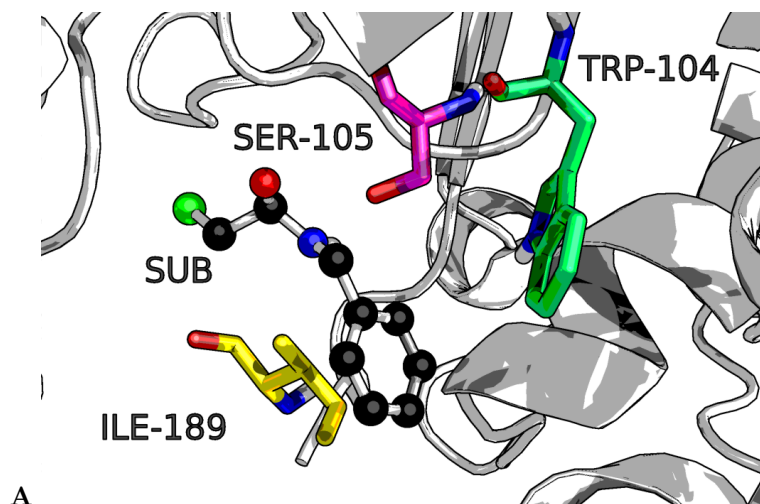


Figure 1.4 CALB active site, some residues omitted for clarity. SUB labels the substrate.

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 position 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 (not shown), the barrier is reduced in one case to around 17.5 kcal/mol and in G39A-A141Q-I189G-L278A, the I189G mutation is among the top three mutants found (6.3 kcal/mol). Conversely, the mutant G39A-A141Q-L278A has a moderately low barrier (10.9 kcal/mol). These results emphasize the fact that the activity modulating effect of specific

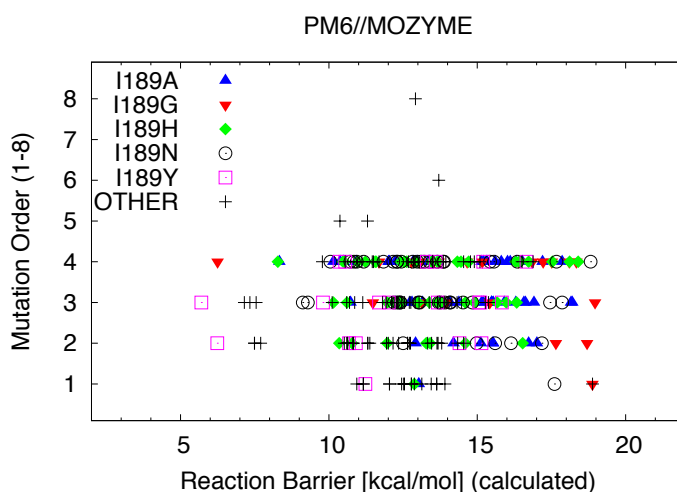


Figure 1.5 Screening of position 189. Each data point corresponds to a mutant with calculated reaction barrier on the x -axis and number of mutations indicated by the value of the y -axis. The data point symbols indicate either of the point mutations I189A/G/H/N or Y. Data points labeled as “Other” do 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 indicated in the graph.

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 not the complete enzyme structure was used in the calculations, due to reasons of exceeding

computational effort, instead only amino acids within 8Å of the substrate 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. 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.

[CONCLUSIONS?] Regarding the initially stated criteria against which to compare a computational screening method, it is noted that the method is sufficiently efficient to screen sets of hundreds of mutants, the predictivity is fairly robust however the method could be improved in terms of user-friendliness.

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

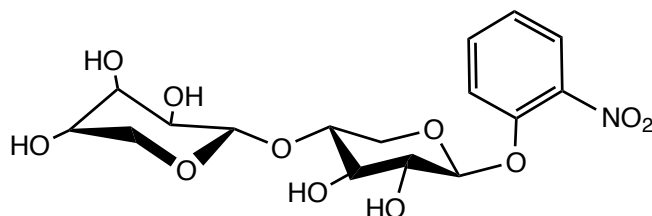


Figure 1.6 *ortho*-nitrophenol- β -xylobioside substrate used in BCX study consisting of two xylose units and ONP.

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 to 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. Future improvements of the

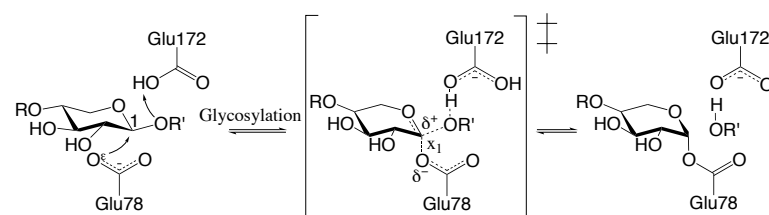


Figure 1.7 Rate determining step in conventional glycosylation. x_1 : constrained reaction coordinate; R : xylose; R' : *ortho*-nitrophenol (ONP). C_1 indicating nucleophilic carbon of first xylose unit[Hediger *et al.* (2013b)].

method might consider later reaction steps such as isomerizations or product release. 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 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 5). 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.⁹ 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. 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. The observations are summarized in Fig. 1.8.

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

⁹The 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.

