

Publishers' page

Publishers' page

Publishers' page

Publishers' page

Dedication Page  
(optional)



## Contents

1. Working title: Calculations	1
1.1 Motivation . . . . .	1
1.2 Introduction . . . . .	2
1.3 Methods . . . . .	3
1.3.1 Idea collection . . . . .	4
1.3.2 Calculation methods . . . . .	5
1.4 Applications . . . . .	6
1.4.1 General overview . . . . .	6
1.4.2 Interesting QM/MM applications . . . . .	12
1.4.3 Requirements for advancement of methods	12
1.5 Conclusions . . . . .	12
1.5.1 Advantages/Disadvantages of using quantum chemistry for modeling enzymatic reactions . . . . .	12

## Chapter 1

# Working title: Calculations

Martin R. Hediger<sup>a</sup>, Harm Otten<sup>b</sup>

<sup>a</sup> *Zürich*

<sup>b</sup> *København*

*m—@—.ch*

*harm.otten@maxlab.lu.se*

### 1.1 Motivation

Let us imagine two companies A and B. Both companies use very similar technical equipment to carry out the biotechnological process  $a \xrightarrow{\text{Enzyme}} b$  using two individual versions of the indicated catalyst. 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.<sup>1</sup>

---

<sup>1</sup>We use the terms *enzyme* and *bio-/catalyst* interchangeably.

---

*Book Title*

Editor Name

Copyright © Year Pan Stanford Publishing Pte. Ltd.

ISBN

[www.panstanford.com](http://www.panstanford.com)



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[Claeysens *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 2013.

By giving an overview and some details in this article, we hope to also, at least in part, answer these questions.

## 1.2 Introduction

We provide an introduction into the topic of enzyme catalysis modeling for interested people from inside and outside of the field. We outline the latest achievements, methods in use and outline required developments to establish computational enzyme engineering not only in the literature, but also in (commercial) practice.

Modeling involving enzymes happens in different forms. In one form, the enzyme is a constant parameter and the variable is the inhibitor. These tasks are usually addressed by docking approaches and studies involving millions of substrates have been published[Zhou and Caffisch (2010)]. In another form, the substrate is constant but the enzyme can vary. The approaches to this task, as it appears, are much less established. The reasons are manifold, primarily it appears that modeling of an enzyme is less

straightforward than a small inhibiting molecule. Again, why is that?

As is known in the community, enzymes behave highly non-predictable. The best reasoning about a mutation (except for elimination of the catalytic residues) will not even allow to decide *qualitatively* if the mutated enzyme will be more or less active. Shifts in pH optima might be predicted more readily[Ludwiczek *et al.* (2013)]. It is proposed therefore that in cases where no *a priori* knowledge about how an enzyme should be engineered is available, the only reasonable strategy is to carry out systematic screening studies. For these to be feasible, either the computer power needs to be very large or the method has to be fast, there is an inherent trade off in this situation. Once initial lead structures are identified, higher level methods can be used to narrow down on most promising candidates.

This review is written for a more general audience. For specialized details, the reader is referred to the primary literature. 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 points where special care needs to be taken.

### 1.3 Methods

A variety of methods has been established for the use to model enzyme catalysis. Depending on the properties of interest, the system is treated differently. Molecular mechanics methods allow to study the structural behavior of the enzyme over a significant time period and provide details about rearrangement of loop motives. The description of chemical reactivity however requires the description of the electronic structure of the system (molecular mechanics models do not allow the description of bond formation and braking processes). In this regime, again a number of specialized methods have become available.

### 1.3.1 Idea collection

Practical applicability of QM based enzyme screening has a couple of problems:

- 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
- *Every method has it's chemistry.* The results of one QM chemistry need to be verified, usually high-level standard reference methods are computationally too demanding to be applicable to a large system
- 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 look like, unfortunately only small community
- Many papers 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: “*We do not think that this approach is that relevant for high-throughput screening for several reasons. First, the amount of calculations required for a single compound is not scalable to thousands or millions of ligands, and second, because the degree of detail provided by the methodology is probably not needed during the screening phase. Instead, we envisage a successful application in the lead optimization phase on possible tens of lead compounds to completely reconstruct and understand their binding pathways before moving forward in the pipeline. The method, although computationally expensive, is certainly feasible if performed on just a few ligands as indicated. Here, we have made use of our volunteer GPU project GPUGRID.net (7) based on ACEMD (6). Extrapolating from the current study, it would be possible to reconstruct the binding pathways of 5 to 10 ligands per week in-house on a moderately sized GPU cluster of 3264 GPUs.*” [Buch et al. (2011)]

