Publishers' page

Publishers' page

Publishers' page

Publishers' page

Dedication Page

(optional)

Contents

Chapter 1

Working title: Calculations

Martin R. Hediger^a, Harm Otten^b

^{a}*Zürich*
^{b}*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 $\xrightarrow{Enzyme}$ 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.[1]

---

[1]We use the terms *enzme* and *bio-/catalyst* interchangeably.

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[Claeyssens *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 Caflisch (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.

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 intimatly on the structure, optimizing two structures differing by 0.1 Åin 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 friendlyness 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 en- visage 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/...

1.4 **Applications**

1.4.1 *Method Hediger*

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

Candida antarctica Lipase B Engineering

Introducing amidase functionality into the lipase of *Candida
antarctica* (CALB).

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.

Bacillus circulans glycoside hydrolase engineering

As a proof of concept and to show general applicability of the
approach outlined for CalB, the glycoside hydrolase of *Bacillus
circulans* is engineered to be catalytically active towards an
artifical substrate. Importantly, we note that the same method as
developed for computational screening of CALB activity could be
directly transfered to the study of BCX activity. The requirements
are that a structure representing a bound intermediate along the

reaction pathway is available. Ideally, the mechanism for the rate determining step is established in the literature, however this is not a critical requirement. If the mechanism is not established, the method can just as well be used to outline competing mechanisms.

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_{cat}) of many mutants to validate against.

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%2Fjournal.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.

10 *Bibliography*

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., DAngelo, 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`.

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