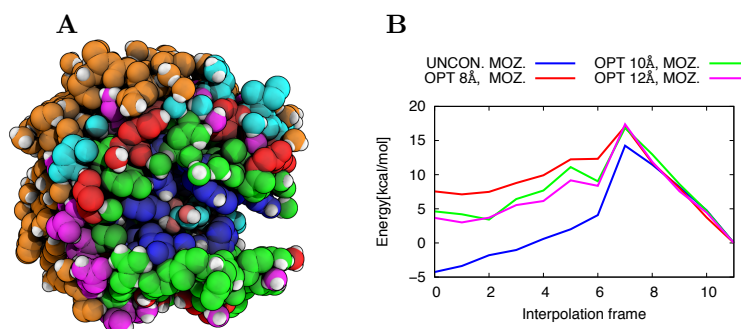


Figure 1.8 **A:** Illustration of optimization layers of BCX with bound ONP substrate. Color coded residue layers have at least one atom within the indicated distance to any atom of the substrate: *green* 8 Å, *red* 10 Å, *purple* 12 Å, *cyan* 14 Å. *Brown* residues remain at their optimized GE geometry, *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 the MOZYME linear scaling method.

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 from transition state theory[Joshi *et al.* (2000)]. This strongly supports that the outlined modeling procedure, the configuration of the quantum chemical method and the included/omitted physical effects provide a meaningful model of the enzyme kinetics. As pointed out above, when applying constraints on remote parts of the structure, 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. Furthermore, 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).¹⁰ 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 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.¹¹ 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 $\{K_l\}$, 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

¹⁰The PYMOL script used for the preparation of the mutants is available at <https://github.com/mzhKU/Enzyme-Screening/blob/master/vsc-bcx.py>.

¹¹All obtained barriers are provided at <http://www.scribd.com/doc/133445214/Supp-Mat-Paper-4>.

double mutants, Fig. 1.9.

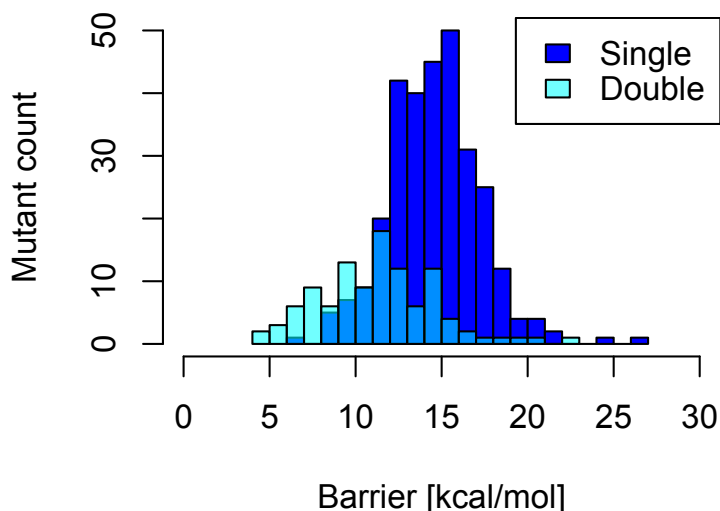


Figure 1.9 Barrier distribution of single and double BCX mutants[Hediger *et al.* (2013b)]. The plot is a histogram of reaction barriers, i.e. how many mutants are identified for a given reaction barrier.

Analysis of the set of mutants indicates that position 127 has 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 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 A. Furthermore, as illustrated in Fig. 1.7, the transition state involves a partially positive charge on C_1 of the substrate. Negative charges on side chains (such

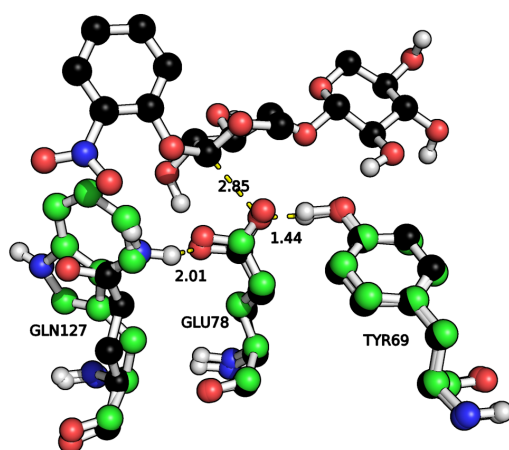


Figure 1.10 Overlay of Q127W mutant (green spheres) and wild type enzyme substrate complex. The Q127W mutation preserves structural integrity but prevents hydrogen bond formation to the nucleophilic E78. The substrate is in black spheres in the upper part of the figure. Distances in Å[Hediger *et al.*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

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 but when in neutral protonation state, it could be acting as a hydrogen bond donor to the catalytically active E172. This rationalization of N35 mutations would be related to a recently proposed *reverse protonation* mechanism[Joshi *et al.* (2000)], which still is the subject of ongoing research.

[Conclusions and Outlook.] 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 be attributed to significantly promoted catalysis by a better fitting of the substrate in the active site.

1.6 Conclusions

The three most significant findings are that firstly, using linear scaling methods and a number of customized software packages it is possible to carry out large scale computational enzyme activity screening studies. Secondly, 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. And 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. These findings can then be further exploited in future enzyme engineering efforts.

Bibliography

- Agresti, J. J., Antipov, E., Abate, A. R., Ahn, K., Rowat, A. C., Baret, J.-C., Marquez, M., Klivanov, 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.
- 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.
- Buch, I., Giorgino, T. and De Fabritiis, G. (2011). Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations, *Proceedings of the National Academy of Sciences* **108**, 25, pp. 10184–10189.
- 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.
- Claeysens, 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*

28 | 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.
- 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.
- 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.
- Rathore, R., Sumakanth, M., Reddy, M. S., Reddanna, P., Rao, A. A., Erion, M. D. and Reddy, M. (2013). Advances in binding free energies calculations: Qm/mm-based free energy perturbation method for drug design, *Current pharmaceutical design* **19**, 26, pp. 4674–4686.
- 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.

30 | 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. (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.
- Zhou, T. and Caffisch, A. (2010). High-Throughput Virtual Screening Using Quantum Mechanical Probes: Discovery of Selective Kinase Inhibitors, *ChemMedChem* **5**, 7, pp. 1007–1014.