### 1.3.2 Calculation methods

Different methods used should be briefly described to make it easier for the reader of the literature to understand what is going on

- AM1/RM1/PM3/PM6/PM7/MOZYME
- DFT/B3LYP/Dispersion corrections/M06/...
- Modeling procedures: Define constraints- how does it affect the calculation (examples in literature?)
- Explain parameters of PM6: MOZYME, gradient convergence and NDDO cutoff

## 1.4 Applications

### 1.4.1 General overview

We have recently presented two studies where new functionality has been introduced into an enzyme active site using quantum chemical methods[Hediger *et al.* (2012, 2013a,b)].

It is currently agreed that in many cases, rational arguments alone are not sufficient to design new reactivity in an active site. Put simply, enzymes do not behave in the way the engineer hopes for. Much rather, the only meaningful strategy proves to be to screen a large number of variant candidates for apparent activity of the desired kind. In a second step, identified lead candidates are subjected to higher level computational methods and wet-lab experiments which then allow to further establish or dismiss the nature of the candidate. Given the large mutational space available, initial *in silico* screening has to be efficient. Hediger *et al.* therefore chose to use semi-empirical methods for the description of the electronic structure of the enzyme-substrate complex. In their approach, 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  $k_{\text{cat}}$  becomes rate limiting, a reasonable assumption under industrial conditions.

**General outline of method.** In both studies, the basic procedure consists of preparing structures for the enzyme substrate complex (“ES”) and the first intermediate (CALB: “TI”<sup>2</sup>, BCX: “GE”<sup>3</sup>) and calculating the reaction barrier between these two pairs of structures, respectively. The reaction barrier is calculated by preparing intermediate structures of the model along the reaction coordinate  $\text{ES}_{\text{CALB}} \rightarrow \text{TI}$  or  $\text{ES}_{\text{BCX}} \rightarrow \text{GE}$  (around 10 intermediate structures, 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 major reason for

---

<sup>2</sup>Tetrahedral intermediate

<sup>3</sup>Glycosyl enzyme

meaningless results is significant structural rearrangement during the geometry optimization.

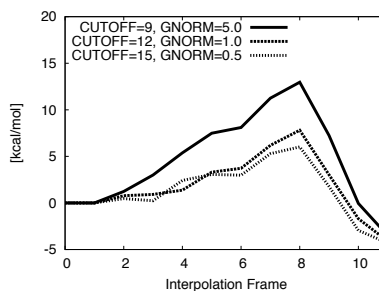
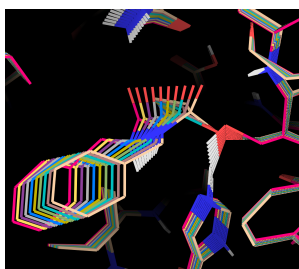
**Requirements.** The most basic requirements are that a structure representing a bound intermediate along the reaction pathway is available. Alternatively, a homology model could be used. Furthermore, ideally the mechanism for the rate determining step is established in the literature. In case no generally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

#### 1.4.1.1 *Candida antarctica* Lipase B (CALB) Engineering

##### UPCOMING

The questions answered in this study are:

- 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?



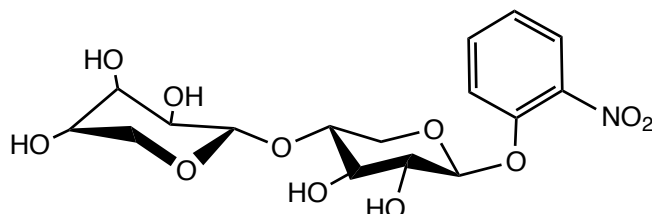
**Figure 1.1** Intermediate structures along reaction coordinate of CALB reaction.

CALB is used as a versatile catalyst in a variety of applications among which enzymatic kinetic resolution of enantiomeric mixtures is of high pharmaceutical importance[Gotor-Fernández *et al.* (2006)]. While CALB is also known for its reactive promiscuity[Svedendahl *et al.* (2008)], if the catalyst can be engineered such as to increase its performance for certain of these applications, possible bottle-necks in production could be widened. About the cutoff: this is only relevant in studies where experimental data for many mutants is available. Then it can be used to assess the lower barrier for the accuracy. Given all approximations, x out of y mutants are correctly predicted. So any method improving on the approximations should provide at least at many correct hits. About the cutoff: this is only relevant in studies where experimental data for many mutants is available. Then it can be used to assess the lower barrier for the accuracy. Given all approximations, x out of y mutants are correctly predicted. So any method improving on the approximations should provide at least at many correct hits. A number of weak points remain:

- Selection of model ambiguous/not full structure
- Not systematically screening all mutants
- No other protein system/enzymatic reaction tested
- Binding effects not considered
- Large rearrangements can not be modeled
- Semi-empirical methods might not be adequate for metals
- Subsequent steps of reaction not modeled (eg. product leaving of active site)
- Different ionization states of residues

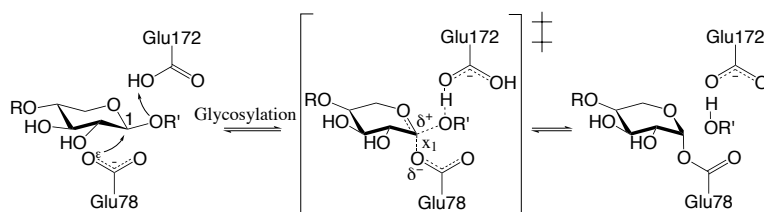
#### 1.4.1.2 *Bacillus circulans* xylanase engineering

**Introduction.** In order to address the weak points summarized above and to show general applicability of the approach outlined for CALB, the xylanase of *Bacillus circulans* (BCX) is engineered to be catalytically active towards an artificial substrate (*ortho*-nitrophenol- $\beta$ -xylobioside), Fig. 1.2. In this study, the mechanism was taken from the literature as reported previously[Joshi *et al.* (2000, 2001)].



**Figure 1.2** Substrate used in BCX study.

The paper follows the following reasoning in order to establish independence of results on modeling procedure and internal consistency within the framework of approximations (both from the quantum chemical model as well as the limited including of physical effects). In a first step, the wild-type (WT) reference reaction barrier height is established. This is a crucial step because all subsequent conclusions about mutants are depending on how good the WT reaction barrier is estimated. Again only the rate determining step is modeled, Fig. 1.3. As outlined in the methods



**Figure 1.3** Conventional glycosylation step.  $x_1$ : constrained reaction coordinate;  $R$ : xylose;  $R'$ : *ortho*-nitrophenol (ONP).  $C_1$  indicating nucleophilic carbon of first xylose unit[Hediger *et al.* (2013b)].

section, the reaction barrier is mapped out by fixing the reaction coordinate at a fixed interval 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 not constrained and



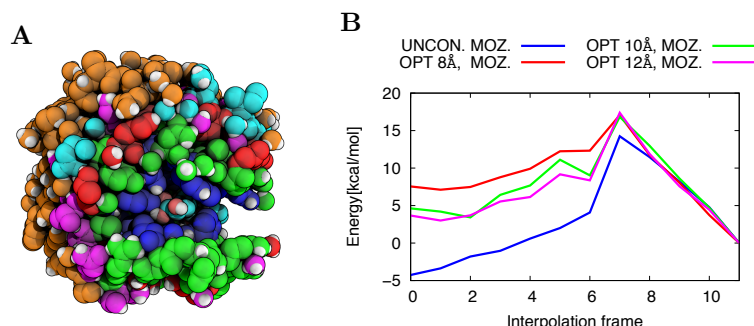
the proton is free to transfer or remain on E172. In simple terms, forcing the quantum chemical model to do something it would not if unconstrained, would result in an increased reaction barrier and therefore it is advisable to keep the number of constraints as low as possible.

**Molecular modeling procedure and calculation parameters.** In a second step, the modeling procedure is worked out. A PDB structure using a covalent 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. There is also an apoenzyme of BCX available, however placing a substrate into an empty active site requires considerable rearrangement of active site residues. Furthermore, ambiguity in how the substrate should be placed in the active site can be reduced using the structure including the inhibitor, which also increases reproducibility of the results. As shown in Fig. 2 of the paper, 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, in the second (“Interpolation 2”), first the GE structure is optimized and only then is the ES structure prepared. The main advantage of the second approach is that a significant amount of computer time can be saved because a large part of the enzyme is already optimized when the ES structure is prepared. The time required for optimizing the structures by these two approaches are found to differ by a factor of six. A similar approach reveals that the greatest calculation efficiency is reached when the structures of the mutated enzymes are prepared from WT structures with an already optimized backbone scaffold.

In the next step, an assessment of parameter dependency of optimized structure energies reveals that within the given framework of approximations, conclusive results are obtained when the gradient convergence criterion is set to 1 kcal/(molÅ) and an NDDO cutoff of 15 Å is used. From such an approach, the reaction barrier is found to be 18.5 kcal/mol which is very close to the experimentally reported value of 17.0 kcal/mol obtained from transition state theory[Joshi *et al.* (2000)].

Then, it is evaluated how the results depend on the size of the model included in the optimization calculations. In the below

figure, the different layers of constraints are illustrated. The

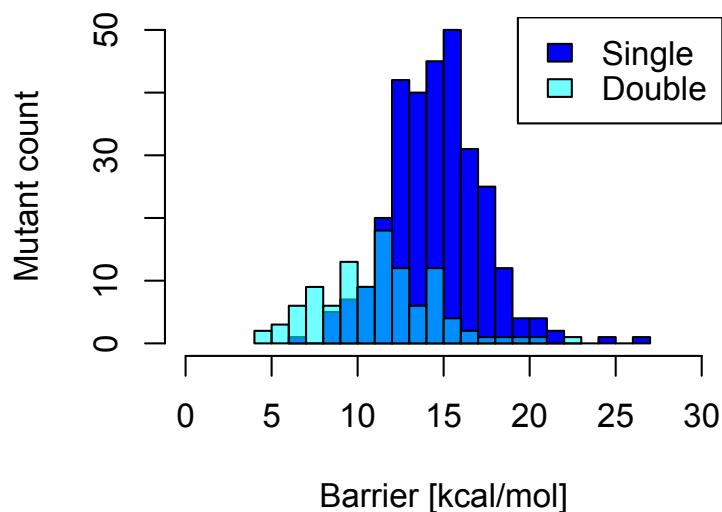


**Figure 1.4** **A:** Illustration of BCX with bound ONP substrate. Color coded residue layers have at least one atom within the respective distance to any atom of the substrate: *green* 8 Å, *red* 10 Å, *purple* 12 Å, *cyan* 14 Å. *Brown* residues are always constrained (they are optimized in the WT model), *blue* are mutated residues. The substrate is visible in the center of the figure. **B:** Calculated reaction barriers for different layers of optimized residues.[Hediger *et al.* (2013b)]

calculated barrier depends on the number of constraints in the model (note, this does not correspond to the constraint used in defining the reaction coordinate). However, it is assumed that the same reduction in reaction barrier carries over to the models of the mutants and therefore cancels out as a source of error.

These experiments allowed to clearly justify the selected quantum chemical model (PM6) and its parametric configuration, the sequence of modeling steps in order to carry out the calculations and the layer thickness of optimized residues around the active site. In the final step, using PYMOL[The PyMOL Molecular Graphics System, Schrödinger, LLC (2010)], systematically all possible single mutants within a given distance of the substrate are prepared<sup>4</sup>. Analysis of reaction barriers reveals that, after combining the best single mutants from all mutated positions, 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.

<sup>4</sup>The PYMOL script used for the preparation of the mutants is available at <https://github.com/mzhKU/Enzyme-Screening/blob/master/vsc-bcx.py>.



**Figure 1.5** Barrier distribution of single and double BCX mutants[Hediger *et al.* (2013b)].

### 1.4.2 Interesting QM/MM applications

#### The cocaine hydrolase guys

Make a cocaine hydrolase better[Gao *et al.* (2006)].

### 1.4.3 Requirements for advancement of methods

Experimental catalytic data ( $k_{\text{cat}}$ ) of many mutants to validate against.

## 1.5 Conclusions

### 1.5.1 Advantages/Disadvantages of using quantum chemistry for modeling enzymatic reactions

- Bond formation/breaking can be modeled
- Large systems might require constraints to reach convergence

*Conclusions* | **13**

- Calculation times depend strongly on model chemistry used
- Not yet as established and convenient to use like e.g. docking software



## Bibliography

- Beilen, J. and Li, Z. (2002). Enzyme technology: an overview, *Current Opinion in Biotechnology* **13**, 4, pp. 338–344.
- 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.
- 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.
- Gao, D., Cho, H., Yang, W., Pan, Y., Yang, G., Tai, H.-H. and Zhan, C.-G. (2006). Computational design of a human butyrylcholinesterase mutant for accelerating cocaine hydrolysis based on the transition-state simulation, *Angewandte Chemie International Edition* **45**, 4, pp. 653–657.
- 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.

## 16 | Bibliography

- 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.
- 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.
- 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.
- 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.
- 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>.
- The PyMOL Molecular Graphics System, Schrödinger, LLC (2010).
- 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.